UvA-DARE (Digital Academic Repository)

Standard errors and confidence intervals of norm statistics for educational and psychological tests

Oosterhuis, H.E.M.; van der Ark, L.A.; Sijtsma, K.

DOI

[10.1007/s11336-016-9535-8](https://doi.org/10.1007/s11336-016-9535-8)

Publication date

2017

Document Version

Final published version

Published in

Psychometrika

[Link to publication](#)

Citation for published version (APA):

Oosterhuis, H. E. M., van der Ark, L. A., & Sijtsma, K. (2017). Standard errors and confidence intervals of norm statistics for educational and psychological tests. *Psychometrika*, 82(3), 559-588. <https://doi.org/10.1007/s11336-016-9535-8>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

STANDARD ERRORS AND CONFIDENCE INTERVALS OF NORM STATISTICS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTS

HANNAH E. M. OOSTERHUIS

TILBURG UNIVERSITY

L. ANDRIES VAN DER ARK

UNIVERSITY OF AMSTERDAM

KLAAS SIJTSMA

TILBURG UNIVERSITY

Norm statistics allow for the interpretation of scores on psychological and educational tests, by relating the test score of an individual test taker to the test scores of individuals belonging to the same gender, age, or education groups, et cetera. Given the uncertainty due to sampling error, one would expect researchers to report standard errors for norm statistics. In practice, standard errors are seldom reported; they are either unavailable or derived under strong distributional assumptions that may not be realistic for test scores. We derived standard errors for four norm statistics (standard deviation, percentile ranks, stanine boundaries and Z-scores) under the mild assumption that the test scores are multinomially distributed. A simulation study showed that the standard errors were unbiased and that corresponding Wald-based confidence intervals had good coverage. Finally, we discuss the possibilities for applying the standard errors in practical test use in education and psychology. The procedure is provided via the R function `check.norms`, which is available in the `mokken` package.

Key words: norm statistics, percentile norms, standard errors for norm statistics, test norms.

1. Introduction

Norm statistics allow for the interpretation of scores on educational and psychological tests, by relating the test score of an individual test taker to the test scores of a group of individuals having, for example, the same gender, age, or education level. Examples of norm statistics frequently used in practice are percentile ranks, linear standard scores such as Z-scores, and normalized standard scores such as stanines (Mertler, 2007, Module 6). A norm statistic obtained from a normative sample should be viewed as a point estimate of the norm in the population (Crawford, Garthwaite, & Slick, 2009), which means that the norm estimate should be accompanied by an indication of estimation precision.

The publication manual of the American Psychological Association (2010, p. 34) also requires that when point estimates are provided, the authors “always include an associated measure of variability (precision), with an indication of the specific measure used (e.g., the standard error)”. In addition, on the same page the publication manual strongly recommends reporting confidence intervals (CIs). CIs, of which Wald-based CIs are the most commonly used, are directly related to the standard errors (SEs). Let $\hat{\theta}$ be the point estimate with standard error $SE_{\hat{\theta}}$, and let α be

Correspondence should be made to Hannah E. M. Oosterhuis, Department of Methodology and Statistics, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands. Email: H.E.M.Oosterhuis@tilburguniversity.edu

the two-tailed p value and $z_{\alpha/2}$ the corresponding Z -score. Then the limits of the $100(1 - \alpha) \%$ Wald-based CI are

$$\hat{\theta} \pm z_{\alpha/2} \cdot SE_{\hat{\theta}}. \quad (1)$$

Hence, CIs for norm scores serve to remind us that norm estimates based on normative samples are fallible and quantify the degree of fallibility that is caused by estimation imprecision (Crawford, Cayley, Lovibond, Wilson, & Hartley, 2011). However, norm developers commonly fail to provide SEs or CIs to quantify estimation precision of the norms (e.g., Aardoom, Dingemans, Slof Op't Land, & Van Furth, 2012; Cavaco et al., 2013; Glaesmer et al., 2012; Goretti et al., 2014; Grande, Romppel, Glaesmer, Petrowski, & Herrmann-Lingen, 2010; Kessels, Montagne, Hendriks, Perrett, & De Haan, 2014; Mond, Hay, Rodgers, & Owen, 2006; Palomo et al., 2011; Sartorio et al., 2013; Shi et al., 2014), because for many norm statistics the SEs are unknown, difficult to derive, or if available not included in popular software. For example, Crawford et al. (2011) described a procedure for obtaining CIs for percentile norms, but it cannot be used for estimating CIs for other norm statistics such as stanine boundaries. For other norm statistics, the SEs are known only under strong assumptions. For example, the SE of the standard deviation assumes data to be normally distributed, but test scores often are discrete and their distribution skewed because the test is too difficult or the attribute measured is rare (test-score distribution skewed to the right), or the test is too easy or the attribute is highly prevalent (test-score distribution skewed to the left).

We derived SEs for the test-score standard deviation, percentile rank scores, the boundaries of the stanines, and Z -scores. The SEs were derived under the mild assumption that raw test scores follow a multinomial distribution. Let a trial have three or more possible outcomes, and let N represent the number of independent identical trials. The probabilities of obtaining a certain outcome follow a multinomial distribution (Agresti, 2013, p. 6). This model can easily be extended to the raw scores obtained by means of psychological and educational tests. For example, let the number of items answered correctly be the raw score on such a test. The administration of the test to a norm sample of N respondents then corresponds to N trials, and each value of the raw score corresponds to a trial outcome. The probabilities of obtaining each raw score then follow a multinomial distribution. Although the method is based on a discrete distribution, the test scores need not be integers. Hence, the method can also be applied to continuous measures such as reaction time or blood pressure. For large N , the multivariate normal central limit theorem (see, e.g., Rao, 1973, p. 128) ascertains that the multinomial distribution is close to the multivariate normal distribution if the model parameters are not near the boundaries of the parameter space (i.e., probabilities are not close to 0 or 1). Hence, the method described in this article can also be used for data that are (approximately) normally distributed.

This article is structured as follows. First, we provide an example that illustrates the use of SEs and CIs for norm statistics in practice. Second, we explain the general framework for deriving SEs for norm statistics. Third, using this general framework we derived the SEs for the test-score standard deviation, the Z -scores, the percentile rank scores, and the stanine boundaries. Fourth, using a simulation study we investigated the bias and the precision of the derived SE estimates and the coverage of the corresponding Wald-based CIs. Finally, we briefly discuss the results and provide computer code to obtain the SEs for norm statistics.

2. An Illustration of Using Norms With and Without Standard Errors

Because estimation precision, quantified by SEs or CIs, usually is not taken into account when test constructors present norm statistics, norm statistics are often presented and interpreted as if

FIGURE 1.

Comparison of raw scores to percentile norms when **a** no measurement or norm sampling error, **b** only measurement error, or **c** both measurement and norm sampling error are taken into account.

they were parameters. Ignoring imprecision of norm statistics may produce wrong conclusions about test performance and can have serious consequences for individual test takers if decisions are based on their test performance. As an example, Fig. 1 shows the scores on Social Skills measured by means of the Preschool and Kindergarten Behavior Scales (PKBS; Merrell, 1994) for three five-year-old boys named Oliver, Jack, and Harry. The horizontal lines at raw scores 59

and 75 correspond to the 5th and 20th percentile ranks of the raw-score distribution, respectively. These percentile ranks were estimated in a norm sample, and according to the PKBS test manual (Merrell, 1994), a score below the 5th percentile rank (i.e., raw score 59) indicates a significant deficit of social skills, a score between the 5th and 20th percentile ranks (i.e., raw scores 59 and 76, respectively) indicates a moderate deficit, and a score above the 20th percentile rank (i.e., raw score 76) indicates average social skills. In what follows, we distinguish the influence of random measurement error on test performance, typical of classical test theory, and the influence of sampling error on norm statistics.

In Fig. 1a, influence of measurement error and sampling error was not taken into account for the individual scores and the norm score, respectively, which means that all values were treated as population values. Oliver, Jack, and Harry had raw scores equal to 56, 82, and 61, respectively. Based on Fig. 1a, we conclude that Oliver has a significant social skills deficit, Jack has average social skills, and Harry has a moderate deficit.

Next, we assess the influence of random measurement error on test performance and follow the classical test model, which assumes that an observable test score (X) can be decomposed into an unobservable true score (T) and unobservable random measurement error (E), so that $X = T + E$. In Fig. 1b, the norm value was treated as a population value, but random measurement error related to individual test scores was taken into account using a 68% CI, or a score band. The score band contained the likely values of the true score and was based on the standard error of measurement (SEM), which indicates the magnitude of measurement error associated with a particular test score and in the classical model it is equal for all test scores. Hence, the width of the score band indicates the degree to which an individual's test score is expected to vary across replications. Several methods for estimating the SEM are available (Brennan & Lee, 1999; Lee, Brennan, & Kolen, 2000). If the score band contained the norm value, the individual's true score was not significantly different from the norm value based on a 68% confidence level. Oliver's score band (i.e., [53.2; 58.8]) was located completely below the 5th percentile rank (i.e., raw score 59), meaning we conclude that Oliver has a significant deficit. Jack's score band (i.e., [79.2; 84.8]) was located above the 20th percentile rank (i.e., raw score 76), meaning we conclude that Jack's skills are average. However, Harry's score band (i.e., [58.2; 63.8]) contained the 5th percentile rank (i.e., raw score 59), meaning we are uncertain whether Harry's deficit is significant or moderate.

In Fig. 1c, in addition to random measurement error for the individual test scores, sampling error was taken into account for the percentile rank values. The horizontal dotted lines correspond to the boundaries of the 95% CIs for the percentile ranks. Using the heuristic rule that an overlap of 25% or less between the CIs of two statistics suggests a significant difference between statistics (Van Belle, 2003, Sect. 2.6), we conclude that Jack's true score differed significantly from the 20th percentile rank, indicating that he has average social skills. On the other hand, we conclude that Oliver's and Harry's true scores did not differ significantly from the 5th percentile rank, which means that for both boys we are not sure whether they suffer from a significant or a moderate deficit.

This example shows that the interpretation of the boys' test results changed depending on whether measurement error and sampling error were taken into account for the individual test results and for the norm values, respectively. Especially for individuals that score close to norm values, taking into account the sampling error of the norm values can have important consequences for the interpretation of their test scores.

Crawford and Howell (1998) have argued that treating norm statistics as population values is justifiable if the sample is large enough. However, there are no clear size requirements for a norm sample. For example, Evers, Lucassen, Meijer, and Sijtsma (2009) have provided Dutch test constructors with practical guidelines for the size of norm groups, but these guidelines lack a sufficient statistical basis. The American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education

(NCME) have provided guidelines for test construction (AERA, APA, & NCME, 1999), without sample size recommendations.

In addition to sample size, precision of a norm estimate also depends on the statistic's location in the norm sample distribution. It is well known that, empirically, extreme test scores show more measurement error variation (Brennan & Lee, 1999; Lee et al., 2000). Hence, norms that are based on extreme test scores are also expected to show more error variation. For example, for equal norm sample size, Oosterhuis, Van der Ark, and Sijtsma (2016) found that raw scores associated with percentile ranks further away from the median were estimated with less precision than percentile ranks closer to the median. Hence, a norm sample might be large enough for precise estimation of the median or the mean, but not for extreme percentile scores. SEs and CIs can be used to quantify the sampling error associated with each norm statistic and might be used to determine whether the sample was large enough to obtain precise estimates for all norm values of interest.

3. A General Framework for Deriving SEs Under a Multinomial Distribution

3.1. A Two-Step Procedure

We used a general framework consisting of two steps to compute the SEs of the norm statistics (e.g., Bergsma, Croon, & Hagnaars, 2009; Kuijpers, Van der Ark, & Croon, 2013a). The first step is to write the norm statistic as a function of the frequencies of the raw scores. Suppose that the raw scores, obtained from administering a test to a sample of N respondents, are collected in an $N \times 1$ vector $\mathbf{x}$. It may be noted that the unique realizations of the raw score need not be integers and need not include all possible realizations. The k unique realizations of the raw score $r_{(1)}, \dots, r_{(k)} (r_{(i)} < r_{(i+1)})$, for $i = 1, \dots, k$, are contained in a $k \times 1$ vector $\mathbf{r}$, and the expected and observed frequencies of the realizations $r_{(1)}, \dots, r_{(k)}$ are collected in a $k \times 1$ vector $\mathbf{m}$ and a $k \times 1$ vector $\hat{\mathbf{m}}$, respectively. Let $\theta = (\theta_1, \dots, \theta_l)'$ denote a vector containing l —possibly mutually dependent—norms, such as the $l = k$ Z-scores or the $l = 8$ boundaries of the stanines. The first step encompasses showing that $\hat{\theta} = \mathbf{g}(\hat{\mathbf{m}})$, where $\mathbf{g}(\cdot)$ is a vector function. For a single norm, such as the mean or the standard deviation, vector $\hat{\theta}$ reduces to scalar $\hat{\theta}$, and $\hat{\theta} = \mathbf{g}(\hat{\mathbf{m}})$ reduces to $\hat{\theta} = g(\hat{\mathbf{m}})$. The technique to obtain $\mathbf{g}(\hat{\mathbf{m}})$ is called the *generalized exp-log notation* and is described later.

In the second step, the delta method (e.g., Agresti, 2013, Sect. 16.1.4; Kendall & Stuart, 1977, pp. 246–247) is used to approximate the variance of the norm statistic. If $\hat{\mathbf{m}}$ is a consistent estimator, it converges to its true value $\mathbf{m}$, and the central limit theorem can be applied to obtain asymptotical normality,

$$(\hat{\mathbf{m}} - \mathbf{m}) \xrightarrow{D} N(0, \mathbf{V}_{\hat{\mathbf{m}}}), \quad (2)$$

where $\mathbf{V}_{\hat{\mathbf{m}}}$ is the covariance matrix of $\hat{\mathbf{m}}$. Let $\mathbf{D}(\mathbf{y})$ be a diagonal matrix with the elements of vector $\mathbf{y}$ on the diagonal, and let $\mathbf{Y}'$ be the transpose of $\mathbf{Y}$. Under a multinomial distribution, let $\hat{\mathbf{V}}_{\hat{\mathbf{m}}}$ be the sample estimate of the variance of $\hat{\mathbf{m}}$,

$$\hat{\mathbf{V}}_{\hat{\mathbf{m}}} = \mathbf{D}(\hat{\mathbf{m}}) - \hat{\mathbf{m}}N^{-1}\hat{\mathbf{m}}' \quad (3)$$

(e.g., Agresti, 2013, p. 6). Using the first two terms of the Taylor series, $\mathbf{g}(\hat{\mathbf{m}})$ can be approximated by

$$\mathbf{g}(\hat{\mathbf{m}}) \approx \mathbf{g}(\mathbf{m}) + \mathbf{G}'(\hat{\mathbf{m}} - \mathbf{m}), \quad (4)$$

where $\mathbf{G}$ is the matrix of first partial derivatives, or the Jacobian, of $\mathbf{g}(\hat{\mathbf{m}})$ with respect to $\hat{\mathbf{m}}$, evaluated at $\mathbf{m}$. Equation 4 implies that the variance of $\mathbf{g}(\hat{\mathbf{m}})$ is approximately

$$\mathbf{V}_{\mathbf{g}(\hat{\mathbf{m}})} \approx \mathbf{G} \left(\hat{\mathbf{V}}_{\hat{\mathbf{m}}} \right) \mathbf{G}'. \quad (5)$$

Therefore, the delta method implies that

$$(\mathbf{g}(\hat{\mathbf{m}}) - \mathbf{g}(\mathbf{m})) \xrightarrow{D} N(0, \mathbf{G} \mathbf{V}_{\hat{\mathbf{m}}} \mathbf{G}'). \quad (6)$$

Based on Eq. 6, the sample estimate of the asymptotic variance of $\mathbf{g}(\hat{\mathbf{m}})$ is

$$\begin{aligned} \hat{\mathbf{V}}_{\mathbf{g}(\hat{\mathbf{m}})} &= \mathbf{G} \hat{\mathbf{V}}_{\hat{\mathbf{m}}} \mathbf{G}' \\ &= \mathbf{G} \left(\mathbf{D}(\hat{\mathbf{m}}) - \hat{\mathbf{m}} N^{-1} \hat{\mathbf{m}}' \right) \mathbf{G}' \\ &= \mathbf{G} \mathbf{D}(\hat{\mathbf{m}}) \mathbf{G}' - \mathbf{G} \hat{\mathbf{m}} N^{-1} \hat{\mathbf{m}}' \mathbf{G}'. \end{aligned} \quad (7)$$

By taking the square root of the diagonal elements of $\hat{\mathbf{V}}_{\mathbf{g}(\hat{\mathbf{m}})}$, the sample estimate of the asymptotic SEs for $\mathbf{g}(\hat{\mathbf{m}})$ is obtained. In some cases, Eq. 7 can be further reduced. Let the constant $t > 0$; if $\mathbf{g}(\hat{\mathbf{m}}) = \mathbf{g}(t\hat{\mathbf{m}})$, then vector function $\mathbf{g}(\hat{\mathbf{m}})$ is homogeneous of order 0. If $\mathbf{g}(\hat{\mathbf{m}})$ is homogeneous of order 0, then $\mathbf{G} \hat{\mathbf{m}} N^{-1} \hat{\mathbf{m}}' \mathbf{G}' = \mathbf{0}$, and Eq. 7 reduces to

$$\hat{\mathbf{V}}_{\mathbf{g}(\hat{\mathbf{m}})} = \mathbf{G} \mathbf{D}(\hat{\mathbf{m}}) \mathbf{G}' \quad (8)$$

(e.g., Bergsma, 1997, Appendix D). For example, if $\mathbf{g}(\hat{\mathbf{m}})$ is homogeneous of order 0, then it does not matter whether absolute frequencies ($\hat{\mathbf{m}}$) or relative frequencies (e.g., proportions $\hat{\mathbf{P}} = \hat{\mathbf{m}}/N$) are used as the argument of $\mathbf{g}$, because for $t = 1/N$, $t\hat{\mathbf{m}} = \hat{\mathbf{P}}$.

3.2. Generalized exp-log Notation

In general, obtaining function $\mathbf{g}(\hat{\mathbf{m}})$ and Jacobian $\mathbf{G}$ requires tedious derivations. The generalized exp-log notation (Bergsma, 1997; Grizzle, Starmer, & Koch, 1969; Kritzer, 1977) is a method that alleviates this problem by rewriting $\mathbf{g}(\hat{\mathbf{m}})$ using a series of q appropriate design matrices $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_q$ (to be explained below), such that

$$\mathbf{g}(\hat{\mathbf{m}}) = \mathbf{A}_q \cdot \exp \left(\mathbf{A}_{q-1} \cdot \log \left(\mathbf{A}_{q-2} \cdot \exp \left(\dots \log \left(\mathbf{A}_3 \cdot \exp \left(\mathbf{A}_2 \cdot \log \left(\mathbf{A}_1 \cdot \hat{\mathbf{m}} \right) \right) \right) \right) \right) \right). \quad (9)$$

The reason for using such complex notation to write $\mathbf{g}(\hat{\mathbf{m}})$ is to make it relatively easy to derive the Jacobian using the chain rule (e.g., Larson & Edwards, 2013, pp. 129–135). The Jacobian for Eq. 9 can be derived using the following steps (Bergsma, 1997, pp. 66–68; Kritzer, 1977; Kuijpers et al., 2013a). First, a series of $q + 1$ functions $\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_q$ are defined, such that

$$\mathbf{g}_0 = \hat{\mathbf{m}}, \quad (10)$$

and, for $i = 1, \dots, q$,

$$\mathbf{g}_i = \log(\mathbf{A}_i \cdot \mathbf{g}_{i-1}), \text{ if } i \text{ is an odd number,} \quad (11)$$

and

$$\mathbf{g}_i = \exp(\mathbf{A}_i \cdot \mathbf{g}_{i-1}), \text{ if } i \text{ is an even number.} \quad (12)$$

The last function in this series is

$$\mathbf{g}(\hat{\mathbf{m}}) = \mathbf{g}_q = \mathbf{A}_q \cdot \mathbf{g}_{q-1}. \quad (13)$$

Second, let $\mathbf{G}_0, \mathbf{G}_1, \dots, \mathbf{G}_q$ denote the matrices of the partial derivatives of $\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_q$, with respect to $\hat{\mathbf{m}}$, respectively. Let $\mathbf{I}_p$ denote an identity matrix of order p . Standard calculus shows that

$$\mathbf{G}_0 = \frac{\partial \mathbf{g}_0}{\partial \hat{\mathbf{m}}'} = \mathbf{I}_k. \quad (14)$$

Furthermore, let $\mathbf{Y}^{-1}$ be the inverse of $\mathbf{Y}$. For $i = 1, \dots, q-1$,

$$\mathbf{G}_i = \frac{\partial \mathbf{g}_i}{\partial \hat{\mathbf{m}}'} = \mathbf{D}^{-1}(\mathbf{A}_i \cdot \mathbf{g}_{i-1}) \cdot \mathbf{A}_i \cdot \mathbf{G}_{i-1}, \text{ if } i \text{ is an odd number,} \quad (15)$$

and

$$\mathbf{G}_i = \mathbf{D}(\exp(\mathbf{A}_i \cdot \mathbf{g}_{i-1})) \cdot \mathbf{A}_i \cdot \mathbf{G}_{i-1}, \text{ if } i \text{ is an even number.} \quad (16)$$

Finally, the last function in the series is

$$\mathbf{G} = \mathbf{G}_q = \mathbf{A}_q \cdot \mathbf{G}_{q-1}. \quad (17)$$

3.3. SEs for Norm Statistics

To illustrate the derivation of SEs for norm statistics, the case of the sample mean is discussed extensively. For other norm statistics, the derivation is only outlined, and details can be found in Appendices 1–4.

3.4. Mean Test Score

Sample mean $\bar{X} = \sum_i X_i / N$ (i indexes persons) can be cast in the generalized exp-log notation (Eq. 5), using three design matrices as follows. Let $\mathbf{1}_p$ be a vector of ones of length p , let $\mathbf{A}_1$ be a $2 \times k$ matrix, defined as $\mathbf{A}_1 = [\mathbf{r} \ \mathbf{1}_k]'$, let $\mathbf{A}_2$ be a 1×2 vector, defined as $\mathbf{A}_2 = [1 \ -1]$, and let $\mathbf{A}_3$ be a 1×1 matrix containing the scalar 1, that is, $\mathbf{A}_3 = [1]$. It follows that $\mathbf{g}_0 = \hat{\mathbf{m}}$ (Eq. 10); next that

$$\mathbf{g}_1 = \log(\mathbf{A}_1 \cdot \hat{\mathbf{m}}) = \log\left([\mathbf{r} \ \mathbf{1}_k]' \cdot \hat{\mathbf{m}}\right) = \log\left(\begin{bmatrix} \mathbf{r}' \cdot \hat{\mathbf{m}} \\ \mathbf{1}_k' \cdot \hat{\mathbf{m}} \end{bmatrix}\right) = \begin{bmatrix} \log(\sum X_i) \\ \log(N) \end{bmatrix} \quad (18)$$

(Eq. 11); then

$$\mathbf{g}_2 = \exp(\mathbf{A}_2 \cdot \mathbf{g}_1) = \exp\left([1 \ -1] \cdot \begin{bmatrix} \log(\sum X_i) \\ \log(N) \end{bmatrix}\right) = \left[\frac{\sum X_i}{N}\right], \quad (19)$$

(Eq. 12); and finally that

$$\mathbf{g}(\hat{\mathbf{m}}) = \mathbf{g}_3 = \mathbf{A}_3 \cdot \mathbf{g}_2 = [1] \cdot \left[\frac{\sum X_i}{N} \right] = \frac{\sum X_i}{N} \quad (20)$$

(Eq. 13), which equals the sample mean. It may be noted that $\mathbf{g}(\hat{\mathbf{m}})$ is homogenous of order 0.

Next, the Jacobian matrices $\mathbf{G}_0$, $\mathbf{G}_1$, $\mathbf{G}_2$, and $\mathbf{G}_3$ can be derived for $\mathbf{g}_0$, $\mathbf{g}_1$, $\mathbf{g}_2$, and $\mathbf{g}_3$, respectively. First, it follows that $\mathbf{G}_0 = \mathbf{I}_k$ (Eq. 14). Second, let $\mathbf{y}/p$ denote the element-wise division of $\mathbf{y}$ by p . Then $\mathbf{G}_1$ is the $2 \times k$ matrix,

$$\mathbf{G}_1 = \frac{\partial \mathbf{g}_1}{\partial \hat{\mathbf{m}}'} = \mathbf{D}^{-1} [\mathbf{A}_1 \cdot \mathbf{g}_0] \cdot \mathbf{A}_1 \cdot \mathbf{G}_0 = \mathbf{D}^{-1} \left[\frac{\sum X_i}{N} \right] \cdot \begin{bmatrix} \mathbf{r}' \\ \mathbf{1}'_k \end{bmatrix} \cdot \mathbf{I} = \begin{bmatrix} \mathbf{r}' / \sum X_i \\ \mathbf{1}'_k / N \end{bmatrix} \quad (21)$$

(Eq. 15). Third, $\mathbf{G}_2$ is the $1 \times k$ vector,

$$\begin{aligned} \mathbf{G}_2 &= \frac{\partial \mathbf{g}_2}{\partial \hat{\mathbf{m}}'} = \mathbf{D} [\exp(\mathbf{A}_2 \cdot \mathbf{g}_1)] \cdot \mathbf{A}_2 \cdot \mathbf{G}_1 = \left[\frac{\sum X_i}{N} \right] \cdot [1 \quad -1] \cdot \begin{bmatrix} \mathbf{r}' / \sum X_i \\ \mathbf{1}'_k / N \end{bmatrix} \\ &= (\mathbf{r}' - \bar{X} \cdot \mathbf{1}'_k) / N \end{aligned} \quad (22)$$

(Eq. 16). Finally, $\mathbf{G}_3$ is the $1 \times k$ vector,

$$\mathbf{G} = \mathbf{G}_3 = \mathbf{A}_3 \cdot \mathbf{G}_2 = [1] \cdot ((\mathbf{r}' - \bar{X} \cdot \mathbf{1}'_k) / N) = (\mathbf{r}' - \bar{X} \cdot \mathbf{1}'_k) / N \quad (23)$$

(Eq. 17).

Because $\mathbf{g}(\hat{\mathbf{m}})$ (Eq. 20) is homogenous of order 0, the variance of $\mathbf{g}(\hat{\mathbf{m}})$ (Eq. 8) is approximated by

$$\begin{aligned} V_{\bar{X}} &\approx \mathbf{GD}(\hat{\mathbf{m}})\mathbf{G}' = [(\mathbf{r}' - \bar{X} \cdot \mathbf{1}'_k) \mathbf{D}(\hat{\mathbf{m}}) (\mathbf{r} - \bar{X} \cdot \mathbf{1}_k)] / N^2 \\ &= \sum_{j=1}^k \hat{m}_j [(r_j - \bar{X}) / N]^2. \end{aligned} \quad (24)$$

Note that Eq. 24 equals the variance of a population of size N divided by N . The approximated SE for the sample mean is then given by

$$S_{\bar{X}} = \sqrt{V_{\bar{X}}}. \quad (25)$$

Note that Eq. 25 equals the standard deviation of a population of size N divided by the square root of N .

3.5. Standard Deviation

The unbiased sample estimator of the standard deviation of the raw scores equals

$$s_X = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2}. \quad (26)$$

Let i and j be index realizations of $\mathbf{r}$, let δ_{ij} be Kronecker's delta ($\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ if $i \neq j$), let $d_i = r_i - \bar{X}$, let $e = \frac{1}{N-1}$, and let $SS = \sum (X_i - \bar{X})^2$. In Appendix 1, we show that the standard error of s_X can be approximated by

$$S_{s_X} \approx 0.5s_X \sqrt{\sum_i \sum_j \left(\frac{d_i^2}{SS} - e \right) \left(\frac{d_j^2}{SS} - e \right) \left(\delta_{ij} \hat{m}_i - \frac{\hat{m}_i \hat{m}_j}{N} \right)}. \quad (27)$$

For large N , Eq. 27 reduces to

$$S_{s_X} \approx 0.5s_X \sqrt{\sum_i \hat{m}_i \left(\frac{d_i^2}{SS} - \frac{1}{N} \right)^2}. \quad (28)$$

Ahn and Fessler (2003) derived a SE estimate for the sample standard deviation under the assumption that the data are normally distributed,

$$\dot{S}_{s_X} = \frac{s_X}{\sqrt{2(N-1)}}. \quad (29)$$

3.6. Percentile Rank Scores

Percentile rank score PR_x is the percentage of individuals in the normative sample that have raw score x or lower. For example, if $X = 20$ is associated with $PR_{20} = 90$, 90% of the individuals in the normative sample have a raw score of 20 or lower. If researchers use percentile ranks as norm statistics, they either associate the raw scores with particular percentile ranks (e.g., 75 or 95) or they present the percentile ranks for each possible raw score. Percentile ranks can also be used as cut-off scores. For example, a job selection test may use the raw score corresponding to the 95th percentile as the cut-off score. Then, individuals are only selected if they belong to the highest 5% of the raw score distribution compared to individuals having, for example, the same age, sex, or education. Let x be a particular raw score on the test of interest, and let $P(\cdot)$ denote a sample proportion. Then PR_x is estimated as

$$\begin{aligned} PR_x &= \{P(X < x) + 1/2 \cdot P(X = x)\} \cdot 100 \\ &= 50 \cdot P(X \leq x) + 50 \cdot P(X < x). \end{aligned} \quad (30)$$

In Appendix 2, we show that the sample estimate of the asymptotic standard error of PR_x equals

$$S_{PR_x} \approx \frac{50}{N} \sqrt{\sum_i \hat{m}_i (\gamma_{xi} - P(X < x) - P(X \leq x))^2}, \quad (31)$$

where $\gamma_{xi} = 0$ if $x < i$, $\gamma_{xi} = 1$ if $x = i$, and $\gamma_{xi} = 2$ if $x > i$.

3.7. Stanines

Stanines are norm scores 1, 2, ..., 9 that correspond to ranges under the normal distribution. The first stanine corresponds to raw scores smaller than 1.75 standard deviations below the mean, and the ninth stanine corresponds to raw scores larger than 1.75 standard deviations above the mean. The remaining seven stanines (stanines 2, 3, ..., 8) have a width of half a standard deviation and divide the range between 1.75 standard deviations below the mean and 1.75 standard deviations above the mean in segments 0.5 standard deviations wide. The raw score mean is located in the middle of the fifth stanine. When stanines are used as norm statistics, the raw scores that correspond to the eight boundaries of the nine stanines are presented. Let $\mathbf{f}$ be an 8×1 vector containing values starting from -1.75 and ending at 1.75 with increments of 0.50 . The elements of $\mathbf{f}$ are denoted f_b ($b = 1, \dots, 8$) and correspond to the eight boundaries of the stanines. On the scale of the raw scores, the b th boundary of the stanines is given by

$$St_b = f_b \cdot s_X + \bar{X}. \quad (32)$$

Let $d_i^* = d_i^2 / SS = (r_i - \bar{X})^2 / [\sum X_i^2 - (\sum X_i)^2 / N]$. In Appendix 3, we show that the standard error of St_b can be approximated by

$$S_{St_b} \approx \sqrt{\sum_{j=1}^k \sum_{i=1}^k \left[\delta_{ij} \hat{m}_j - \frac{\hat{m}_i \hat{m}_j}{N} \right] \cdot \left[\frac{f_b s_X}{2} \cdot (d_i^* - e) + \frac{d_i}{N} \right] \cdot \left[\frac{f_b s_X}{2} \cdot (d_j^* - e) + \frac{d_j}{N} \right]}. \quad (33)$$

For large N , Eq. 33 reduces to

$$S_{St_b} \approx \sqrt{\sum_{i=1}^k \hat{m}_i \cdot \left[\frac{f_b s_X}{2} \cdot \left(d_i^* - \frac{1}{N} \right) + \frac{d_i}{N} \right]^2}. \quad (34)$$

3.8. Z-Scores

Z-scores are expressed in standard deviations relative to the group mean. For instance, if the mean raw score equals 15 and the standard deviation equals 2, a raw score equal to 18 corresponds to a Z-score equal to $(18 - 15)/2 = 1.5$. When Z-scores are used as norm statistics, either a selection of the raw scores or all raw scores are presented with the corresponding Z-scores. The Z-score corresponding to X_h is computed as

$$Z_h = (X_h - \bar{X}) / s_X = X_h / s_X - \bar{X} / s_X. \quad (35)$$

In Appendix 4, we show that the standard error of Z_h can be approximated by

$$S_{Z_h} \approx \sqrt{\sum_{j=1}^k \sum_{i=1}^k t \left[-\bar{X} \cdot \left(\frac{r_i}{\sum X} - \frac{1}{N} - u_i \right) - r_h u_i \right] \cdot \left[-\bar{X} \cdot \left(\frac{r_j}{\sum X} - \frac{1}{N} - u_j \right) - r_h u_j \right]}, \quad (36)$$

where $t = \frac{1}{s_X^2} \left(\delta_{ij} \hat{m}_j - \frac{\hat{m}_i \hat{m}_j}{N} \right)$ and $u_i = 0.5 (d_i^* - e)$. For large N , Eq. 36 reduces to

$$S_{Z_h} \approx \sqrt{\sum_i \frac{\hat{m}_i}{s_X^2} \cdot \left(-\bar{X} \cdot \left(\frac{r_i}{\sum X_i} - \frac{1}{N} + u_i \right) - r_h u_i \right)^2}. \quad (37)$$

4. Simulation Study

We performed a simulation study to investigate the bias and the precision of the SE estimates of the standard deviation, the percentile rank scores, the stanine boundaries, and the Z-scores (Eqs. 23, 26, 29, and 32, respectively), as well as the coverage probability of Wald CIs based on the SE estimates of these norm statistics for different sample and test characteristics.

4.1. Simulation Model

We simulated data using the 2-parameter logistic model (2PLM; Birnbaum, 1968; Van der Linden & Hambleton, 1997). The 2PLM describes response probabilities to each item j (with scores $x = 0, 1$) by means of a logistic function with a slope parameter α_j and a location parameter β_j . The probability of a score of $X_j = 1$ on item j equals

$$P(X_j = 1 | \theta) = \frac{\exp[\alpha_j(\theta - \beta_j)]}{1 + \exp[\alpha_j(\theta - \beta_j)]}. \quad (38)$$

Slope parameter $\alpha_j = 0.85, 0.95, 1.05, \dots, 1.75$, where $j = 1, 2, \dots, 10$. For tests consisting of more than 10 items, the first ten slope parameters $\alpha_1, \dots, \alpha_{10}$ had the same values as slope parameters $\alpha_{11}, \dots, \alpha_{20}$ and $\alpha_{21}, \dots, \alpha_{30}$, etcetera. Location parameter $\beta_j = -2.25, -1.75, -1.25, \dots, 2.25$, where $j = 1, 2, \dots, 10$. For tests consisting of more than 10 items, the first ten location parameters $\beta_1, \dots, \beta_{10}$ had the same values as location parameters $\beta_{11}, \dots, \beta_{20}$ and $\beta_{21}, \dots, \beta_{30}$, etcetera.

First, θ values were randomly drawn from a standard normal distribution. Second, each θ value was inserted in Eq. 38, and the resulting probabilities were used to generate item score vectors by means of random draws from a multinomial distribution. The raw score for each θ value was then computed as $X = \sum X_j$ and could range from 0 to J , the total number of items in the test. The simulated item scores were discrete; hence, the resulting raw test scores were also discrete. Given that item location was symmetrically dispersed around 0, and that θ was sampled from a normal distribution, the raw scores were approximately normally distributed, especially for a larger number of items. The simulation study consisted of $Q = 10,000$ replications, which guaranteed the stability of the results.

4.2. Dependent Variables

Bias of the SEs. To estimate the bias of the SEs, we first computed the standard deviation of the norm estimates across the Q replications. Let $\bar{\theta} = \frac{1}{Q} \sum_{q=1}^Q \hat{\theta}_q$. Then

$$s_{\hat{\theta}} = \sqrt{\frac{1}{Q-1} \sum_{q=1}^Q (\hat{\theta}_q - \bar{\theta})^2}. \quad (39)$$

Standard deviation $s_{\hat{\theta}}$ estimates the variance of the norm statistic across replications and serves as a gold standard for the SE. Let $\hat{S}_{\hat{\theta}_q}$ denote the estimated SE of the q th norm statistic. Then, the bias of the SE estimator equals

$$bias.se = \frac{1}{Q} \sum_{q=1}^Q (\hat{S}_{\hat{\theta}_q} - s_{\hat{\theta}}). \quad (40)$$

RMSE of the SEs. The root mean square error (RMSE) was used as a measure of the precision of the SE estimator of a statistic. To estimate the RMSE of the SEs, we first computed $s_{\hat{\theta}}$ (Eq.

39). The RMSE is then computed as follows:

$$RMSE_{\hat{\theta}} = \sqrt{\frac{1}{Q} \sum_{q=1}^Q (\hat{S}_{\hat{\theta}_q} - s_{\hat{\theta}})^2}. \quad (41)$$

The bias and the RMSE of the standard deviation and the stanine boundaries could not always be readily compared across design cells because the bias and the RMSE of these norm statistics are scale dependent. To alleviate this problem, we divided the bias and the RMSE of the SEs of these norm statistics by the average standard deviation of the simulated scores; that is, we divided by

$$\bar{s}_Y = \frac{1}{Q} \sum_{q=1}^Q s_{Y_q}, \quad (42)$$

where s_{Y_q} is the standard deviation of the simulated raw scores in the q th replication. Hence, bias and RMSE are expressed on a scale with mean 0 and standard deviation 1. Dividing the bias and the RMSE by $\bar{s}_Y$ guaranteed a fair comparison of the bias and the RMSE of the standard deviation and the stanine boundaries for different sample and test characteristics. For the bias and the RMSE of the percentile rank scores, expressed on a scale from 0 to 100, and Z-scores expressed on a scale with mean 0 and variance 1, such a correction was not necessary.

Coverage of the 95% CI. To investigate the coverage of the 95% CI, in each replication we first constructed a Wald-based CI (Eq. 1) for $\hat{\theta}_q$. The 95% coverage was defined as the proportion of replications in which CI contained $\bar{\theta}$. Given that the lowest and highest scores might not be observed in all replicated samples, only replications in which a particular raw score was observed were used to compute the coverage.

4.3. Independent Variables

In the simulation study, the test and norm sample characteristics were varied as follows.

Number of items (J) The number of items was equal to either $J = 10, 30$, or 50 . These values were chosen to cover the range of test lengths often found in practice (Oosterhuis et al., 2016).

Sample size (N) The sample size was equal to either $N = 500, 1000, 1500, 2000$, or 2500 . These values were chosen to cover sample sizes often found in practice (Oosterhuis et al., 2016).

5. Results

For tests with 10, 30, and 50 items, $\bar{s}_Y$ (Eq. 42) was equal to 2.076, 5.530, and 8.966, respectively. These values were used to adjust the bias and the RMSE of the SEs of the standard deviation and the stanine boundaries.

5.1. Standard Deviation

For none of the combinations of sample size and test length, the SE of the standard deviation (Eq. 27) showed bias, whereas the SE of the standard deviation derived by Ahn & Fessler (2003; Eq. 29), $\hat{S}_{S_X}$ (Eq. 29), was positively biased (Table 1). The bias of $\hat{S}_{S_X}$ decreased as sample size increased. Precision of our SE was higher than the precision of $\hat{S}_{S_X}$ (Table 1). For both SEs, precision increased as sample size increased. The precision of both SEs did not depend on test length. The coverage of the 95% CIs that were based on our SE was close to 0.95 for all combinations of sample size and test length, whereas the coverage of the CIs based on $\hat{S}_{S_X}$ was too high (Table 1).

TABLE 1.
Standardized bias, standardized RMSE, and coverage probability of 95% CIs of the estimated SEs of the standard deviation.

Items	N	Bias		RMSE		Coverage	
		S_{s_X}	$\hat{S}_{S_X}^*$	S_{s_X}	$\hat{S}_{S_X}^*$	S_{s_X}	$\hat{S}_{S_X}^*$
10	500	−0.0001	0.0037	0.0010	0.0038	0.949	0.973
	1000	−0.0001	0.0026	0.0005	0.0026	0.951	0.974
	1500	−0.0001	0.0021	0.0003	0.0021	0.947	0.971
	2000	0.0000	0.0019	0.0002	0.0019	0.950	0.97
	2500	0.0000	0.0017	0.0002	0.0017	0.951	0.976
30	500	−0.0004	−0.0033	0.0012	0.0034	0.946	0.970
	1000	−0.0002	0.0024	0.0006	0.0024	0.949	0.972
	1500	0.0000	0.0021	0.0004	0.0021	0.944	0.971
	2000	−0.0001	0.0016	0.0003	0.0017	0.948	0.971
	2500	0.0000	0.0016	0.0002	0.0016	0.949	0.973
50	500	−0.0002	0.0035	0.0011	0.0036	0.951	0.975
	1000	−0.0001	0.0024	0.0006	0.0025	0.952	0.973
	1500	−0.0001	0.0020	0.0004	0.0020	0.952	0.975
	2000	0.0001	0.0019	0.0003	0.0019	0.950	0.973
	2500	−0.0001	0.0014	0.0003	0.0015	0.95	0.972

* SE based on Ahn & Fessler (2003, Eq. 29).

5.2. Percentiles

For a test consisting of 10 items, for none of the sample sizes, the SEs of the percentile rank scores (Eq. 31) showed bias. On the other hand, for tests consisting of ≥ 30 items, the lowest and highest raw scores showed small positive bias (Fig. 2). For these extreme raw scores, bias decreased as sample size increased. For $N = 500$, precision of the SE estimates only increased as test length increased (Fig. 2). For other sample sizes, test length did not influence estimation precision. The RMSE of the SEs first increased as scores became more extreme, but decreased for the most extreme scores. For all conditions and percentile rank scores, the coverage of the 95% CIs was generally close to 0.95, except when the raw scores were closer to the extremes of the scale (Fig. 3). For these extreme raw scores, the coverage was lower than 0.95. Furthermore, a sharp increase of the coverage was visible for the most extreme raw scores.

The increase of the RMSE and the decrease of the coverage for scores further away from the mean can be explained by the decrease of the observed scores further away from the mean, whereas the decrease of the RMSE and the increase of the coverage for the most extreme scores were caused by the proximity of the corresponding percentile ranks to 0 and 100, respectively. Considering that the extreme raw scores had a low prevalence in the samples, in absolute terms they did not change much from sample to sample. This resulted in a lower RMSE and a higher coverage.

We further investigated which percentile rank scores were associated with the extreme raw scores that corresponded to low coverage. We found that the coverage of the CIs was close to 0.95 when $N > 500$ and the population percentile rank scores were $\geq 1\%$ or $\leq 99\%$. For a test consisting of 10 items (Fig. 3), percentile ranks $\geq 1\%$ and $\leq 99\%$ corresponded to the raw scores of 2 and 10, respectively. For tests containing 30 or 50 items (Fig. 3), the raw scores for percentile ranks $\geq 1\%$ and $\leq 99\%$ were equal to 4 and 28, or 7 and 45, respectively. For $N = 500$, the coverage of the CIs was close to 0.95 when the population percentile rank score was $\geq 2.5\%$ and $\leq 97.5\%$. These percentile rank scores corresponded to raw scores 2 and 9, 6 and 26, and 9 and 43 for tests containing 10, 30, and 50 items, respectively.

FIGURE 2.

Bias and RMSE of percentile rank SEs for $N = 500$ (dotted), 1000 (dot-dashed), 1500 (long dashed), 2000 (two-dashed), or 2500 (dashed).

5.3. Stanine Boundaries

For none of the combinations of sample size and test length, the SEs of the stanine boundaries (Eq. 33) showed bias (Fig. 4). Precision of the SE estimates increased as sample size increased (Fig. 4). Precision was lower for the lowest and highest stanine boundaries, which can be explained by the smaller number of observations located further away from the mean. For all conditions and stanine boundaries, the coverage of the 95% CIs was close to 0.95 (Fig. 5).

5.4. Z-scores

For none of the combinations of sample size and test length, the SEs of the Z-score estimates (Eq. 36) showed bias (Fig. 6). Precision of the SE estimates increased as sample size increased (Fig. 6). For all conditions and Z-scores, the coverage of the 95% CIs was close to 0.95 (Fig. 5).

FIGURE 3.

Coverage of 95% CIs for percentile ranks for $N = 500$ (dotted), 1000 (dot-dashed), 1500 (long dashed), 2000 (two-dashed), or 2500 (dashed).

6. Discussion

We derived SEs for the standard deviation, percentile rank scores, stanine boundaries, and Z-scores, assuming that the data were multinomially distributed. This is a reasonable assumption, which does not restrict the applicability of the SEs. The SEs were derived using the delta method in conjunction with the generalized exp-log notation. A simulation study suggested that for all practical purposes the SEs of the standard deviation, the stanine boundaries, and the Z-scores were unbiased, whereas the SE of the standard deviation proposed by Ahn and Fessler (2003, Eq. 29) was found to be biased. We also found that Wald-based CIs based on the SEs of the standard deviation, the stanines, and the Z-scores had good coverage.

SEs of the percentile ranks were unbiased, unless the statistic was close to 0 or 100. We found that small samples (i.e., $N < 1000$) resulted in unbiased SE estimates and CIs with good coverage if the percentile ranks were $\geq 2.5\%$ and $\leq 97.5\%$. SEs or CIs for more extreme percentile ranks (e.g., the 99th percentile rank) were unbiased and had good coverage if a larger sample (i.e., $N \geq 1000$) was used to obtain unbiased estimates. It is well known that the Wald CI may have poor coverage for small samples and for parameters close to the boundary of the parameter space (e.g., Agresti & Min, 2001). Score and profile likelihood CIs for contingency tables (Lang, 2008; also see Agresti, 2012, Sect. 2.3.3) are range preserving, and in the future, they may provide an alternative for norm statistics close to the boundary of the parameter space. Currently, these CIs are not ready to be used in the application at hand as models for norm statistics have not yet been developed. If the intended use of the test requires the distinction between extreme percentile ranks (i.e., $< 2.5\%$ and $> 97.5\%$), the poor coverage of CIs for extreme percentile ranks can be alleviated by increasing the size of the normative sample. This will result in narrower CIs, so that the boundaries of the CI are less likely to cross the parameter boundaries. Alternatively, if a larger normative sample is not feasible, the methods described by Crawford et al. (2011) can be used to derive SEs. However, these methods can only be used to calculate SEs for percentile ranks, which means that test developers still need to employ a different method to estimate SEs for other norm statistics, such as stanine boundaries. Given that the Wald-based CIs performed very well in most cases and given the computational complexity in score and profile likelihood CIs, we recommend using Wald CIs.

The same simulation study also suggested that the precision of the SE estimates of the percentiles ranks, the stanine boundaries, and the Z-scores increased as the number of observations for the corresponding raw scores increased. Hence, the norm estimates that are of interest for the intended use of the test should be based on a sufficient number of observations. For example, if extreme percentile ranks or Z-scores are of interest, a larger sample is required, whereas the SE of the standard deviation can be precisely estimated in relatively small norm samples.

FIGURE 4.

Bias and RMSE of stanine boundary SEs for $N = 500$ (dotted), 1000 (dot-dashed), 1500 (long dashed), 2000 (two-dashed), or 2500 (dashed).

The procedure to derive SEs assumes that a simple sampling design was used. However, normative samples are often based on stratified sampling, which means that the SEs should be adjusted to represent the sampling design. For a range of different sampling designs, Lehtonen and Pahkinen (2004, pp. 61–64) provide an overview of methods to adjust SEs of point estimates. These methods are based on the design effect of the type of sampling that was used to obtain the normative sample and can be readily applied to the SEs derived by our procedure.

The delta method only uses the first partial derivatives to approximate the function of a random variable, which means that this method assumes that the function is linear in the expected range of the parameter. If the function is highly nonlinear and the data show considerable variation, the delta method might not result in a close approximation of the variance of the function (Cooch &

FIGURE 5.

Coverage of 95% CIs for stanine boundaries and Z-scores for $N = 500$ (dotted), 1000 (dot-dashed), 1500 (long dashed), 2000 (two-dashed), or 2500 (dashed).

White, 2015, Appendix B). For percentile rank scores, the delta method can be readily used to approximate the SEs of percentile rank scores, because these statistics are linear combinations of the observed raw score frequencies. On the other hand, the standard deviation, the stanine boundaries, and the Z-scores are nonlinear transformations. Incorporating higher-order Taylor expansions in the derivation of the corresponding SE might result in closer approximations of the SEs. However, such higher-order Taylor expansions very quickly result in lengthy and complicated derivations, which according to the results of our simulation study are unnecessary. Furthermore, the delta method is based on the central limit theorem and therefore cannot be applied to small samples. However, using small samples is not advocated in establishing norms for psychological and educational tests.

The methodology proposed in this article can also be applied to other statistics that can be written as a function of the observed frequencies of the unique realizations of the scores. For example, SEs have been derived for Cronbach's alpha (Kuijpers, Van der Ark, & Croon, 2013b) and coefficient H (Kuijpers et al., 2013a; Van der Ark, Croon, & Sijtsma, 2008). Further research is necessary to determine whether the methodology used in this article can be applied to other functions, such as coefficients, association measures, or goodness-of-fit statistics.

We recommend test developers to provide CIs for norm values that are presented in norm tables. Individual test results can then be compared to the CIs of the norms, instead of the point estimates. As a result, sampling error of norm values is taken into account and, especially for test scores close to the norm values, the comparison of test scores to the norms is less likely to result in erroneous decisions. Applied researchers can easily access the procedure for obtaining SEs and CIs for the mean, standard deviation, percentile ranks scores, stanine boundaries, and Z-scores via the *R* (R Core Team, 2015) function *check.norms*, which is available in the *mokken* package (Van der Ark, 2012).

FIGURE 6.

Bias and RMSE of Z-score SEs for $N = 500$ (dotted), 1000 (dot-dashed), 1500 (long dashed), 2000 (two-dashed), or 2500 (dashed).

Acknowledgments

The funding was provided by Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NL) (Grant No. 406-12-013).

Appendix 1

The sample estimate of the standard deviation, s_X , can be written using the generalized exp-log notation (Eq. 5) as

$$s_X = \mathbf{A}_5 \cdot \exp \left(\mathbf{A}_4 \cdot \log \left(\mathbf{A}_3 \cdot \exp \left(\mathbf{A}_2 \cdot \log \left(\mathbf{A}_1 \cdot \hat{\mathbf{m}} \right) \right) \right) \right). \quad (43)$$

Let $\mathbf{y}^{(2)}$ be the vector containing the squares of the elements in $\mathbf{y}$. Then, the $4 \times k$ matrix $\mathbf{A}_1$ equals

$$\mathbf{A}_1 = [\mathbf{r} \ \mathbf{r}^{(2)} \ \mathbf{1}_k \ \mathbf{1}_k]', \quad (44)$$

the 4×4 matrix $\mathbf{A}_2$ equals

$$\mathbf{A}_2 = \begin{bmatrix} 2 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \quad (45)$$

the 2×4 matrix $\mathbf{A}_3$ equals

$$\mathbf{A}_3 = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \quad (46)$$

the 1×2 vector $\mathbf{A}_4$ equals

$$\mathbf{A}_4 = \left[\frac{1}{2} \ -\frac{1}{2} \right], \quad (47)$$

and $\mathbf{A}_5$ equals the scalar 1. It follows that $\mathbf{g}_0 = \hat{\mathbf{m}}$, the 4×1 vector $\mathbf{g}_1$ equals

$$\mathbf{g}_1 = \log(\mathbf{A}_1 \cdot \mathbf{g}_0) = \log([\mathbf{r} \ \mathbf{r}^{(2)} \ \mathbf{1}_k \ \mathbf{1}_k]' \cdot \hat{\mathbf{m}}) = \begin{bmatrix} \log(\sum X_i) \\ \log(\sum X_i^2) \\ \log(N) \\ \log(N) \end{bmatrix}, \quad (48)$$

the 4×1 vector $\mathbf{g}_2$ equals

$$\mathbf{g}_2 = \exp(\mathbf{A}_2 \cdot \mathbf{g}_1) = \exp\left(\begin{bmatrix} 2 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} \log(\sum X_i) \\ \log(\sum X_i^2) \\ \log(N) \\ \log(N) \end{bmatrix}\right) = \begin{bmatrix} \frac{(\sum X_i)^2}{\sum X_i^2} \\ \frac{\sum X_i^2}{N} \\ 1 \end{bmatrix}, \quad (49)$$

the 2×1 vector $\mathbf{g}_3$ equals

$$\mathbf{g}_3 = \log(\mathbf{A}_3 \cdot \mathbf{g}_2) = \log\left(\begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} \frac{(\sum X_i)^2}{\sum X_i^2} \\ \frac{\sum X_i^2}{N} \\ 1 \end{bmatrix}\right) = \begin{bmatrix} \log(SS) \\ \log(N-1) \end{bmatrix}, \quad (50)$$

where $SS = \sum X_i^2 - \frac{(\sum X_i)^2}{N}$, and

$$\mathbf{g}_4 = \exp(\mathbf{A}_4 \cdot \mathbf{g}_3) = \exp\left(\left[\frac{1}{2} \ -\frac{1}{2}\right] \cdot \begin{bmatrix} \log(SS) \\ \log(N-1) \end{bmatrix}\right) = \sqrt{\frac{SS}{N-1}}, \quad (51)$$

because $\mathbf{A}_5 = \mathbf{1}$, $\mathbf{g}(\hat{\mathbf{m}}) = \mathbf{g}_5 = \mathbf{g}_4 = s_X$.

After some tedious algebra, it may be verified that $\mathbf{G}_0 = \mathbf{I}_k$,

$$\begin{aligned} \mathbf{G}_1 &= \mathbf{D}^{-1} [\mathbf{A}_1 \cdot \mathbf{g}_0] \cdot \mathbf{A}_1 \cdot \mathbf{G}_0 \\ &= \mathbf{D}^{-1} \begin{bmatrix} \sum_{i=1}^N X_i \\ \sum_{i=1}^N X_i^2 \\ N \\ N \end{bmatrix} \cdot [\mathbf{r} \ \mathbf{r}^{(2)} \ \mathbf{1}_k \ \mathbf{1}_k]' \cdot \mathbf{I} = \begin{bmatrix} \mathbf{r}' \\ \frac{\sum X_i}{\sum X_i^2} \\ \frac{\mathbf{r}^{(2)'} }{\sum X_i^2} \\ \frac{\mathbf{1}_k'}{N} \\ \frac{\mathbf{1}_k'}{N} \end{bmatrix}, \end{aligned} \quad (52)$$

$$\begin{aligned} \mathbf{G}_2 &= \mathbf{D} [\exp(\mathbf{A}_2 \cdot \mathbf{g}_1)] \cdot \mathbf{A}_2 \cdot \mathbf{G}_1 \\ &= \mathbf{D} \begin{bmatrix} \frac{(\sum X_i)^2}{\sum X_i^2} \\ \sum X_i^2 \\ N \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{r}' \\ \frac{\sum X_i}{\sum X_i^2} \\ \frac{\mathbf{r}^{(2)'} }{\sum X_i^2} \\ \frac{\mathbf{1}_k'}{N} \\ \frac{\mathbf{1}_k'}{N} \end{bmatrix} = \begin{bmatrix} 2\mathbf{r}'\bar{X} - \bar{X}^2 \\ \mathbf{r}^{(2)'} \\ \mathbf{1}_k' \\ \mathbf{0}_k' \end{bmatrix}. \end{aligned} \quad (53)$$

Then, $\mathbf{G}_3$ is a $2 \times k$ matrix,

$$\begin{aligned} \mathbf{G}_3 &= \mathbf{D}^{-1} [\mathbf{A}_3 \cdot \mathbf{g}_2] \cdot \mathbf{A}_3 \cdot \mathbf{G}_2 \\ &= \mathbf{D}^{-1} \begin{bmatrix} SS \\ N-1 \end{bmatrix} \cdot \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} 2\mathbf{r}'\bar{X} - \bar{X}^2 \\ \mathbf{r}^{(2)'} \\ \mathbf{1}_k' \\ \mathbf{0}_k' \end{bmatrix} = \begin{bmatrix} \mathbf{d}^{(2)'} \\ \frac{SS}{\mathbf{e}'} \end{bmatrix}, \end{aligned} \quad (54)$$

where the $k \times 1$ vector $\mathbf{d} = \mathbf{r} - \bar{X}$ and $\mathbf{e} = \mathbf{1}_k/(N-1)$,
and $\mathbf{G}_4$ is the $1 \times k$ vector

$$\begin{aligned} \mathbf{G}_4 &= \mathbf{D} [\exp(\mathbf{A}_4 \cdot \mathbf{g}_3)] \cdot \mathbf{A}_4 \cdot \mathbf{G}_3 \\ &= \mathbf{D} [s_X] \cdot \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{d}^{(2)'} \\ \frac{SS}{\mathbf{e}'} \end{bmatrix} = 0.5s_X \left(\frac{\mathbf{d}^2}{SS} - \mathbf{e} \right). \end{aligned} \quad (55)$$

Finally, one can derive that $\mathbf{G} = \mathbf{G}_5 = \mathbf{G}_4$.

Because the sample estimate of the standard deviation is obtained by dividing the sum of squared deviation scores by $N-1$, $\mathbf{g}(\hat{\mathbf{m}})$ (Eq. 51) is not homogeneous of order 0. As a result, the asymptotic variance is approximated by $V_{s_X} \approx \mathbf{G}\mathbf{D}(\hat{\mathbf{m}})\mathbf{G}' - \mathbf{G}\hat{\mathbf{m}}N^{-1}\hat{\mathbf{m}}'\mathbf{G}'$ (cf. Eq. 2). Inserting the Jacobian $\mathbf{G}$ (Eq. 55) in Eq. 2 results in the sample estimate of the asymptotic variance of s_X ,

$$V_{s_X} \approx .25s_X^2 \cdot \sum_i \sum_j \left(\frac{d_i^2}{SS} - e \right) \left(\frac{d_j^2}{SS} - e \right) \left(\delta_{ij}\hat{m}_i - \frac{\hat{m}_i\hat{m}_j}{N} \right). \quad (56)$$

For large N ,

$$V_{s_X} \approx .25s_X^2 \cdot \sum_i \hat{m}_i \left(\frac{d_i^2}{SS} - \frac{1}{N} \right)^2. \quad (57)$$

Appendix 2

The sample estimates of the percentile ranks, collected in vector $\mathbf{PR} = (PR_{r_1} \cdots PR_{r_k})'$ (Eq. 26), can be written as

$$\mathbf{PR} = \mathbf{A}_3 \cdot \exp(\mathbf{A}_2 \cdot \log(\mathbf{A}_1 \cdot \hat{\mathbf{m}})). \quad (58)$$

Let $\mathbf{L}_p$ be a lower triangular matrix of ones. The $(k+1) \times k$ matrix $\mathbf{A}_1$ equals

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{L}_k \\ \mathbf{1}'_k \end{bmatrix}, \quad (59)$$

and the $k \times (k+1)$ matrix $\mathbf{A}_2$ equals

$$\mathbf{A}_2 = [\mathbf{I}_k - \mathbf{1}_k]. \quad (60)$$

Let $\mathbf{A}_3$ be a lower bidiagonal $k \times k$ matrix in which each nonzero element equals 50; that is,

$$\mathbf{A}_3 = 50 \left(\mathbf{I}_k + \begin{bmatrix} \mathbf{0}'_{k-1} & 0 \\ \mathbf{I}_{k-1} & \mathbf{0}_{k-1} \end{bmatrix} \right). \quad (61)$$

It follows that $\mathbf{g}_0 = \hat{\mathbf{m}}$. Let $\hat{\mathbf{m}}^* = (\sum_{i=1}^1 \hat{m}_i, \sum_{i=1}^2 \hat{m}_i, \dots, \sum_{i=1}^k \hat{m}_i)$ be a vector of cumulative frequencies; that is, $\hat{\mathbf{m}}^* = \mathbf{L}_k \cdot \hat{\mathbf{m}}$, and let $\mathbf{P}^* = \hat{\mathbf{m}}^*/N = (P(X \leq r_1), \dots, P(X \leq r_k))'$. Function $\mathbf{g}_1$ is a $(k+1) \times 1$ vector,

$$\mathbf{g}_1 = \log(\mathbf{A}_1 \cdot \hat{\mathbf{m}}) = \log\left(\begin{bmatrix} \mathbf{L}_k \\ \mathbf{1}'_k \end{bmatrix} \cdot \hat{\mathbf{m}}\right) = \log\left[\begin{matrix} \hat{\mathbf{m}}^* \\ N \end{matrix}\right], \quad (62)$$

$\mathbf{g}_2$ is a $k \times 1$ vector,

$$\mathbf{g}_2 = \exp(\mathbf{A}_2 \cdot \mathbf{g}_1) = \exp\left(\begin{bmatrix} \mathbf{I}_k - \mathbf{1}_k \end{bmatrix} \cdot \log\left[\begin{matrix} \hat{\mathbf{m}}^* \\ N \end{matrix}\right]\right) = \mathbf{P}^*, \quad (63)$$

and $\mathbf{g}_3$ is a $k \times 1$ vector,

$$\mathbf{g}(\hat{\mathbf{m}}) = \mathbf{g}_3 = \mathbf{A}_3 \cdot \mathbf{g}_2 = 50 \left(\mathbf{I}_k + \begin{bmatrix} \mathbf{0}'_{k-1} & 0 \\ \mathbf{I}_{k-1} & \mathbf{0}_{k-1} \end{bmatrix} \right) \cdot \mathbf{P}^* = \mathbf{PR}. \quad (64)$$

It follows that $\mathbf{G}_0 = \mathbf{I}$, the $(k+1) \times k$ matrix $\mathbf{G}_1$ equals

$$\mathbf{G}_1 = \mathbf{D}^{-1} [\mathbf{A}_1 \cdot \mathbf{g}_0] \cdot \mathbf{A}_1 \cdot \mathbf{G}_0 = \mathbf{D}^{-1} \left(\begin{bmatrix} \hat{\mathbf{m}}^* \\ N \end{bmatrix} \right) \cdot \begin{bmatrix} \mathbf{L}_k \\ \mathbf{1}'_k \end{bmatrix}, \quad (65)$$

the $k \times k$ matrix $\mathbf{G}_2$ equals

$$\mathbf{G}_2 = \mathbf{D} [\exp(\mathbf{A}_2 \cdot \mathbf{g}_1)] \cdot \mathbf{A}_2 \cdot \mathbf{G}_1 = \mathbf{D}(\mathbf{P}^*) \cdot [\mathbf{I}_k - \mathbf{1}_k] \cdot \mathbf{D}^{-1} \left(\begin{bmatrix} \hat{\mathbf{m}}^* \\ N \end{bmatrix} \right) \cdot \begin{bmatrix} \mathbf{L}_k \\ \mathbf{1}'_k \end{bmatrix}, \quad (66)$$

and the $k \times k$ matrix $\mathbf{G}_3$ equals

$$\begin{aligned} \mathbf{G} &= \mathbf{G}_3 = \mathbf{A}_3 \cdot \mathbf{G}_2 \\ &= 50 \left(\mathbf{I}_k + \begin{bmatrix} \mathbf{0}'_{k-1} & 0 \\ \mathbf{I}_{k-1} & \mathbf{0}_{k-1} \end{bmatrix} \right) \cdot \mathbf{D}([\mathbf{P}^*]) \cdot [\mathbf{I}_k \ -\mathbf{1}_k] \cdot \mathbf{D}^{-1} \left(\begin{bmatrix} \hat{\mathbf{m}}^* \\ N \end{bmatrix} \right) \cdot \begin{bmatrix} \mathbf{L}_k \\ \mathbf{1}'_k \end{bmatrix}. \end{aligned} \quad (67)$$

Some tedious algebra shows that the elements of $\mathbf{G}$ equal

$$G_{gi} = \frac{50}{N} \cdot \begin{cases} 2 - P_{g-1}^* - P_g^* & \text{if } g > i \\ 1 - P_{g-1}^* - P_g^* & \text{if } g = i \\ 0 - P_{g-1}^* - P_g^* & \text{if } g < i \end{cases}, \quad (68)$$

where $P_x^* = P(X \leq x)$. Because the percentile ranks are homogeneous of order 0, the asymptotic variance is approximated by $V_{\mathbf{PR}} \approx \mathbf{GD}(\hat{\mathbf{m}})\mathbf{G}'$ (cf. Eq. 3). The sample estimates of the elements of the asymptotic covariance matrix of $\mathbf{PR}$ are

$$V_{\mathbf{PR}_{gh}} \approx \frac{2500}{N^2} \sum_i \hat{m}_i (\gamma_{gi} - P_{g-1}^* - P_g^*) (\gamma_{hi} - P_{h-1}^* - P_h^*) \quad (69)$$

for $g, h = 1, \dots, k$.

Appendix 3

The sample estimates of the eight boundaries of the stanines (Eq. 28) can be written using the generalized exp-log notation (Eq. 5) as follows:

$$\mathbf{St}_b = \mathbf{A}_5 \cdot \exp(\mathbf{A}_4 \cdot \log(\mathbf{A}_3 \cdot \exp(\mathbf{A}_2 \cdot \log(\mathbf{A}_1 \cdot \hat{\mathbf{m}})))) \quad (70)$$

Let $\mathbf{A}_1$ be the $4 \times k$ matrix,

$$\mathbf{A}_1 = [\mathbf{r} \ \mathbf{r}^{(2)} \ \mathbf{1}_k \ \mathbf{1}_k]', \quad (71)$$

let $\mathbf{A}_2$ be the 5×4 matrix

$$\mathbf{A}_2 = \begin{bmatrix} 2 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \quad (72)$$

let $\mathbf{A}_3$ be the 3×5 matrix

$$\mathbf{A}_3 = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad (73)$$

let $\mathbf{A}_4$ be the 2×3 matrix,

$$\mathbf{A}_4 = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (74)$$

and let $\mathbf{A}_5$ be the 8×2 matrix

$$\mathbf{A}_5 = [\mathbf{f} \ \mathbf{1}_8]. \quad (75)$$

It follows that $\mathbf{g}_0 = \hat{\mathbf{m}}$, $\mathbf{g}_1$ is the 4×1 vector

$$\mathbf{g}_1 = \log(\mathbf{A}_1 \cdot \mathbf{g}_0) = \log\left(\left[\mathbf{r} \ \mathbf{r}^{(2)} \ \mathbf{1}_k \ \mathbf{1}_k\right]' \cdot \hat{\mathbf{m}}\right) = \log\left(\begin{bmatrix} \sum X_i \\ \sum X_i^2 \\ N \\ N \end{bmatrix}\right), \quad (76)$$

$\mathbf{g}_2$ is the 5×1 vector

$$\mathbf{g}_2 = \exp(\mathbf{A}_2 \cdot \mathbf{g}_1) = \exp\left(\begin{bmatrix} 2 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \cdot \log\left(\begin{bmatrix} \sum X_i \\ \sum X_i^2 \\ N \\ N \end{bmatrix}\right)\right) = \begin{bmatrix} \frac{(\sum X_i)^2}{\sum X_i^2} \\ \frac{\sum X_i^2}{N} \\ \bar{X} \\ 1 \end{bmatrix}, \quad (77)$$

$\mathbf{g}_3$ is the 3×1 vector

$$\begin{aligned} \mathbf{g}_3 &= \log(\mathbf{A}_3 \cdot \mathbf{g}_2) = \log\left(\begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{(\sum X_i)^2}{\sum X_i^2} \\ \frac{\sum X_i^2}{N} \\ \bar{X} \\ 1 \end{bmatrix}\right), \\ &= \log\left(\begin{bmatrix} SS \\ N - 1 \\ \bar{X} \end{bmatrix}\right), \end{aligned} \quad (78)$$

$\mathbf{g}_4$ is the 2×1 vector

$$\mathbf{g}_4 = \exp(\mathbf{A}_4 \cdot \mathbf{g}_3) = \exp\left(\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \log\left(\begin{bmatrix} SS \\ N - 1 \\ \bar{X} \end{bmatrix}\right)\right) = \begin{bmatrix} s_X \\ \bar{X} \end{bmatrix}, \quad (79)$$

and $\mathbf{g}_5$ is the 8×1 vector

$$\mathbf{g}_5 = \mathbf{A}_5 \cdot \mathbf{g}_4 = [\mathbf{f} \ \mathbf{1}_8] \cdot \begin{bmatrix} s_X \\ \bar{X} \end{bmatrix} = \mathbf{g}(\hat{\mathbf{m}}) = \mathbf{St}_b. \quad (80)$$

Next, $\mathbf{G}_0 = \mathbf{I}_k$. $\mathbf{G}_1$ is the $4 \times k$ matrix,

$$\mathbf{G}_1 = \mathbf{D}^{-1} [\mathbf{A}_1 \cdot \mathbf{g}_0] \cdot \mathbf{A}_1 \cdot \mathbf{G}_0 = \mathbf{D}^{-1} \begin{bmatrix} \sum X_i \\ \sum X_i^2 \\ N \\ N \end{bmatrix} \cdot [\mathbf{r} \ \mathbf{r}^{(2)} \ \mathbf{1}_k \ \mathbf{1}_k]' = \begin{bmatrix} \frac{\mathbf{r}'}{\sum X_i} \\ \frac{\mathbf{r}^{(2)'}}{\sum X_i^2} \\ \frac{\mathbf{1}'_k}{N} \\ \frac{\mathbf{1}'_k}{N} \end{bmatrix}. \quad (81)$$

Then $\mathbf{G}_2$ is the $5 \times k$ matrix

$$\begin{aligned} \mathbf{G}_2 &= \mathbf{D} [\exp(\mathbf{A}_2 \cdot \mathbf{g}_1)] \cdot \mathbf{A}_2 \cdot \mathbf{G}_1 \\ &= \mathbf{D} \left(\begin{bmatrix} \frac{(\sum X_i)^2}{N} \\ \sum X_i^2 \\ N \\ \bar{X} \\ 1 \end{bmatrix} \right) \begin{bmatrix} 2 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{\mathbf{r}'}{\sum X_i} \\ \frac{\mathbf{r}^{(2)'}}{\sum X_i^2} \\ \frac{\mathbf{1}'_k}{N} \\ \frac{\mathbf{1}'_k}{N} \end{bmatrix} = \begin{bmatrix} 2\mathbf{r}'\bar{X} - \bar{X}^2 \\ \mathbf{r}^{(2)'} \\ \mathbf{1}'_k \\ \frac{\mathbf{d}'_k}{N} \\ \mathbf{0}'_k \end{bmatrix}, \end{aligned} \quad (82)$$

$\mathbf{G}_3$ is the $3 \times k$ matrix

$$\begin{aligned} \mathbf{G}_3 &= \mathbf{D}^{-1} [\mathbf{A}_3 \cdot \mathbf{g}_2] \cdot \mathbf{A}_3 \cdot \mathbf{G}_2 \\ &= \mathbf{D}^{-1} \left(\begin{bmatrix} SS \\ N - 1 \\ \bar{X} \end{bmatrix} \right) \cdot \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2\mathbf{r}'\bar{X} - \bar{X}^2 \\ \mathbf{r}^{(2)'} \\ \mathbf{1}'_k \\ \frac{\mathbf{d}'_k}{N} \\ \mathbf{0}'_k \end{bmatrix} = \begin{bmatrix} \mathbf{d}^{*'} \\ \mathbf{e}' \\ \frac{\mathbf{d}'_k}{\sum X_i} \end{bmatrix}, \end{aligned} \quad (83)$$

where $\mathbf{d}^* = \frac{\mathbf{d}^{(2)}}{SS}$. Then $\mathbf{G}_4$ is the $2 \times k$ matrix

$$\begin{aligned} \mathbf{G}_4 &= \mathbf{D} [\exp(\mathbf{A}_4 \cdot \mathbf{g}_3)] \cdot \mathbf{A}_4 \cdot \mathbf{G}_3 \\ &= \begin{bmatrix} s_X & 0 \\ 0 & \bar{X} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{d}^{*'} \\ \mathbf{e}' \\ \frac{\mathbf{d}'_k}{\sum X_i} \end{bmatrix} = \begin{bmatrix} \frac{s_X}{2} \cdot (\mathbf{d}^{*'} - \mathbf{e}') \\ \frac{\mathbf{d}'_k}{N} \end{bmatrix}, \end{aligned} \quad (84)$$

and $\mathbf{G}_5$ is the $8 \times k$ matrix

$$\mathbf{G}_5 = \mathbf{G} = \mathbf{A}_5 \cdot \mathbf{G}_4 = [\mathbf{f} \ \mathbf{1}_8] \begin{bmatrix} \frac{s_X}{2} \cdot (\mathbf{d}^{*'} - \mathbf{e}') \\ \frac{\mathbf{d}'_k}{N} \end{bmatrix} = \frac{\mathbf{f} \cdot s_X}{2} \cdot (\mathbf{d}^{*'} - \mathbf{e}') + \frac{\mathbf{1}_8 \cdot \mathbf{d}'}{N}. \quad (85)$$

The stanine boundaries are not homogeneous of order 0, because $\mathbf{St}_b$ is obtained using s_X . As a result, the asymptotic variance is approximated by $\mathbf{V}_{\mathbf{St}} \approx \mathbf{G} \mathbf{V}_{\hat{\mathbf{m}}} \mathbf{G}'$ (cf. Eq. 2). Some tedious algebra shows that the sample estimates of the elements of the asymptotic covariance matrix of $\mathbf{V}_{\mathbf{St}}$ are

$$V_{St_{gh}} \approx \sum_{j=1}^k \sum_{i=1}^k t \cdot \left[\frac{f_g \cdot s_X}{2} \cdot (d_i^* - e) + \frac{d_i}{N} \right] \cdot \left[\frac{f_h \cdot s_X}{2} \cdot (d_j^* - e) + \frac{d_j}{N} \right], \quad (86)$$

where $t = \delta_{ij} \hat{m}_j - \frac{\hat{m}_i \hat{m}_j}{N}$. For large N , Eq. 86 reduces to

$$V_{St_{gh}} \approx \sum_{i=1}^k \hat{m}_i \cdot \left[\frac{f_g \cdot s_X}{2} \cdot \left(d_i^* - \frac{1}{N} \right) + \frac{d_i}{N} \right] \cdot \left[\frac{f_h \cdot s_X}{2} \cdot \left(d_i^* - \frac{1}{N} \right) + \frac{d_i}{N} \right]. \quad (87)$$

Appendix 4

The sample estimate of the k standardized scores corresponding to $\mathbf{r}$, collected in a $k \times 1$ vector $\mathbf{z}$ (Eq. 31), can be written as

$$\mathbf{z} = \mathbf{A}_7 \cdot \exp \left(\mathbf{A}_6 \cdot \log \left(\mathbf{A}_5 \cdot \exp \left(\mathbf{A}_4 \cdot \log \left(\mathbf{A}_3 \cdot \exp \left(\mathbf{A}_2 \cdot \log \left(\mathbf{A}_1 \cdot \hat{\mathbf{m}} \right) \right) \right) \right) \right) \right). \quad (88)$$

Let $\mathbf{A}_1$ be the $(k+2) \times k$ matrix

$$\mathbf{A}_1 = [\mathbf{I}_k \ \mathbf{1}_k \ \mathbf{1}_k]'. \quad (89)$$

Let $\oplus$ indicate the direct sum, for example $\mathbf{X} \oplus \mathbf{Y} = \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{Y} \end{bmatrix}$. Then $\mathbf{A}_2$ is the $(k+1) \times (k+2)$ matrix

$$\mathbf{A}_2 = \mathbf{I}_k \oplus [1 \ -1], \quad (90)$$

$\mathbf{A}_3$ is the $(k+4) \times (k+1)$ matrix

$$\mathbf{A}_3 = [\mathbf{r} \ \mathbf{r}^{(2)} \ \mathbf{1}_k \ \mathbf{1}_k]' \oplus \mathbf{r}, \quad (91)$$

$\mathbf{A}_4$ is the $(5+k) \times (4+k)$ matrix

$$\mathbf{A}_4 = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 2 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \oplus \mathbf{I}_k, \quad (92)$$

$\mathbf{A}_5$ is the $(k+3) \times (k+5)$ matrix

$$\mathbf{A}_5 = \mathbf{1} \oplus [-1 \ 1] \oplus [1 \ -1] \oplus \mathbf{I}_k, \quad (93)$$

$\mathbf{A}_6$ is the $(1+k) \times (3+k)$ matrix

$$\mathbf{A}_6 = \begin{bmatrix} 1 & -\frac{1}{2} & \frac{1}{2} & \mathbf{0}'_k \\ \mathbf{0}_k & -\frac{1}{2} \cdot \mathbf{1}_k & \frac{1}{2} \cdot \mathbf{1}_k & \mathbf{I}_k \end{bmatrix}, \quad (94)$$

and $\mathbf{A}_7$ is the $k \times (k + 1)$ matrix

$$\mathbf{A}_7 = \begin{bmatrix} -\mathbf{1}_k & \mathbf{I}_k \end{bmatrix}. \quad (95)$$

It follows that $\mathbf{g}_0 = \hat{\mathbf{m}}$, the $(k + 2) \times 1$ vector $\mathbf{g}_1$ equals

$$\mathbf{g}_1 = \log(\mathbf{A}_1 \cdot \hat{\mathbf{m}}) = \log\left(\begin{bmatrix} \mathbf{I}_k & \mathbf{1}_k & \mathbf{1}_k \end{bmatrix}' \cdot \hat{\mathbf{m}}\right) = \log\left(\begin{bmatrix} \hat{\mathbf{m}}' & N & N \end{bmatrix}'\right), \quad (96)$$

the $(k + 1) \times 1$ vector $\mathbf{g}_2$ equals

$$\mathbf{g}_2 = \exp(\mathbf{A}_2 \cdot \mathbf{g}_1) = \exp\left(\mathbf{I}_k \oplus \begin{bmatrix} 1 & -1 \end{bmatrix} \cdot \log\left(\begin{bmatrix} \hat{\mathbf{m}}' & N & N \end{bmatrix}'\right)\right) = \begin{bmatrix} \hat{\mathbf{m}}' & 1 \end{bmatrix}', \quad (97)$$

the $(k + 4) \times 1$ vector $\mathbf{g}_3$ equals

$$\mathbf{g}_3 = \log(\mathbf{A}_3 \cdot \mathbf{g}_2) = \log\left(\left(\begin{bmatrix} \mathbf{r} & \mathbf{r}^{(2)} & \mathbf{1}_k & \mathbf{1}_k \end{bmatrix}' \oplus \mathbf{r}\right) \cdot \begin{bmatrix} \hat{\mathbf{m}}' & 1 \end{bmatrix}'\right) = \log \begin{bmatrix} \sum X_i \\ \sum X_i^2 \\ N \\ N \\ \mathbf{r} \end{bmatrix}, \quad (98)$$

the $(k + 5) \times 1$ vector $\mathbf{g}_4$ equals

$$\begin{aligned} \mathbf{g}_4 &= \exp(\mathbf{A}_4 \cdot \mathbf{g}_3) \\ &= \exp\left(\left(\left(\begin{bmatrix} 1 & 0 & 0 & -1 \\ 2 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \oplus \mathbf{I}_k\right) \cdot \log \begin{bmatrix} \sum X_i \\ \sum X_i^2 \\ N \\ N \\ \mathbf{r} \end{bmatrix}\right)\right) = \begin{bmatrix} \bar{X} \\ \frac{(\sum X_i)^2}{\sum X_i^2} \\ N \\ N \\ 1 \\ \mathbf{r} \end{bmatrix}, \end{aligned} \quad (99)$$

the $(k + 3) \times 1$ vector $\mathbf{g}_5$ equals

$$\begin{aligned} \mathbf{g}_5 &= \log(\mathbf{A}_5 \cdot \mathbf{g}_4) \\ &= \log\left(\begin{bmatrix} 1 \oplus \begin{bmatrix} -1 & 1 \end{bmatrix} \oplus \begin{bmatrix} 1 & -1 \end{bmatrix} \oplus \mathbf{I}_k \end{bmatrix} \cdot \begin{bmatrix} \bar{X} \\ \frac{(\sum X_i)^2}{\sum X_i^2} \\ N \\ N \\ 1 \\ \mathbf{r} \end{bmatrix}\right) = \log \begin{bmatrix} \bar{X} \\ SS \\ N - 1 \\ \mathbf{r} \end{bmatrix}, \end{aligned} \quad (100)$$

the $(k + 1) \times 1$ vector $\mathbf{g}_6$ equals

$$\begin{aligned} \mathbf{g}_6 &= \exp(\mathbf{A}_6 \cdot \mathbf{g}_5) \\ &= \exp\left(\begin{bmatrix} 1 & -\frac{1}{2} & \frac{1}{2} & \mathbf{0}_k' \\ \mathbf{0}_k & -\frac{1}{2} \cdot \mathbf{1}_k & \frac{1}{2} \cdot \mathbf{1}_k & \mathbf{I}_k \end{bmatrix} \cdot \log \begin{bmatrix} \bar{X} \\ SS \\ N - 1 \\ \mathbf{r} \end{bmatrix}\right) = \begin{bmatrix} \bar{X} \\ \frac{s_X}{s_X} \\ \frac{\mathbf{r}}{s_X} \end{bmatrix}, \end{aligned} \quad (101)$$

and the $k \times 1$ vector $\mathbf{g}_7$ equals

$$\mathbf{g}_7 = \mathbf{g}(\hat{\mathbf{m}}) = \mathbf{A}_7 \cdot \mathbf{g}_6 = \begin{bmatrix} -\mathbf{1}_k & \mathbf{I}_k \end{bmatrix} \cdot \begin{bmatrix} \frac{\bar{X}}{s_X} \\ \frac{s_X}{\mathbf{r}} \\ \frac{\mathbf{r}}{s_X} \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ s_X \end{bmatrix} = \mathbf{z}. \quad (102)$$

Next, $\mathbf{G}_0 = \mathbf{I}_k$, the $(k+2) \times k$ matrix $\mathbf{G}_1$ equals

$$\begin{aligned} \mathbf{G}_1 &= \mathbf{D}^{-1} [\mathbf{A}_1 \cdot \mathbf{g}_0] \cdot \mathbf{A}_1 \cdot \mathbf{G}_0 \\ &= \mathbf{D}^{-1} \left([\hat{\mathbf{m}}' \ N \ N]' \right) \cdot [\mathbf{I}_k \ \mathbf{1}_k \ \mathbf{1}_k]' = \begin{bmatrix} \mathbf{D}^{-1}(\hat{\mathbf{m}}) \\ \mathbf{1}'_k/N \\ \mathbf{1}'_k/N \end{bmatrix}, \end{aligned} \quad (103)$$

the $(k+1) \times k$ matrix $\mathbf{G}_2$ equals

$$\begin{aligned} \mathbf{G}_2 &= \mathbf{D} [\exp(\mathbf{A}_2 \cdot \mathbf{g}_1)] \cdot \mathbf{A}_2 \cdot \mathbf{G}_1 \\ &= \mathbf{D} \left([\hat{\mathbf{m}} \ 1]' \right) \cdot [\mathbf{I}_k \oplus [1 \ -1]] \cdot \begin{bmatrix} \mathbf{D}^{-1}(\hat{\mathbf{m}}) \\ \mathbf{1}'_k/N \\ \mathbf{1}'_k/N \end{bmatrix} = \begin{bmatrix} \mathbf{I}_k \\ \mathbf{0}'_k \end{bmatrix}, \end{aligned} \quad (104)$$

the $(k+4) \times k$ matrix $\mathbf{G}_3$ equals

$$\mathbf{G}_3 = \mathbf{D}^{-1} \left(\begin{bmatrix} \sum X_i \\ \sum X_i^2 \\ N \\ N \\ \mathbf{r} \end{bmatrix} \right) \cdot \left[[\mathbf{r} \ \mathbf{r}^{(2)} \ \mathbf{1}_k \ \mathbf{1}_k]' \oplus \mathbf{r} \right] \cdot \begin{bmatrix} \mathbf{I}_k \\ \mathbf{0}'_k \end{bmatrix} = \begin{bmatrix} \mathbf{r}' / \sum X_i \\ \mathbf{r}^{(2)'} / \sum X_i^2 \\ \mathbf{1}'_k / N \\ \mathbf{1}'_k / N \\ \mathbf{0}_{k \times k} \end{bmatrix}, \quad (105)$$

the $(k+5) \times k$ matrix $\mathbf{G}_4$ equals

$$\begin{aligned} \mathbf{G}_4 &= \mathbf{D} [\exp(\mathbf{A}_4 \cdot \mathbf{g}_3)] \cdot \mathbf{A}_4 \cdot \mathbf{G}_3 \\ &= \mathbf{D} \left(\begin{bmatrix} \bar{X} \\ \frac{(\sum X_i)^2}{N} \\ \sum X_i^2 \\ N \\ 1 \\ \mathbf{r} \end{bmatrix} \right) \cdot \left[\begin{bmatrix} 1 & 0 & 0 & -1 \\ 2 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \oplus \mathbf{I}_k \right] \cdot \begin{bmatrix} \mathbf{r}' / \sum X_i \\ \mathbf{r}^{(2)'} / \sum X_i^2 \\ \mathbf{1}'_k / N \\ \mathbf{1}'_k / N \\ \mathbf{0}_{k \times k} \end{bmatrix} = \begin{bmatrix} \mathbf{d}' \\ \mathbf{r}' 2\bar{X} - \bar{X}^2 \\ \mathbf{r}^{(2)'} \\ \mathbf{1}'_k \\ \mathbf{0}_{(k+1) \times k} \end{bmatrix}, \end{aligned} \quad (106)$$

the $(k + 3) \times k$ matrix $\mathbf{G}_5$ equals

$$\begin{aligned}\mathbf{G}_5 &= \mathbf{D}^{-1} [\mathbf{A}_5 \cdot \mathbf{g}_4] \cdot \mathbf{A}_5 \cdot \mathbf{G}_4 \\ &= \mathbf{D}^{-1} \begin{bmatrix} \bar{X} \\ SS \\ N-1 \\ \mathbf{r} \end{bmatrix} \cdot [1 \oplus [-1 \ 1] \oplus [1 \ -1] \oplus \mathbf{I}_k] \cdot \begin{bmatrix} \mathbf{d}' \\ \mathbf{r}' 2\bar{X} - \bar{X}^2 \\ \mathbf{r}^{(2)'} \\ \mathbf{1}'_k \\ \mathbf{0}_{(k+1) \times k} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\mathbf{r}'}{\sum \bar{X}_i} - \frac{1}{N} \\ \mathbf{d}^{*'} \\ \mathbf{e}' \\ \mathbf{0}_{k \times k} \end{bmatrix},\end{aligned}\quad (107)$$

and the $(k + 1) \times k$ matrix $\mathbf{G}_6$ equals

$$\begin{aligned}\mathbf{G}_6 &= \mathbf{D} [\exp(\mathbf{A}_6 \cdot \mathbf{g}_5)] \cdot \mathbf{A}_6 \cdot \mathbf{G}_5 \\ &= \mathbf{D} \left(\begin{bmatrix} \bar{X} \\ s_X \\ \frac{\mathbf{r}'}{s_X} \end{bmatrix} \right) \cdot \begin{bmatrix} 1 & -\frac{1}{2} & \frac{1}{2} & \mathbf{0}'_k \\ \mathbf{0}_k & -\frac{1}{2} \cdot \mathbf{1}_k & \frac{1}{2} \cdot \mathbf{1}_k & \mathbf{I}_k \end{bmatrix} \cdot \begin{bmatrix} \frac{\mathbf{r}'}{\sum \bar{X}_i} - \frac{1}{N} \\ \mathbf{d}^{*'} \\ \mathbf{e}' \\ \mathbf{0}_{k \times k} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\bar{X}}{s_X} \cdot \left[\frac{\mathbf{r}'}{\sum \bar{X}} - \frac{\mathbf{1}'_k}{N} - 0.5 (\mathbf{d}^{*'} - \mathbf{e}') \right] \\ \frac{-0.5\mathbf{r}}{s_X} \cdot (\mathbf{d}^{*'} - \mathbf{e}') \end{bmatrix}.\end{aligned}\quad (108)$$

Finally, the $k \times k$ matrix $\mathbf{G}_7$ equals

$$\begin{aligned}\mathbf{G} &= \mathbf{A}_7 \cdot \mathbf{G}_6 = [-\mathbf{1}_k \ \mathbf{I}_k] \cdot \begin{bmatrix} \frac{\bar{X}}{s_X} \cdot \left[\frac{\mathbf{r}'}{\sum \bar{X}} - \frac{\mathbf{1}'_k}{N} - 0.5 (\mathbf{d}^{*'} - \mathbf{e}') \right] \\ \frac{-0.5\mathbf{r}}{s_X} \cdot (\mathbf{d}^{*'} - \mathbf{e}') \end{bmatrix} \\ &= \frac{\mathbf{1}_k}{s_X} \cdot \left\{ -\bar{X} \left[\frac{\mathbf{r}'}{\sum \bar{X}} - \frac{\mathbf{1}'_k}{N} - 0.5 (\mathbf{d}^{*'} - \mathbf{e}') \right] - 0.5\mathbf{r} (\mathbf{d}^{*'} - \mathbf{e}') \right\},\end{aligned}\quad (109)$$

with elements

$$G_{ij} = \frac{1}{s_X} \cdot \left\{ -\bar{X} \left[\frac{r_j}{\sum \bar{X}} - \frac{1}{N} - 0.5 (d_j^* - e) \right] - 0.5r_i (d_j^* - e) \right\}.\quad (110)$$

The Z-scores are not homogeneous of order 0, because $\mathbf{z}$ is obtained using s_X . As a result, the asymptotic variance is approximated by $\mathbf{V}_z \approx \mathbf{G}\mathbf{V}_{\hat{\mathbf{m}}}\mathbf{G}'$ (cf. Eq. 2). Some tedious algebra shows that the sample estimate of the asymptotic covariance matrix of $\mathbf{V}_z$ is

$$V_{Z_{gh}} \approx \sum_{j=1}^k \sum_{i=1}^k t \left[-\bar{X} \cdot \left(\frac{r_i}{\sum \bar{X}} - \frac{1}{N} - u_i \right) - r_g u_i \right] \cdot \left[-\bar{X} \left(\frac{r_j}{\sum \bar{X}} - \frac{1}{N} - u_j \right) - r_h u_j \right],\quad (111)$$

where $u_i = 0.5 [d_i^* - e]$. For large N ,

$$V_{Z_{gh}} \approx \sum_i \frac{\hat{m}_i}{s_X^2} \left[-\bar{X} \left(\frac{r_i}{\sum X} - \frac{1}{N} - u_i \right) - r_g u_i \right] \cdot \left[-\bar{X} \left(\frac{r_j}{\sum X} - \frac{1}{N} - u_j \right) - r_h u_j \right]. \quad (112)$$

References

- Aardoom, J. J., Dingemans, A. E., Landt, M. S. C., & Van Furth, E. F. (2012). Norms and discriminative validity of the Eating Disorder Examination Questionnaire (EDE-Q). *Eating Behaviors*, 13, 305–309. doi:10.1016/j.eatbeh.2012.09.002.
- AERA, Apa, & NCME., (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Agresti, A. (2012). *Analysis of ordinal categorical data* (2nd ed.). Hoboken, NJ: Wiley.
- Agresti, A. (2013). *Categorical data analysis* (3rd ed.). Hoboken, NJ: Wiley.
- Agresti, A., & Min, Y. (2001). On small-sample confidence intervals for parameters in discrete distributions. *Biometrics*, 57(963), 971. doi:10.1111/j.0006-341X.2001.00963.x.
- Ahn, S., & Fessler, A. (2003). Standard errors of mean, variance, and standard deviation estimators. Technical Report. Ann Arbor, MI: EECS Department, University of Michigan: July 2003. <http://www.eecs.umich.edu/~fessler/papers/files/tr/stderr.pdf>.
- American Psychological Association. (2010). *Publication Manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- Bergsma, W. P. (1997). *Marginal models for categorical data*. Tilburg: Tilburg University Press.
- Bergsma, W. P., Croon, M. A., & Hagenars, J. A. (2009). *Marginal models for dependent, clustered and longitudinal categorical data*. New York, NY: Springer.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores* (pp. 453–479). Reading, MA: Addison-Wesley.
- Brennan, R. L., & Lee, W.-C. (1999). Conditional scale-score standard errors of measurement under binomial and compound binomial assumptions. *Educational and Psychological Measurement*, 56, 5–24. doi:10.1177/0013164499591001.
- Cavaco, S., Gonçalves, A., Pinto, C., Almeida, E., Gomes, F., Moreira, I., et al. (2013). Trail making test: Regression-based norms for the Portuguese population. *Archives of Clinical Neuropsychology*, 28, 189–198. doi:10.1093/arclin/acs115.
- Cooch, E., & White, G. (2015). *Program MARK: A gentle introduction* (14th ed.). Fort Collins, CO: Colorado State University.
- Crawford, J., Cayley, C., Lovibond, P. F., Wilson, P. H., & Hartley, C. (2011). Percentile norms and accompanying interval estimates from an Australian general adult population sample for self-report mood scales (BAI, BDI, CRS-D, CES-D, DASS, DASS-21, STAI-X, STAI-Y, SRDS, and SRAS). *Australian Psychologist*, 46, 3–14. doi:10.1111/j.1742-9544.2010.00003.x.
- Crawford, J. R., Garthwaite, P. H., & Slick, D. J. (2009). On percentile norms in neuropsychology: Proposed reporting standards and methods for quantifying the uncertainty over the percentile ranks of test scores. *The Clinical Neuropsychologist*, 23, 1173–1195.
- Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, 12, 482–486. doi:10.1076/clin.12.4.482.7241.
- Evers, A., Lucassen, W., Meijer, R. R., & Sijtsma, K. (2009). *COTAN assessment system for the quality of tests*. Amsterdam: Nederlands Instituut van Psychologen.
- Glaesmer, H., Rief, W., Martin, A., Mewes, R., Brähler, E., Zenger, M., & Hinz, A. (2012). Psychometric properties and population-based norms of the Life Orientation Test Revised (LOT-R). *British Journal of Health Psychology*, 17, 432–445. doi:10.1111/j.2044-8287.2011.02046.x.
- Goretti, B., Niccolai, C., Hakiki, B., Sturchio, A., Falautano, M., Eleonora, M., et al. (2014). The Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS): Normative values with gender, age and education corrections in the Italian population. *BMC Neurology*, 14, 171–176. doi:10.1186/s12883-014-0171-6.
- Grande, G., Rompel, M., Glaesmer, H., Petrowski, K., & Herrmann-Lingen, C. (2010). The type-D scale (DS14): Norms and prevalence of type-D personality in a population-based representative sample in Germany. *Personality and Individual Differences*, 48, 935–939. doi:10.1016/j.paid.2010.02.026.
- Grizzle, J. E., Starmer, C. F., & Koch, G. G. (1969). Analysis of categorical data for linear models. *Biometrics*, 25, 489–504. doi:10.2307/2528901.
- Kendall, M., & Stuart, A. (1977). *The advanced theory of statistics, distributional theory* (4th ed., Vol. 1). New York, NY: Macmillan.
- Kessels, R. P., Montagne, B., Hendriks, A. W., Perrett, D. I., & De Haan, E. H. (2014). Assessment of perception of morphed facial expression using the Emotion Recognition Task: Normative data from healthy participants aged 8–75. *Journal of Neuropsychology*, 8, 75–93. doi:10.1111/jnp.12009.
- Kritzer, H. M. (1977). Analyzing measures of association derived from contingency tables. *Sociological Methods and Research*, 5, 35–50. doi:10.1177/004912417700500401.

- Kuijpers, R. E., Van der Ark, L. A., & Croon, M. A. (2013a). Standard errors and confidence intervals for scalability coefficients in Mokken scale analysis using marginal models. *Sociological Methodology*, 43, 42–69. doi:[10.1177/0081175013481958](https://doi.org/10.1177/0081175013481958).
- Kuijpers, R. E., Van der Ark, L. A., & Croon, M. A. (2013b). Testing hypotheses involving Cronbach's alpha using marginal models. *British Journal of Mathematical and Statistical Psychology*, 66, 503–520. doi:[10.1111/bmsp.12010](https://doi.org/10.1111/bmsp.12010).
- Lang, J. B. (2008). Score and profile likelihood confidence intervals for contingency table parameters. *Statistics in Medicine*, 27, 5975–5990. doi:[10.1002/sim.3391](https://doi.org/10.1002/sim.3391).
- Larson, R., & Edwards, B. (2013). *Calculus* (10th ed.). Boston, MA: Cengage Learning, Brooks/Cole.
- Lee, W.-C., Brennan, R. L., & Kolen, M. J. (2000). Estimators of conditional scale-score standard errors of measurement: A simulation study. *Journal of Educational Measurement*, 37, 1–20. doi:[10.1111/j.1745-3984.2000.tb01073.x](https://doi.org/10.1111/j.1745-3984.2000.tb01073.x).
- Lehtonen, R., & Pahkinen, E. (2004). *Practical methods for design and analysis of complex surveys* (2nd ed.). West Sussex: Wiley.
- Merrell, K. W. (1994). *Preschool and Kindergarten Behavior Scales. Test manual*. Brandon, VT: Clinical Psychology Publishing Company.
- Mertler, C. A. (2007). *Interpreting standardized test scores: Strategies for data-driven instructional decision making*. Thousand Oaks, CA: Sage.
- Mond, J. M., Hay, P. J., Rodgers, B., & Owen, C. (2006). Eating Disorder Examination Questionnaire (EDE-Q): Norms for young adult women. *Behaviour Research and Therapy*, 44, 53–62. doi:[10.1016/j.brat.2004.12.003](https://doi.org/10.1016/j.brat.2004.12.003).
- Oosterhuis, H. E. M., Van der Ark, L. A., & Sijtsma, K. (2016). Sample size requirements for traditional and regression-based norms. *Assessment*, 23, 191–202. doi:[10.1177/1073191115580638](https://doi.org/10.1177/1073191115580638).
- Palomo, R., Casals-Coll, M., Sánchez-Benavides, G., Quintana, M., Manero, R. M., Rognoni, T., et al. (2011). Spanish normative studies in young adults (NEURONORMA young adults project): Norms for the Rey-Osterrieth Complex Figure (copy and memory) and Free and Cued Selective Reminding Test. *Neurología*, 28, 226–235. doi:[10.1016/j.nrl.2012.03.008](https://doi.org/10.1016/j.nrl.2012.03.008).
- R Core Team (2015). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>.
- Rao, R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York, NY: Wiley.
- Sartorio, F., Bravini, E., Vercelli, S., Ferriero, G., Plebani, G., Foti, C., & Franchignoni, F. (2013). The functional dexterity test: Test-retest reliability analysis and up-to-date reference norms. *Journal of Hand Therapy*, 26, 62–68. doi:[10.1016/j.jht.2012.08.001](https://doi.org/10.1016/j.jht.2012.08.001).
- Shi, J., Wei, M., Tian, J., Snowden, J., Zhang, X., Li, T., et al. (2014). The Chinese version of story recall: A useful screening tool for mild cognitive impairment and Alzheimer's disease in the elderly. *BMC Psychiatry*, 14, 71–80. doi:[10.1186/1471-244X-14-71](https://doi.org/10.1186/1471-244X-14-71).
- Van Belle, G. (2003). *Statistical rules of thumb* (2nd ed.). Hoboken, NJ: Wiley.
- Van der Ark, L. A. (2012). New developments in Mokken Scale Analysis in R. *Journal of Statistical Software*, 48(5), 1–27. doi:[10.18637/jss.v048.i05](https://doi.org/10.18637/jss.v048.i05).
- Van der Ark, L. A., Croon, M. A., & Sijtsma, K. (2008). Mokken scale analysis for dichotomous items using marginal models. *Psychometrika*, 73, 183–208. doi:[10.1007/s11336-007-9034-z](https://doi.org/10.1007/s11336-007-9034-z).
- Van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York, NY: Springer.

Manuscript Received: 8 JUL 2015

Final Version Received: 4 AUG 2016

Published Online Date: 14 NOV 2016