Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Suggestion of confidence interval methods for the Cronbach alpha in application to complex survey data

by Jihnhee Yu, Ziqiang Chen, Kan Wang and Mine Tezal

Release date: December 17, 2019

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2019
Vol. 45, No. 3, pp. 465-484
Statistics Canada, Catalogue No. 12-001-X

Suggestion of confidence interval methods for the Cronbach alpha in application to complex survey data

Jihnhee Yu, Ziqiang Chen, Kan Wang and Mine Tezal[1]

Abstract

We discuss a relevant inference for the alpha coefficient (Cronbach, 1951) - a popular ratio-type statistic for the covariances and variances in survey sampling including complex survey sampling with unequal selection probabilities. This study can help investigators who wish to evaluate various psychological or social instruments used in large surveys. For the survey data, we investigate workable confidence intervals by using two approaches: (1) the linearization method using the influence function and (2) the coverage-corrected bootstrap method. The linearization method provides adequate coverage rates with correlated ordinal values that many instruments consist of; however, this method may not be as good with some non-normal underlying distributions, e.g., a multi-lognormal distribution. We suggest that the coverage-corrected bootstrap method can be used as a complement to the linearization method, because the coverage-corrected bootstrap method is computer-intensive. Using the developed methods, we provide the confidence intervals for the alpha coefficient to assess various mental health instruments (Kessler 10, Kessler 6 and Sheehan Disability Scale) for different demographics using data from the National Comorbidity Survey Replication (NCS-R).

Key Words: Clustered data; Complex survey; Coverage-correction method; Influence function; Linearization.

1 Introduction

In this paper, we propose methods to incorporate the survey designs in confidence intervals for the alpha coefficient (Cronbach, 1951) based on the large sample approximation (linearization) and the "double" bootstrap approach. These methods have not been investigated in the related literature, even though the alpha coefficient is widely used in psychology and other relevant research areas. For a practical application of these methods, we analyze mental health instruments data from the National Comorbidity Survey Replication (NCS-R), a survey conducted between 2001 and 2003 intended to measure the prevalence of mental disorders (Kessler, Berglund, Chiu, Demler, Heeringa, Hiripi, Jin, Pennell, Walters, Zaslavsky and Zheng, 2004). In the analysis, we show the feasibility of the confidence interval method for the alpha coefficient on a survey data set.

A great deal of psychological and sociological research uses assessment instruments (i.e., questionnaires) to obtain quantitative information for a population of interest. Ideally, the different items in one instrument measure the same concepts to achieve a high internal consistency. The alpha coefficient, also known as Cronbach's alpha (henceforth referred to as α) is a popular statistic (e.g., a quick search of PubMed with the keywords "Cronbach alpha" and "scale" from the years of 2012-2016 brings up more than 700 publications) that is widely used to measure the internal consistency reliability of various instruments.

Let x denote the p-variate column vector of the observations indicating p items from an instrument, and let Σ indicate the corresponding covariance matrix. The value α is defined as

1. Jihnhee Yu, Department of Biostatistics, University at Buffalo, State University of New York, NY 14214, U.S.A. E-mail: jinheeyu@buffalo.edu; Ziqiang Chen and Kan Wang, Department of Biostatistics, University at Buffalo, State University of New York, NY 14214, U.S.A.; Mine Tezal, Department of Oral Biology, University at Buffalo, State University of New York, NY 14214, U.S.A.

$$\alpha = p/(p-1)\left(1 - \operatorname{tr}\Sigma / \mathbf{1}^T \Sigma \mathbf{1}\right),$$

where $\mathbf{1}$ is the conforming column vector consisting of 1, and tr indicates the trace of a matrix. The value α shows the ratio between the sum of the covariances and the sum of variances and covariances, thus a high value for α suggests that the items are highly correlated within the instrument. The theoretical values of α range from 0 to 1, where a higher value is considered to be more desirable. The estimator of α (denoted by $\hat{\alpha}$) is defined as

$$\hat{\alpha} = p/(p-1)\left(1 - \operatorname{tr}\hat{\Sigma} / \mathbf{1}^T \hat{\Sigma} \mathbf{1}\right),$$

where $\hat{\Sigma}$ is a consistent estimator of Σ. The estimator $\hat{\alpha}$ can take any value less than or equal to 1, including negative values.

In the literature, many confidence interval strategies for α can be found (e.g., van Zyl, Neudecker and Nel, 2000; Yuan, Guarnaccia and Hayslip, 2003; Kistner and Muller, 2004; Bonett and Wright, 2015), but discussions regarding the applications for complex survey data where observations in the data can have unequal weights due to stratifications and multistage cluster sampling (Lohr, 1999) are largely lacking.

This paper is structured as follows: In Section 2, we propose strategies for obtaining the confidence intervals of α using the linearization method and the coverage-corrected bootstrap method. In Section 3, simulation results are presented based on scenarios of stratified multi-stage cluster sampling and unequal probability sampling scenarios. In Section 4, the developed methods are applied to analyze the NCS-R data sets, and the results comparing different demographics are reported. The Section 5 is devoted to the concluding remarks.

2 Design-based confidence intervals for α

In this section, we discuss two methods to obtain the confidence interval for α, the confidence interval based on the linearization method using the influence function (Deville, 1999; Demnati and Rao, 2004) and the coverage-corrected bootstrap method (Hall, Martin and Schucany, 1989). In this discussion, we consider strategies to deal with stratification, since stratification is a common feature in surveys and may decrease the magnitude of the variances for the statistics of interest (Lohr, 1999). We note that the sampling design for the NCS-R used stratification (more details in Section 4). Later, in Section 3, we show that the linearization will be sufficient for most practical cases (e.g., scales with ordinal responses); however, the coverage rate may not be satisfactory with some non-normal distributions. The coverage-corrected bootstrap method when applied to survey data is proposed as a possible alternative to the linearization method in those cases (Section 2.2).

2.1 Linearization

A symmetric confidence interval can be obtained based on the normal approximation of an estimator for a finite population (Hájek, 1981; Sen, 1995). The linearization method is applied for the variance estimation of complex statistics. In a survey sampling setting, we consider a population index set $U = \{1, \ldots, N\}$ with population size N. A random sample S of size n is selected from U by a sampling design $p(s) = \Pr\{S = s\}$ for all $s \subset U$. The value w_k denotes the sampling weight associated with the index $k \in s$. For probability sampling, the sampling weight for index k is the inverse of the first order inclusion probability, i.e., $w_k = [\Pr\{k \in s\}]^{-1}$. For each unit k of the population U, there is a point (or observation) x_k of $\mathbf{R}^p$, a p-dimensional real space. In a similar manner to Deville (1999), let us consider the population U that is represented by the measure M as having a mass of $1/N$ in each of the points x_k. In this way, we have $\int 1 dM = 1$ and $\int y dM = N^{-1} \sum_{k \in U} y_k$ for any vector value $y_k = y(x_k)$, where we define the integral of a vector as the integral of each component of the vector. The measure $\hat{M}$ is the estimator of M allocating a weight w_k / N to any point x_k, $k \in s$ and 0 to any other points. Following some conventional notation (e.g., Cochran, 1977), let $\int y dM = \bar{Y}$. Also let $\int y d\hat{M} = \hat{\bar{Y}}$. The influence function of a "functional" T is defined as

$$IT(M;\, x) = \lim_{t \to 0} \frac{T(M + t\delta_x) - T(M)}{t},$$

where δ_x denotes the added unit mass at point x (Deville, 1999), and the functional T (Krätschmer, Schied and Zähle, 2012) maps a measure to a set (e.g., the real line). The examples of the functional include $\bar{Y}$ and $\hat{\bar{Y}}$. Note that this classical definition of the influence function (Hampel, Ronchetti, Rousseeuw and Stahel, 1986; Davison and Hinkley, 1997) is slightly different from that of Deville (1999) where he defines a measure M to satisfy $\int y dM = \sum_{k \in U} y_k$. Let us define the linearized value $z_k = IT(M;\, x_k)$. Let $T(\hat{M})$ indicate the substitution estimator of $T(M)$ by replacing M by $\hat{M}$. Assume that the postulate of Deville (1999), i.e., $n^{-1/2} N^{-1} (\hat{X} - X)$ has a zero-mean multi-normal distribution as a limit, where X and $\hat{X}$ are the population total and the total estimator for general observation x_k, and N and n tend toward infinity. This fact leads to $\int x d\hat{M} - \int x dM = O_p(n^{-1/2})$. Assuming that T can be derived for any direction of an increase, a similar argument to Deville (1999) gives rise to the result

$$\frac{T(\hat{M}) - T(M)}{N^c} = \frac{1}{N^{c+1}} \sum_{k \in U} z_k (w_k - 1) + o(n^{-1/2}), \tag{2.1}$$

for some positive value c. Equation (2.1) results in the asymptotic variance of $T(\hat{M})$

$$\text{Avar}\{T(\hat{M})\} = \text{var}(\hat{\bar{Z}}). \tag{2.2}$$

If $T = \int x dM(x)$, then the influence function at x_k $(k \in U)$ is

$$IT(M;\, x_k) = \lim_{t \to 0} \frac{\sum_{i \in U} x_i / N + t x_k - \sum_{i \in U} x_i / N}{t} = x_k. \tag{2.3}$$

For a complex statistic as the functions of simple statistics, we have the influence function

$$I(f(T)) = D(f)\, IT, \tag{2.4}$$

where f is a differentiable function on the space of values for T and $D(f)$ is the matrix of the partial derivatives of f (Deville, 1999). In many cases, the linearized value z_k includes parameters to be estimated. Let $\hat{z}_k$ indicate the approximation of z_k using some statistics estimated by the sample. Deville (1999) notes that with a fixed and finite number of estimated parameters, the variance estimators based on $\hat{z}_k$ and z_k are equivalent by an asymptotically negligible quantity.

Now, we obtain the linearized value for α as follows. Consider a data set

$$\mathbf{X} = (x_1, \ldots, x_n)^T,$$

where x_k is a p-variate observation indicating p items in an instrument and n is the sample size. Let σ_{ij} and $\hat{\sigma}_{ij}$ $(i,\ j = 1, \ldots, p)$ denote the $(i,\ j)^{\text{th}}$ elements of Σ and $\hat{\Sigma}$ as defined in Section 1, respectively. Specifically, we define $\sigma_{ij} = \sum_{k=1}^{N} \left({}_i x_k - {}_i \bar{X}_k\right)\left({}_j x_k - {}_j \bar{X}_k\right) \big/ (N-1)$ (Lohr, 1999), where ${}_i x_k$ and ${}_i \bar{X}_k$ are i^{th} element of x_k and its population mean, respectively. For simple random sampling without replacement (SRSWOR), we define $\hat{\sigma}_{ij} = \sum_{k \in s} \left({}_i x_k - {}_i \hat{\bar{X}}_k\right)\left({}_j x_k - {}_j \hat{\bar{X}}_k\right) \big/ (n-1)$, where n is the size of the sample s and ${}_i \hat{\bar{X}}_k$ is the sample mean of ${}_i x_k$ (Lohr, 1999). For obtaining $\hat{\sigma}_{ij}$ for more complicated sampling methods including unequal probability sampling, we refer to Swain and Mishra (1994) and Patel and Bhatt (2016). In survey sampling, the sampling weights are used for correcting the disproportionality of the sample regarding the target population of interest (Pfeffermann, 1993). With the complex sampling designs often used in practice, failure to consider the sampling designs may provide biased inferences. For more of a discussion of the role of sampling weights, we refer to Pfeffermann (1993). For the variance estimation of survey data, the linearization method can be applied as in formula (2.2) incorporating the sampling weights. Following conventional notations of the vectorization of a matrix, let $\text{vech}(\mathbf{A})$ be the column vector of nonduplicated elements of the matrix $\mathbf{A}$, $\text{vec}(\mathbf{A})$ be the column vector composed of the columns of $\mathbf{A}$. Let $\hat{t}$ indicate the collection of statistics as the components of $\text{vech}^T(\hat{\Sigma})$ and t indicate the collection of corresponding parameters. Specifically, we let $t = \left(\text{vech}^T\left(\sum_{k \in U} x_k x_k^T / N\right), \left(\sum_{k \in U} x_k / N\right)^T\right)$. Also, let the matrix K_p indicate a transition matrix that satisfies the relationship $\text{vech}(\mathbf{A}) = K_p^T \text{vec}(\mathbf{A})$, which borrows the transition matrix expression from van Zyl et al. (2000). We propose a linearized value for $\text{Var}(\hat{\alpha})$ as

$$z_k = \frac{p}{p-1} \frac{1}{\left(\mathbf{1}^T \hat{\Sigma} \mathbf{1}\right)^2} \left\{\mathbf{1}^T \hat{\Sigma} \mathbf{1} \,\text{vec}^T\left(\mathbf{I}_p\right) - \text{tr}\left(\hat{\Sigma}\right)\left(2\text{vec}^T\left(\mathbf{1}\mathbf{1}^T\right) - \text{vec}^T\left(\mathbf{I}_p\right)\right)\right\} K_p \,\hat{\mathbf{J}}\, u_k,\ k = 1, \ldots, n, \quad (2.5)$$

where a Jacobean matrix $\mathbf{J} = \partial \text{vech}^T(\Sigma) / \partial t$, $\hat{\mathbf{J}} = \left[\partial \text{vech}^T(\Sigma) / \partial t\right]_{\Sigma = \hat{\Sigma},\ t = \hat{t}}$, $u_k = \left(\text{vech}^T\left(x_k x_k^T\right),\ x_k^T\right)^T$ and $\mathbf{I}_p$ is the $p \times p$ identity matrix. We can now obtain the linearized value (2.5).

Derivation of (2.5): We consider the variance of $p/(p-1)\ \text{tr}\left(\hat{\Sigma}\right) / \mathbf{1}^T \hat{\Sigma} \mathbf{1}$ since its variance is the same as $\text{Var}(\hat{\alpha})$. Let $\alpha^* = p/(p-1)\ \text{tr}(\Sigma) / \mathbf{1}^T \Sigma \mathbf{1}$. Also, let $\text{vec}^T(\Sigma) = \left(\sigma_{11}, \ldots, \sigma_{p1}, \sigma_{12}, \ldots, \sigma_{p2}, \sigma_{13}, \ldots, \sigma_{pp}\right)$ and $\text{vech}^T(\Sigma) = \left(\sigma_{11}, \ldots, \sigma_{p1}, \sigma_{22}, \ldots, \sigma_{p2}, \sigma_{33}, \ldots, \sigma_{pp}\right)$, a p^2-vector and a $p(p+1)/2$-vector, respectively. Then, we have

$$\frac{\partial \alpha^*}{\partial t} = \frac{\partial \alpha^*}{\partial \mathrm{vech}^T(\Sigma)} \mathbf{J} = \frac{p}{p-1} \frac{\partial}{\partial \mathrm{vech}^T(\Sigma)} \left\{ \frac{\mathrm{tr}(\Sigma)}{\mathbf{1}^T \Sigma \mathbf{1}} \right\} \mathbf{J}. \tag{2.6}$$

Now, in (2.6), we can show

$$\begin{aligned} \frac{\partial}{\partial \mathrm{vech}^T(\Sigma)} \left\{ \frac{\mathrm{tr}(\Sigma)}{\mathbf{1}^T \Sigma \mathbf{1}} \right\} &= \frac{1}{\mathbf{1}^T \Sigma \mathbf{1}} \frac{\partial \mathrm{tr}(\Sigma)}{\partial \mathrm{vech}^T(\Sigma)} - \frac{\mathrm{tr}(\Sigma)}{\left(\mathbf{1}^T \Sigma \mathbf{1}\right)^2} \frac{\partial \mathbf{1}^T \Sigma \mathbf{1}}{\partial \mathrm{vech}^T(\Sigma)} \\ &= \frac{1}{\left(\mathbf{1}^T \Sigma \mathbf{1}\right)^2} \left\{ \mathbf{1}^T \Sigma \mathbf{1} \,\mathrm{vec}^T\left(\mathbf{I}_p\right) - \mathrm{tr}(\Sigma) \left(2 \mathrm{vec}^T\left(\mathbf{1}\mathbf{1}^T\right) - \mathrm{vec}^T\left(\mathbf{I}_p\right)\right)\right\} K_p. \end{aligned} \tag{2.7}$$

Using (2.6) and (2.7), we can obtain

$$\frac{\partial \alpha^*}{\partial t} = \frac{p}{p-1} \frac{1}{\left(\mathbf{1}^T \Sigma \mathbf{1}\right)^2} \left\{ \mathbf{1}^T \Sigma \mathbf{1} \,\mathrm{vec}^T\left(\mathbf{I}_p\right) - \mathrm{tr}(\Sigma)\left(2\mathrm{vec}^T\left(\mathbf{1}\mathbf{1}^T\right) - \mathrm{vec}^T\left(\mathbf{I}_p\right)\right)\right\} K_p \,\mathbf{J}. \tag{2.8}$$

Note that the expression (2.8) is a vector that consists of the derivatives of α^* with respect to the components of t. Each element in (2.8) is multiplied by the influence function corresponding to the statistics in t as in (2.4). This is accomplished by multiplying (2.8) by $u_k = \left(\mathrm{vech}^T\left(x_k x_k^T\right), x_k^T\right)^T, k = 1, \ldots, n,$ which is obtained by using (2.3). Now substituting Σ by $\hat{\Sigma}$ leads to the linearized value (2.5).

The formula for the new value (2.5) is easily implemented in the computer code using commonly available computer software. The relevant R code is available in the Supplementary Material.

We note that, in application to survey sampling, the estimate $\hat{\Sigma}$ should be obtained properly by incorporating the survey design. The variance is estimated by $\widehat{\mathrm{Var}}\left(\hat{\hat{Z}}\right)$, where $\widehat{\mathrm{Var}}$ indicates an operation to obtain the variance incorporating the weights and survey design properly, e.g., the Sen-Yates-Grundy variance estimator (Sen, 1953; Yates and Grundy, 1953), an unbiased variance estimator for the Horvitz-Thompson estimator (Horvitz and Thompson, 1952) under designs with fixed sample sizes (e.g., Särndal, Swensson and Wretman, 1992) or the variance estimator for sampling with the replacement as a conservative approximation (Wolter, 1985). Specifically, in this paper, the variance for the NCS-R data is estimated as

$$\widehat{\mathrm{Var}}\left(\hat{\hat{Z}}\right) = \sum_{h=1}^{H} \widehat{\mathrm{Var}_h}\left(\hat{\hat{Z}}\right), \tag{2.9}$$

where $\widehat{\mathrm{Var}_h}\left(\hat{\hat{Z}}\right)$ indicates the design-specific variance estimator for stratum $h\,(h = 1, \ldots, H)$. Once z_k values are obtained, standard statistical software for survey sampling such as R package "survey" (Lumley, 2004) can be used for the calculation of (2.9).

Now, consider a case that x is a random variable following a distribution and that an observation is a realization of the random variable; in addition, a sample of size n is obtained according to the random variable. In this specific case, we do not consider the finite population, where the design-based variance estimation is suitable as shown in the previous discussion. In a random variable setting, let $\hat{\hat{Z}}$ indicate the estimator with the measure $\hat{M}$ as the empirical distribution function (Fernholz, 1991). Employing the concept of a robust statistical inference based on the influence function (Davison and Hinkley, 1997), the sample variance for the population can be calculated by

$$\widehat{\text{Var}}\,(\hat{\hat{Z}}) = n^{-1}\sum_{k=1}^{n}(z_k - \overline{z})^2 \Big/ (n-1), \tag{2.10}$$

where z_k is the linearized value (2.5) obtained from the statistic $\hat{\alpha}$ based on the sample (size n) and $\overline{z}$ is the sample mean of z_k $(k = 1, \ldots, n)$. We also note that the formula (2.10) is not constructed for infinite populations in survey methodology, where the finite population is seen as a realization from an infinite population. In that case, the outcomes of a statistical model give rise to the values of the characteristics of interest in the finite population, thus the model-based variance estimation is appropriate (Binder and Roberts, 2009). The formula (2.10) can be used for a general data analytical setting, where observations are considered as realizations of a random variable.

2.2 The coverage-corrected bootstrap method

The linearization provides reasonable estimates for the confidence intervals; however, in some cases, the coverage rate may not be satisfactory when the underlying distributions are non-normal (see Section 3). In these cases, some computer-intensive approaches such as the double bootstrap method, which is also called the coverage-corrected bootstrap may be implemented (Hall et al., 1989). We primarily discuss the double bootstrap method instead of the typical “single” bootstrap method (DiCiccio and Romano, 1988) since we observe that the single bootstrap method may not be satisfactory with non-normal underlying distributions (e.g., lognormal distribution) in terms of the coverage rate (Table 3.3).

For adjusting the bootstrap weight, the rescaling method referred to as the Rao-Wu bootstrap (Rao and Wu, 1988) is a popular approach for analyzing a lot of survey data, e.g., from Statistics Canada surveys (Mach, Saïdi and Pettapiece, 2007). The Rao-Wu bootstrap method is based on the assumption of sampling with a replacement, but is often employed for sampling without a replacement as well, when the first-stage sampling fraction is negligible (Mach et al., 2007). Herein, we propose implementing the coverage-corrected bootstrap method using the weight adjustment from Rao and Wu (1988). Among the various bootstrap confidence interval techniques (e.g., for these varieties, see Hwang, 1995), we consider the percentile bootstrap interval, which is a strictly nonparametric bootstrap approach (Hall, Martin and Schucany, 1989).

The coverage rates of the bootstrap confidence intervals can be corrected by incorporating additional bootstrap procedures. Because of bootstrapping the bootstrap sample, this kind of a procedure is referred to as the double bootstrap method (Martin, 1992). It is known that this method reduces the coverage error of two-sided confidence intervals by a factor of the order n^{-1} compared to the single bootstrap or normal-theory confidence intervals (Martin, 1992). Suppose $\hat{l}$ and $\hat{u}$ are the lower and upper bounds of the percentile bootstrap confidence interval using the original data. As proposed by Hall et al. (1989), the $100\,(1-q)\,\%$ coverage-corrected bootstrap confidence interval can be defined as $\left(\hat{l} - \delta,\, \hat{u} + \delta\right),$ where a positive value of δ satisfies

$$1 - q = \Pr\left\{\hat{\alpha} \in \left(\hat{l}^{\,*} - \delta,\, \hat{u}^{*} + \delta\right)\right\}. \tag{2.11}$$

The values $\hat{l}^*$ and $\hat{u}^*$ indicate the lower and upper bounds, respectively, of the confidence interval obtained by bootstrapping a resampled data set. The probability in the right-hand side of equation (2.11) is empirically evaluated as shown in the following steps.

Step 1: For each bootstrap sample $i\,(i = 1, \ldots, B)$, we obtain the intervals based on second-time resamples, $\left(\hat{l}_i^*, \hat{u}_i^*\right)$.

Step 2: We search t satisfying $\min\left\{t\colon \left|1 - q - \widehat{\Pr}\left\{\hat{\alpha} \in \left(\hat{l}^* - \delta, \hat{u}^* + \delta\right)\right\}\right| \geq 0\right\}$ where $\widehat{\Pr}$ indicates the empirical probability.

Step 3: The confidence interval is obtained by $\left(\max\left(\hat{l} - \delta, 0\right), \min\left(\hat{u} + \delta, 1\right)\right)$.

We use $\left(\max\left(\hat{l} - \delta, 0\right), \min\left(\hat{u} + \delta, 1\right)\right)$ since the true α is assumed to be between 0 and 1.

In the analysis, we have to resample the data without disrupting the survey design structure. The bootstrap is carried out within each stratum, and all observations in the same cluster should be kept together in a resampled data set (Lohr, 1999). For each resampled data set, new weights need to be obtained (Rao, Wu and Yue, 1992). Specifically, let n_h indicate the sample size of the primary sampling unit (PSU) in stratum $h\,(h = 1, \ldots, H)$. Suppose we resample n_h^* clusters for each stratum. Then, the rescaled weight for observation k in the resample is

$$w_k^{(b)} = w_k \left\{\left(1 - \sqrt{\frac{n_h^*}{n_h - 1}}\right) + \sqrt{\frac{n_h^*}{n_h - 1}}\,\frac{n_h}{n_h^*}\,m_k\right\}, \tag{2.12}$$

where m_k is the number of repetitions of the PSU that observation k belongs to and w_k is the original weight of observation k (Rao et al., 1992; Mach, Dumais and Robinson, 2005; Mach et al., 2007). When $n_h^* = n_h - 1$, the bootstrap weight becomes $w_k^{(b)} = w_k \left\{\frac{n_h}{n_h - 1}\,m_k\right\}$, which is a conventional bootstrap weight (Lohr, 1999). This procedure is repeated to obtain a total of B bootstrap samples. For the actual data analysis, we use $B = 500$ following the common practice of Statistics Canada surveys (Canadian Community Health Survey - Annual Component, 2007). To obtain the estimates, the stratification or cluster structure is no longer considered since the bootstrap weights take into account the survey design structure (Lohr, 1999). The percentile interval will be obtained based on the B values of the estimates of α. For each resample, α is estimated based on the sample variance and covariance matrix incorporating the weights. To obtain the coverage-corrected confidence interval, we carry out the additional bootstrap with each bootstrap sample in a similar manner to what was explained above. In the simulation and data analysis we use 200 bootstrap samples for the second round of bootstrapping. The relevant R code is provided in the Supplementary Material.

3 Simulation

We investigate the performance of the proposed methods in two scenarios; stratified two-stage cluster sampling and single-stage unequal probability sampling.

For stratified two-stage cluster sampling, the finite population is generated using three strata where each stratum includes 200 PSUs and 50 secondary sampling units (SSUs) totaling 30,000 SSUs. The underlying distributions that are used include the multi-normal distribution, multi-lognormal distribution and correlated ordinal data categorized from multi-lognormal distribution variables. The cases of $p = 5$ and $p = 10$ are considered. Different means are used for the different strata. The observations are correlated within a PSU. See the footnote of Table 3.1 for the detailed parameter information. Simple random sampling is carried out at the first-stage and second-stage, respectively, within each stratum. Thus, the appropriate weights are calculated per stratum as $(N_h M_h)/(n_h m_h)$ for each individual (SSU), where N_h, M_h, n_h and m_h are the number of PSUs per stratum, the number of SSUs per PSU, first-stage sample size per stratum, and second-stage sample size, respectively. Since the population is finite, the true value of α is known from the generated population.

For unequal probability sampling (Table 3.2), we generate a population of 30,000, where the underlying distributions of the data are the multi-normal distribution, multi-lognormal distribution and correlated ordinal data categorized from the multi-lognormal distribution variables similar to the cases found in Table 3.1. See the footnote of Table 3.2 for the detailed parameter information. Each individual i is assigned a random number x_i from the Binomial (20, 0.5) distribution, achieving the semblance of SSU sizes per PSUs. For sampling, the first-order inclusion probability is proportional to size x_i (probability proportional to size sampling). Thus, the weight for an individual i is obtained as $n^{-1}\sum_k x_k / x_i$, where n is the sample size. The sample selection procedure uses the systematic sampling technique that considers first-order inclusion probabilities. For the linearization method, the variance is estimated using the usual estimator for with-replacement sampling (Mach et al., 2007) as a conservative approximation of the methods for without-replacement sampling (Wolter, 1985). Since the sampling fraction is negligible in the simulation, the finite population correction is not incorporated. The 95% confidence interval is obtained based on the normal approximation.

Table 3.1 (stratified two-stage cluster sampling) and Table 3.2 (single-stage unequal probability sampling) show the coverage rates and average widths of the confidence intervals based on the proposed linearization method and the coverage-corrected bootstrap method (1,000 simulations per scenario). The linearization method and the coverage-corrected methods are evaluated using same simulated data sets. For the coverage-corrected method, we use $B = 200$ for the first bootstrap, $B = 200$ for the second bootstrap. The linearized method shows the coverage rates as being close to the target confidence level for the multi-normal distributions and correlated ordinal data in most scenarios. We note that, in the random variable settings, the confidence intervals based on a normal approximation work well with various ordinal data once the variance is correctly obtained (Maydeu-Olivares, Coffman and Hartmann, 2007). Our simulation results show that the normal approximation works well with the ordinal data in finite population settings as well. When the underlying distribution is the multi-lognormal distribution, the coverage rates of the confidence intervals based on the normal approximation may be somewhat lower than the target coverage rate, but they improve with increasing sample sizes. For the multi-lognormal distribution, the coverage-corrected bootstrap method using the weight adjustment by Rao and Wu (1988) shows substantially improved coverage rates comparing to the linearized method. In comparison to the linearization method, the coverage-corrected bootstrap method has slightly increased widths, and the coverage rates are reasonably close to the target confidence level for most cases in Tables 3.1 and 3.2.

We also note that for the stratified sampling cases with relatively low α values, we can identify cases that the coverage-corrected method provides less-than-desirable coverage rates, with the multi-normal or ordinal data indicating that the coverage-corrected method is not a panacea for interval estimation. Here, the linearization method is a reasonable choice over the coverage-corrected method if the underlying distribution is ordinal or normal.

Table 3.1
(Stratified two-stage cluster sampling). The coverage rates (CR) and average widths (Width) of 95% confidence intervals based on the linearization method and coverage-corrected method (Double Bt). The values of npsu and nssu are the sample sizes for PSUs and SSUs within a PSU, respectively. Two α values indicate α for $p = 5$ and $p = 10$, respectively

Method	Distribution	(npsu, nssu)	α	$p = 5$		$p = 10$	
				CR	Width	CR	Width
Linearization	Multi-normal	(10, 20)	0.91, 0.91	0.941	0.024	0.946	0.023
		(20, 20)	0.90, 0.91	0.934	0.017	0.962	0.016
		(10, 20)	0.56, 0.67	0.941	0.121	0.944	0.095
		(20, 20)	0.56, 0.67	0.953	0.084	0.961	0.064
	Multi-lognormal	(10, 20)	0.85, 0.85	0.904	0.067	0.902	0.059
		(20, 20)	0.86, 0.85	0.908	0.054	0.935	0.049
		(10, 20)	0.51, 0.53	0.913	0.163	0.924	0.154
		(20, 20)	0.51, 0.55	0.933	0.118	0.928	0.108
	Correlated ordinal	(10, 20)	0.85, 0.87	0.939	0.043	0.938	0.035
		(20, 20)	0.85, 0.87	0.939	0.030	0.954	0.025
		(10, 20)	0.48, 0.53	0.934	0.147	0.928	0.130
		(20, 20)	0.48, 0.60	0.955	0.103	0.955	0.077
Double Bootstrap	Multi-normal	(10, 20)	0.91, 0.91	0.959	0.026	0.960	0.025
		(20, 20)	0.90, 0.91	0.954	0.019	0.964	0.017
		(10, 20)	0.56, 0.67	0.939	0.120	0.909	0.084
		(20, 20)	0.56, 0.67	0.955	0.084	0.942	0.059
	Multi-lognormal	(10, 20)	0.85, 0.85	0.945	0.080	0.959	0.071
		(20, 20)	0.86, 0.85	0.948	0.063	0.963	0.057
		(10, 20)	0.51, 0.53	0.947	0.186	0.942	0.163
		(20, 20)	0.51, 0.55	0.955	0.125	0.942	0.109
	Correlated ordinal	(10, 20)	0.85, 0.87	0.964	0.047	0.955	0.038
		(20, 20)	0.85, 0.87	0.950	0.033	0.960	0.026
		(10, 20)	0.48, 0.53	0.937	0.148	0.919	0.121
		(20, 20)	0.48, 0.60	0.957	0.104	0.942	0.073

The values of α are based on the generated finite populations in all scenarios. For multi-normal data, the mean vectors consist of values of 1, 1.05 and 1.1 for strata 1, 2, and 3, respectively, and the common covariance within PSUs in addition to the covariance within multivariate data is 0.05. The covariance matrix has diagonal elements of 1 and the common off-diagonal elements to produce relevant α values. Multi-lognormal data are exponential of multi-normal data with the same mean and covariate structures. In the covariance matrix, common off-diagonal values are selected to produce relevant α values. For the correlated ordinal data, we first generate the multi-lognormal data with the same structures described above, then categorize them to 0, 1, 2 and 3 for values ≤ 2, $2 <$ values ≤ 10, $10 <$ values ≤ 15, and values > 10, respectively.

Table 3.2
(Single-stage unequal probability sampling). The coverage rates (CR) and average widths (width) of 95% confidence intervals based on the linearization method and coverage-corrected method (Double Bt). The values of n indicate the sample sizes for PSUs. Two α values indicate α for $p = 5$ and $p = 10$, respectively

Method	Distribution	n	α	$p = 5$		$p = 10$	
				CR	Width	CR	Width
Linearization	Multi-normal	100	0.90, 0.90	0.942	0.063	0.936	0.058
		200	0.90, 0.90	0.921	0.044	0.956	0.042
		100	0.50, 0.51	0.942	0.317	0.936	0.291
		200	0.50, 0.50	0.921	0.219	0.956	0.210
	Multi-lognormal	100	0.85, 0.84	0.816	0.116	0.853	0.104
		200	0.85, 0.85	0.870	0.103	0.901	0.083
		100	0.47, 0.47	0.851	0.346	0.887	0.312
		200	0.48, 0.47	0.911	0.264	0.935	0.253
	Correlated ordinal	100	0.84, 0.86	0.926	0.110	0.923	0.086
		200	0.84, 0.86	0.930	0.078	0.947	0.063
		100	0.43, 0.43	0.938	0.368	0.945	0.335
		200	0.43, 0.42	0.942	0.260	0.948	0.245
Double Bt	Multi-normal	100	0.90, 0.90	0.961	0.073	0.950	0.068
		200	0.90, 0.90	0.953	0.049	0.965	0.047
		100	0.50, 0.51	0.958	0.361	0.951	0.241
		200	0.50, 0.50	0.948	0.335	0.962	0.232
	Multi-lognormal	100	0.85, 0.84	0.912	0.166	0.943	0.138
		200	0.85, 0.85	0.940	0.136	0.948	0.107
		100	0.47, 0.47	0.954	0.436	0.946	0.382
		200	0.48, 0.47	0.946	0.318	0.965	0.295
	Correlated ordinal	100	0.84, 0.86	0.940	0.134	0.937	0.103
		200	0.84, 0.86	0.937	0.090	0.956	0.066
		100	0.43, 0.43	0.954	0.428	0.946	0.388
		200	0.43, 0.42	0.949	0.287	0.957	0.271

The values of α are based on the generated finite populations in all scenarios. For multi-normal data, the mean vectors consist of values of 1. The covariance matrix has diagonal elements of 1 and the common off-diagonal elements to produce relevant α values. Multi-lognormal data are exponential of multi-normal data with the same mean and covariate structures. Common off-diagonal values are selected to produce relevant α values. For the correlated ordinal data, we first generated the multi-lognormal data with the same structures described above, then categorize them to 0, 1, 2 and 3 for values ≤ 2, $2 <$ values ≤ 10, $10 <$ values ≤ 15, and values > 10, respectively.

Thus, we conclude that, for general ordinal data, which are typical responses for most assessment instruments, the linearization method will be satisfactory to obtain the confidence intervals. When the instruments consist of continuous data and some skewed distributions are observed, the coverage-corrected bootstrap method will generally provide more accurate confidence intervals than the normal approximation.

It may be of interest to compare the performance of the proposed confidence interval methods to other existing confidence interval methods in a random variable setting since the proposed methods can be applied to these settings, as shown in (2.10). Table 3.3 presents the comparisons of the coverage rates and widths of various confidence interval methods based on the data generated from a random variable. The existing confidence interval methods can be categorized to either using an analytical distribution based on the multi-normal distribution, or using a large sample approximation for the normal distribution of $\hat{\alpha}$ or a transformation of $\hat{\alpha}$. For the existing methods, we consider three normal-based confidence intervals and a bootstrap method, i.e., confidence intervals based on the exact F distribution using the normal data (van Zyl, Neudecker and Nel, 2000; Kistner and Muller, 2004), a large sample approximation of $\log(1-\hat{\alpha})/2$ (van Zyl et al., 2000), a large sample approximation of $\hat{\alpha}$ based on the "distribution-free" standard error estimate (Yuan et al., 2003; Maydeu-Olivares et al., 2007), and the percentile bootstrap confidence interval with a single bootstrap (DiCiccio and Romano, 1988). These techniques are compared to the confidence intervals based on the linearization method and the coverage-corrected bootstrap method. The data are generated from the multi-normal distribution, multi-lognormal distribution, and the correlated ordinal data similar to the simulations in the previous tables. The values of α in Table 3.3 are for the random variables. In general, the results seem similar to those of finite population cases. The existing confidence interval methods, as well as the linearization method, perform unsatisfactorily with the lognormal data, yet their coverage rates are close to the target confidence levels using the ordinal data and normal distributions when the sample sizes increase. The coverage-corrected bootstrap method shows a coverage rate close to the confidence level with a lognormal distribution while providing wider confidence interval widths than the other methods. In the case of the multi-normal distribution, the coverage-corrected bootstrap method seems to have higher coverage rates than the target confidence level. In comparison with the single bootstrap method, the coverage-corrected method increases the coverage rates by 1 to 3% overall for the multi-lognormal distribution cases.

Table 3.3
The coverage rates and widths of 95% confidence intervals based on F distribution (F dist), the asymptotic distribution of the transformed $\hat{\alpha}$ (Asymp1), the asymptotic distribution by Yuan et al. (Asymp2), the linearization method (Linearization), the percentile bootstrap method with single bootstrap (Single Bt) and the coverage corrected method (Double Bt). In the first column, p, low α, and high α values are shown in the parentheses

Distribution	Approach	n	$p = 5$				$p = 10$			
			Low α		High α		Low α		High α	
			CR	Width	CR	Width	CR	Width	CR	Width
Multi-normal (5, 0.5, 0.9) (10, 0.5, 0.9)	F dist	50	0.955	0.461	0.955	0.092	0.960	0.429	0.960	0.086
		100	0.954	0.319	0.954	0.064	0.943	0.298	0.943	0.060
		200	0.948	0.222	0.948	0.044	0.954	0.208	0.042	0.954
	Asymp1	50	0.954	0.471	0.954	0.094	0.956	0.440	0.956	0.088
		100	0.947	0.322	0.947	0.064	0.939	0.302	0.939	0.060
		200	0.947	0.223	0.947	0.045	0.959	0.209	0.959	0.042
	Asymp2	50	0.937	0.432	0.937	0.086	0.931	0.407	0.931	0.081
		100	0.948	0.311	0.948	0.062	0.943	0.293	0.943	0.059
		200	0.945	0.218	0.945	0.044	0.953	0.205	0.953	0.041
	Linearization	50	0.937	0.441	0.937	0.088	0.937	0.415	0.937	0.083
		100	0.948	0.315	0.948	0.062	0.944	0.296	0.944	0.059
		200	0.946	0.219	0.946	0.044	0.953	0.206	0.953	0.041
	Single Bt	50	0.936	0.490	0.936	0.098	0.935	0.465	0.935	0.093
		100	0.944	0.334	0.944	0.067	0.939	0.314	0.939	0.063
		200	0.944	0.227	0.944	0.045	0.944	0.227	0.965	0.043
	Double Bt	50	0.959	0.498	0.960	0.107	0.959	0.484	0.960	0.103
		100	0.958	0.355	0.960	0.072	0.954	0.336	0.954	0.068
		200	0.954	0.238	0.954	0.048	0.954	0.238	0.974	0.045

The values of α are theoretical values except cases of correlated ordinal data. The α values for the correlated ordinal data are obtained based on 60,000 simulations. Structures of the mean vector and covariance matrix follow those explained in Table 3.2.

Table 3.3 (continued)
The coverage rates and widths of 95% confidence intervals based on F distribution (F dist), the asymptotic distribution of the transformed $\hat{\alpha}$ (Asymp1), the asymptotic distribution by Yuan et al. (Asymp2), the linearization method (Linearization), the percentile bootstrap method with single bootstrap (Single Bt) and the coverage corrected method (Double Bt). In the first column, p, low α, and high α values are shown in the parentheses

Distribution	Approach	n	p = 5				p = 10			
			Low α		High α		Low α		High α	
			CR	Width	CR	Width	CR	Width	CR	Width
Multi-lognormal (5, 0.47, 0.85) (10, 0.47, 0.84)	F dist	50	0.919	0.487	0.829	0.151	0.928	0.457	0.860	0.140
		100	0.888	0.337	0.763	0.101	0.884	0.317	0.813	0.095
		200	0.862	0.235	0.727	0.070	0.906	0.221	0.782	0.066
	Asymp1	50	0.921	0.497	0.827	0.155	0.923	0.469	0.859	0.143
		100	0.884	0.341	0.759	0.103	0.888	0.321	0.809	0.097
		200	0.858	0.237	0.720	0.070	0.909	0.223	0.787	0.066
	Asymp2	50	0.837	0.410	0.805	0.146	0.870	0.406	0.844	0.132
		100	0.874	0.338	0.825	0.119	0.883	0.318	0.854	0.108
		200	0.903	0.267	0.853	0.097	0.927	0.244	0.876	0.086
	Linearization	50	0.842	0.419	0.814	0.149	0.878	0.415	0.850	0.135
		100	0.877	0.342	0.828	0.120	0.885	0.321	0.862	0.109
		200	0.903	0.269	0.858	0.098	0.928	0.245	0.879	0.086
	Single Bt	50	0.929	0.472	0.887	0.174	0.930	0.464	0.889	0.158
		100	0.928	0.362	0.883	0.133	0.929	0.337	0.887	0.119
		200	0.932	0.274	0.900	0.102	0.941	0.251	0.917	0.090
	Double Bt	50	0.943	0.524	0.944	0.221	0.950	0.504	0.930	0.199
		100	0.950	0.422	0.935	0.170	0.951	0.385	0.938	0.150
		200	0.955	0.318	0.943	0.126	0.954	0.283	0.948	0.109
Correlated ordinal (5, 0.84, 0.54) (10, 0.91, 0.70)	F dist	50	0.941	0.424	0.926	0.149	0.950	0.256	0.931	0.075
		100	0.931	0.292	0.929	0.102	0.939	0.177	0.904	0.052
		200	0.938	0.203	0.917	0.071	0.956	0.123	0.933	0.036
	Asymp1	50	0.945	0.432	0.919	0.152	0.947	0.262	0.927	0.077
		100	0.930	0.295	0.922	0.103	0.938	0.179	0.907	0.053
		200	0.934	0.204	0.914	0.071	0.954	0.124	0.936	0.036
	Asymp2	50	0.922	0.432	0.911	0.144	0.928	0.242	0.920	0.074
		100	0.928	0.289	0.933	0.108	0.931	0.177	0.918	0.055
		200	0.940	0.205	0.931	0.077	0.950	0.125	0.947	0.039
	Linearization	50	0.928	0.402	0.916	0.147	0.932	0.247	0.923	0.075
		100	0.929	0.292	0.936	0.109	0.936	0.178	0.925	0.056
		200	0.940	0.206	0.931	0.078	0.950	0.126	0.950	0.039
	Single Bt	50	0.927	0.447	0.901	0.163	0.921	0.275	0.898	0.084
		100	0.928	0.308	0.929	0.116	0.935	0.189	0.908	0.059
		200	0.938	0.213	0.935	0.080	0.949	0.131	0.942	0.041
	Double Bt	50	0.950	0.476	0.927	0.189	0.945	0.310	0.934	0.101
		100	0.943	0.334	0.945	0.131	0.951	0.208	0.937	0.069
		200	0.955	0.227	0.948	0.087	0.956	0.140	0.959	0.045

The values of α are theoretical values except cases of correlated ordinal data. The α values for the correlated ordinal data are obtained based on 60,000 simulations. Structures of the mean vector and covariance matrix follow those explained in Table 3.2.

4 Application

In this section, we provide detailed information regarding the NCS-R survey and subgroup analysis using the data sets. The relevance of the instruments may vary based on the different demographic groups studied, and thus a relatively low reliability in a certain group would be an indication that the instrument items may need some adjustments for that group. Using the data from the NCS-R, we investigate the changes of α using the Kessler 10 (K10, Kessler, Andrews, Colpe, Hiripi, Mroczek, Normand, Walters and Zaslavsky, 2002), the Kessler 6 (K6, Kessler et al., 2002) and the Sheehan Disability Scale (SDS, Sheehan, Harnett-Sheehan and Raj, 1996). More details about these scales are explained in Section 4.1.

4.1 The data

The NCS-R is a mental health survey for a nationally representative sample of English-speaking noninstitutionalized household residents in the United States (Kessler et al., 2004) and it uses the fully structured World Health Organization's (WHO) World Mental Health Survey version of the Composite International Diagnostic Interview (WMH-CIDI) (Byers, Yaffe, Covinsky, Friedman and Bruce, 2010). Using computer-assisted personal interviews, the NCS-R was carried out to obtain further information not fully covered in the previous baseline National Comorbidity Survey (NCS). A total of 9,282 participants 18 years and older completed the Part I interview, and a subsample of 5,692 participants completed the Part II instruments. The data sets are publicly accessible and downloadable on the ICPSR (Inter-university Consortium for Political and Social Research) website (https://www.icpsr.umich.edu/icpsrweb). The NCS-R is based on a stratified multi-stage probability sample design (42 strata where each stratum has two PSUs, totaling 84 PSUs), and the sample weights are provided in the data to reflect the survey design. Each PSU consists of metropolitan statistical areas or counties (Kessler et al., 2004). The final weights in the NCS-R data are adjusted for nonresponses to the survey instruments. Weights accounting for the designs of the different parts of the surveys (i.e., Parts I and II) are provided, respectively, in the NCS-R data. The weights are normalized to have a sum equal to 9,282 for Part I and 5,692 for Part II (mean weight = 1), respectively. In this case, the weights do not represent the inverse of the selection probabilities. Due to this and the fact that the sample size is quite small compared to the total population of interest, the finite population correction is not considered in the data analysis. Incorporating these weights corrects the overrepresentation of "racial minorities, females, residents of the Midwest, people with 13+ years of education, and residents of metropolitan areas" (Kessler et al., 2004).

The 10-item Kessler psychological distress scale or the K10 is an instrument used to assess the distress level of people (Kessler et al., 2002), and the K6 is an abbreviated set of six items from the K10. Both the K10 and K6 are considered effective scales for screening mental disorders (Brouwer, Cornelius, van der Klink and Groothoff, 2013). The K10 for 30-day symptoms is included in the Part II instruments. It is composed of 10 questions of a self-reported assessment of psychological distresses in the worst month of the past year for each interviewee. The questions ask feelings such as tiredness, nervousness, hopelessness, and so forth. All 10 questions produce an ordinal data scoring of 1 (all of the time) to 5 (none of the time). The final total score ranges from 10 to 50 with the higher scores showing more distress. The K10 values in the NCS-R have missing data, and the weights given by the NCS-R adjust for survey nonresponses, but they do not adjust for items with missing data. Although these missing data may compromise the unbiasedness of the weighted estimation (Alegria, Jackson, Kessler and Takeuchi, 2007), we use only completed data and remedial approaches such as weighting class adjustment or imputation of the data are not considered in our analysis.

The SDS assesses functional impairment associated with mental disorders (Sheehan et al., 1996). The SDS in the NCS-R assesses disorder-specific role impairments (Sheehan et al., 1996; Druss, Hwang, Petukhova, Sampson, Wang and Kessler, 2009). It consists of four questions evaluating the disruption of activities associated with home, work, social and close relationship using 0 to 10 scales, with higher scores showing more severe impairment. In this paper, among the SDS scales of various mental disorders, we use the SDS for the participants with chronic conditions as a Part II instrument. Since the SDS is disorder-specific, it has missing data. For the data analysis, we use only complete data.

4.2 Subsample analysis

For the subgroups, a domain analysis may be applied. Suppose that a domain indicator function I_k^d $(d = 1, \ldots, D)$ has a value of 1 if the unit k is in a domain d (i.e., $k \in s_d$) and 0 otherwise. Then, the statistics of the domain are estimated by modifying the weight as $w_k^{(d)} = w_k I_k^d$. The procedures used to obtain the estimates and the corresponding variance or covariance are carried out with the modified weights. Since the sample size is not fixed but is rather treated as an estimate, an estimator such as the sample mean and sample variance can be considered as the ratio estimator, i.e., both the numerator and the denominator are estimated, and the variance of the estimator is obtained accordingly. However, when the sample size is large, thus the ratio between the domain sample size and the whole sample size is close to the true population ratio, it is known that the variance of the ratio estimator is approximately the same as that of the estimator with the fixed sample size using only the subgroup of interest, making "little difference in practice" regarding those estimators (Lohr, 1999, page 79). The negligible difference between the domain estimator and the estimator using only the subsample can be easily shown using the variance estimator in an unequal probability sampling with replacement setting. Let $\hat{\bar{Y}}_d$ indicate the domain estimator of the mean (Lohr, 1999) for single-stage sampling, i.e., $\hat{\bar{Y}}_d = \sum_{k=1}^n w_k I_k^d y_k \Big/ \sum_{k=1}^n w_k I_k^d = \sum_{k \in s_d} w_k y_k \Big/ \sum_{k \in s_d} w_k$, where the last term uses only the subsample. Now, for the variance estimator of $\hat{\bar{Y}}_d$ (Paben, 1999; SAS/STAT user's guide, 2010), we can show

$$\hat{V}\left(\hat{\bar{Y}}_d\right) = \sum_{k=1}^n \left\{ \frac{w_k I_d \left(y_k - \hat{\bar{Y}}_d\right)}{\sum_{l=1}^n w_l I_d} \right\}^2 \frac{n}{n-1} \simeq \sum_{k \in s_d} \left\{ \frac{w_k \left(y_k - \hat{\bar{Y}}_d\right)}{\hat{N}_d} \right\}^2 \frac{n_d}{n_d - 1}, \tag{4.1}$$

where n_d is the sample size of s_d. Here, the right-hand side of equation (4.1) uses the observation only in domain s_d. Based on this fact, the variance for a subgroup is obtained based only on the data from the subgroup of interest in this paper.

When implementing the bootstrap method, we use $n_h^* = 2,$ which produces all the positive weights in (2.12). In the subsample analysis, the bootstrap sample may contain only one PSU per stratum. In this case, the variance cannot be estimated. If we have multiple strata with one PSU, we combine those strata. If we have only one stratum with one PSU, we merge that stratum with another stratum arbitrarily. The rationale of this practice is that the variance incorporating strata is usually smaller than that without strata, thus such a practice may produce a wider (more conservative) confidence interval.

4.3 Results

The estimates of α and their confidence intervals for the whole participants are shown in Table 4.1. The table presents the confidence intervals using the coverage-corrected percentile method and the confidence interval using the linearization method for each instrument. Between the K10 and K6, it appears that the K10 has a higher α estimate. This may be explained by the fact that the removed items from the K10 are highly correlated with the remaining items in the K6, thus removing these items results in a reduced $\hat{\alpha}$ value. The coverage-corrected percentile method shows confidence intervals that are close to the linearization method, while slightly wider. Considering the ease of calculation, when an analysis deals with instruments with ordinal data, the results of the similar confidence intervals in Table 4.1 may indicate that a normal approximation using the proper variance estimation may be satisfactory for the investigated instruments, which do not include the skewed continuous data that we examined in Tables 3.1 and 3.2.

The subgroup analysis is shown in Table 4.2, where $\hat{\alpha}$ and the confidence intervals are presented for different groups by age, gender and marriage status. The age groups are defined as young (34 years and under), middle aged (35-64 years), and old aged (65 years and over) per the available literature (e.g., Sunderland, Hobbs, Anderson and Andrews, 2012), where the cut-off points for the age groups are decided by epidemiological studies and the traditional definition of old age. The marriage status is defined by grouping married and unmarried (including divorced, separated, widowed and never married). Both the coverage-corrected bootstrap method and the linearization method provide comparable confidence intervals while the coverage-corrected bootstrap produces a slightly wider confidence interval. Considering that the coverage-corrected method is computationally intensive, the linearization method may be preferred when the instruments consist of ordinal scales.

Table 4.1
Estimates of α and their 95% confidence intervals (CI) for overall sample

Instrument	$\hat{\alpha}$	Cov-Correct CI	Linearization CI	n
K10	0.901	(0.893, 0.911)	(0.893, 0.909)	2,378
K6	0.840	(0.829, 0.857)	(0.827, 0.852)	3,442
SDS	0.867	(0.852, 0.883)	(0.853, 0.880)	3,983

Table 4.2
Estimates of α and their 95% confidence intervals (CI) for subgroups

Instrument	Subgroups	$\hat{\alpha}$	Cov-Correct CI	Linearization CI	*n*
K10	Female	0.898	(0.880, 0.914)	(0.882, 0.914)	869
	Male	0.902	(0.896, 0.912)	(0.895, 0.910)	1,509
	Young age	0.888	(0.875, 0.900)	(0.875, 0.900)	890
	Middle age	0.913	(0.902, 0.925)	(0.902, 0.924)	1,281
	Old age	0.862	(0.827, 0.894)	(0.830, 0.893)	207
	Married	0.895	(0.882, 0.910)	(0.882, 0.907)	1,232
	Unmarried	0.902	(0.892, 0.913)	(0.892, 0.912)	1,146
K6	Female	0.824	(0.805, 0.849)	(0.803, 0.844)	1,288
	Male	0.848	(0.835, 0.866)	(0.835, 0.861)	2,154
	Young age	0.830	(0.810, 0.855)	(0.810, 0.849)	1,268
	Middle age	0.856	(0.842, 0.875)	(0.841, 0.870)	1,847
	Old age	0.773	(0.728, 0.821)	(0.725, 0.820)	327
	Married	0.823	(0.807, 0.844)	(0.806, 0.840)	1,805
	Unmarried	0.851	(0.833, 0.875)	(0.832, 0.869)	1,637
SDS	Female	0.874	(0.854, 0.895)	(0.853, 0.896)	1,589
	Male	0.861	(0.844, 0.880)	(0.847, 0.876)	2,394
	Young age	0.837	(0.805, 0.866)	(0.808, 0.866)	1,159
	Middle age	0.883	(0.870, 0.898)	(0.871, 0.896)	2,296
	Old age	0.849	(0.779, 0.903)	(0.796, 0.901)	555
	Married	0.886	(0.870, 0.903)	(0.871, 0.900)	2,286
	Unmarried	0.841	(0.818, 0.864)	(0.820, 0.861)	1,697

To this end, we conclude this section with a discussion of the results of the subgroups. Sizable differences in $\hat{\alpha}$ between the groups are found in the age groups with the K10 and K6 and marital status in the SDS. There are no overlaps of the confidence intervals between the middle and old-age groups in the K10 and K6. This indicates that the questions in the K10 and K6 may be relatively less consistent among the old-age group than the middle-age group. For the SDS, there is also no overlap of the confidence intervals between the married and the unmarried groups. That is, the consistency of the questions is substantially lower for the unmarried group than for the married group. We speculate that the SDS items include the impairment of a certain area that may be more relevant to the married group than the unmarried group (e.g., a disruption of activities associated with home, work, social and close relationship).

5 Concluding remarks

We explained how to obtain the confidence intervals of α in survey sampling through the linearization and coverage-corrected bootstrap methods. Through the simulation study in the setting of multi-stage cluster sampling and unequal probability sampling, the linearization method showed the workable property in terms of the coverage rate in the case of the multi-normal distribution or correlated ordinal data. When dealing with some problematic continuous data such as the multi-lognormal distribution, the coverage-corrected bootstrap method showed better performance than the linearization method in terms of the coverage rates. The discussed interval estimation methods were applied to the NCS-R data set. The application demonstrated that both the interval estimation methods provide workable options to carry out an inference of α incorporating the survey design.

We conclude this section by noting the following recommendations. First, in the case of an unknown continuous and skewed distribution, the coverage-corrected confidence interval is a safe way to provide a confidence interval whose actual confidence level may be close to the nominal confidence level. Second, if the data are discrete with a large sample size, the normal approximation using the linearization method may provide satisfactory coverage rates and be preferred because of the easiness of computation.

Acknowledgements

The authors are grateful to the Associate Editor and two reviewers for comments and suggestions that led to a substantial improvement in this paper.

References

Alegria, M., Jackson, J.S., Kessler, R.C. and Takeuchi, D. (2007). *Collaborative Psychiatric Epidemiology Surveys (CPES)*, 2001-2003 [United States]. Retrieved from: http://doi.org/10.3886/ICPSR20240.v8.

Binder, D.A., and Roberts, G. (2009). Design-and model-based inference for model parameters. In *Handbook of Statistics*, Elsevier, 29, 33-54.

Bonett, D.G., and Wright, T.A. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of Organizational Behavior*, 36, 3-15.

Brouwer, S., Cornelius, B.L.R., van der Klink, J.J.L. and Groothoff, J.W. (2013). The performance of the K10, K6 and GHQ-12 to screen for present state DSM-IV disorders among disability claimants. *BMC Public Health*, 13(1), 128.

Byers, A.L., Yaffe, K., Covinsky, K.E., Friedman, M.B. and Bruce, M.L. (2010). High occurrence of mood and anxiety disorders among older adults: The national comorbidity survey replication. *Archives of General Psychiatry*, 67(5), 489-496.

Canadian Community Health Survey – Annual Component (CCHS) (2007). Retrieved from: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=29539.

Cochran, W. (1977). *Survey Sampling*. New York: John Wiley & Sons, Inc.

Cronbach, L.J. (1951). Coefficient alpha and the interval structure of tests. *Psychometrika*, 16, 297-334.

Davison, A.C., and Hinkley, D.V. (1997). *Bootstrap Methods and their Application*. New York: Cambridge University Press.

Demnati, A., and Rao, J.N.K. (2004). Linearization variance estimators for survey data. *Survey Methodology*, 30, 1, 17-26. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2004001/article/6991-eng.pdf.

Deville, J.-C. (1999). Variance estimation for complex statistics and estimators: Linearization and residual techniques. *Survey methodology*, 25, 2, 193-203. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1999002/article/4882-eng.pdf.

DiCiccio, T.J., and Romano, J.P. (1988). A review of bootstrap confidence intervals. *Journal of the Royal Statistical Society, Series B (Methodological)*, 338-354.

Druss, B.G., Hwang, I., Petukhova, M., Sampson, N.A., Wang, P.S. and Kessler, R.C. (2009). Impairment in role functioning in mental and chronic medical disorders in the United States: Results from the National Comorbidity Survey Replication. *Molecular Psychiatry*, 14(7), 728-737.

Fernholz, L.T. (1991). Almost sure convergence of smoothed empirical distribution functions. *Scandinavian Journal of Statistics*, 18(3), 255-262.

Hájek, J. (1981). *Sampling from a Finite Population*. New York: Dekker.

Hall, P., Martin, M.A. and Schucany, W.R. (1989). Better nonparametric bootstrap confidence intervals for the correlation coefficient. *Journal of Statistical Computation and Simulation*, 33(3), 161-172.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: John Wiley & Sons, Inc., 114.

Horvitz, D.G., and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260), 663-685.

Hwang, J.T. (1995). Fieller's problems and resampling techniques. *Statistica Sinica*, 5(1), 161-171.

Kessler, R.C., Andrews, G., Colpe, L.J., Hiripi, E., Mroczek, D.K., Normand, S.L.T., Walters, E.E. and Zaslavsky, A.M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32,(6) 959-976. http://dx.doi.org/10.1017/S0033291702006074.

Kessler, R.C., Berglund, P., Chiu, W.T., Demler, O., Heeringa, S., Hiripi, E., Jin, R., Pennell, B.E., Walters, E.E., Zaslavsky, A. and Zheng, H. (2004). The US National Comorbidity Survey Replication (NCS-R): Design and field, procedures. *International Journal of Methods in Psychiatric Research*, 13(2), 69-92.

Kistner, E.O., and Muller, K.E. (2004). Exact distributions of intraclass correlation and Cronbach's alpha with Gaussian data and general covariance. *Psychometrika*, 69(3), 459-474.

Krätschmer, V., Schied, A. and Zähle, H. (2012). Qualitative and infinitesimal robustness of tail-dependent statistical functionals. *Journal of Multivariate Analysis*, 103(1), 35-47.

Lohr, S.L. (1999). *Sampling: Design and Analysis, 1st Edition*. Pacific Grove, CA: Duxbury Press.

Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, 9(1), 1-19.

Mach, L., Dumais, J. and Robinson, A.A. (2005). A study of the properties of a bootstrap variance estimator under sampling without replacement. In *Arlington (Va): Federal Committee on Statistical Methodology (FCSM) Research Conference*.

Mach, L., Saïdi, A. and Pettapiece, R. (2007). Study of the properties of the Rao-Wu bootstrap variance estimator: What happens when assumptions do not hold. In *Proceedings of the Survey Methods Section, SSC Annual Meeting*.

Martin, M.A. (1992). On the double bootstrap. In *Computing Science and Statistics*, *Statistics of Many Parameters: Curves, Images, Spatial Models*, (Eds., C. Page and R. LePage), 78-78. New York: Springer.

Maydeu-Olivares, A., Coffman, D.L. and Hartmann, W.M. (2007). Asymptotically distribution-free (ADF) interval estimation of coefficient alpha. *Psychological Methods*, 12(2), 157-176.

Paben, S. (1999). Comparison of variance estimation methods for the National Compensation Survey. In *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Patel, P.A., and Bhatt, S. (2016). A model-based estimation of finite population variance under PPS sampling. *Imperial Journal of Interdisciplinary Research*, 2(4).

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review/Revue Internationale de Statistique,* 317-337.

Rao, J.N.K., and Wu, C.F.J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83(401), 231-241.

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18, 2, 209-217. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1992002/article/14486-eng.pdf.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Springer Series in Statistics. Model Assisted Survey Sampling.* New York: US.

SAS Institute (2010). *SAS/STAT User's Guide: Version 9.2.* Cary, NC.

Sen, A.R. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 5, 119-127.

Sen, P.K. (1995). The Hájek asymptotics for finite population sampling and their ramifications. *Kybernetika*, 31(3), 251-68.

Sheehan, D.V., Harnett-Sheehan, K. and Raj, B.A. (1996). The measurement of disability. *International Clinical Psychopharmacology*, 11(3), 89-95.

Sunderland, M., Hobbs, M.J., Anderson, T.M. and Andrews, G. (2012). Psychological distress across the lifespan: Examining age-related item bias in the Kessler 6 Psychological Distress Scale. *International Psychogeriatrics*, 24(2), 231-242.

Swain, A.K.P.C., and Mishra, G. (1994). Estimation of finite population variance under unequal probability sampling. *Sankhyā: The Indian Journal of Statistics, Series B,* 374-388.

Van Zyl, J.M., Neudecker, H. and Nel, D.G. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika*, 65(3), 271-280.

Wolter, K. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

Yates, F., and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society B*, 15, 253-261.

Yuan, K.-H., Guarnaccia, C.A. and Hayslip, B. (2003). A study of the distribution of sample coefficient alpha with the Hopkins symptom checklist: Bootstrap versus asymptotics. *Educational and Psychological Measurement*, 63(1), 5-23.