RTI Press

Methods Report

ISSN 2378-7813

February 2020

Better Coverage Intervals for Estimators from a Complex Sample Survey

Phillip S. Kott

RTI INTERNATIONAL

RTI Press publication MR-0041-2002

RTI International is an independent, nonprofit research organization dedicated to improving the human condition. The RTI Press mission is to disseminate information about RTI research, analytic tools, and technical expertise to a national and international audience. RTI Press publications are peer-reviewed by at least two independent substantive experts and one or more Press editors.

Suggested Citation

Kott, Phillip S. (2020). *Better Coverage Intervals for Estimators from a Complex Sample Survey*. RTI Press Publication No. MR-0041-2002. Research Triangle Park, NC: RTI Press. https://doi.org/10.3768/rtipress.2020.mr.0041.2002

This publication is part of the RTI Press Methods Report series..

RTI International
3040 East Cornwallis Road
PO Box 12194
Research Triangle Park, NC
27709-2194 USA

Tel: +1.919.541.6000
E-mail: rtipress@rti.org
Website: www.rti.org

Contents

About the Author

Phillip S. Kott, PhD, is a senior research statistician in RTI International's Center of Excellence for Complex Data Analysis.

RTI Press Associate Editor

Darryl Creel

Abstract

Coverage intervals for a parameter estimate computed using complex survey data are often constructed by assuming the parameter estimate has an asymptotically normal distribution and the measure of the estimator's variance is roughly chi-squared. The size of the sample and the nature of the parameter being estimated render this conventional "Wald" methodology dubious in many applications.
I developed a revised method of coverage-interval construction that "speeds up the asymptotics" by incorporating an estimated measure of skewness. I discuss how skewness-adjusted intervals can be computed for ratios, differences between domain means, and regression coefficients.

Acknowledgments

Several reviewers made valuable comments which improved the quality of the presentation.

Introduction

Statisticians are interested in estimating intervals likely to contain a parameter. Wald intervals are most commonly used for this purpose. A hypothesis test for the location of the parameter can be conducted using Wald intervals.

Suppose $\hat{\theta}$ is a nearly (i.e., asymptotically) unbiased estimator for a parameter θ estimated from a complex probability sample. The one-sided Wald intervals at the α (e.g., 95%) level for θ are

$$\theta \le \hat{\theta} + \Phi^{-1}(\alpha)\sqrt{v} \text{ and } \theta \ge \hat{\theta} - \Phi^{-1}(\alpha)\sqrt{v}\,, \tag{1}$$

where v is a good estimator for V, the variance of $\hat{\theta}$, under either probability-sampling theory or a reasonable model, and $\Phi(.)$ is the cumulative distribution function of a standard normal distribution. It is well-known that when the sample size is large enough, both inequalities hold for roughly α percent of samples drawn using the same sampling design as the probability survey. That is because the random variable $\hat{\theta}$ is asymptotically normal under mild conditions, and those conditions are assumed to hold.

A symmetric two-sided α percent Wald interval easily derivable from Equation 1 is

$$\hat{\theta} - \Phi^{-1}([1+\alpha]/2)\sqrt{v} \le \theta \le \hat{\theta} + \Phi^{-1}([1+\alpha]/2)\sqrt{v}\,.$$

In this paper, we will focus on one-sided intervals because creating a symmetric two-sided interval from two one-sided intervals is easily done, as just demonstrated. For similar reasons, we will not discuss hypothesis tests derived from coverage intervals.

Often, the sample size in an application will not be nearly large enough for a one-sided Wald interval to contain ("cover") θ with the frequency suggested by the asymptotic theory. We will use the term "coverage interval" here rather than "confidence interval" because one rarely has confidence that the true value of θ falls within the designated interval at least α percent of the time across repeated realizations of the random variable $\hat{\theta}$ (as it would were $\hat{\theta}$ normally distributed and $\Phi(.)$ replaced with the appropriate Student's *t*-distribution).

Kott and Liu (2010) proposed using skewness-adjusted one-sided coverage intervals in place of the Wald intervals to "speed up the asymptotics" (i.e., be roughly correct for samples that are not very large). The next section describes those intervals. The section after that looks at intervals based on a stratified simple random sample and a stratified multistage sample with attention to intervals for a ratio and for the difference between two domain means. The next section shows how skewness adjustments can affect estimates produced from a stratified cluster sample. It is followed by a section providing additional comments on a range of topics, from coverage intervals for regression coefficients, to potentially useful approximations when constructing skewness-adjusted intervals is impractical, to intervals based on estimates computed with calibrated weights. Finally, the paper offers some concluding remarks.

Much of the research into the impact of slow asymptotic normality on coverage has concentrated on proportions either estimated from an independent and identically distributed (*iid*) sample (e.g., Clopper & Pearson, 1934; Hall, 1982; Newcombe, 1998; Brown, Cai, & Dasgupta, 2001; Cai, 2005) or a complex probability sample (e.g., Korn & Graubard, 1998; Kott & Liu, 2009; Franco, Little, Lewis, & Slud, 2014). Kott, Andersson, & Nerman (2001) demonstrated the close relationship between the two-sided version of the Anderson-Nerman interval and the Wilson (score) coverage interval for a proportion (Andersson & Nerman, 2000). Kott (2017) showed the relationship between the Wilson interval and the logistic-transformation approach to creating coverage intervals for proportions (described in Liu & Kott, 2009, and elsewhere).

Here, we will look at more-general estimators computed from complex probability samples; in particular, we focus on estimators for ratios, differences between domain means, and regression coefficients. Critical to this endeavor will be estimating the third central moment of $\hat{\theta}$. We will use probability sampling (design-based) theory in the investigation. Analogous conclusions using models-based assumptions are straightforward.

Skewness-Adjusted Intervals

Kott and Liu (2010) propose the following α-percent one-sided skewness-adjusted coverage intervals for a parameter θ estimated from a complex probability sample by $\hat{\theta}$:

$$\theta \le \hat{\theta} + \delta + \sqrt{z^2 v + \delta^2} \text{ and } \theta \ge \hat{\theta} + \delta - \sqrt{z^2 v + \delta^2}, \quad (2)$$

where

$$\delta = \frac{1}{6}(1 - z^2)\frac{m_3}{v} + \frac{z^2}{2}b, \quad (3)$$

$z = \Phi^{-1}(\alpha)$, m_3 is a nearly unbiased estimator for the third central moment of $\hat{\theta}$: $M_3 = E[(\hat{\theta} - \theta)^3]$ and b is a nearly unbiased estimator for $B = \frac{E[v(\hat{\theta} - \theta)]}{v}$.

The $\left(\frac{z^2}{2}\right)b$ term on the right-hand side of Equation (3) derives from replacing v in the asymptotically normal Wald pivotal, $(\hat{\theta} - \theta)/\sqrt{v}$, with the more asymptotically efficient $v_b = v - b(\hat{\theta} - \theta)$, and then solving the resulting quadratic inequalities for $\hat{\theta}$. Andersson and Nerman (2000) suggested this after observing that the variance of $v_\beta = v - \beta(\hat{\theta} - \theta)$, a conceptual estimator for V, is minimized when $\beta = B$.

The remainder of the right-hand side of Equation 3 has the opposite sign of $\left(\frac{z^2}{2}\right)b$ when $z > 1$. It comes from an Edgeworth expansion of $\hat{\theta}$, which is Kott and Liu's contribution to the intervals in Equation 2. A similar result for estimators based on *iid* samples can be found in Abramovitch and Singh (1985). Note that although a nonzero δ in Equation 2 expands the size of Kott and Liu's skewness-adjusted coverage interval slightly, its primary impact is to move the interval, which can be in either a positive or negative direction.

If $b \approx m_3 / v$, which is true in many contexts (as we shall see in the next section), then, Kott and Liu noted,

$$\delta \approx \left(\frac{1}{6} + \frac{z^2}{3}\right)\frac{m_3}{v}. \quad (4)$$

When Equation 4 holds, the coverage intervals in Equation 2 can be expressed as

$$\theta \le \hat{\theta} + \left(\frac{1}{6} + \frac{z^2}{3}\right)\hat{\tau}\sqrt{v} + z\sqrt{v}\sqrt{1 + \delta^2/(z^2 v)} \text{ and}$$

$$\theta \ge \hat{\theta} + \left(\frac{1}{6} + \frac{z^2}{3}\right)\hat{\tau}\sqrt{v} - z\sqrt{v}\sqrt{1 + \delta^2/(z^2 v)},$$

or

$$\theta \le \hat{\theta} + \left\{\left(\frac{1}{6} + \frac{z^2}{3}\right)\hat{\tau} + z\left[1 + \frac{1}{z^2}\left(\frac{1}{6} + \frac{z^2}{3}\right)^2 \hat{\tau}^2\right]^{1/2}\right\}\sqrt{v} \text{ and}$$

$$\theta \ge \hat{\theta} + \left\{\left(\frac{1}{6} + \frac{z^2}{3}\right)\hat{\tau} - z\left[1 + \frac{1}{z^2}\left(\frac{1}{6} + \frac{z^2}{3}\right)^2 \hat{\tau}^2\right]^{1/2}\right\}\sqrt{v}, \quad (5)$$

where $\hat{\tau} = m_3 / v^{3/2}$ is the estimated skewness for $\hat{\theta}$ and $\tau = M_3 / V^{3/2}$ is the skewness measure $\hat{\tau}$ is estimating.

Kott and Liu conjecture that $|\tau|$ should be less than 1 for their skewness-adjusted intervals to be effective. For Wald coverage intervals, τ needs to be close to 0. Cochran (1977, p. 42) suggests $|\tau|$ should be less than 0.2. (Cochran's original suggestion is for simple random sampling. We expand it here.)

The skewness, τ, of an estimator tends to decrease in absolute value as the sample size increases (the central limits theorem tells us that it converges to zero as the sample size grows arbitrarily large). For an estimated proportion, p, under either simple random sampling with replacement or an *iid* model, $m_3 = p(1-p)(1-2p)/[(n-1)(n-2)]$, $b = (1-2p)/(n-1)$, and $\hat{\tau} \approx [(1-2p)/np(1-p)]^{1/2}$.

When v is not too close to zero, the following modification on Equation 5 drops terms of a smaller asymptotic order (rendering $\hat{\tau}^2 \approx 0$):

$$\theta \le \hat{\theta} + \left\{\left(\frac{1}{6} + \frac{z^2}{3}\right)\hat{\tau} + z\right\}\sqrt{v} \text{ and}$$

$$\theta \ge \hat{\theta} + \left\{\left(\frac{1}{6} + \frac{z^2}{3}\right)\hat{\tau} + z\right\}\sqrt{v},$$

or

$$\theta \le \hat{\theta} + \left(\frac{1}{6} + \frac{z^2}{3}\right)b + z\sqrt{v} \text{ and}$$

$$\theta \ge \hat{\theta} + \left(\frac{1}{6} + \frac{z^2}{3}\right)b + z\sqrt{v}. \quad (6)$$

These are the one-sided Wald intervals shifted by $\left(\frac{1}{6} + \frac{z^2}{3}\right)\hat{\tau}\sqrt{v} = \left(\frac{1}{6} + \frac{z^2}{3}\right)b$.

Two Common Complex Survey Frameworks

Stratified Simple Random Sampling

In a population sample-survey design, the population U is divided H into mutually exclusive strata, where U_h denotes the population of stratum h, h=1,…, H. Each stratum h contains N_h units. Let S_h be a simple random sample of n_h units selected without replacement from U_h, and $S = \cup_{h\in H} S_h$.

Next, suppose we are interested in constructing one-sided coverage intervals for a finite-population total or mean based on a stratified simple random sample. The former can be expressed as $T_y = \Sigma^H \sum_{k\in U_h} y_k$, where y_k is the variable of interest for element k. The corresponding population mean is $\bar{Y} = \dfrac{T_y}{N} = \Sigma^H W_h \bar{Y}_h$, where $N = \Sigma^H N_h$, $W_h = \dfrac{N_h}{N}$, and $\bar{Y}_h = N_h^{-1} \Sigma_{U_h} y_k$. An unbiased estimator for the finite-population total T_y using probability sampling theory is $\hat{T}_y = Nt_y$, where $t_y = \bar{y} = \Sigma^H W_h \bar{y}_h$ and $\bar{y}_h = n_h^{-1} \Sigma_{S_h} y_k$.

Kott and Liu (2010) showed that when every stratum has at least three sampled members (i.e., $n_h \geq 3$), one can construct one-sided coverage intervals for $\bar{Y}$ based on probability-sampling theory by setting

$$v = \Sigma_{h=1}^H W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{\sum_{k\in S_h} (y_k - \bar{y}_h)^2}{n_h(n_h - 1)},$$

$$m_3 = \Sigma_{h=1}^H W_h^3 \left(1 - \frac{n_h}{N_h}\right)\left(1 - \frac{2n_h}{N_h}\right) \frac{\sum_{k\in S_h} (y_k - \bar{y}_h)^3}{n_h(n_h - 1)(n_h - 2)}, \text{ and}$$

$$b = \frac{\Sigma_{h=1}^H W_h^3 \left(1 - \frac{n_h}{N_h}\right)^2 \frac{\sum_{k\in S_h} (y_k - \bar{y}_h)^3}{n_h(n_h - 1)(n_h - 2)}}{v} \tag{7}$$

in Equations 2 and 3. The first two equalities in Equation 7 provide unbiased estimators for the second and third central moments of $\theta = \bar{y}$, and b is a consistent estimator for B under mild conditions we assume to hold (e.g., that those central moments exist). When all the $2n_h / N_h$ are small enough to be ignored, $b \approx m_3 / v$, and Equation 3 can be approximated by Equation 4. Surprisingly, when $2n_h / N_h = 1$, stratum h has no impact on m_3. When $2n_h < N_h$, the impact of stratum h on m_3 is in the opposite direction of $\sum_{k\in S_h} (y_k - \bar{y}_h)^3$.

It is important to realize that when one n_h is less than 3, Equation 7 becomes useless. Useful approximations for m_3 and b become necessary. We discuss some in a later section ("Some Simple Approximations"), although there are other possibilities beyond those discussed.

The ratio of two totals, T_x / T_z, can be estimated in a consistent manner using data from a without-replacement stratified simple random sample by $\hat{t}_{x/z} = \hat{t}_x / \hat{t}_z$. In other words, the difference between $\hat{t}_{x/z}$ and T_x / T_z tends to zero in probability as the sample size grows arbitrarily large under mild conditions (i.e., the population and sample design are such that the relative mean squared errors of $\hat{t}_x$, $\hat{t}_z$, $\hat{t}_{x^2}$, $\hat{t}_{z^2}$, and $\hat{t}_{xz}$ are all $\mathrm{O}(1/n)$, where $n = \sum^H n_h$, while T_z is positive and converges asymptotically to a positive value). The variance and third central moment of $\hat{t}_{x/z}$ can be estimated using Equation 7 with each y_k replaced by the linearized term $e_k = [x_k - (\hat{t}_{x/z})z_k] / \hat{t}_z$, which is asymptotically indistinguishable from $u_k = [x_k - (t_{x/z})z_k] / \hat{t}_z$. Observe that $\hat{t}_{x/z} - t_{x/z} = \Sigma^H W_h \bar{u}_h$.

A ratio estimator of special interest is the estimator of a domain mean. If $d_k = 1$ for an element in the domain and 0 otherwise, then the estimated mean of y-values in the domain has the form $\hat{t}_{x/z} = \hat{t}_x / \hat{t}_z$, where $z_k = d_k$ and $x_k = d_k y_k$.

A Stratified Multistage Sample

Consider now constructing a coverage interval for a population mean based on stratified multistage sample when a nearly unbiased estimator for that parameter can be put in the form

$$\hat{t} = \sum_{h=1}^{H} \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{t}_{hi}, \tag{8}$$

where there are n_h primary sampling units (PSUs) in stratum h, and each $\hat{t}_{hi}$ for a PSU i in stratum h is a nearly unbiased estimator for the same value. We will make the common (but often inaccurate) assumption that the PSUs were selected randomly but with replacement, while any subsampling was done using probability sampling principles.

A univariate component of an estimated linear regression coefficient can also be put into the form of

https://doi.org/10.3768/rtipress.2020.mr.0041.2002

Equation 8. We focus now on the difference between two domain means estimated using data from the same sample, S. Each element in S had a value y_k and a sampling weight w_k attached to it, so the estimated different in domains means can be expressed as follows:

$$\frac{\sum_{k\in S} w_k y_k d_k^{(1)}}{\sum_{k\in S} w_k d_k^{(1)}} - \frac{\sum_{k\in S} w_k y_k d_k^{(2)}}{\sum_{k\in S} w_k d_k^{(2)}} = \bar{y}_{(1)} - \bar{y}_{(2)},$$

where $d_k^{(a)} = 1$ when k is in Domain a and 0 otherwise. Here:

$$\hat{t}_{hi} \approx u_{hi} = \sum\nolimits_{k\in S_{hi}} w_k y_k \left(\frac{d_k^{(1)}}{\hat{N}_1} - \frac{d_k^{(2)}}{\hat{N}_2} \right),$$

where S_{hi} is the set of sampled elements in PSU i of stratum h, and $\hat{N}_a = \sum\nolimits_{q\in S} w_q d_q^{(a)}$ is the estimated population size of Domain a, that is, N_a. Observe that the

$$\hat{t}_{hi} \approx u_{hi} = \sum\nolimits_{k\in S_{hi}} w_k y_k \left(\frac{d_k^{(1)}}{N_1} - \frac{d_k^{(2)}}{N_2} \right)$$

are independent under probability sampling theory for PSUs in the same stratum (recall we are assuming with-replacement sampling in the first stage of sample selection).

When all $n_h \geq 3$, the following equalities can be used in Equations 2 and 3:

$$v = \sum_{h=1}^{H} \frac{N_h^2}{n_h} \sum_{i=1}^{n_h} \frac{\left(e_{hi} - \bar{e}_h\right)^2}{(n_h - 1)},\quad m_3 = \sum_{h=1}^{H} \frac{N_h^3}{n_h} \sum_{i=1}^{n_h} \frac{(e_{hi} - \bar{e}_h)^3}{(n_h - 1)(n_h - 2)},$$

$$\text{and } b = \frac{m_3}{v}, \tag{9}$$

where each e_{hi} has the following linearized expression:

$$e_{hi} = n_h \sum\nolimits_{k\in S_{hi}} w_k \left(\frac{d_k^{(1)}}{\hat{N}_1}[y_k - \bar{y}_{(1)}] - \frac{d_k^{(2)}}{\hat{N}_2}[y_k - \bar{y}_{(2)}] \right).$$

We ignore finite-population correction when comparing domain means because an analyst is usually interested in whether there is an underlying process causing the domain means to be different, not the actual difference between the means in the finite population. Strictly speaking, this assumes models are generating the domain models, but, following a probability-sampling framework, those domains are vaguely specified.

An Example

In this section, we will look at computing one-sided coverage intervals for two sets of parameters for the MU284 population from Särndal, Swensson, & Wretman (1992), available at http://lib.stat.cmu.edu/datasets/mu284. The population consists of 284 Swedish administrative municipalities separated into 50 clusters with eight strata. We collapse the final two strata, creating seven strata total. We divide the population into two domains. The 26 municipalities with a 1985 population of more than 64,000 are in Domain 1, and the remaining 258 municipalities are in Domain 2. We are interested in constructing coverage intervals for (1) the arithmetic average across municipalities in 1985 of the municipal taxation per person within each domain and (2) the fraction of municipalities within each domain with more tax receipts than 9 million kronor per 1,000 persons in 1985. We are also interested in constructing coverage intervals for the differences between the domains.

We suppose a cluster sample of three clusters per stratum ($n_h = 3$) are selected from the MU284 population via simple random sampling with replacement. Letting y_k be either the tax revenue per person in municipality k or a (0/1) indicator of whether that ratio is greater than 9 million kronor per 1,000 persons, we define

$$e_{hi} = N_h \sum_{k\in S_{hi}} \left(\frac{d_k^{(a)}}{\hat{N}_1}[y_k - \bar{y}_{(a)}] \right) \text{ for Domain } a \text{ (}a = 1 \text{ or } 2\text{)},$$

and

$$e_{hi} = N_h \sum_{k\in S_{hi}} \left(\frac{d_k^{(1)}}{\hat{N}_1}[y_k - \bar{y}_{(1)}] - \frac{d_k^{(2)}}{\hat{N}_2}[y_k - \bar{y}_{(2)}] \right)$$

for the difference between the domains, where N_h is the number of clusters in stratum h; S_{hi} is the sample of municipalities in cluster i of stratum h (in this example, S_{hi} is every municipality in the cluster); $d_k^{(a)} = 1$ when municipality k is in Domain a, and 0 otherwise; $\bar{y}_{(a)}$ is the estimated mean of the y–values in Domain a; and $\hat{N}_a$ is the estimated number of municipalities in Domain a.

For constructing coverage intervals in this example, we replace v, m_3, and $\hat{\tau}$ in Equation 9 by what they estimate:

$$V = \sum_{h=1}^{7} N_h^2 \frac{\sum_{i=1}^{N_h} (e_{hi} - \bar{E}_h)^2}{n_h (N_n - 1)},$$

$$M_3 = \sum_{h=1}^{7} N_h^3 \frac{N_h \sum_{i=1}^{N_h} (e_{hi} - \bar{E}_h)^3}{n_h^2 (N_h - 1)(N_h - 2)}, \text{ and} \qquad (10)$$

$$\tau = M_3 / V^{3/2},$$

where $\bar{E}_h = \sum_{i=1}^{N_h} e_{hi} / N_h$. By using these replacements, we produce coverage intervals around the respective estimates close to what the average from an infinite number of simulations would produce.

Table 1 compares one-sided 95% and 99% Wald coverage intervals to skewness-adjusted coverage intervals computed with Equation 10 replacing 7. The bounds in the table are also the bounds of two-sided 90% and 98% coverage intervals, respectively. The table displays the estimands (targets) for context. In practice, the intervals are computed from the sample rather than the population and are added to estimates, which are likewise computed from the sample.

The symmetric Wald and asymmetric skewness-adjusted intervals in Table 1 tend to be closer to each other in Domain 2 than in Domain 1. The larger sample size in Domain 2 reduces the impact of skewness adjustment. The sizes of the coverage intervals for the differences tend to be dominated by the smaller Domain 1 samples.

Additional Comments

Coverage Intervals for a Regression Coefficient

Suppose $\boldsymbol{\beta}$ is a vector of regression coefficients of y_k on $\mathbf{x}_k = (x_{k1,\ldots,}x_{kJ})^T$. Each component j of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ can be expressed or approximated as

$$\hat{\beta}_j - \beta_j \approx \sum_{k \in S} w_k \boldsymbol{g}_j{}^T \boldsymbol{x}_k [y_k - f(\boldsymbol{x}_k{}^T \hat{\boldsymbol{\beta}})],$$

so that

$$e_{hi} = n_h \sum_{k \in S_{hi}} w_k \boldsymbol{g}_j{}^T \boldsymbol{x}_k [y_k - f(\boldsymbol{x}_k{}^T \hat{\boldsymbol{\beta}})], \qquad (11)$$

where $f(\theta) = \theta$ for linear regression and $f(\theta) = 1/[1 + \exp(-\theta)]$ for logistic regression, and $\boldsymbol{g}_j$ is the j^{th} row of the matrix

$$\mathbf{A} = \left[\sum_{a \in S} w_a f'\left(\boldsymbol{x}_a{}^T \hat{\boldsymbol{\beta}}\right) \boldsymbol{x}_a \boldsymbol{x}_a^T \right]^{-1}.$$

We assume matrix $\mathbf{A}$ to exist while the components of $N\mathbf{A}$ have finite limits as the population of size N grows arbitrarily large. Strictly speaking, the estimator for a logistic regression coefficient is determined by solving the weighted estimating equation

$$\sum_{k \in S} w_k \boldsymbol{x}_k \left\{ y_k - \left[1 + exp(\boldsymbol{x}_k{}^T \hat{\boldsymbol{\beta}}) \right]^{-1} \right\} = \mathbf{0},$$

which cannot be expressed in the form of Equation (8). Nevertheless, the variance and third

Table 1. Idealized coverage intervals

Estimates	Target	τ	One-sided 95% intervals (compared with estimate)			One-sided 99% intervals (compared with estimate)		
				Skewness-adjusted			Skewness-adjusted	
			Wald	Lower	Upper	Wald	Lower	Upper
Municipal tax per person (in 1,000s)								
Domain 1	6.88	-0.60	± 1.19	-1.66	0.72	± 1.68	-2.54	0.82
Domain 2	6.90	-0.09	± 0.19	-0.20	0.18	± 0.27	-0.29	0.25
Difference	-0.02	-0.62	± 1.15	-1.61	0.68	± 1.62	-2.48	0.76
Fraction above 9 million per thousand persons								
Domain 1	0.15	0.66	± 0.17	-0.10	0.24	± 0.24	-0.10	0.37
Domain 2	0.02*	0.48	± 0.02	-0.01	0.03	± 0.03	-0.02*	0.04
Difference	0.13	0.69	± 0.16	-0.09	0.24	± 0.23	-0.10	0.37

* Computed to another digit, the target is 0.023, while the lower bound is -0.018.

Notes: The average lower Wald interval is bound from below by the estimate minus the unsigned value in the table. The average lower skewness-adjusted interval is bound from below by the estimate plus the lower value in the table. The average upper Wald interval and skewness-adjusted interval are bound from above; the bounds are the estimate plus the unsigned value or plus the upper value, respectively.

central moment of the estimator can be measured by implementing Equations 9 and 11.

Although the e_{hi} in Equation 11 are not independent within strata, each is almost equal to a u_{hi}, in which the $\hat{\beta}$ on the right-hand size of Equation 11 is replaced by $\boldsymbol{\beta}$ and the components of N$\mathbf{g}_j$ by their asymptotic limits. The u_{hi} are independent in a probability-sampling sense under the assumption that the first-stage sample is drawn with replacement.

Observe that we can only create coverage intervals for one regression coefficient at a time. A coverage interval for a univariate linear combination or regression coefficients can be created using the same principles. To test a vector of r regression coefficients, one may need to use a conservative r-dimensional Bonferroni box rather than a Wald-based coverage ellipsoid.

Some Simple Approximations

There is a practical problem in computing m_3, and consequently $\hat{\tau}$ using either Equation 7 or 9: there is no available software routine to do so. Even if there were or a statistician wanted to program the equations, there may not be three PSUs in every stratum. Unlike collapsing strata for variance estimation, the direction of the potential bias of $\hat{\tau}$ can be positive or negative when the population means of the strata being combined are different (the population means are either the expected value of the y_k in each stratum in Equation 7 or the expected values of the u_{hi} corresponding to the e_{hi} for a particular h in Equation 9). Consequently, strata collapsed together should have equal (or nearly equal) expected population means.

A key to skewness-adjusted coverage intervals, especially when finite-population correction can be ignored, is the estimated value $b = m_3 / v$. From the last section, assuming a large sample, the value of this term for the difference between proportions estimated for two distinct domains from a simple random sample is approximately

$$b = \frac{m_3}{v} \approx \frac{p_1(1-p_1)(1-2p_1)/n_1^{\,2} - p_2(1-p_2)(1-p_2)/n_2^{\,2}}{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2},$$

where p_a is the estimated proportion in Domain a based on n_a sampled elements being in Domain a.

When $p_1 = p_2$, this collapses to

$$b \approx (1-2p_1)\left(\frac{1}{n_1} - \frac{1}{n_2}\right).$$

These results are similar (for large n_a) to what we would get looking at the estimate of the two proportions as coming from independent simple random samples (which they are *conditionally* given their respective realized domain sample sizes): the estimated variance of the difference equals the *sum* of the individual estimated variances, whereas the estimated third central moment of the difference is the *difference* between the individual estimated third central moments.

To many, these results might suggest that when assessing the difference between proportions in two distinct domains estimated using a complex probability sample, one simply multiplies the domain sample sizes above by their respective design effects. Such a practice is not recommended, however, for two reasons. One, the design effect captures the effect of clustering and stratification on the variance of an estimator, not on its third central moment. Two, the unequal weighting effect, another component of the design effect, is not the same for the third central moment of an estimator and its variance.

A wiser procedure might be to estimate $B = M_3 / V$ for an estimated proportion $p = \sum_{k\in S} w_k y_k \big/ \sum_{k\in S} w_k$, where p estimates the fraction of the population with $y_k = 1$ rather than 0, with

$$b_{simple} = \frac{\sum_{k\in S} w_k^{\,3}}{\sum_{k\in S} w_k \sum_{k\in S} w_k^{\,2}}(1-2p), \tag{12}$$

and then inserting $\hat{\tau}_{simple} = b_{simple} \big/ \sqrt{v}$ into Equation 5 or 6. This estimate ignores the impact of stratification and clustering on b. When estimating the proportion in a domain, the S in Equation 12 and in the definition of p are replaced by the subset of the sample within the domain. For the difference between two domain means, $\hat{\tau}_{simple} = b_{simple} \big/ \sqrt{v}$ with

$$b_{simple} = \frac{p_1(1-p_1)(1-2p_1)/\tilde{n}_1^{\,2} - p_2(1-p_2)/\tilde{n}_2^{\,2}}{p_1(1-p_1)/n_1^{*} + p_2(1-p_2)/n_2^{*}}$$

where

$$\tilde{n}_a^{\,2} = \frac{\left(\sum_{S_a} w_k\right)^3}{\sum_{S_a} w_k^{\,3}}, \text{ and } n_a^{*} = \frac{\left(\sum_{S_a} w_k\right)^2}{\sum_{S_a} w_k^{\,2}} \tag{13}$$

For a more-general population or domain mean of a *y–variable*, one can replace $\hat{\tau}$ in Equation 5 or 6 with $\hat{\tau}_{simple} = b_{simple} / \sqrt{v}$, where

$$b_{simple} = \frac{\sum_{k \in S} w_k^3 (y_k - \bar{y})^3}{\sum_{k \in S} w_k \sum_{k \in S} w_k^2 (y_k - \bar{y})^2} \quad (14)$$

and

$$\bar{y} = \sum_S w_k y_k / \sum_S w_k.$$

Table 2 assesses Equations 12 through 14 with the examples from Table 1, replacing terms of the form $\sum_{k \in S_a} w_k^{\,b} z_k$ with $\sum_{k \in U_a} w_k^{\,b-1} z_k$. The approximations in the table may not be perfect, but they are closer to the true $B = M_3 / V$ than 0, the value implied when Wald intervals are constructed.

Table 2. Approximating *B*

Estimates	*B*	B_{simple} (Equation [14])	B_{simple} (Equation [12] or [13])
Municipal tax per person (in 1,000s)			
Domain 1	-0.436	-0.486	
Domain 2	-0.011	-0.020	
Difference	-0.436	-0.477	
Fraction above 9 million per thousand persons			
Domain 1	0.068	0.073	0.067
Domain 2	0.006	0.008	0.010
Difference	0.069	0.072	0.066

Calibration Weighting and the Jackknife

So far, we have implicitly assumed either that w_k is the inverse of the probability of selected element k into the sample or that the n_h elements in stratum h were selected with equal probability in each stratum. We have ignored the effect of coverage error and unit nonresponse on the respondent sample ultimately used in estimation.

Under simple random sampling without replacement, it is common to assume that the list from which the sample has been drawn is complete and without duplication and that elements in the same stratum are either equally likely to respond (treating response as a phase of probability sampling) or have a common means whether or not they respond. This allows one to use the formulae in Equations 2 through 6, replacing the sample and stratum samples with the analogous respondent samples; n_h is redefined accordingly.

For multistage sampling calibration, weighting can be used either to adjust for nonresponse or undercoverage (Kott, 2006) or simply to reduce the standard error of the estimates (Deville & Särndal, 1992). When w_k is a calibrated weight, the e_{hi} in Equation (8) are weighted sums of calibration residuals. For example, when estimating a total, the e_{hi} are the weighted totals within PSU *hi* of

$$e_k = y_k - \mathbf{z}_k^{\,T} \left[\sum_{a \in S} d_a \xi'(\mathbf{z}_a^{\,T}\mathbf{q}) \mathbf{z}_a \mathbf{z}_a^{\,T} \right]^{-1} \sum_{a \in S} d_a \xi'(\mathbf{z}_a^{\,T}\mathbf{q}) \mathbf{z}_a y_a,$$

where $\mathbf{z}_k$ is the vector of calibration variables, d_k is the weight of k before calibration, the calibrated weight of k is $w_k = d_k \xi(\mathbf{z}_k^{\,T}\mathbf{q})$, $\xi(.)$ is the weight-adjustment function connecting the weight before calibration to the calibrated weights (e.g., $\xi(\theta)$ can be $1+\theta$, $\exp(\theta)$, or $1+\exp(\theta)$), and $\mathbf{q}$ is chosen so that $\sum_{k \in S} w_k \xi(\mathbf{z}_z^{\,T}\mathbf{q}) = T_{\mathbf{z}}$ is the vector of population totals for the components of $\mathbf{z}_k$ or an estimate for that vector calculated using information outside the respondent sample (e.g., from the frame; a larger, more-efficient sample than *S*; or the full sample before nonresponse).

Calibration weighting often removes much of the impact of stratification and clustering from an estimated mean. For example, calibration by region can reduce the impact of stratification by geographical units, whereas calibration by race and ethnicity can reduce the impact of clustering within neighborhoods. As a result, estimating the skewness of an estimated proportion or mean using Equations 10 or 11 may not be unreasonable, although it would often be better to replace the y_k with a calibrated residual. Moreover, when estimating domain means, the impact of calibration weighting, like that of stratification and clustering, is diminished, except for any impact caused by increased variability of the weights themselves. Calibration weighting makes the use of Equations 12, 13, or 14 within the intervals in Equation 5 or 6 more viable.

If calibrated jackknife weights have been constructed to compute v for an estimator $\hat{t}$, then these weights can also be used in estimating the third central moment of $\hat{t}$:

$$m_{3(J)} = \sum_{h=1}^{H} \frac{(n_h - 1)^2}{n_h (n_h - 2)} \sum_{i=1}^{n_h} (\hat{t} - \hat{t}_{(hi)})^3 ,$$

where $\hat{t}$ is computed with calibrated weights, $\hat{t}_{(hi)}$ is computed with the calibrated weights for the stratum *h*, PSU *i* jackknife replicate, and

$$\left[1 - \frac{1}{(n_h - 1)^2}\right]^{-1} = \frac{(n_h - 1)^2}{n_h (n_h - 2)} .$$

Noting that

$$\sum_{i=1}^{n_h} (\hat{t} - \hat{t}_{(hi)}) \approx \hat{t}_{hi} - \frac{1}{n_h - 1} \sum_{\substack{j=1 \\ j \neq i}}^{n_h} \hat{t}_{hj} ,$$

$\dfrac{(n_h - 1)^2}{n_h (n_h - 2)}$ in $m_{3(J)}$ is analogous to

$$\left[1 + \frac{1}{(n_h - 1)}\right]^{-1} = \frac{(n_h - 1)}{n_h}$$

in the jackknife variance estimator:

$$v_{(J)} = \sum_{h=1}^{H} \frac{(n_h - 1)}{n_h} \sum_{i-1}^{n_h} (\hat{t} - \hat{t}_{(hi)})^2 .$$

Estimating the variance and third central moment of an estimator whose weights incorporate more than one calibration step can be difficult using the linearization methods applied previously. Computing jackknife measures is much simpler.

Some Concluding Remarks

Statisticians who base their inferences on probability samples often claim that those inferences are model-free. One deviation from that claim is the assumption that the sample under study is large enough that an estimator based on the sample is approximately normally distributed. This assumption, which is often only made implicitly, justifies constructing Wald intervals for the parameter being estimated and conducting hypothesis tests based on those intervals.

This asymptotic assumption may not be justified in practice. Many alternatives to Wald intervals have been suggested in the literature for when the estimand is a proportion based on a simple random sample, Kott and Liu (2010) provide the means for constructing skewness-adjusted coverage intervals for estimators other than proportions. Moreover, these estimators can be based on more complex designs than simple random sampling. . We follow up on those intervals, describing interval estimates for ratios, differences between domain means, and regression coefficients based on either a stratified simple random sample or a stratified multistage probability sample employing with-replacement sampling at the first stage.

The practical stumbling block to using skewness-adjusted intervals is that the statisticians must estimate the third central moment of their estimator. This cannot be done if any stratum contains less than three PSUs. Several approximations have been offered here instead.

Simulations assessing the viability of those approximations are still needed, as are simulations of the skewness-adjusted intervals themselves with realistic data (a few simulations are presented in Kott and Liu [2009], for stratified simple random samples). Here, the estimator's population-based third central moment was used in place of its estimate in the examples. In practice, a statistician will usually have to estimate a third central moment. The impact of this estimation, which may not be very stable, needs to be evaluated.

It may be that using skewness-adjusted intervals or their approximations is roughly equivalent to using Wald intervals for a particular application. If that is the case, then assuming asymptotic normality may be justified. The skewness-adjusted intervals discussed here can be used to make that justification.

References

Abramovitch, L., & Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Annals of Statistics*, *13*(1), 116–132. https://doi.org/10.1214/aos/1176346580

Andersson, P., & Nerman, O. (2000). A balanced adjusted confidence interval procedure applied to finite population sampling. Presented at the Second International Conference of Establishment Surveys, Buffalo, NY.

Brown, L. D., Cai, T. T., & Dasgupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, *16*(2), 101–133. https://doi.org/10.1214/ss/1009213286

Cai, T. T. (2005). One-sided confidence intervals in discrete distributions. *Journal of Statistical Planning and Inference*, *131*, 63–88.

Clopper, C., & Pearson, E. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, *26*(4), 404–413. https://doi.org/10.1093/biomet/26.4.404

Cochran, W.G. (1977). *Sampling techniques (3rd ed.)*. New York, NY: John Wiley & Sons.

Deville, J.-C., & Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, *87*(418), 376–382. https://doi.org/10.1080/01621459.1992.10475217

Franco, C., Little, R. Lewis, T., & Slud, E. (2014). Coverage properties of confidence intervals for proportions in complex sample surveys. *Proceedings of the ASA Survey Research Methods Section.* http://www.math.umd.edu/~slud/s770/SurveyConfidenceIntervals/JSSAM-2017-065-FINAL.pdf

Hall, P. (1982). Improving the normal approximation when constructing one-sided confidence intervals for binomial or Poisson parameters. *Biometrika*, *69*(3), 647–652. https://doi.org/10.1093/biomet/69.3.647

Korn, E., & Graubard, B. (1998). Confidence intervals for proportions with small expected number of positive counts estimated from survey data. *Survey Methodology*, *24*, 193–201.

Kott, P. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, *32*, 133–142.

Kott, P. (2017). A note on Wilson coverage intervals for proportions estimated from complex samples. *Survey Methodology*, *43*, 235–240.

Kott, P., Andersson, P., & Nerman, O. (2001). Two-sided coverage intervals for small proportions based on survey data. Presented at Federal Committee on Statistical Methodology Research Conference, Washington, DC. https://pdfs.semanticscholar.org/ff0a/0a1af716681771d49f555c35304b002d9ff3.pdf

Kott, P., & Liu, Y. (2009). One-sided coverage intervals for a proportion estimated from a stratified simple random sample. *International Statistical Review*, *77*(2), 251–265. https://doi.org/10.1111/j.1751-5823.2009.00081.x

Kott, P., & Liu, Y. (2010). Speeding up the asymptotics when constructing one-sided coverage intervals with survey data. *Metron*, *68*(2), 137–151. https://doi.org/10.1007/BF03263531

Liu, Y., & Kott, P. (2009). Evaluating alternative one-sided coverage intervals for a proportion. *Journal of Official Statistics*, *25*, 569–588.

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine*, *17*(8), 857–872. https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E

Särndal, C., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. New York, NY: Springer.

www.rti.org/rtipress

RTI Press publication MR-0041-2002