A Bayesian solution for a statistical auditing problem

Glen Meeden*
School of Statistics
University of Minnesota
Minneapolis, MN 55455

July 2002

*Research supported in part by NSF Grant DMS 9971331

SUMMARY

Auditors often consider a stratified finite population where each unit is classified as either acceptable or in error. Based on a random sample the auditor may be required to give an upper confidence bound for the number of units in the population that are in error. In other cases the auditor may need to give a p value for the hypothesis that at least 5% of the units in the population are in error. Frequentist methods for these problems are not straightforward and can be difficult to compute. Here we give a noninformative Bayesian solution for these problems. This approach is easy to implement and is seen to have good frequentist properties.

Key Words: Noninformative Bayes, Finite population sampling, Dichotomous variable, Statistical auditing.

1 Introduction

Recently Wendell and Schmee (1996) discussed an interesting sampling problem that arises in auditing. Often in practice stratified finite populations are encountered where each unit is either in compliance or noncompliance with certain regulations. Typically the units which are in noncompliance or in error occur infrequently. One may desire an upper confidence bound for the the number of units in the population which are in error or a p-value for the hypothesis that at least 5% of the units in the population are in noncompliance. A small p-value for a sample would suggest that almost all units are in compliance and the auditor could attest that the population contains fewer than 5% units in error without any more sampling. They noted that because of the small number of units in error in these populations the usual large sample approach gives poor answers for such problems. They suggested some alternative methods. These new methods however are difficult to apply especially when the number of strata is larger than four.

In this note we consider a noninformative Bayesian approach to this problem. If a unit is in error or noncompliance we assign it the value 1 otherwise it is assigned the value 0. Hence the population total is just the total number of units in error. We propose a simple hierarchical Bayesian model which is appropriate for a population with a dichotomous characteristic where one of the two values appears very infrequently. For this model given a sample one can simulate from the posterior distribution copies of the entire population and for each simulated copy find the population total. Given these simulated values one can find approximately the point, say t^*, such that the posterior probability given the sample that the population total is less than or equal to t^* is equal to 0.95. Hence t^* is a 0.95 upper credible bound for the population total. We then consider various populations and present simulation results that demonstrate that these 0.95 Bayesian upper credible bounds have a frequentist coverage probability of approximately 95% and the corresponding posterior probability can be interpreted as a p-value. In section 2 we present the Bayesian model. In section 3 we compare our bound to the bound of Wendell and Schmee (1996). In section 4 we study the behavior of our bound by simulation for a variety of populations. Section 5 concludes with a brief summary.

2 The Bayesian model

We begin by considering the simple case where the population is not stratified. We start with some notation. Let $\mathcal{U}$ denote the finite population which consists of N units labeled $1, 2. \ldots, N$. Attached to unit i let y_i be 1 if the ith unit is in error or noncompliance and 0 otherwise. $y = (y_1, \ldots, y_N)$ is assumed to belong to $\mathcal{Y}$ the set of 2^N vectors of length N which consist of 0's and 1's. A subset s of $\{1, 2, \ldots, N\}$ of size n is a sample. Let $s = \{i_1, \ldots, i_n\}$ be the labels of the units appearing in the sample and $z_s = \{z_{i_1}, \ldots, z_{i_n}\}$ where z_{i_j} is the observed value of unit i_j. Finally we let $z = (s, z_s)$, the set of labels in the sample along with their values denote a typical observed data point.

Next we define a simple hierarchical Bayes model on $\mathcal{Y}$. The three stages are as follows:

$$
\begin{aligned}
y|\theta, w &\sim \text{independent Bernoulli}(\theta) \\
\theta|w &\sim \text{Beta}(w, 1-w) \\
w &\sim \text{Uniform on } \Lambda
\end{aligned}
$$

where Λ is a set of K distinct real numbers strictly between zero and one. Note the joint probability density function is given by

$$
\begin{aligned}
\pi(y, \theta, w) &= \pi(w)\pi(\theta|w)\pi(y|\theta, w) \\
&= \frac{1}{K} \frac{\theta^{w-1}(1-\theta)^{1-w-1}}{\Gamma(w)\Gamma(1-w)} \prod_{i=1}^{N} \theta^{y_i}(1-\theta)^{1-y_i} \qquad (1) \\
&= \frac{1}{K\Gamma(w)\Gamma(1-w)} \theta^{c(y)+w-1}(1-\theta)^{N-c(y)+1-w-1}
\end{aligned}
$$

where $c(y) = \sum_{i=1}^{N} y_i$ is the number of units in error in the vector y. Note we are following standard Bayesian practice by letting π denote either an arbitrary marginal or conditional probability function or density function. From this it follows that

$$
\pi(y, w) = \int_0^1 \frac{\theta^{c(y)+w-1}(1-\theta)^{N-c(y)+1-w-1}}{K\Gamma(w)\Gamma(1-w)} \, d\theta \qquad (2)
$$

Next we must find the posterior distribution of y given a observed data point $z = (s, z_s)$. If the sample $s = \{i_1, \ldots, i_n\}$ we let $y(s) = (y_{i_1}, \ldots, y_{i_n})$.

Then $c(y(s))$ and $c(z_s)$ are defined analogously to $c(y)$. That is, they are just the total number of 1's appearing in each vector. Note if y is a vector which is consistent with the observed data point z then $y(s) = z_s$ and $c(y(s)) = c(z_s)$. If y is not consistent with z then it will have zero posterior probability given that z is observed. So we assume that y is consistent with z.

We begin by finding $\pi(y(s), w)$, the joint marginal of $y(s)$ and w. Hence for each $i \notin s$ we must sum equation (2) over $y_i = 0$ and $y_i = 1$. To do this we just move the summation inside the integral and sum conditional on a value of θ. Remembering equation (1) we see that this yields

$$\pi(y(s), w) = \frac{\Gamma(c(y(s)) + w)\Gamma(n - c(y(s)) + 1 - w)}{K\Gamma(w)\Gamma(1 - w)\Gamma(n + 1)}$$

$$\pi(w|y(s)) = \pi(w|z) \propto \frac{\Gamma(c(y(s)) + w)\Gamma(n - c(y(s)) + 1 - w)}{\Gamma(w)\Gamma(1 - w)} \tag{3}$$

Now following standard calculations for Bayesian models in finite population sampling (see for example Ghosh and Meeden (1997) page 11) we have

$$\begin{aligned}
\pi(y|z) &\propto \sum_w \int_0^1 \prod_{j \notin s} \theta^{y_j}(1 - \theta)^{1 - y_j} \\
&\qquad \times \theta^{c(y(s))}(1 - \theta)^{n - c(y(s))} f(\theta|w, 1 - w)\, d\theta \\
&\propto \int_0^1 \prod_{j \notin s} \theta^{y_j}(1 - \theta)^{1 - y_j} \\
&\qquad \times \left\{ \sum_w \frac{\Gamma(c(y(s)) + w)\Gamma(n - c(y(s)) + 1 - w)}{\Gamma(w)\Gamma(1 - w)} \right. \\
&\qquad\qquad \left. \times f(\theta|c(y(s)) + w, n - c(y(s)) + 1 - w) \right\} d\theta \\
&= \int_0^1 \left\{ \prod_{j \notin s} \theta^{y_j}(1 - \theta)^{1 - y_j} \right\} \times \sum_w \pi(w|z)\pi(\theta|z, w)\, d\theta
\end{aligned} \tag{4}$$

where $f(\theta|\alpha, \beta)$ is a beta density function with parameters α and β.

It remains to select the set of possible values for Λ. Now under our model for a given unit i we have

$$E(y_i) = E(E[y_i|\theta, w]) = E(\theta) = E(E[\theta|w]) = E(w)$$

This means that $E(w)$ is the prior probability that a typical unit takes on the value one.

Since for this problem the frequentist approach is so difficult to apply our goal was to find a simple Bayesian model which would yield an approximate 95% upper confidence bound under repeated sampling. The main populations of interest are those with approximately 5% of their units in error. Now for such populations we are quite likely to observe a z with $c(z_s) = 0$. For such samples we want a prior distribution which will give some sensible amount of posterior probability to the event that other units in the population could have the value one. But at the same time our model must not grossly inflate this possibility or be over committed to some specific prior guess for the unknown proportion of units in error in the population. The form of our model allows for this possibility and after some experimentation we selected

$$\Lambda = \{.0001 + .01499k \; : \; \text{for } k = 0, 1, \ldots, 9, 10\} \tag{5}$$

For this choice we have $E(w) = 0.07505$. Note that increasing the values of Λ will give probability models which give higher probability to vectors with more ones in them. But as we see in equation (4) when $c(z_s) > 1$ our choice of Λ will have little affect on the posterior because the possible values of w are so small. For such samples this posterior will behave very much like the Polya posterior. This is a noninformative stepwise Bayesian model that yields procedures with good frequentist properties. For more details see Ghosh and Meeden (1997). Lo (1988) calls it the Bayesian bootstrap for finite population sampling and it is related to the Bayesian bootstrap of Rubin (1981).

This choice of Λ given in equation (5) was the one used for all the simulations presented in the next section. Using the expression for $\pi(y|z)$ in equation (4) it is easy to simulate copies of the entire population given an observed data point z of sample size n. First one must select a value for w from the set Λ according to the distribution $\pi(w|z)$. Then using this value one selects a θ from $\pi(\theta|z, w)$ which is a $\text{Beta}(c(y(s)) + w, n - c(y(s)) + 1 - w)$ distribution. Then using this value of θ one observes $N - n$ independent Bernoulli trials with probability θ of getting a one.

If our population consists of several strata then within each stratum we use the same model and assume independence across the strata. Hence to get a simulated copy of a stratified population with observations from each strata we just simulated copies of each stratum using the above and then combine them.

3 Comparison to other methods

We assume that the population consists of L strata with N_i units in stratum i. Let $N = N_1 + \cdots + N_L$. We assume that a simple random sample without replacement of size $n_i > 0$ will be taken from stratum i for $i = 1, 2, \ldots, L$. Following the notation of the previous section we let $z_i = (s_i, z_{s_i})$ be the observed data for stratum i. Let M_i denote the number of units in error for stratum i with $M = (M_1, \ldots, M_L)$. Let $M_t = M_1 + \cdots + M_L$, the total number of units in error in the population and $p_t = M_t/N$. Using the sample we wish to find an approximate 95% upper confidence bound for p_t or a p-value for testing $H : p_t \geq 0.05$ against $K : p_t < 0.05$.

Given the observed data in each stratum our Bayesian model yields a posterior distribution for the unobserved units in the stratum. As we noted previously it is easy to simulate independent copies of the entire stratum from this posterior. By the assumption of independence across the strata it is easy to simulate independent copies of the entire population given the observed data. For each such copy we can find the total number of units in error. So for our model it is easy to study the posterior distribution of the population total through simulation.

Given a sample let U_b be the 95% quantile of the posterior distribution of the population total. As we just noted this will be found approximately by simulation and is an upper bound for M_t. Equivalently $p_b = U_b/N$ would be an upper bound for $p_t = M_t/N$. On the other hand for the testing problem one possible test is to reject $H : p_t \geq 0.05$ when $p_b < 0.05$. Note if p_b is an approximate 95% upper confidence bound for p_t then this test should be an approximate level 0.05 test. Also note that $p_b < 0.05$ is equivalent to the posterior probability of the null hypothesis H being less than 0.05. We let p_H denote this probability. Moreover it could be the case that p_H could be interpreted as a p-value for this testing problem. We will investigate this possibility a bit later.

Wendell and Schmee (1996) present the following solutions to these problems. Given the observed data the usual unbiased point estimate of the number of errors in the population is

$$\hat{M}_{st} = \sum_{i=1}^{L} N_i \frac{c(z_{s_i})}{n_i}$$

Then for the observed data and for a fixed vector M, consistent with the data and for which $M_t = .05N$, they find the probability, say p_M, of all possible data points which yield an estimate of the number of errors in the population no larger than that given by the observed data. Then among all such vectors M they find the one which maximizes p_M and this value is denoted by $p_{\max}$ and is the stated p-value for the observed data. This amounts to selecting a value of the nuisance parameter M which yields the most conservative interpretation of the observed data. This procedure is attractive to auditors. This p-value can be used to construct a test in the usual way and an upper confidence bound can be found for p_t and hence for M_t by inverting this hypothesis test.

We begin by comparing $p_{\max}$ to p_H, the posterior probability of $H : p_t \geq 0.05$. In Table 1, which is similar to Table 2 of Wendell and Schmee we compare p_H to $p_{\max}$ and p_{norm}. Note p_{norm} is the p-value based on usual normal theory and is known to perform poorly here. From the table we see that $p_H > p_{norm}$ in every case and $p_H < p_{\max}$ in every case save one.

Put Table 1 about here

Next we will compare the upper 95% bounds for the two methods. Following Wendall and Schmee (1996) we denote by U_{st} the 95% upper confidence bound for M_t obtained from the tests based on the p-value given by $p_{\max}$. Recall the upper 0.95 credible bound for the population total for our Bayesian model is denoted by U_b. This calculation of U_b depends on the choice of Λ given in equation (5). To help see how sensitive our model is to the choice of Λ we also found this upper 0.95 credible bound for

$$\Lambda' = \{.0005 + .01995k \ : \ \text{for } k = 0, 1, \ldots, 9, 10\} \tag{6}$$

Under this choice $E(w) = 0.10025$ This upper bound is denoted by $U_{b'}$. For selected cases the bounds are give in Table 2, which is similar to Table 3 of Wendell and Schmee. We see from the table that $U_{st} > U_b$ in every case but one. Also U_b and $U_{b'}$ do not differ by much but when then do $U_{b'}$ is slightly larger as was expected.

Put Table 2 about here

Now by construction U_{st}, the 95% upper bound of Wendall and Schmee, must exceed M_t at least 95% percent of the time in repeated sampling. The only possible criticism of U_{st}, beyond the computation difficulties, is that it could be too large for certain populations. On the other hand we have seen that the 0.95 Bayesian credible bound presented here, U_b tends to be smaller than U_{st}. If simulations indicate that it tends to exceed the true value, M_t, approximately 95% of the time then it should be a reasonable alternate. It would be of interest to compare U_{st} and U_b directly for some simple examples. But even in the case of just three strata computing U_{st} in a large simulation study is quite difficult. Since U_b is very easy to compute we will just study its behavior for a variety of populations. We will check through repeated sampling how often p_b exceeds the true population proportion, $p_t = M_t/N$.

4 Some simulation results

It would be of interest to prove results about the frequentist coverage properties of the bounds based on the Bayesian analysis proposed here. That however seems quite difficulty so we will use simulation to study its behavior.

To do this we have selected two groups of nine populations each. In the first group each population has three strata of size (300,200,100) respectively for a total size of 600. The first three populations in this group have 30 units in error, i.e. $p_t = 30/600 = .05$. The next three have 20 units in error and the last three have 45 units in error. Table 3 gives the these nine populations with the number of units in error within each strata.

In the second group $(500, 400, 300, 200, 100)$ is the size of the five strata for the populations giving a total population size of 1500. The first three populations in this group have 75 units in error, i.e. $p_t = 75/1500 = .05$. The next three have 45 units in error and the last three have 105 units in error. Table 4 gives these nine populations with the number of units in error within each strata.

Recall that for auditors a very important special case is populations that have 5% of their units in error. For this reason in each group of nine populations we have selected three populations where exactly three of them have 5% of their units in error, three more with slightly less that 5% in error and three more with slightly more than 5% in error. For each set of three populations we have selected different distributions of the units in error among the strata. However the actual locations of the units in error within each strata were selected at random. We believe that this collection of populations forms a good test for the Bayesian model proposed here.

Put Tables 3 and 4 about here

For each of these 18 populations we selected 500 random samples under two different sampling plans. The sampling plans and the results are given in Tables 5 and 6. For each population and sampling plan the third column of the tables gives the average value of p_b for the 500 samples. These values behave as one would expected. The next column gives the relative frequency that p_b exceeded $p_t = M_t/M$, the true proportion of units in error in the population. These numbers for the most part are quite close to 0.95. Note that this is especially true for second group of populations with five strata. This indicates that for all these populations and sampling plans p_b is an approximate 95% upper confidence bound. The last column gives the relative frequency that $p_b < 0.05$. For the test that rejects $H : p_t \geq 0.05$ when $p_b < 0.05$ this will be approximately the probability of making the Type I error when $p_t \geq 0.05$ and the power of the test when $p_t < 0.05$. Again these numbers appear to be reasonable.

Put Table 5 and 6 about here

As we have seen rejecting $H : p_t \geq 0.05$ when $p_b < 0.05$ is equivalent to rejecting it when the posterior probability of $H : p_t \geq 0.05$ is less than 0.05. More generally we could consider testing $H : p_t \geq p_t^*$ against $K : p_t < p_t^*$ where p_t^* is some specified value. A possible test is to reject H when the posterior probability of H, denoted by p_H say, is less than α. If this test is approximately a level α test (which we have see to be the case) and if the distribution of p_H is approximately uniform(0,1) when p_t^* is the true proportion then the posterior probability p_H could be interpreted as a p-value or level of significance. To check on the distribution of p_H for the nine populations in Table 4 we calculated the mean and variance of p_H for 500 random samples. The results are given in Table 7. Recall that the variance of a uniform(0,1) distribution is $1/12 \doteq 0.083$. These results indicate that it is not unreasonable to interpret the posterior probability p_H as a p-value.

Put Table 7 about here

In this note we have been mainly concerned with populations where the number of items in error is quite small. As we noted earlier we would expect these methods to work well if the number of items in error were larger. To demonstrate this fact we constructed six more populations of size 1500 where three of them had 150 units in error and three had 300 units in error. For these populations, given in Table 8, and two different sampling plans we took 500 random samples and computed p_b The results are in Table 9. We see that the probability of p_b exceeding the true proportion of units in error in the population is very near 0.95. We also computed the mean and variance of p_H for the 500 random samples.

Put Tables 8 and 9 about here

It is well known that in general a p-value cannot be given a Bayesian interpretation. So it might be somewhat of a surprise that the posterior probability studied here behaves approximately like a p-value. However Casella and Berger (1987) showed that for certain common one-sided testing problems it is possible to give frequentist p-values a Bayesian interpretation. Here we are arguing in the other direction, i.e. producing a Bayesian model with a posterior probability that can be interpreted, approximately as a p-value. Presumably this occurs because the hypothesis of interest is one sided.

5 Summary

Simulations for other populations were carried out and gave results much like those presented here. In particular it should be emphasized that having a large number of strata poses no problems for these methods. For example we considered several populations with 10 strata. In fact the frequentist properties tend to improve a bit as the number of strata increase. Moreover this approach is straightforward to apply. See the appendix to learn how one can go on line and simulate from the posterior discussed here. In summary, for the problem of testing $H : p_t \geq 0.05$ against $K : p_t < 0.05$ at level $\alpha = 0.05$ or finding an upper 95% confidence bound for p_t our Bayesian model yields procedures with good frequentist properties even for problems where standard frequentist approaches are very difficult or impossible to apply.

References

[1] George Casella and Roger P. Berger. Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *Journal of the American Statistical Association*, 82:106–11, 1987.

[2] Malay Ghosh and Glen Meeden. *Bayesian methods for finite population sampling*. Chapman & Hall, London, 1997.

[3] Albert Lo. A Bayesian bootstrap for a finite population. *Annals of Statistics*, 16:1684–1695, 1988.

[4] Donald Rubin. The Bayesian bootstrap. *Annals of Statistics*, 9:130–134, 1981.

[5] John P. Wendell and Josef Schmee. Exact inference for proportions from a stratified finite population. *Journal of the American Statistical Association*, 91:825–830, 1996.

Appendix

The bounds presented here are easy to find approximately through simulation. A program has been written in R which allows one to simulate values of M_t for a given sample of outcomes. A interested reader can use this program on *Rweb* at the author's web site

`http://www.stat.umn.edu/~glen/`

Once there click on the "Rweb functions" link and select "simulateT.html". Then you can construct simple examples to see how the methods presented here work in practice. To use the program one needs to specify the set Λ which in the program is denote by "grd".

Table 1: Comparison of $p$max, p_H and p_{norm} for selected cases where $N = (N_1, \ldots, N_L)$, $n = (n_1, \ldots, n_L)$ and $obs = (c(z_{s_1}), \ldots, c(z_{s_L}))$ is the number of units in error in the sample.

N	M_t	n	obs	$p_{\max}$	p_H	p_{norm}
(200, 100)	15	(50, 25)	(0, 0)	.01194	.00068	0
(200, 100)	15	(50, 50)	(1, 0)	.07232	.026	.00077
(2000, 1000)	150	(50, 50)	(1, 0)	.09958	.026	.00268
(300, 200)	25	(75, 50)	(1, 1)	.02918	.015	.00027
(300, 200)	25	(100, 100)	(3, 2)	.03514	.042	.00469
(500, 500)	50	(100, 50)	(2, 1)	.08662	.029	.00424
(5000, 5000)	500	(100, 50)	(2, 1)	.10908	.029	.00676
(100, 100, 100)	15	(25, 25, 25)	(0, 0, 0)	.01205	.0011	0
(300, 200, 100)	30	(50, 50, 50)	(1, 1, 0)	.03775	.022	.00103
(3000, 2000, 1000)	300	(50, 50, 50)	(1, 1, 0)	.04902	.022	.00255
(300, 200, 100)	30	(75, 50, 25)	(1, 1, 0)	.00931	.0054	.00004
(500, 300, 200)	50	(50, 50, 50)	(2, 1, 0)	.12760	.082	.04756
(500, 300, 200)	50	(75, 50, 25)	(2, 1, 0)	.02768	.019	.00137
(5000, 3000, 2000)	500	(75, 50, 25)	(2, 1, 0)	.03639	.019	.00272

Table 2: Comparison of U_{st}, U_b and $U_{b'}$ for selected cases where $N = (N_1, \ldots, N_L)$, $n = (n_1, \ldots, n_L)$ and $obs = (c(z_{s_1}), \ldots, c(z_{s_L}))$ is the number of units in error in the sample.

N	n	obs	U_{st}	U_b	$U_{b'}$
(200, 100)	(50, 25)	(0, 0)	10	3	4
(200, 100)	(50, 50)	(1, 0)	16	13	13
(2000, 1000)	(50, 50)	(1, 0)	181	129	132
(300, 200)	(75, 50)	(1, 1)	22	20	20
(300, 200)	(100, 100)	(3, 2)	23	24	24
(500, 500)	(100, 50)	(2, 1)	56	45	45
(5000, 5000)	(100, 50)	(2, 1)	599	446	453
(100, 100, 100)	(25, 25, 25)	(0, 0, 0)	10	4	5
(300, 200, 100)	(50, 50, 50)	(1, 1, 0)	28	25	25
(3000, 2000, 1000)	(50, 50, 50)	(1, 1, 0)	298	249	254
(300, 200, 100)	(75, 50, 25)	(1, 1, 0)	22	20	21
(500, 300, 200)	(50, 50, 50)	(2, 1, 0)	61	56	56
(500, 300, 200)	(75, 50, 25)	(2, 1, 0)	45	42	43
(5000, 3000, 2000)	(75, 50, 25)	(2, 1, 0)	471	422	429

Table 3: Nine populations of size $N = 600$ with strata sizes $(300, 200, 100)$ with the number of units in error in each stratum.

population	number of units in error		
	stratum 1	stratum 2	stratum 3
pop6h30a	15	10	5
pop6h30b	10	10	10
pop6h30c	5	10	15
pop6h20a	11	6	3
pop6h20b	7	7	6
pop6h20c	3	6	11
pop6h45a	24	14	7
pop6h45b	15	15	15
pop6h45c	7	14	24

Table 4: Nine populations of size $N = 1500$ with strata sizes $(500, 400, 300, 200, 100)$ with the number of units in error in each stratum.

population	number of units in error				
	stratum 1	stratum 2	stratum 3	stratum 4	stratum 5
pop15h75a	25	20	15	10	5
pop15h75b	15	15	15	15	15
pop15h75c	5	10	15	20	25
pop15h45a	15	12	9	6	3
pop15h45b	9	9	9	9	9
pop15h45c	3	6	9	12	15
pop15h105a	35	28	21	14	7
pop15h105b	21	21	21	21	21
pop15h105c	7	14	21	28	35

Table 5: The average value of the 0.95 upper credible bound, $p_b = U_b/N$, for the true population proportion of units in error for nine populations of size $N = 600$ for 500 random samples with sample size $n = (n_1, n_2, n_3)$. Also the relative frequency which p_b exceeds the true population proportion of units in error and is less than the value 0.05.

population	n	Ave Val p_b	Rel Freq $p_b >$ true	Rel Freq $p_b < 0.05$
pop6h30a	(50,50,50)	0.0836	0.926	0.074
	(75,50,25)	0.0834	0.968	0.032
pop6h30b	(50,50,50)	0.0833	0.948	0.052
	(75,50,25)	0.0824	0.968	0.052
pop6h30c	(50,50,50)	0.0793	0.966	0.034
	(75,50,25)	0.0799	0.966	0.034
pop6h20a	(50,50,50)	0.0627	0.950	0.266
	(75,50,25)	0.0608	0.918	0.236
pop6h20b	(50,50,50)	0.0608	0.912	0.300
	(75,50,25)	0.0610	0.930	0.222
pop6h20c	(50,50,50)	0.0582	0.950	0.318
	(75,50,25)	0.0587	0.904	0.248
pop6h45a	(50,50,50)	0.116	0.942	0.0
	(75,50,25)	0.113	0.960	0.002
pop6h45b	(50,50,50)	0.115	0.954	0.004
	(75,50,25)	0.113	0.972	0.002
pop6h45c	(50,50,50)	0.108	0.952	0.0
	(75,50,25)	0.111	0.970	0.006

Table 6: The average value of the 0.95 upper credible bound, $p_b = U_b/N$, for the true population proportion of units in error for nine populations of size $N = 1500$ for 500 random samples with sample size $n = (n_1, n_2, n_3, n_4, n_5)$. Also the relative frequency which p_b exceeds the true population proportion of units in error and is less than the value 0.05.

population	n	Ave Val p_b	Rel Freq $p_b >$ true	Rel Freq $p_b < 0.05$
pop15h75a	(50,50,50,50,50)	0.0775	0.952	0.048
	(82,66,50,34,18)	0.0758	0.970	0.030
pop15h75b	(50,50,50,50,50)	0.0751	0.960	0.040
	(82,66,50,34,18)	0.0732	0.940	0.060
pop15h75c	(50,50,50,50,50)	0.0714	0.944	0.056
	(82,66,50,34,18)	0.0731	0.954	0.046
pop15h45a	(50,50,50,50,50)	0.0524	0.930	0.452
	(82,66,50,34,18)	0.0503	0.960	0.506
pop15h45b	(50,50,50,50,50)	0.0498	0.942	0.518
	(82,66,50,34,18)	0.0498	0.958	0.498
pop15h45c	(50,50,50,50,50)	0.0483	0.950	0.568
	(82,66,50,34,18)	0.0491	0.946	0.522
pop15h105a	(50,50,50,50,50)	0.1000	0.950	0.004
	(82,66,50,34,18)	0.0981	0.962	0.0
pop15h105b	(50,50,50,50,50)	0.0980	0.938	0.0
	(82,66,50,34,18)	0.0968	0.960	0.0
pop15h105c	(50,50,50,50,50)	0.0936	0.946	0.0
	(82,66,50,34,18)	0.0947	0.944	0.002

Table 7: For 500 random samples the mean and variance of the posterior probability that the population mean is greater than or equal to the true population mean for nine populations of size $N = 1500$ with sample sizes $n = (n_1, n_2, n_3, n_4, n_5)$.

population	n	Mean	Variance
pop15h75a	(50,50,50,50,50)	0.47	0.076
	(82,66,50,34,18)	0.50	0.074
pop15h75b	(50,50,50,50,50)	0.48	0.071
	(82,66,50,34,18)	0.47	0.079
pop15h75c	(50,50,50,50,50)	0.48	0.074
	(82,66,50,34,18)	0.49	0.077
pop15h45a	(50,50,50,50,50)	0.47	0.083
	(82,66,50,34,18)	0.49	0.077
pop15h45b	(50,50,50,50,50)	0.46	0.077
	(82,66,50,34,18)	0.49	0.076
pop15h45c	(50,50,50,50,50)	0.49	0.078
	(82,66,50,34,18)	0.49	0.077
pop15h105a	(50,50,50,50,50)	0.46	0.078
	(82,66,50,34,18)	0.48	0.073
pop15h105b	(50,50,50,50,50)	0.47	0.077
	(82,66,50,34,18)	0.47	0.075
pop15h105c	(50,50,50,50,50)	0.45	0.074
	(82,66,50,34,18)	0.47	0.075

Table 8: Six more populations of size $N = 1500$ with strata sizes $(500, 400, 300, 200, 100)$ with the number of units in error in each stratum.

population	number of units in error				
	stratum 1	stratum 2	stratum 3	stratum 4	stratum 5
pop15h150a	50	40	30	20	10
pop15h150b	30	30	30	30	30
pop15h150c	10	20	30	40	50
pop15h300a	100	80	60	40	20
pop15h300b	60	60	60	60	60
pop15h300c	30	50	70	90	60

Table 9: The average value of the 0.95 upper credible bound, $p_b = U_b/N$, for the true population proportion of units in error for six populations of size $N = 1500$ for 500 random samples with sample size $n = (n_1, n_2, n_3, n_4, n_5)$. Also the relative frequency which p_b exceeds the true population proportion of units in error and the mean and variance of the posterior probability that the population mean is greater than or equal to the true population mean.

population	n	Ave Val p_b	Rel Freq $p_b >$ true	Mean	Var
pop15h150a	(50,50,50,50,50)	0.136	0.940	0.48	0.081
	(82,66,50,34,18)	0.132	0.962	0.48	0.079
pop15h150b	(50,50,50,50,50)	0.132	0.940	0.46	0.077
	(82,66,50,34,18)	0.131	0.962	0.48	0.074
pop15h150c	(50,50,50,50,50)	0.127	0.944	0.45	0.077
	(82,66,50,34,18)	0.127	0.952	0.45	0.075
pop15h300a	(50,50,50,50,50)	0.244	0.954	0.46	0.080
	(82,66,50,34,18)	0.238	0.956	0.45	0.071
pop15h300b	(50,50,50,50,50)	0.239	0.924	0.46	0.082
	(82,66,50,34,18)	0.236	0.940	0.44	0.073
pop15h300c	(50,50,50,50,50)	0.235	0.942	0.44	0.073
	(82,66,50,34,18)	0.235	0.948	0.46	0.075