TWO-SIDED COVERAGE INTERVALS FOR SMALL PROPORTIONS BASED ON SURVEY DATA

Phillip S. Kott, Per Gösta Andersson, and Olle Nerman
National Agricultural Statistics Service, Linköping University,
and Chalmers University of Technology

Abstract

The standard two-sided Wald coverage interval for a small proportion, P, may perversely include negative values. One way to correct this anomaly when analyzing data from a simple random sample is to compute an asymmetric Wilson (or score) coverage interval. This approach has proven not only theoretically satisfying but empirically effective. Some have suggested computing an *ad hoc* Wilson-like coverage interval for P when it is weighted or is estimated with complex sample data. We propose an alternative, theoretically motivated, approach to two-sided coverage-interval construction. In the case where the population P is unweighted and the data from a simple random sample, the coverage interval generated by the new pivotal is asymptotically identical to the Wilson coverage interval. A modest empirical evaluation shows that our coverage intervals are slightly better than those derived from the ad hoc Wilson approach and much better than standard Wald intervals. Better yet is a model-based Wilson approach, but in our study the model was correct.

Key Words: Wald coverage interval, Wilson coverage interval, Asymptotic, Asymmetric.

Background

Statisticians, especially those dealing with biological and environmental data, are increasingly being asked to estimate very small (or very large) proportions. Although it is often not difficult to develop a point estimate for such proportions, statistically defensible coverage-interval determination can be more problematic.

Our focus here will be in constructing a two-sided coverage interval for a population proportion. We consider first the estimation of an unweighted proportion based on a simple random sample employing an idea in use since at least Wilson (1927). Later, we extend the method to an estimator derived from complex sampling data using a recent proposal by Andersson and Nerman (2000).

Let p be an estimated proportion (i.e., $\sum_S y_k/n$, where y_k is either 0 or 1), P be the population proportion p is estimating ($\sum_U y_k/N$), z be the relevant Normal-value (for a two-sided 95% coverage interval, z = 1.96), and n be the size of the sample, S. The size of the population, U, is N, which we will assume is much greater than n, so much so that finite population correction can be ignored.

Assuming p is approximately normal, a two-sided coverage interval for p contains all P values such that

$$|p - P|/\{P(1-P)/n\}^{1/2} \le z$$

or $|p - P| \le z\{P(1-P)/n\}^{1/2}$. Consequently, $(p-P)^2 \le z^2P(1-P)/n$ and $p^2 - 2pP + P^2 \le P(z^2/n) - P^2(z^2/n)$, which implies

$$(1 + z^2/n)P^2 - (2p + z^2/n)P + p^2 \leq 0.$$

The extreme points of which are

$$P = p + (z^2/[2n])(1 - 2p)/(1 + z^2/n) \pm z\downarrow \{[p(1-p)/n] + z^2/(4n^2)\}/(1 + z^2/n). \quad (1)$$

Agresti and Coull (1998) call this the "score" coverage interval and point out it is better at determining coverage than the conventional two-sided Wald interval, $P = p \pm z\downarrow [p(1-p)/(n-1)]$. They go on to argue that the score coverage interval usually exhibits better coverage properties than the exact Clopper-Pearson confidence interval (1934) as well. This is because an exact confidence interval assures that at least the nominal coverage rate will obtain for all values of P. As a result, it can be (and is) conservative for particular values of P. That is why we use the modifier "coverage" instead of "confidence" in the title of this paper. Our goal is to determine coverage intervals for particular values of P.

Additional empirical evidence in favor of the score approach can be found in Newcombe (1998). In the latest version of WesVar (Westat, 2000), the interval in equation (1) is called the "Wilson" coverage interval, terminology Agresti and Coull echo, and we will adopt.

It will prove helpful to drop $O_p(n^{-3/2})$ terms from the right hand side of equation (1) yielding

$$P \approx p + (z^2/[2n])(1 - 2p) \pm z\downarrow \{[p(1-p)/n] + z^2/(4n^2)\}. \quad (2)$$

The $z^2/(4n^2)$ term within the square root sign has *not* been dropped because $p(1-p)$ can be as low as 0.

In this paper, we will restrict our attention to two-sided intervals and assume that $p - P$ is approximately normal. It is well know that the approximate normality of $p - P$ breaks down faster for one-sided intervals than for two-sided intervals.

Observe that the center of the coverage interval for P in equation (2) is only p when $p = \frac{1}{2}$; otherwise, it is somewhere between $\frac{1}{2}$ and p when $p > \frac{1}{2}$ or somewhere between p and $\frac{1}{2}$ when $p < \frac{1}{2}$. Observe further that, when $p = 0$, the coverage interval is $[0, z^2/n]$, and when $p = 1$, the coverage interval is $[1 - z^2/n, 1]$.

Since the randomization variance of p is $v_{srs} = p(1-p)/(n-1)$, equation (2) can be approximated as

$$P \approx p + (z^2/[2n])(1 - 2p) \pm z\downarrow \{v_{srs}^2 + z^2/(4n^2)\}. \quad (3)$$

The conventional Wald coverage interval for a proportion under simple random sampling is derived from the assumption that the *pivotal* statistic, $t = (p-P)/v_{srs}$, is approximately N(0, 1). The Wilson coverage interval replaces v in the pivotal with $V_{srs} = P(1-P)/n$. Observe that as n grows arbitrarily large, equation (3) collapses into the two-sided Wald interval: $P = p \pm zv_{srs}$.

Extension to a Complex Sample Design

Kott and Carr (1997) proposed that equation (1) be generalized to complex samples by replacing n with n*, the *effective* sample size. By definition, $n^* = p(1-p)/v^2$, where v is the estimated randomization variance of p under the complex sample design. S

Kott and Carr's *ad hoc* solution, however, has little theoretical basis. In fact, the reason why Wilson coverage intervals are effective may be that $p - P$ is uncorrelated with $P(1-P)/n$ since the latter is a constant. In contrast, $p - P$ can be correlated with $P(1-P)/n^*$ through n*.

Andersson and Nerman (2000) take an apparently different tack. They suggest v in the pivotal be replaced with $v' = \sqrt{\{ v^2 - (p-P)\, E[v^2(p-P)]/V^2 \}}$ because v'^2 is uncorrelated with $p - P$. It can be shown that v' has less variance than v^2 as well. In practice, Andersson and Nerman point out, $E[v^2(p-P)]/V^2$ must be estimated from the sample. Replacing $E[v^2(p-P)]/V^2$ by a consistent estimator under simple random sampling, $(1-2p)/n$ (see Appendix A), and solving $(p-P)^2/v_{srs}'^2 \le z^2$ for P leads to

$$\begin{aligned} P &= p + (z^2/[2n])(1-2p) \pm z\sqrt{\{v_{srs}^2 + z^2(1-2p)^2/(4n^2)\}} \\ &= p + (z^2/[2n])(1-2p) \pm z\sqrt{\{v_{srs}^2 + z^2(1-4p[1-p])/(4n^2)\}} \\ &= p + (z^2/[2n])(1-2p) \pm z\sqrt{\{v_{srs}^2 + z^2/(4n^2)\}} + O(n^{-3/2}) \\ &\approx p + (z^2/[2n])(1-2p) \pm z\sqrt{\{v_{srs}^2 + z^2/(4n^2)\}}, \end{aligned}$$

which is equation (3) above (we say $z^2(1-4p[1-p])/(4n^2) \approx z^2/(4n^2)$ because the term only matters asymptotically compared to v_{srs}^2 when $p(1-p)$ is very small).

The preceding argument suggests the following generalization of equation (1) allowing a slightly broader definition of P (and p) and a complex sampling design:

$$\begin{aligned} P &\approx p + (z^2/2)e[(p-P)v^2] \pm z\sqrt{\{v^2 + ([z^2/2]e[(p-P)v^2])^2\}} \\ &\approx p + (z^2/[2n'])(1-2p) \pm z\sqrt{\{v^2 + z^2/[2n']^2\}}, \end{aligned} \qquad (4)$$

where $P = (\sum_U x_k y_k) / (\sum_U x_k)$, $p = (\sum_S x_k y_k/\pi_k) / (\sum_S x_k/\pi_k) = \sum_S w_{ik} y_k$, y_k is again either 0 or 1, $x_k \ge 0$, π_k is the selection probability of element k, $w_k = (x_k/\pi_k) / (\sum_S x_j/\pi_j)$, v is a randomization variance estimator for p under the complex design, $e[(p-P)v^2]$ is an estimator for $E[(p-P)v^2]$ with $O_P(1/\sqrt{n})$ relative error, and $n' = (1-2p)v^2/e[(p-P)v^2]$ when $p \ne \frac{1}{2}$, and ∞ otherwise. (Technical note: we assume $v/[p(1-p)] = O_P(1/n)$ to move from the first to the second equation of (4)).

Equation (4) essentially replaces n* in Kott and Carr's *ad hoc* two-sided Wilson interval by n'. Both can be looked at as generalizations of equation (3).

Estimating $E[(p - P)v^2]$

Consider a general stratified, multistage sample, where h (= 1, ..., H) index the strata, and j =(1, ..., n_h) the primary sampling units (PSUs). Let S_{hj} denote the set of sampled elements in PSU j of stratum h, and S_h be the set of PSU in stratum h. Unstratified samples are covered in this general framework when H =1 and single-stage surveys when each PSU j contains a single element.

The linearization variance estimator for p is

$$v_L^2 = \sum_{h=1}^{H} (n_h/[n_h - 1]) \sum_{j \in S_h} [u_{hj} - (\sum_{i \in S_h} u_{hi})/n_h]^2, \qquad (5)$$

where $u_{hj} = \sum_{k \in S_{hj}} w_k(y_k - p)$. For future use, we define $U_{hj} = \sum_{k \in S_{hj}} w_k(y_k - P)$ and $w_{hj} = \sum_{k \in S_{hj}} w_k$.
Technically v_L^2 estimates mean squared error of p because p has a slight (randomization) bias of an estimator for P, but under mild conditions (which we assume the hold) that bias is an asymptotically ignorable component of the mean squared error of p. Formally, $p - P = O_p(1/\sqrt{n_I})$, where $n_I = \sum^H n_h$, while $E(p - P) = O(1/n_I)$.

The linearization variance estimator is consistent under mild conditions when the PSU's within each stratum are sampled *with* replacement. This is because each u_{hj} is approximately equal to $U_{hj} = u_{hj} + w_{hj}(p - P)$, and the U_{hj} can be treated as independent random variables. For a formal discussion of these matters, see Rao and Wu (1985), where it is also shown see that stratified jackknife and balanced-repeated-replication (BRR) variance estimators are asymptotically identical to v_L^2 in equation (5).

We will assume here that the PSU's are either drawn with replacement or that there are so many PSU's in a stratum the distinction between with and without replacement sampling vanishes. An alternative, model-based treatment – where P in the definition of U_{hi} is replaced by a model parameter, and the U_{hi} are assumed independent – is straightforward, but left for the reader. The U_{hj} in a sampled stratum have a common mean, say μ_h. Since $p - P = \sum^H \sum_j U_{hj}$, $E(p - P) = \sum^H n_h\mu_h \approx 0$.

Let $D_{hj} = U_{hj} - \mu_h$. We can rewrite $p - P$ (approximately) as $\sum^H \sum_j D_{hj}$, which is the sum of independent random variables each with mean zero. Since $u_{hj} = D_{hj} + \mu_h + w_{hj}(P - p)$, the variance estimator v_L can be written as

$$v_L^2 \approx \sum_{h=1}^{H} (n_h/[n_h - 1]) \sum_{j \in S_h} [D_{hj} - (\sum_{i \in S_h} D_{hi})/n_h]^2,$$

$$= \sum \{ \sum D_{hj}^2 - (\sum_{i \neq j \in S_h} D_{hj}D_{hi})^2/(n_h - 1)\},$$

It is now easy to see that $E[(p - P)v_L^2] \approx \sum^H \sum_j E(D_{hj}^3)$, which in turn has the consistent estimator:

$$e_L = \sum_{h=1}^{H} \{n_h^2/([n_h - 1][n_h - 2])\} \sum_{j \in S_h} [u_{hj} - (\sum_{i \in S_h} u_{hi})/n_h]^3, \quad (6)$$

Too see why e_L is consistent, note first that nothing is lost asymptotically be replacing each u_{hi} by D_{hi} in equation (6). Then, observe that

$$E\{\sum_j [D_{hj} - (\sum_i D_{hi})/n_h]^3 = \sum_j E(D_{hj}^3)(1 - 3/n_h + 3/n_h^2 - 1/n_h^2)$$

$$= \sum_j E(D_{hj}^3)(n_h - 2)(n_h - 1)/n_h^2.$$

It is important to realize that e_L can not be computed when $n_h = 2$. This means that under quite-common sampling designs having only two sampled PSU's per stratum, e_L can not be calculated. That is one reason to ignore the strata and estimate $E[(p - P)v_L^2]$ with

$$e_C = \{n_I^2/([n_I - 1][n_I - 2])\} \sum_{h=1}^{H} \sum_{j \in S_h} u_{hj}^3. \quad (7)$$

This collapsed-stratum method is consistent when the μ_h are all equal to zero. Unlike the analogously-defined collapsed-stratum variance estimator (which looks like equation (6) with u_{hj}^2 replacing u_{hj}^3 and $n_I/(n_I - 1)$ replacing $n_I^2/([n_I - 1][n_I - 2])$), the bias of e_C when the μ_h are not all zero can be positive or negative. Observe that $u_{hj} \approx U_{hj}$, and $E(U_{hj}^3) = E(D_{hj}^3) + 3\mu_h E(D_{hj}^2) + \mu_h^3$. The latter two terms are the source for the potential bias in e_C. Recall that $\sum^H n_h \mu_h \approx 0$. Thus, strata for which μ_h and μ_h^3 are both positive will be to some extent off set by strata for which both are negative.

Not only is the asymptotic bias in e_C likely to be negligible, but this estimator will be considerably more efficient than e_L when there are many strata and few sampled PSU's per stratum.

When all $n_h > 2$, it is possible to develop and consistent estimator for $E[(p - P)v_J^2]$ to use with a stratified jackknife. That is beyond the scope of this endeavor. Since balanced repeated replication requires n_h to be 2, no BRR analogue to e_J exists.

A Modest Empirical Study

Although the focus in the text has been on estimating proportion under a complex sample design, the modest empirical study we are about to described assumes a simple random sample. What complicates matters is that the population P is not a simple proportion but a weighted proportion. In particular, $P = (\sum_U x_k y_k)/(\sum_U x_k)$, where the x_k are generated from a Chi-square distribution and the y_k from a Bernoulli *independently* of the x_k. The randomization-based estimator for P is simply $p = (\sum_S x_k y_k)/(\sum_S x_k) = \sum_S w_k y_k$. (Under the model generating the data, a better estimator would be $\rho = \sum_S y_k/n$, but we will ignore this inconvenient fact.)

We evaluated 50,000 sets of samples generated from the superpopulation creating P. In fact, we measured p as an estimator of the superpopulation parameter rather that P. That distinction is trivial when U is much larger than S, and we assume it is. We considered eight scenarios. We let P (actually the superpopulation parameter) be 0.1 or 0.05, n be 100 or 200, and the x_k have either 1 or 5 degrees of freedom.

We computed the fraction of times a nominal 95% interval actually covered P. We determined the coverage intervals four different ways:

Two-sided Wald Interval: $P \approx p \pm zv$
Ad-Hoc Wilson (AHW) Interval: $P \approx p + (z^2/[2n^*])(1 - 2p) \pm z\downarrow \{v^2 + z^2/([2n^*]^2)\}$
The Andersson-Nerman (AN) Interval: $P \approx p + (z^2/[2n'])(1 - 2p) \pm z\downarrow \{v^2 + z^2/[2n']^2\}$
Model-based Wilson (MbW) Interval: $P \approx p + (z^2/[2n_M])(1 - 2p) \pm z\downarrow \{p(1 - p)/n_M + z^2/[2n_M]^2\}$,

where $z = 1.96$, $v^2 = \sum_S w_k^2 (y_k - p)^2$, $n^* = p(1 - p)/v$, and $n' = (1 - 2p)v^2/e[(p - P)v^2]$ when $p \neq ½$, n otherwise, and $n_M = 1 / \sum_S w_k^2$. The last method applies a Wilson-like derivation using the model-based variance estimator $v_M^2 = P(1-P)/n_M$ in place of v^2.

The results are given in Table 1. The Andersson-Nerman intervals cover much better than the two-sided Wald intervals and slightly better than the *ad-hoc* Wilsons (essentially the interval in Kott and Carr, although there is a slight difference analogous to the change from equation (1) to (2)). It does not do as well as we would have liked, however. The model-based Wilson intervals have the best coverages.

Some thought reveals what has happened. The model-based Wilson interval uses a variance estimator in the denominator of it is pivotal that has absolutely no variance under the model. By contrast, the Anderson-Nerman interval uses a variance estimator that is asymptotically identical to

$$v'^2 = v^2 - (p - P)\, E[v^2(p - P)]/V^2.$$

Appendix B show that v'^2 has less variance under the simple Bernoulli model than the Wald's v^2, but that variance is positive when any of the w_k are unequal to the others.

As the variance of the variance estimator decreases, the coverage rate increases. This suggests we treat the pivotal, $(p - P)/v'$, as if it had a Student's t distribution with $D = 2/\{Var_M(v'^2) / [E_M(v'^2)]^2\}$ degrees of freedom. In practice, D needs to be estimated, most obviously (see equation (B.1) of Appendix B) by

$$d = 2p(1 - p)\, (\sum w_i^2)^2 / \{(1 - 2p)^2\, [\sum w_k^4 - \{\sum w_k^3\}^2 / \sum w_k^2]\}. \quad (8)$$

A t-adjusted version of Andersson-Nerman (tAN) coverage interval replaces the z-score in the AN interval by the corresponding value from a Student's t distribution with d degrees of freedom. The results in Table 1 shows that t-adjusted Andersson-Nerman intervals can be conservative (i.e., have coverages over 95%), especially when the x_k are Chi-squared with 1 degree of freedom. It appears the lower the average value of d, the wider and more conservative the coverage interval.

Of course, we can get better coverages, easier, and with shorten average lengths (measured on the table as percentages of the unit interval) using model-based Wilson intervals. The problem with those, however, is that they depend on the y_k being independent of the x_k. If that were not true or nearly true, we suspect a model-based Wilson interval would cover considerably less well. Although the t-adjustment in a t-adjusted Andersson-Nerman interval relies on the same model as a model-based Wilson, the former may prove preferable because it collapses into the model-free Anderson-Nerman interval as d grows large. Consequently, at the very least, the d calculation serves to warn us when the sample size is not large enough to justify using a purely randomization-based AN interval.

More empirical work is needed to see how well the t-adjusted Andersson-Nerman and model-based Wilson methods work when the model fails. Our suspicion is that the former is less sensitive to model failure. Nevertheless, when the sample size is small, in particular when t^2/n' is not ignorably small, or $p = 0$ (so that $d = 0$), one has no choice but to use the purely model-based method.

It should be noted that for a two-sided hypothesis test of whether $P = P_0$, the t-adjusted Andersson-Nerman method can be made a bit sharper by replacing the p in equation (8) by P_0, which is more consistent with the null hypothesis.

It is possible to estimate effective degrees of freedom and use the t-adjusted Andersson-Nerman coverage interval when p (and v') is based on a sample drawn using a complex design, although we will not discuss how here. When doing so, it may be advisable to compute d values under a number of alternative models, ones that allow for stratum and/or clustering effects. Randomization-based estimation of D (and thus d) is likely to prove impractical. See Kott (1994).

Finally, a word about one-sided coverages. The lower (upper) coverage displayed in Table 1 is the percentage of times a 97.5% one-sided coverage interval is wholly below (above) P. The Wald method is notoriously bad at determining one-sided intervals. Even our best two-sided Wald (when $P = 0.1$, $n = 200$, and the x_k are Chi-squared with 5 degrees of freedom) is linked to one-sided intervals that are either much too conservative (an "upper coverage" of 1% instead of 2.5%) or too lax (a lower coverage of 6.1%). The model-based Wilson method is not supposed to have good one-sided coverages, and it doesn't. In contrast to the Wald, it tends to be conservative for lower coverages and lax for upper coverages. This is the situation noted in the literature for the Wilson method when p is unweighted. The results are a bit more muddied for the *ad-hoc* Wilson and the two Andersson-Nerman intervals. Nevertheless, it does *not* seem to be advisable to use these methods for one-sided interval construction.

Table 1. Comparing the Actual Coverages of Some Two-Sided 95% Coverage Intervals

When the x_k are Chi-square with 1 degree of freedom

	Wald	AHW	MbW	AN	tAN*		Wald	AHW	MbW	AN	tAN
						P=0.05					
			n=100						n=200		
Coverage	75.6	81.6	94.6	83.9	98.1		83.2	88.3	95.7	89.3	96.9
Lower Coverage	24.2	15.1	0.6	14.4	0.6		16.5	8.8	0.0	7.7	0.7
Upper Coverage	0.1	3.3	4.8	1.7	1.3		0.3	2.9	4.3	3.0	2.4
Average Length	11.8	12.8	16.0	15.3	33.1		9.2	9.8	10.9	11.1	15.7
Average d					10						14
						P=0.10					
			n=100						n=200		
Coverage	84.0	88.6	95.6	91.3	97.0		88.5	91.6	95.6	93.0	95.4
Lower Coverage	15.5	8.3	0.0	7.4	1.8		10.9	5.7	0.6	4.6	2.4
Upper Coverage	0.5	3.1	4.4	1.3	1.2		0.6	2.7	3.8	2.4	2.2
Average Length	17.9	18.3	20.2	22.1	26.8		13.4	13.6	14.3	15.4	16.7
Average d					24						34

When the x_k are Chi-square with 5 degrees of freedom

	Wald	AHW	MbW	AN	tAN		Wald	AHW	MbW	AN	tAN
						P=0.05					
			n=100						n=200		
Coverage	86.2	91.4	95.6	92.0	95.7		90.5	93.4	95.2	937	94.8
Lower Coverage	13.4	5.3	0.5	4.8	1.2		8.9	3.4	0.1	2.8	1.9
Upper Coverage	0.4	3.3	3.9	3.1	3.1		0.6	3.2	3.7	3.5	3.3
Average Length	9.3	10.0	10.5	11.0	12.2		6.9	7.2	7.3	7.6	7.9
Average d					32						56
						P=0.10					
			n=100						n=200		
Coverage	90.6	93.5	95.4	94.6	95.5		92.9	94.4	95.2	94.9	95.2
Lower Coverage	8.6	3.5	1.1	2.7	1.9		6.1	2.8	1.6	2.1	2.9
Upper Coverage	0.8	3.0	3.5	2.6	2.6		1.0	2.8	3.2	2.9	1.9
Average Length	13.3	13.6	13.9	14.8	15.3		9.6	9.7	9.8	10.2	10.3
Average d					78						135

* For those cases where d = 0 (because p =0), we let the coverage interval be the entire unit interval. This only occurred when P = 0.05 and n = 100.

Appendix A: Estimating $E[(p-P)v^2]$ under Simple Random Sampling

In these appendix, we assume $P = \sum_U y_k /N$, N is so large than finite population correction can be ignored, and the sample can be treated as if it were drawn with replacement. Furthermore the variance of $p = \sum_S y_k /n$ is estimated with $v^2 = n^{-1}\{ \sum_S [y_k^2 - (\sum_S y_i)/n]^2\}/(n-1)$.

Let $D_k = y_k - P$, so that $p - P = \sum_S D_k /n$. Since we are assuming the sample was effectively drawn with replacement, the D_k are (effectively) independent and identically distributed random variables.

The variance estimator, v^2, can be rewritten as

$$\begin{aligned} v^2 &= n^{-1}\{ \textstyle\sum_S [D_k^2 - (\sum_S D_i)/n]^2\}/(n-1) \\ &= n^{-2}\{ \textstyle\sum_S D_k^2 - \sum_{S(i,\neq k)} D_i D_k /(n-1)\} \end{aligned}$$

Consequently, $E(v^2) = n^{-2} \sum_S E(D_k^2) = n^{-1} E(D_k^2)$, and $E[(p-P)v^2] = n^{-2} E(D_k^3)$. Since Dh_k has a probability of P of equaling $(1-P)$, and a probability of $(1-P)$ of equaling $-P$,

$$E(D_k^2) = P(1-P)^2 + (1-P)^2P = P(1-P),$$

$$E(D_k^3) = P(1-P)^3 - (1-P)P^3 = P(1-P)(1-2P),$$

and

$$E(D_k^4) = P(1-P)^4 + (1-P)P^4 = P(1-P)(1-3P+3P^2)$$

It is now obvious that $E(v^2) = P(1-P)/n$, as we know, and $E[(p-P)v^2] = P(1-P)(1-2P)/n^2$. Since the former has the unbiased estimator $v_{srs}^2 = p(1-p)/(n-1)$, a consistent estimator for the latter would be $e[(p-P)v^2] = p(1-p)(1-2p)/[n(n-1)]$. Moreover, a consistent estimator for $E[(p-P)v^2]/V^2$ is $(1-2p)/n$.

Appendix B: The Variance of v'^2 Under the Bernouli Model

Letting P and p be weighted, and v be the randomization-based estimator for the variance of p when finite population correction is ignored. The model variance of $v'^2 = v^2 - (p-P)\, E[v^2(p-P)]/V^2$ under the simple Bernouli model is

$$\begin{aligned} Var_M(v'^2) &= Var_M\{ v^2 - (p-P)\, E[v^2(p-P)]/V^2 \} \\ &= Var_M(v^2) - \{E[v^2(p-P)]\}^2/V^2 \\ &\approx Var_M(v^2) - (E[\textstyle\sum \{w_k^3[y_k - p]^3\}])^2/E(v^2), \\ &\approx E_M[\textstyle\sum\{w_k^4([y_k - p]^4 - [P\{1-P\}]^2)\}] - [\sum w_k^3]^2\,[P(1-P)(1-2P)]^2/[\sum w_k^2 P(1-P)] \\ &= P(1-P)(1-4P+4P^2)[\textstyle\sum w_k^4 - \{\sum w_k^3\}^2/\sum w_k^2] \\ &= P(1-P)(1-2P)^2\,[\textstyle\sum w_k^4 - \{\sum w_k^3\}^2/\sum w_k^2]. \end{aligned} \tag{B.1}$$

This value collapses to zero when all the w_k are equal, but is positive otherwise.

Observe that through similar reasoning the model variance of v^2 is $P(1-P)(1-2P)^2 \sum w_k^4$.

References

Agresti, A. and Coull, B.A. (1998). Approximate is Better than "Exact" for Interval Estimation with Binomial Proportions. American Statistician, 52, 119-126.

Anderson. P.G. and Nerman, O (2000). A Balanced Adjusted Confidence Interval Procedure Applied to Finite Population Sampling. Presented at the International Conference of Establishment Surveys, Buffalo, NY.

Clopper, C. J. and Pearson, E. S. (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*, 26, 404-413.

Kott, P.S., (1994). A Hypothesis Test of Linear Regression Coefficients with Survey Data. *Survey Methodology*, **20**, 159-164.

Kott, P.S. and Carr, D.A. (1997). Developing an Estimation Strategy for a Pesticide Data Program. *Journal of Official Statistics*, **13**, 367-383.

Newscombe, R.G. (1998). Two-Sided Confidence Intervals for the Single Proportion: Comparisons of Seven Methods. *Statistics in Medicine*, **17**, 857-872.

Rao, J.N.K. and Wu C.F.J. (1985). Inferences from Stratified Samples: Second-Order Analysis of Three Methods for Nonlinear Statistics. *Journal of the American Statistical Association* **80**, 620-630.

Wetstat (2000). *A User's Guide to WesVar, Version 4.* Rockville, MD.

Wilson, E.B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association*, **22**, 209, 212.