

The Degrees of Freedom of a Variance Estimator in a Probability Sample

Phillip S. Kott



RTI Press publication MR-0043-2008

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, P. S. (2020). *The Degrees of Freedom of a Variance Estimator in a Probability Sample*. RTI Press Publication No. MR-0043-2008. Research Triangle Park, NC: RTI Press. <https://doi.org/10.3768/rtipress.2020.mr.0043.2008>

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

©2020 RTI International. RTI International is a trade name of Research Triangle Institute. RTI and the RTI logo are U.S. registered trademarks of Research Triangle Institute.



This work is distributed under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 license (CC BY-NC-ND), a copy of which is available at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

<https://doi.org/10.3768/rtipress.2020.mr.0043.2008>

www.rti.org/rtipress

Contents

About the Author	i
Acknowledgments	ii
Abstract	ii
Introduction	1
Background	1
The Population and Estimands of Interest	3
Results	4
Discussion	6
References	7

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

Cynthia Bland

Abstract

Inferences from probability-sampling theory (more commonly called “design-based sampling theory”) often rely on the asymptotic normality of nearly unbiased estimators. When constructing a two-sided confidence interval for a mean, the ad hoc practice of determining the degrees of freedom of a probability-sampling variance estimator by subtracting the number of its variance strata from the number of variance primary sampling units (PSUs) can be justified by making usually untenable assumptions about the PSUs. We will investigate the effectiveness of this conventional and an alternative method for determining the effective degrees of freedom of a probability-sampling variance estimator under a stratified cluster sample.

Introduction

Many surveys are based on stratified multi-stage probability samples. When analyzing probability-sample data, one usually needs to take the complex nature of the sample design into account because sampled elements from the same strata or cluster can be correlated. Moreover, the item values of a sampled element may be related to its probability of selection. Probability-sample theory provides a convenient way of analyzing probability-sample data in a rigorous manner. (The more commonly used but misleading term “design-based theory” is not employed here because models are often used, at least implicitly, in motivating the probability-sampling design.)

One common failing of probability-sampling theory is its reliance on an ad hoc method for determining the degrees of freedom (*dof*) of its variance estimators. Ingram et al. (2018, p. 5) describe this method in their recommendation for analyses of National Center of Health Statistics record-level survey data: “the recommended number of degrees of freedom . . . is the number of PSUs minus the number of sampling strata.” This is the default *dof* measure employed in the probability-sampling software language SUDAAN (RTI International, 2012) and in the probability-sampling procedures in SAS (SAS Institute Inc., 2015).

Here we investigate using an alternative method of *dof* calculation derived by assuming a contrafactual model for the sample data. We focus on estimating means, although generalizations to other statistics will be briefly discussed in the final section. We will look at full-population and domain-mean estimates computed from a stratified cluster samples drawn from a particular population and evaluate how well the conventional and the model-based alternative *dof* measures fare at constructing two-sided confidence intervals (CIs) for the estimands.

We first provide some technical background. We then describe our population and conduct the evaluation. Finally, we offer some caveats, conclusions, and speculations.

Background

Suppose we are estimating the mean of a finite population of item values using data from a complex probability sample. For simplicity, we will assume that there is no nonresponse and that the sample has been drawn from a complete population using a stratified multistage sample. In particular, the population U consisting of M elements has been divided in mutually exclusive and exhaustive primary sampling units (PSUs), and the PSUs themselves have been grouped into H mutually exclusive and exhaustive strata. A probability sample of $n_h \geq 2$ PSUs is selected in each stratum h ($h = 1, \dots, H$), and a probability sample of elements is selected within each sampled PSU. By “probability sample,” we mean every PSU in stratum h has a known positive probability of selection (or expected number of selections) into the sample of PSUs, and every element in each sampled PSU i has a known positive probability of selection into the sample of elements from that PSU. The selection of elements from PSUs can have any number of stages.

Let S_{1h} denote the sample of the n_h PSUs selected from stratum h , S_1 the union of $n = \sum^H n_h$ sampled PSUs, and S the sample of elements drawn from within those n PSUs. The subscript 1 in S_{1h} and S_1 refers to the first stage of sampling in what could be many stages of sample selection.

Let $k \in U$ denote an element in the population and π_k its probability of selection into S : the product of the probability of selection of the PSU (call it i) containing k into S_1 and the probability of k 's selection into the element sample drawn from within PSU i . That element sample is denoted by S_i .

The sampling (or design) weight of element k is denoted $d_k = 1/\pi_k$. A nearly unbiased estimator for a population mean $T_y = \sum_{k \in U} y_k / M$ under probability-sampling theory is

$$t_y = \frac{\sum_{k \in S} d_k y_k}{\sum_{k \in S} d_k}. \quad (1)$$

An estimator like t_y is *nearly unbiased* when its relative bias tends to zero as the number of sampled

PSUs in S_1 grows arbitrarily large under mild conditions we assume to hold. See, for example, Isaki and Fuller (1982) for a discussion of those conditions.

We will also assume the mild conditions under which the contribution of the bias of t_y to its mean squared error tends to zero as the number of sampled PSUs in S_1 grows arbitrarily large. The reader should understand that, in what follows, the mild conditions we assume to hold remain in effect.

One conventionally estimates the variance/mean squared error of t_y under probability-sampling theory by treating the random selection of PSUs within each stratum as if it were done with replacement, even though that is rarely the case in practice. We follow that conventional treatment here because it greatly simplifies variance estimation.

The conventional probability-sampling variance estimator for t_y is

$$v = \sum_{h=1}^H \frac{n_h}{n_h-1} \left[\sum_{i \in S_{1h}} e_{i+}^2 - \frac{1}{n_h} \left(\sum_{i \in S_{1h}} e_{i+} \right)^2 \right], \tag{2}$$

$$\text{where } e_{i+} = \frac{\sum_{k \in S_i} d_k (y_k - t_y)}{\sum_{k \in S} d_k} \tag{3}$$

Observe that e_{i+} is nearly equal to $u_{i+} = \sum_{k \in S_i} d_k (y_k - t_y) / M$, and the u_{i+} are independent random variables under the stratified with-replacement selection of PSUs. “Nearly equal” means that the ratio of the two, e_{i+}/u_{i+} , converges to 1 as the number of PSUs in S_1 grows arbitrary large.

If the u_{i+} were normally distributed and each had the same variance, then v would nearly have a (central) chi-squared distribution with $n - H$ degrees of freedom. (It would be exactly such a distribution if the e_{i+} were replaced by the u_{i+} .) As a result, the statistic $z = (t_y - T_y) / v^{1/2}$, which is used to make inferences about of t_y as an estimate for of T_y , would nearly have a Student t distribution with $n - H$ degrees of freedom.

This is where the conventional practice of determining the degrees of freedom for a statistic derived from a stratified multistage probability sample comes from. Many have noted that even

under the ideal conditions assumed here—a stratified multistage probability sample drawn from a complete population, where there is no nonresponse, at least two PSUs chosen with replacement from each stratum, and a probability sample of elements selected within each sampled PSUs—the conventional practice is rarely justified. See, for example, Valliant and Rust (2010). That is because the PSU-level “residuals” captured in the u_{i+} seldom have the same variance and may not be normally distributed.

The probability-sampling variance estimation in Equation (2) captures any loss in efficiency of the estimator t_y due to the individual y_k within each PSU being correlated (whether there are additional stages of sampling) as well as any gains in efficiency due to stratification. The variance estimate also captures any loss efficiency due to the d_k being unequal. Were all the y_k in U normal, independent, and identically distributed, T_y could be efficiently estimated by a simple mean of the m y_k -values in S , the variance of which would be efficiently estimated by

$$v_I = \frac{1}{m(m-1)} \left[\sum_{k \in S} y_k^2 - \frac{1}{m} \left(\sum_{k \in S} y_k \right)^2 \right],$$

which has a chi-squared distribution with $m - 1$ degrees of freedom. Heuristically, in addition to the v usually being larger than v_I , the smaller $n - H$ nominal degrees of freedom in v (compared with $m-1$) is the price one pays for using t_y and v for making inferences about T_y . That price may not be steep enough, as we shall see.

The degrees of freedom of a chi-squared variance estimator is 2 divided by its relative variance. Accordingly, it seems reasonable to employ a Satterthwaite approximation (1946) to measure the *effective degrees of freedom (edof)* for a non-chi-squared variance estimator, like v ; namely, $edof(v) = 2/relvar(v)$. Kott (1994) proposed computing such *edof* for v by treating each y_k as it is were normal, independent, and identically distributed. Assuming this contrafactual model makes computing the relative variance of v straightforward, and truly provides a measure of the price one pays for using v in place of v_I when drawing inferences about a population mean.

Here we will look at estimates of domain means as well as population means. A domain mean has the form $T_{y,g} = \sum_{k \in U} g_k y_k / \sum_{k \in U} g_k$, where $g_k = 1$ when element k is in the domain and 0 otherwise. Its nearly-unbiased estimator under mild conditions is $t_{y,g} = \sum_{k \in S} d_k g_k y_k / \sum_{k \in S} d_k g_k$, and its conventional probability-sampling variance estimator has the form of Equation (2) with the e_{i+} in Equation (3) replaced by

$$e_{i+} = \frac{\sum_{k \in S_i} d_k g_k (y_k - t_{y,g})}{\sum_{k \in S} d_k g_k}$$

A population mean and its estimator can be expressed as a domain mean and its estimator by letting $g_k = 1$ for every $k \in U$.

Setting $w_k = d_k g_k$, the *edof* measure for the probability-sampling variance estimator of a domain mean derived under the contrafactual assumption that the y_k in the domain are normal, independent, and identically distributed, is

$$edof(v) = \frac{\left(\sum_{k \in S} w_k^2 \right)^2}{\sum_{h=1}^H \sum_{i \in S_{1h}} \left(\sum_{k \in S_i} w_k^2 \right)^2 + \sum_{h=1}^H \frac{1}{(n_h - 1)^2} \left\{ \left[\sum_{i \in S_{1h}} \left(\sum_{k \in S_i} w_k^2 \right) \right]^2 - \sum_{i \in S_{1h}} \left(\sum_{k \in S_i} w_k^2 \right)^2 \right\}}, \quad (4)$$

(Kott, 1994).

A little algebra reveals that if $\sum_{k \in S_i} w_k^2$ for each PSU were the same, the right-hand side of Equation (4) would collapse to

$$edof(v) = \frac{n}{1 + \frac{1}{n} \sum_{h=1}^H \frac{n_h}{(n_h - 1)}}$$

which is close to $n - H$. Exact equality obtains when all $n_h = n/H$. Observe that the contrafactual *edof* in Equation (4) tends to decrease as the variability of the $\sum_{k \in S_i} w_k^2$ across PSUs, whether due to the variability of the w_k^2 or the sizes of the PSUs, increases.

The Population and Estimands of Interest

We now will compare how well the contrafactual *edof* in Equation (4) and the nominal degrees do at measuring “the true degrees of freedom” of a probability sampling variance estimator given a stratified cluster sample drawn for the MU281 population (Ohlsson, n.d., based on data from Appendix B of Särndal et al., 1992). Determining the true degrees of freedom is elusive. Instead, we define the *empirical degrees of freedom* for a two-sided q percent CI as the degrees of freedom of a chi-squared distribution with the same q percent CI as that produced by the sample data.

The MU281 population is a collection of Swedish municipalities with three outliers removed. The population is divided into 50 clusters from 8 regional strata. After combining strata 7 and 8, the clusters and strata are further divided (based on 1975 population counts) into 108 PSUs within 20 design strata. A stratified cluster sample of 4 PSUs is sampled with

replacement from each design stratum via simple random sampling. All the municipalities in each sampled PSU are the elements of the sample. By sampling with replacement, the variance estimator in Equation (1) should produce nearly unbiased estimates of the true mean squared error.

This sampling process is repeated 10,000 times twice. The first time to estimate the full population means of five variables. The second time to estimate domain mean for the same five variables. The five variables (with their per municipality estimands in parentheses) are

- RMT85, 1985 revenue from municipal taxes in millions of kronor (189.15)
- ME84, 1984 number of municipal employees (1,381.26)
- P85, 1985 population in thousands (25.0285)

- REV84, 1984 real estate values in millions of kronor (2694.83)
- RMT/P85, a binary variable equal to 100 when the ratio RTM85/P85 is greater than 8, 0 otherwise (18.5053, which is the estimate the percent of municipalities with $RMT/P85 \geq 8$).

All exhibit some degree of intra-stratum and intra-cluster correlation. These can be measured indirectly by treating the population as an equally weighted stratified with-replacement cluster sample and comparing alternative standard-error estimates for the means assuming: no strata or clustering, clustering but no strata, strata but no clustering, and both strata and clustering. This was done, but the results are not shown.

The domain used in the analysis is the set of those municipalities where at least half of the municipal council were Social Democrats in 1982. Roughly 37.01 percent of municipalities in the MU281 population are in this domain. The domain estimands are 196.5, 1449, 25.4808, 2751.33, and 25, respectively. (Analyses of other domain means gave comparable results and are not shown here.)

Results

For each of our simulated samples, the sampling weights in Equation (1) are $d_k = N_{[k]}/4$, where k now denotes a single selection of an element, and $N_{[k]}$ is the number of PSUs (before sampling) in the stratum containing the PSU and thus the element. The sampling weights vary because the number of PSUs in each stratum before sampling vary.

The set S is now the set of sampled element selections, and S_{1h} and S_I are defined analogously. Under with-replacement sampling, a PSU can be selected more than once; each selection is a separate member of S_I , leading to potentially multiple memberships for its component elements in S .

We denote the results computed from each $r = 1, \dots, 10,000$ simulated sample with the subscript (r), and define the following empirical summary statistics (recall a full-population mean can be expressed as a domain mean):

The empirical relative bias of $t_{y,g}$ is
$$ERB(t_{y,g}) = \frac{\sum_{r=1}^{10,000} \frac{(t_{y,g(r)} - T_{y,g})}{T_{y,g}}}{10,000} \times 100\%.$$

The empirical variance of $t_{y,g}$ is
$$EV(t_{y,g}) = \frac{\sum_{r=1}^{10,000} t_{y,g(r)}^2 - \frac{1}{10,000} \left(\sum_{r=1}^{10,000} t_{y,g(r)} \right)^2}{9,999}.$$

The empirical relative bias of v is
$$ERB(v) = \frac{\sum_{r=1}^{10,000} \frac{(v(r) - EV(t_{y,g}))}{T_{y,g}}}{10,000} \times 100\%.$$

The empirical variance of v is
$$EV(v) = \frac{\sum_{r=1}^{10,000} v(r)^2 - \frac{1}{10,000} \left(\sum_{r=1}^{10,000} v(r) \right)^2}{9,999}.$$

The empirical relative variance of v is
$$ERV(v) = \frac{EV(v)}{\left(\sum_{r=1}^{10,000} \frac{v(r)}{10,000} \right)^2}.$$

The empirical Satterthwaite DOF of v is
$$ESD(v) = 2/ERV(v).$$

In the previous expressions, v is the probability-sampling variance of $t_{y,g}$.

The nominal *dof* of v is 60 (80 PSUs minus 20 strata), whereas the predicted *edof* are computed using Equation (4), which varies from sample to sample, but it is same for all population variables depending on whether a full-population of domain mean is being estimated. Table 1 displays the summary statistics for the *edof*(v). All summaries in this paper were computed using the PROC MEANS procedure within the SAS software (SAS Institute Inc., 2015). Estimated probability-

sampling means and variances were computed using the DESCRPT procedure in the SUDAAN language (RTI International, 2012) before being summarized using SAS.

Tables 2 and 3 display many of the empirical results of the simulation including those of the two-sided 95th and 90th percentiles of the test statistic $z = |t_{y(r)} - T_y|/\nu^{1/2}$, denoted by z_{95} and z_{90} . Each of these percentiles has its own empirical *dof* (i.e., the degrees of freedom that would produce that value of z were the u_{i+} normal, independent, and identically distributed). These are denoted $ED_{95}(\nu)$ and $ED_{90}(\nu)$ respectively (they were computed using the TINV function in SAS).

For comparison purposes, the z_{95} and z_{90} values for the nominal 60 degrees of freedom (rounding to two decimal places) are 2.00 and 1.67. For 31.2 degrees of freedom (the mean of the full-population predicted *edof* in Table 1), they are 2.04 and 1.70. For 20.6 the mean of the domain prediction *edof*), they are

2.08 and 1.72. Under normality (infinite degrees of freedom), z_{95} is 1.96, and z_{90} is 1.645.

Tables 2 and 3 also show that, unlike their predictions, the empirical degrees of freedom, whether measured by *ESD*, ED_{95} , or ED_{90} depend on the variable. That is likely because each variable has a different amount of intra-stratum and intra-PSU correlation. The predictions in Table 1, while stable, are not that good. Still, they are clearly better than the nominal measure (60) or infinity. Surprisingly, although developed to predict the Satterthwaite degrees of freedom, which are only supposed to be approximations, the *edof* appear to be closer to the empirical confidence-interval degrees of freedom (ED_{95} and ED_{90}). Also, surprising, is that although the *edof* from the domain means is smaller than that of the population mean, empirically the *dof* measures for domain means are usually smaller—but are sometimes higher—than their full-population analogs, depending on the variable and measure.

Table 1. Summary statistics for the (contrafactual) effective degrees of freedom measure (*edof*(ν))

	MEAN	MEDIAN	MIN	P5	Q1	Q3	P95	MAX
Population	31.21	31.20	25.31	28.46	30.09	32.39	33.96	37.56
Domain	20.58	20.53	14.29	17.44	19.21	21.85	23.94	24.92

Table 2. Some empirical results for the population variables (*edof*(ν) \approx 31.2)

Variable	$ERB(t_y)$	$ERB(\nu)$	$ESD(\nu)$	z_{95}	$ED_{95}(\nu)$	z_{90}	$ED_{90}(\nu)$
RMT85	0.05	0.96	44.60	2.03	36.1	1.69	37.5
ME84	-0.07	-0.47	23.61	2.10	18.0	1.73	18.3
P85	-0.17	-2.21	33.35	2.12	15.8	1.72	22.5
REV84	0.09	1.23	19.87	2.06	23.9	1.71	23.5
PRT/P85	-0.06	0.78	13.97	2.19	11.5	1.77	13.2

Notes: *EDX* = the empirical degrees of freedom of the value zX at the X th percentile; *ERB* = the empirical relative bias; *ESD* = the empirical Satterthwaite (approximate) degrees of freedom; zX = the X percentile of the absolute difference between the estimator and estimand divided by the estimated standard error of the estimator.

Table 3. Some empirical results for the domain variables (*edof*(ν) \approx 20.6)

Variable	$ERB(t_{y,g})$	$ERB(\nu)$	$ESD(\nu)$	z_{95}	$ED_{95}(\nu)$	z_{90}	$ED_{90}(\nu)$
RMT85	0.07	-0.10	26.84	2.10	18.0	1.74	17.5
ME84	0.12	0.19	26.32	2.09	20.1	1.73	18.0
P85	0.07	-2.01	27.46	2.09	19.9	1.73	18.2
REV84	-0.05	-0.91	32.39	2.10	18.0	1.72	22.0
PRT/P85	-0.04	-1.06	8.14	2.42	6.3	1.89	7.1

Notes: *EDX* = the empirical degrees of freedom of the value zX at the X th percentile; *ERB* = the empirical relative bias; *ESD* = the empirical Satterthwaite (approximate) degrees of freedom; zX = the X percentile of the absolute difference between the estimator and estimand divided by the estimated standard error of the estimator.

There is no obvious explanation why the empirical measures of the degrees of freedom for RMT/P85 were so much lower than for the other four variables investigated. If anything, the estimated population and domain means for this variable were less influenced by stratification and clustering than the other four (not shown). How stratification, clustering, and weight variability affect empirical *dof* is an area for future study.

Discussion

That the predicted *edof* measure produced by Equation (4) is not a complete success should not surprise anyone. It is based on the contrafactual assumption that ν can be treated as a chi-squared distribution and that the variable whose mean is being estimated has no intra-stratum nor intra-cluster correction. None of those assumptions are correct for the variables in the MU281 data set being analyzed. In addition, the Satterthwaite formula, $edof(\nu) = 2/relvar(\nu)$, is only an approximation when ν is not chi-squared. Finally, an estimated mean can itself have a skewness and kurtosis that affect building appropriate CIs around them. These characteristics are extremely difficult to measure with a probability sample. Valliant and Rust (2010) attempt incorporating stratum-level kurtosis terms into a degrees of freedom measure with little success. Nearly unbiased kurtosis estimation requires four sampled PSUs in every stratum. Kott (2020) discusses the potential impact of skewness on CIs, an impact greater on an interval's symmetry than its size, which is why only two-sided intervals were evaluated here. A nearly unbiased skewness measure requires at least three sampled PSUs in every stratum.

Despite Equation (4) failing to predict the empirical *edof*, however they are measured, employing the equation when making statistical inferences is a realistic alternative to subtracting the number of sampled PSUs from the number of strata, which is the conventional practice. The equation provides a useful measure of the *potential* loss in efficiency from using a probability-sampling variance estimator. Although it is based on simple model assumptions, those contrafactual assumptions are more reasonable that

then implicit assumption justifying the conventional measure of the *edof*.

Some speculations about extensions beyond those situations covered in the body of the text are in order. It may often be reasonable to use the contrafactual *edof* measure in Equation (4) when the weights in a weighted estimator like $t_{y,g}$ contain adjustments for nonresponse and coverage and when the design strata or design PSUs in Equation (2) are replaced by variance strata or PSUs (i.e., simplifications used in practice when, for example, there are less than two design PSUs in a design stratum). This assertion assumes that the difference $t_{y,g} - T_{y,g}$ can be put in the form $\sum_{k \in S} w_k e_k$. For example, with linear calibration weighting (Kott, 2015), $e_k = y_k - \mathbf{x}_k (\sum_{j \in S} d_j \mathbf{x}_j \mathbf{x}_j^T)^{-1} \sum_{j \in S} d_j y_j$, where \mathbf{x}_k is a vector of calibration variables, and $w_k = d_k [1 + (\sum_{j \in U} \mathbf{x}_j - \sum_{j \in S} d_j \mathbf{x}_j) (\sum_{j \in S} d_j \mathbf{x}_j \mathbf{x}_j^T)^{-1} \mathbf{x}_k]$. Using Equation (4) is most reasonable under the simple linear model for y_k , $y_k = \mathbf{x}_k^T \boldsymbol{\beta} + \varepsilon_k$, where the ε_k (which are nearly equal to the e_k) are independent and identically distributed normal random variables.

The error in a single regression coefficient can also be put in the form $\sum_{k \in S} w_k e_k$, where the w_k are more complicated expressions than the sampling weights. The w_k for the j^{th} coefficient of a probability-sampling linear regression estimator $\mathbf{b} = (\sum_{j \in S} d_j \mathbf{x}_j \mathbf{x}_j^T)^{-1} \sum_{j \in S} d_j \mathbf{x}_j y_j$ is the j^{th} row of $(\sum_{j \in S} d_j \mathbf{x}_j \mathbf{x}_j^T)^{-1} d_k \mathbf{x}_k$, and $e_k = y_k - \mathbf{x}_k \mathbf{b}$, which is nearly equal to ε_k in the linear model, $y_k = \mathbf{x}_k^T \boldsymbol{\beta} + \varepsilon_k$.

Similarly, the w_k for the j^{th} coefficient of a of a probability-sampling logistic-regression estimator or to any \mathbf{b} that solves an estimating equation like $\sum_{k \in S} d_k \mathbf{x}_k [y_k - f(\mathbf{x}_k^T \mathbf{b})] = 0$, where $f(\cdot)$ is scalar and monotonic, is the j^{th} row of $(\sum_{j \in S} d_j f'(\mathbf{x}_j^T \mathbf{b}) \mathbf{x}_j \mathbf{x}_j^T)^{-1} d_k \mathbf{x}_k$, and $e_k = y_k - f(\mathbf{x}_k^T \mathbf{b})$. This suggests that using Equation (4) as a reasonable, if imperfect, measure of the *edof* for the variance estimator of a regression coefficient, linear or otherwise. Because this measure will usually vary across the coefficients of a regression estimator, using a Bonferroni-adjusted t test (1974), rather than a probability-sampling F test (for examples, see RTI International, 2012, p. 217) may be advisable for (imperfectly) testing a joint hypothesis involving multiple coefficients.

Some readers may be wondering about the *edof* measure for the variance estimator of the estimated difference between two domain means. Expressing the estimated difference between two domain means

as a linear regression coefficient is simple and has been discussed previously in this report.

References

- Ingram, D. D., Malec, D. J., Makuc, D. M., Kruszon-Moran, D., Gindi, R. M., Albert, M., ... Sengupta, M., & the National Center for Health Statistics. (2018). National Center for Health Statistics guidelines for analysis of trends. *Vital and Health Statistics. Series 2, Data Evaluation and Methods Research*, (179), 1–71.
- Isaki, C., & Fuller, W. (1982). Survey design under the regression super-population model. *Journal of the American Statistical Association*, 77(377), 89–96. <https://doi.org/10.1080/01621459.1982.10477770>
- Kott, P. (2015). Calibration weighting in survey sampling. *Wiley Interdisciplinary Reviews: Computational Statistics*, 8(1), 39–53. <https://doi.org/10.1002/wics.1374>
- Kott, P. (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>
- Kott, P. S. (1994). Hypothesis testing of linear regression coefficients with survey data. *Survey Methodology*, 20, 159–164.
- Ohlsson, E. (n.d.). *The MU284 population* [Data set]. <http://lib.stat.cmu.edu/datasets/mu284>
- RTI International. (2012). *SUDAAN language manual, release 11.0*. Research Triangle Park, NC: RTI International.
- Särndal, C.-E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. New York, NY: Springer-Verlag.
- SAS Institute Inc. (2015). *SAS/STAT® 14.1 user's guide*. Cary, NC: SAS Institute Inc.
- Valliant, R., & Rust, K. (2010). Degrees of freedom approximations and rules-of-thumb. *Journal of Official Statistics*, 26, 585–602.

RTI International is an independent, nonprofit research institute dedicated to improving the human condition. We combine scientific rigor and technical expertise in social and laboratory sciences, engineering, and international development to deliver solutions to the critical needs of clients worldwide.

www.rti.org/rtipress

RTI Press publication MR-0043-2008