Stockholm University

Master thesis

Department of Statistics

Masteruppsats, Statistiska institutionen

Stratified score interval for a weighted sum of proportions with application on the ICT usage in Swedish enterprises survey

Fredrik Engleryd

Masteruppsats 30 högskolepoäng, vt 2015

Supervisor: Per Gösta Andersson

Abstract

This thesis addresses the issue of interval estimation of large proportions based on a stratified random sample. Interval estimation for large (or small) proportions can be problematic when the standard Wald interval is used. As an alternative to the Wald interval a modified version of a stratified score interval based on a method proposed by Yan and Su (2010) will be presented. The alternative stratified score interval is evaluated through simulations and then applied on the ICT usage in Swedish enterprises survey. The alternative stratified score interval performs better than the standard Wald interval in terms of coverage and can also handle the situation when the estimated proportions are exactly one unlike the standard Wald interval. For interval estimation of proportions in the ICT usage in Swedish enterprises the stratified score interval appears to be a possible alternative method.

Acknowledgements

I would like to express my appreciation to my supervisors Per Gösta Andersson at the department of statistics at Stockholm University and Ingela Düring at the Method unit for Enterprise, Organisation, Real estate and Environment Statistics at Statistic Sweden for their support during this work.

I would also like to thank the rest of the staff at the process department at Statistic Sweden for good company and support. I will miss the lunches and of course the free fruit.

Contents

1. Introduction

Interval estimation for a single population proportion *p* is a popular and well documented subject in statistics. The classic large sample approximate interval, or the Wald interval, is under constant fire and criticized for being inadequate and have poor coverage probability in general and especially for proportions near 0 and 1. As poor coverage was not enough the confidence interval can stretch outside the boundaries [0, 1]. In the case of proportions at exactly zero or one the result is a zero width interval. Compared to most of the alternative methods for confidence interval around a proportion the Wald interval is however simple and versatile.

A confidence Interval procedure for a weighted sum of proportions, like in a stratified sample is not as straightforward. Moving from a single proportion to a weighted sum of proportions seems to decrease the alternative methods in the same way as the coverage probability drops for the Wald interval approaching [0, 1]. In a setting with a finite population things get somewhat more complicated and the available methods are sparse.

The Information and Communication Technology (ICT) usage in Swedish enterprises is a survey conducted by Statistic Sweden on behalf of Eurostat. The survey aims to investigate the usage of information technology in Swedish enterprises. The survey mainly consists of yes and no questions and hence shares of certain attributes are the unit of measurement. The ICT usage in Swedish enterprises is a stratified random sample and the proportions to be estimated comprises of several different stratums. The proportion of interest here is the share of enterprises that use computers. As suspected the estimated proportions get very high. The estimated shares of enterprises that use computers reach a hundred percent in most stratums. Interval estimation in this setting becomes problematic if the standard Wald interval is used. The result is many zero width intervals, intervals that exceed one and poor coverage. Since 2014 the question has actually been dropped from the survey but there are several questions rendering high estimates. Since it´s not likely that the use of information technology is going to decrease, the problem of interval estimation of large proportions will be a future issue in the ICT usage in enterprises. The study aims to investigate the possibility of an alternative procedure for interval estimation for a weighted sum of large proportions that performs better than the standard Wald procedure.

The strategy here is to use an approach for interval estimation for a weighted sum of proportions proposed by Yan and Su (2010). The authors propose an interval based on inverting the score test, Wilson (1927). A modified version of their method taking the finite population in account will be applied to the ICT usage of enterprises 2013. This approach assumes that the proportions are in fact equal. A strong assumption that certainly not is possible to make in many situations. In this case there is at least some justification for this approach since the proportions are extreme in the majority of stratums. The method might not be applicable in general, that is for all combinations of aggregated proportions across stratums in the survey. Only two sided confidence intervals are considered.

The outline of the thesis is the following

In section 2 some of the many alternative methods of constructing confidence intervals for a single proportion are presented and their respective performance will be discussed based on previous knowledge. The standard Wald interval and the score interval will be given some extra attention since these are the methods considered when moving from the single parameter case to the multi-parameter case. In connection to this the evaluation criteria's for the performance of a confidence interval is discussed.

In section 3 the ICT survey will be examined in more detail. Some of the findings regarding the usage of computers in Swedish enterprises 2013 will be presented. The Notation for stratified random sampling without replacement will also be introduced in this section along with the estimation procedures used in the survey for the stratified proportions.

In section 4 some of the previous research for interval estimation of a weighted sum of proportions is briefly described. The modified version of the stratified score interval proposed by Yan and Su (2010) that take the finite population in account is described in more detail in this section. The modified stratified score interval will be assessed through simulations that focus on large proportions. The true proportions are assumed to be equal.

In section 5 the method is applied on the ICT usage in Swedish enterprises. The confidence intervals for the standard Wald interval, that is currently used for interval estimation in the survey, and the stratified score interval are calculated and compared for a sample of the aggregated proportions estimated in the survey.

Finally in section 6 the conclusions will be presented along with some suggestions on further development of a method proposed by Yan and Su (2010) used for the situation when the proportions cannot be assumed to be equal.

The figures, simulations and calculation of confidence intervals are done in R. The r code is found in the appendix.

2. Confidence intervals for a single proportion

In the following section some of the most common alternative confidence intervals for a proportion will be briefly described. Some extra attention will be given to the Wilson (1927) score interval and the standard Wald interval. There are no shortages of alternative methods for constructing confidence intervals for proportions. There exist a number of comparative studies that investigates the performance of the Wald interval and alternative intervals. For example Brown, Cai and DasGupta (2001), Agresti and Coull (1998) and Newcombe (1998a). More or less all of the alternative procedures can be said to perform better than the standard Wald interval in terms of coverage and some of them do not have the problem of boundaries exceeding 1 or going below 0 or zero width intervals when *p* =[0, 1].

There are several ways to evaluate a confidence procedure. The most common would be coverage probability and length. The following definitions of coverage and length are only of interest in this section when discussing the performance of confidence intervals for a single binomial parameter. In the stratified case the evaluation is based on random generated intervals.

The width of an interval is often measured as expected length. The expected length, $E_{n,p}(length(CI))$, is expressed as,

$$\sum_{x=0}^{n}(U(x,n) - L(x,n))\binom{n}{x}p^x(1-p)^{n-x} \qquad (2.1)$$

(Brown,Cai and DasGupta, 2001)

Only nonrandomized intervals are considered in the single parameter case. The coverage probability, $P_p(p \in CI)$ is therefore defined as,

$$C(p,n) = \sum_{x=0}^{n} I(x,p)\binom{n}{x}p^x(1-p)^{n-x}, 0 < p < 1, \qquad (2.2)$$

(Feng, 2006)

If the interval contains *p when X=x* then $I(x,p)$ = 1 and otherwise 0, Agresti and Coull (1998). Since the binomial distribution is discrete, an exact nominal confidence level 1-α is not possible to attain (Newcombe, 1998a). The coverage probability will vary depending on the parameter value. There are two alternatives, approximate and exact intervals, to choose from depending on how strict level of coverage is desired. An exact method, like the Clopper-Pearson interval, guarantees that the coverage probability is at least at nominal level and should be preferred if a minimum level of 1-α is wanted. If the nominal 1-α level is taken as an average then an approximate interval would suffice. An interval that is easy to implement in practice may also be of importance and approximate intervals are often easier to calculate.

All the methods described in this chapter concerns confidence intervals around a single proportion. The methods are strictly only applicable if a simple random sample is taken. The observations are assumed to be iid, independently and identically distributed.
If *X1,……, Xn* are IID Bernoulli random variables then $X = \sum x_i$ follows a binomial distribution with parameters *n* and *p.* The maximum likelihood estimator for *p* is given by $\hat{p} = \sum x_i /n$, where $x_i = 1$ (Casella and Berger,2002).

2.2 Wald and score intervals

The most frequently used confidence interval for a proportion is the Wald interval based on the normal approximation. The interval is obtained by inverting a general large sample normal test, or the Wald test. The Wald test is based on a statistic of the following form (Casella and Berger, 2002)

$$\hat{\theta} - \theta/\widehat{se}(\hat{\theta}) \quad (2.3)$$

Where $\hat{\theta}$ is the maximum likelihood estimate of θ and $\widehat{se}(\hat{\theta})$ is the estimated standard error.

Then by the central limit theorem, if the sample size *n*, is sufficiently large, the estimator of the proportion, $\hat{p}$, is approximately normally distributed with mean p and variance $var(\hat{p}) = p(1-p)/n$ (Casella and Berger, 2002).

This means that for *p*, $0 < p < 1$

$$\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}} \rightarrow n(0,1) \quad (2.4)$$

A *100(1 - α/2) %* standard Wald confidence interval is thus given by,

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (2.5)$$

Where $z_{\alpha/2}$ is the critical region from a standard normal distribution and $\hat{p} = \sum x_i /n$.

The score interval was first introduced by Wilson (1927) and in literature it is sometimes referred to as the Wilson score interval (Brown et al, 2001). Here it will simply be called the score interval. The score interval is just as the standard Wald an approximate interval and is given by an inversion of the score test for *p* see Casella and Berger (2002, p.495). The difference between the Wald and the score test is that the former is based on the estimated standard error and the latter is based on the null standard error (Brown et al, 2001). The confidence interval is obtained by inverting the test in (2.3), replacing the estimated standard error with the true standard error.

$$\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \rightarrow n(0,1) \quad (2.6)$$

For *p*, $0 < p < 1$,

After inverting the score test for *p*, $\left|\frac{\hat{p}-p}{\sqrt{p(1-P)/n}}\right| \leq z_{\alpha/2}$, the following quadratic equation must be solved for *p* in order to obtain an expression for the score interval.

$$(\hat{p}-p)^2 = {z^2}_{\alpha/2}\frac{p(1-p)}{n} \quad (2.7)$$

Expressed in a another form the above equation becomes

$$\left(1+\frac{z^2_{\alpha/2}}{n}\right)p^2 - \left(2\hat{p}+\frac{z^2_{\alpha/2}}{n}\right)p + \hat{p}^2 = 0 \quad (2.8)$$

Solving the equation by using the quadratic formula,

$p = \frac{-b\pm\sqrt{b^2-4ac}}{2a}$, where a = $\left(1+\frac{z^2_{\alpha/2}}{n}\right)$, $b = -\left(2\hat{p}+\frac{z^2_{\alpha/2}}{n}\right)$ and $c = \hat{p}^2$ (2.9)

the lower and upper bounds are given by,

$$L = \frac{\hat{p}+z^2_{\alpha/2}\ /2n}{1+z^2_{\alpha/2}\ /n} - \frac{z_{\alpha/2}}{1+z^2_{\alpha/2}\ /n}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z^2_{\alpha/2}}{4n^2}} \quad (2.10)$$

$$U = \frac{\hat{p}+z^2_{\alpha/2}\ /2n}{1+z^2_{\alpha/2}\ /n} + \frac{z_{\alpha/2}}{1+z^2_{\alpha/2}\ /n}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z^2_{\alpha/2}}{4n^2}} \quad (2.11)$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution and $\hat{p} = \frac{\sum x_i}{n}$.

Brown, Cai and DasGupta (2001) do quite a thorough study on the subject and compare the performance of the Wald interval and some alternative methods, including the score interval. Most likely one of the most cited papers regarding alternative confidence intervals for proportions. The score interval is recognized as an interval with good properties that works well for most situations, that is, for nearly all sample sizes and values of *p* (Agresti and Coull ,1998). Both Agresti and Coull (1998) and Brown, Cai and DasGupta (2001) recommend the score interval for practical use. Brown, Cai and DasGupta (2001) propose the score interval specially for small n, below 40. They do not express the same enthusiasm for the Wald interval which they refer to as “persistently chaotic” among other things and never recommend it to be used, (Brown, Cai and DasGupta , 2001).

The coverage probability of the score interval is more or less always close to the nominal level, even for small n and proportions near 0 or 1 (Agresti and Coull, 1998). In figure 1 the coverage probability for the standard Wald interval and the score interval are plotted for a fixed *n = 10 and 100.* The nominal level is set to 95%. As can be seen from figure 1 the coverage probability, the y-axis, of the Wald interval has much more serious drops in coverage when n equals 10 than compared to the score interval. The coverage of the score interval fluctuates around the 95 % and rarely goes below 90 % coverage. The coverage for the Wald interval gets really poor when *p* is about 0.8 and 0.2. A larger sample size improves the coverage for both the Wald and the score as expected but the Wald interval still has rather poor coverage even when *n* equals 100, for *p* close to 0 and 1.

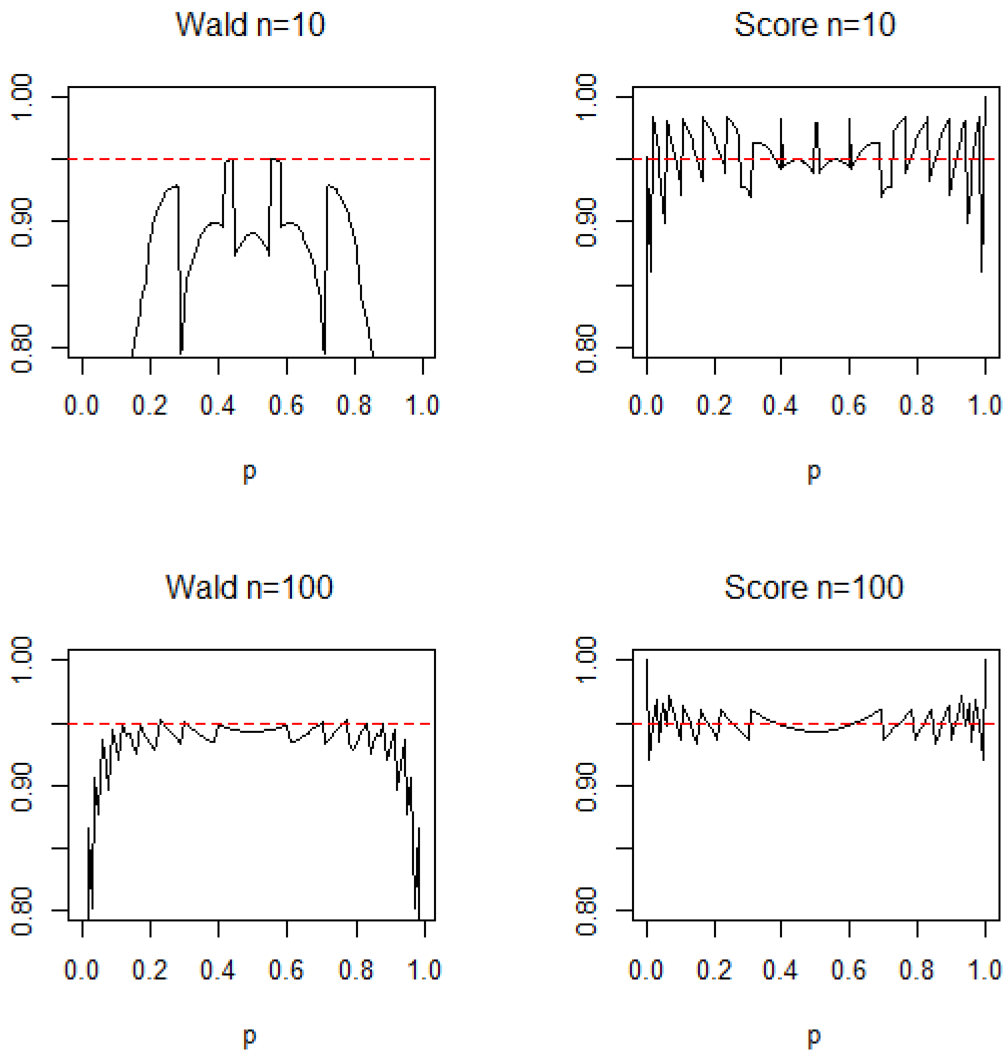

Figure 1. Coverage probability of Wald and score intervals for n=10 and 100.

However the coverage probability for the score interval has two serious drops for some values of *P* very close to 0 or 1. This can be seen in the figure 2, where *p* is plotted as a sequence only for values close to 1. These drops in coverage probability will never disappear. They will however get closer to 0 and one as the sample size increases. For some *p* close to 0 and 1 the minimum coverage will be 0.835 for a 95% confidence interval (Agresti and Coull, 1998). They showed that for instance if n = 10 there is a minimum coverage of 83.5 % at *p* = 0.982 and *p* = 0.018. Taking the whole parameter space in consideration the region where this drop occurs is quite small (Agresti and Coull,1998). The coverage for the Wald interval for the same regions is very poor. The serious drop in coverage starts somewhere when *p* is about 0.96-0.98 when *n*=100.

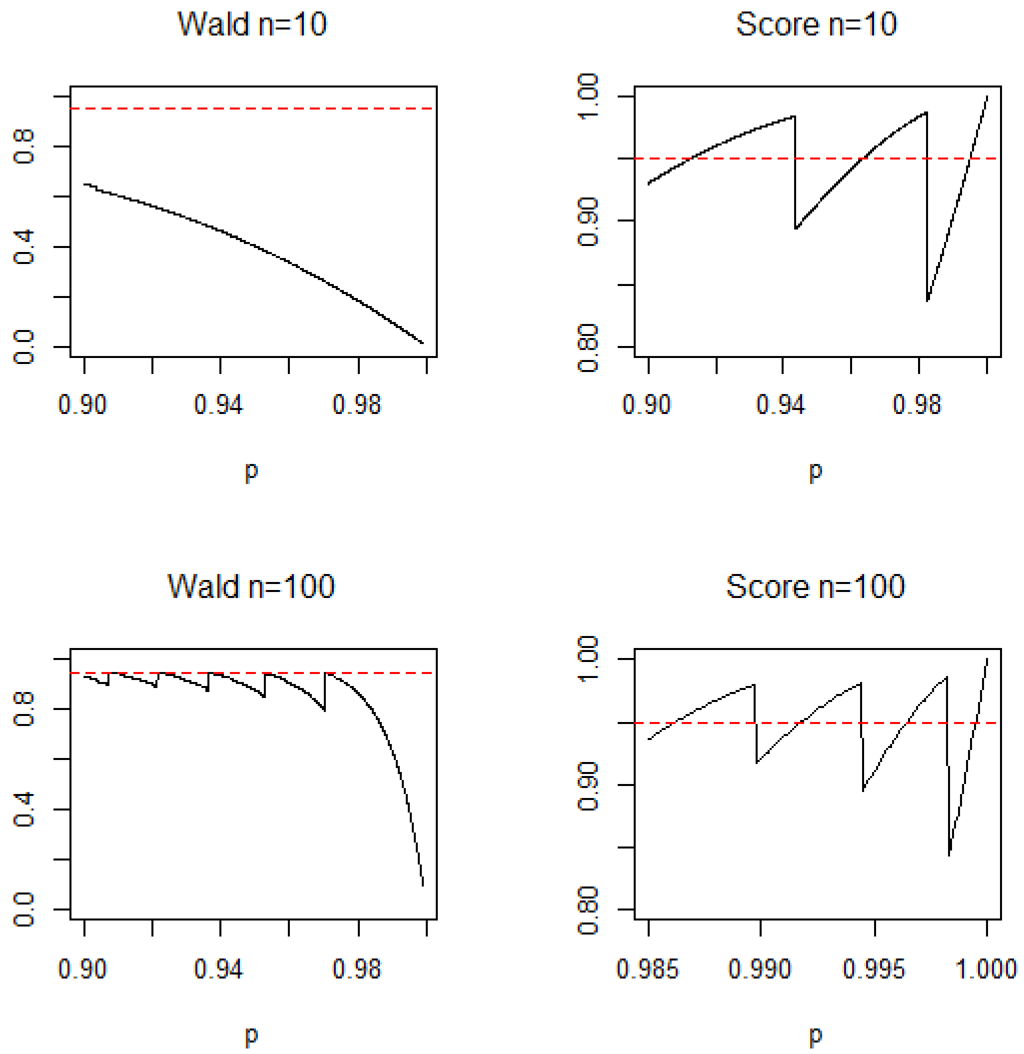

Figure 2. Coverage probability of Wald and score intervals for n=10 and 100 for values of p from 0.9 to 1. In the figure in the lower right corner p ranges from 0.985 to 1.

One important feature of the score interval is the fact that it never results in a zero width interval when the estimated proportion is 0 or 1. A necessity in this case since the data is somewhat extreme in the ICT usage in enterprises survey. In figure 3 and 4 the standard error for the Wald interval and the term that affects the length of the score interval are plotted over a sequence of *p* for n=10 and 100. The standard error for the Wald interval is

$\sqrt{\hat{p}(1-\hat{p}/n}$ and the term for the score is $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z_{1-\alpha/2}^2}{4n^2}}/(1+z_{\alpha/2}^2\ /n)$. The term in the score interval never converges to zero for extreme values of *p*. The second term in the square root ensures that a zero width interval never occurs.

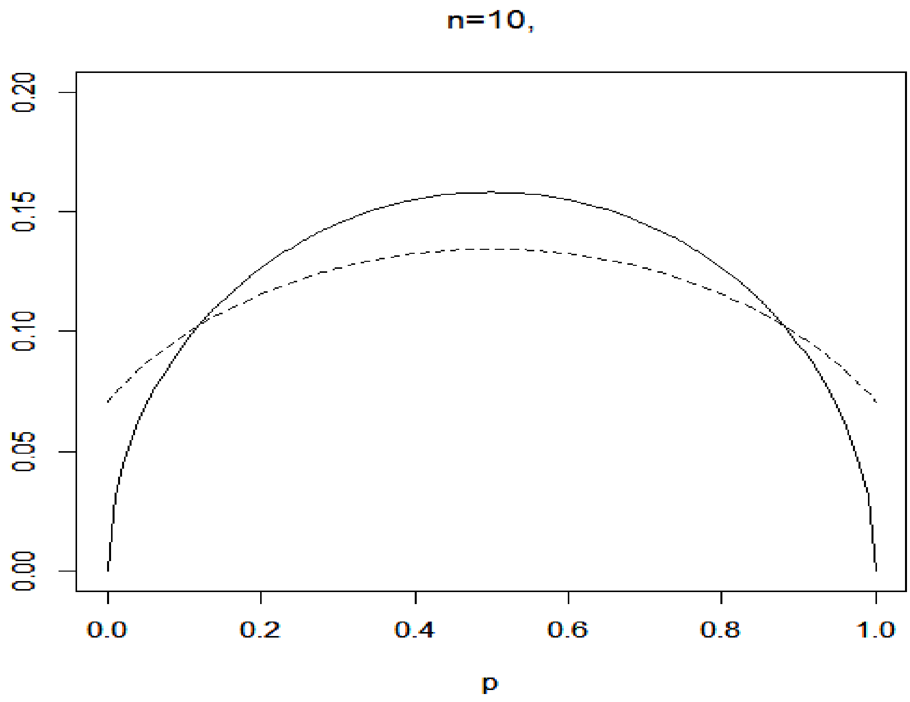

Figure 3. Standard error for the Wald interval and the term that affects the length of the score interval plotted over a sequence of p for n=10.

The difference between the curves will be lower when the sample size gets larger but the standard error for the Wald interval always converges to zero no matter the sample size and hence the result is a zero width interval for proportions at exactly 0 or 1.

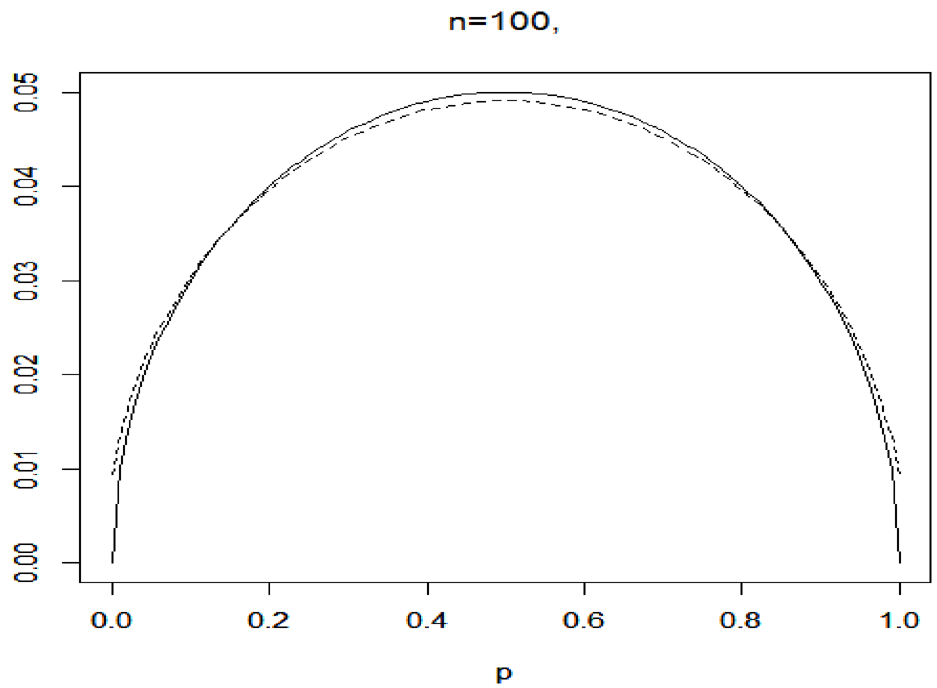

Figure 4. Standard error for the Wald interval and the term that affects the length of the score interval plotted over a sequence of p for n=100.

Unlike the Wald interval the score interval is symmetric around the midpoint, $\frac{\hat{p}+z_{1-\alpha/2}^{2}\ /2n}{1+z_{1-\alpha/2}^{2}\ /n}$.

This is a weighted average which falls between $\hat{p}$ and 0.5. This will shift the interval closer to 0.5. The effect is less when n increases (Agresti and Coull,1998). The score interval is only symmetric around $\hat{p}$ when it is exactly 0.5. The level of asymmetry will not be measured here. The only intervals that are compared and evaluated are the Wald and the score interval. One is symmetric around $\hat{p}$ and the other one is not. It is the asymmetry that makes the score interval a good candidate for estimation of proportions close or at [0, 1]. The Wald intervals poor performance when $\hat{p}$ is near the limits of 0 and 1 is because it is always symmetric around $\hat{p}$ and ignoring that the binomial distribution is skewed (Agresti and Coull, 1998).

The score interval has in general better coverage than the Wald interval but is not necessary wider. If $\hat{p}$ lies within 0.15 and 0.85 the score interval is narrower than the Wald interval for any *n* and any confidence level (Agresti and Coull, 1998).

A continuity correction could be incorporated in both the Wald and score interval. For closed form expressions of the lower and upper limits for the score interval, see Newcombe (1998a). Incorporating a continuity correction will make the score interval similar to that of a conservative. In other words, the coverage probability will be above the nominal level. A Wald interval with continuity correction will improve the coverage but will lead to more instances of overshoot for *p* close to [0, 1].

Besides the unfortunate decrease in coverage probability for some *p* in the vicinity of zero and 1 the score interval has good properties compared to the Wald interval. In the case with a single parameter the choice between the Wald and the score interval is obvious for parameters close to the limits [0, 1]. It is also an interval that is fairly easy to calculate because closed form expression exists for the lower and upper limits. If a strict 1-α nominal level is needed other methods must be considered.

2.3 Adjusted Wald interval

As the name implies the interval is based on the Wald interval and has the similar well-known form. It's also sometimes referred to as the Agresti – Coull interval due to its first appearance in Agresti and Coull (1998). The method consists of adding to successes and two failures and then uses the same formula for constructing the confidence interval as for the standard Wald interval. The point estimate, $\hat{p} = x/n$, then becomes $\tilde{p} = (x + 2)/(n + 4)$.

The adjusted Wald confidence interval is defined as

$$\tilde{p} = \pm 1.96\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} \qquad (2.12)$$

(Agresti and Coull, 1998)

The midpoint of the adjusted Wald interval resembles the midpoint of the Wilson interval. The midpoint of the Wilson interval $(x + z^2/2)/(n + z^2) \approx (x + 2)/(n + 4)$ when $z^2 = 1.96^2$ (Agresti and Coull,1998). The intervals are centered on nearly the same value when the confidence level is 95%. Figure 5 shows the coverage probability when n=10. The coverage probability is rather good. In fact it´s has better coverage than the Wilson interval and is slightly more conservative (Agresti and Coull, 1998). Brown, Cai and DasGupta(2001) recommended the adjusted Wald interval for $n \geq 40$ due to its simple form and good performance in general. They also recognize that it behaves well for small n as did Agresti and Coull (1998). One of the drawbacks of the method is that the interval can stretch below 0 and above 1, just like the standard Wald. Despite its good coverage probabilities and simple form it's not be the best choice for very small or very large proportions.

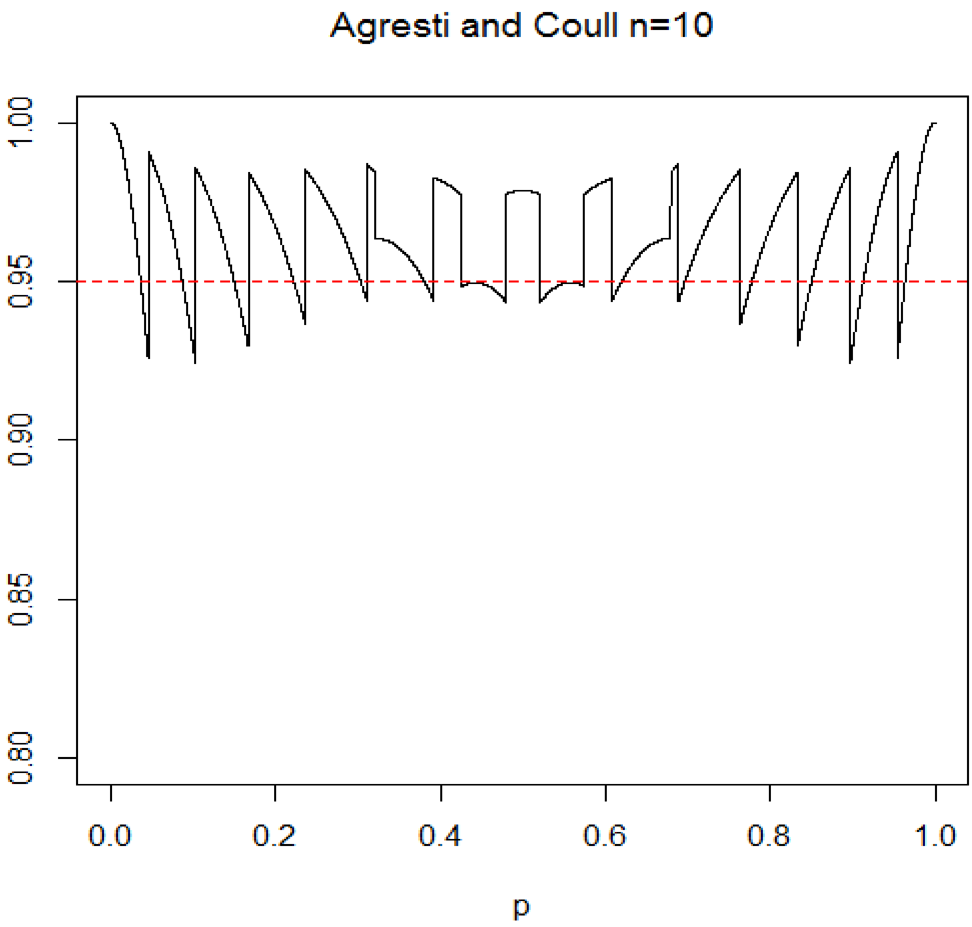

Figure 5. Coverage probability of Agresti and Coull interval for n=10.

2.4 Transformation methods

The logit interval and the arcsin interval are two variance stabilizing transformations for the binomial distribution commonly used (Feng, 2006).

A 100(1-α)% confidence logit interval is obtained by first letting, $\hat{\lambda} = log\left(\frac{\hat{p}}{1-\hat{p}}\right)$

It can then be shown by the delta method that $(\hat{\lambda} - \lambda)/\sqrt{\hat{V}(\hat{\lambda})} \rightarrow N(0.1)$

where $\hat{V}(\hat{\lambda}) = 1/(n\hat{p}(1-\hat{p})$,

The endpoints of the interval are

$$L = \frac{e^{\hat{\lambda}_L}}{1+e^{\hat{\lambda}_L}}, \text{ where } \hat{\lambda}_L = \hat{\lambda} - z_{\alpha/2}\sqrt{\hat{V}(\hat{\lambda})} \quad (2.12)$$

$$U = \frac{e^{\hat{\lambda}_U}}{1+e^{\hat{\lambda}_U}}\text{, where } \hat{\lambda}_U = \hat{\lambda} + z_{\alpha/2}\sqrt{\hat{V}(\hat{\lambda})} \quad (2.13)$$

(Liu and Kott, 2007)

The interval tends to be quite long and performance, in terms of coverage probability, is poor when *p* is close to 0 or 1 (Brown et al, 2001).

The arcsin interval is obtained in a similar fashion. Let $\hat{\delta} = arcsin\sqrt{\hat{p}}$,
and again by the delta method it can be shown that $(\hat{\delta} - \delta)/\sqrt{\hat{V}(\hat{\delta})} \rightarrow N(0.1)$
where $\hat{V}(\hat{\delta}) = 1/4n$
The upper and lower limits are given by

$$L = sin^2\left[arcsin^2 - z_{\alpha/2}/2\sqrt{n}\right] \quad (2.14)$$

$$U = sin^2\left[arcsin^2 + z_{\alpha/2}/2\sqrt{n}\right] \quad (2.15)$$

(Liu and Kott, 2007)

For proportions close to 0 and 1 the coverage probability drops considerably however the interval performs adequate when *p* is distant from the limits (Brown et.al, 2001).

2.5 Binomial likelihood ratio interval, (LRT)

The Binomial likelihood ratio interval is formed by inverting the likelihood ratio statistic. An approximate 1-α likelihood interval for a proportion is then given by

$$p: -2log\left(\frac{p^y(1-p)^{n-y}}{\hat{p}^y(1-\hat{p})^{n-y}}\right) \leq \chi^2_{1,\alpha} \quad (2.16)$$

where $y = \sum_{i=1}^{n} x_i$

(Casella and Berger, 2002)

This is perhaps one of the most common methods of constructing confidence interval but it is more difficult to calculate than some other approximate intervals like the Wald or score interval (Brown et al, 2001). Casella and Berger (2002) compare the performance of the likelihood ratio interval with the Wald and Wilson´s score interval. They found that the length of the approximate LRT interval was the shortest and Wilson´s score interval the longest at least for small n. For small n the coverage probability of the LRT interval was found to be inadequate. It should be noted that they compared the LRT method with a continuity corrected score interval which has coverage probability above nominal level, similar to a conservative interval.

2.6 The Clopper-Pearson interval

The exact or conservative method for calculating a confidence interval for a binomial parameter is often represented by the Clopper-Pearson interval. The Clopper-Pearson method is the standard and most common way of obtaining an interval if one prefers to avoid an approximation (Agresti and Coull, 1998). The interval is constructed by inverting the equal-tailed binomial tests of $H_0: p = p_0$ against the alternative hypothesis of $H_1: p \neq p_0$.

To get the endpoints of the Clopper-Pearson interval the solutions to the following equations are calculated.

$$\sum_{k=x}^{n} \binom{n}{k} \mathrm{p}_0^{\mathrm{k}} (1-p_0)^{n-k} = \alpha/2 \qquad (2.17)$$

and

$$\sum_{k=0}^{x} \binom{n}{k} \mathrm{p}_0^{\mathrm{k}} (1-p_0)^{n-k} = \alpha/2 \qquad (2.18)$$

When $x = 1,2 \dots, n-1$, the Clopper-Pearson interval can be expressed as

$$\left[1 + \frac{n-x+1}{xF_{2x,(n-x+1),1-\alpha/2}}\right]^{-1} < p < \left[1 + \frac{n-x}{xF_{2(x+1),2(n-x),\alpha/2}}\right]^{-1} \qquad (2.19)$$

$F_{a,b,c}$ is the1-c quantile from the *F* distribution with a and b degrees of freedom (Agresti and Coull, 1998).

The lower endpoint in the interval above is the α/2 quantile of a beta distribution with parameters *x* and *n-x+1*. The upper endpoint then is a 1- α/2 quantile of a beta distribution with parameters *x+1* and *n-x.* In the situation when *x* = 0 the lower bound will be 0 and if *x* = *n* the upper bound will be 1 (Agresti and Coull, 1998). The nominal level of this interval is guaranteed to be at least 1- α. However the nominal level is not exactly 1- α. The actual coverage probability is in fact always larger than the nominal confidence level and quite a lot when n is small.

Figure 6 shows the coverage probabilities for the Clopper - Pearson interval for fixed *n* =(10, 25, 50 and 100) and different *p*. When *n* gets larger the coverage probabilities gets closer to the nominal level. Because of the unnecessary long intervals that the Clopper-Pearson method inevitably produces, unless n is quite large, it is not recommended for practical use (Agresti and Coull, 1998, Brown et al, 2001). The Clopper-Pearson interval is symmetrical around $\hat{p}$ which the approximations previously discussed aren´t, except for the Wald interval. It is not exactly symmetrical though. The error probabilities in each tail cannot be precisely 2.5 % (Reizigel, 2003). It should be mentioned that there are other exact methods if a confidence level that is equal or larger than 1- α is demanded, like the Stern interval for example (Casella, 1986).

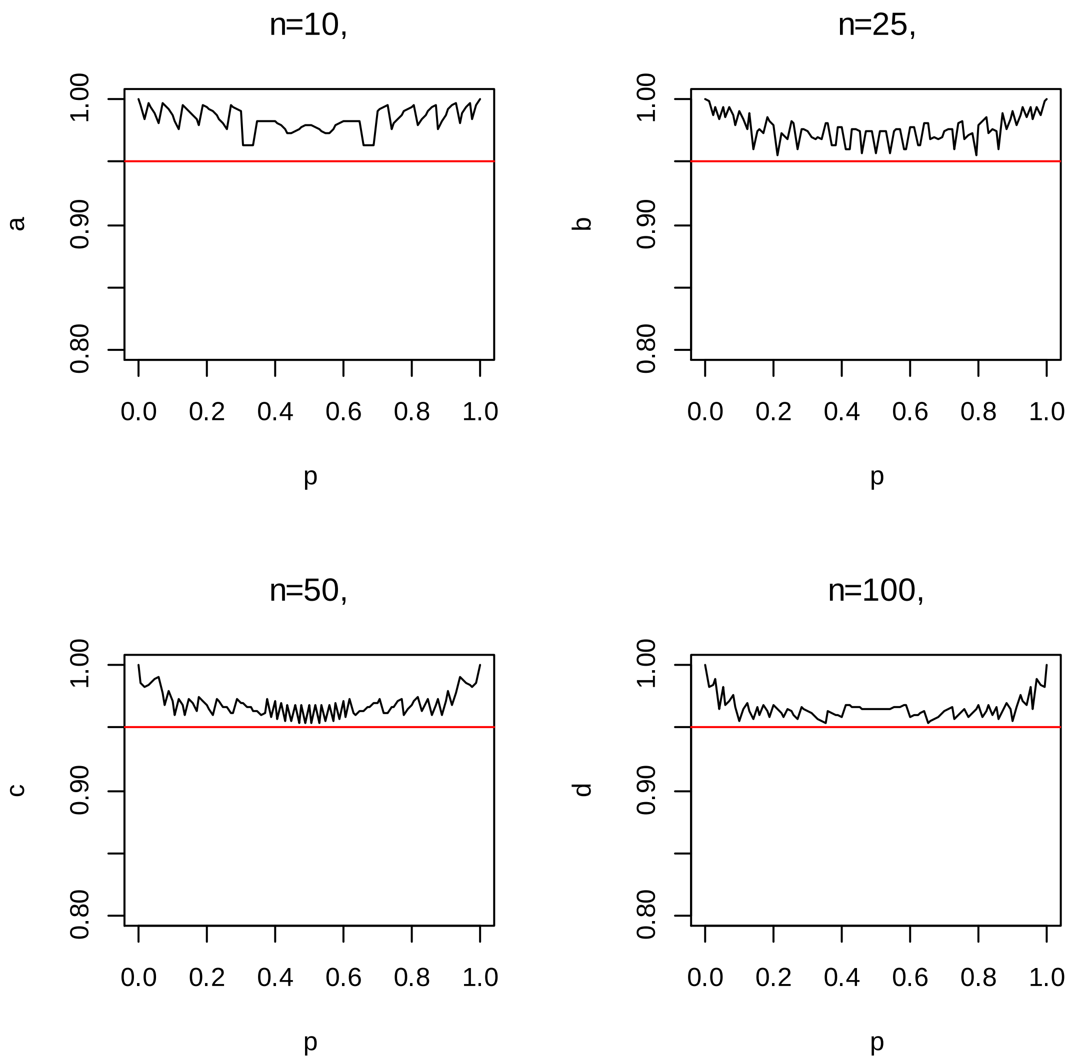

Figure 6.Coverage probability for the Clopper–Pearson interval for n = 10, 25, 50 and 100.

2.7 Jeffreys interval

As an alternative to the frequentist methods one could tackle this problem with a Bayesian approach. A natural prior for the binomial parameter *p* is the beta distribution. If $X \sim Binomial(n, p)$ and $p \sim Beta(a, b)$ the posterior distribution of *p* is $Beta(X + a, n - X + b)$, (Feng, 2006).

The non-informative Jeffreys prior is $Beta(1/2, 1/2)$.

The lower and upper bound for the Jeffreys prior 100(1-α) % equally tailed interval is given by

$$L_J(x) = B(\alpha/2; X + 1/2\,, n - X + 1/2) \quad (2.20)$$

$$U_J(x) = B(\alpha/2; X + 1/2\,, n - X + 1/2) \quad (2.21)$$

When *x= 0 or n* the lower and upper endpoints must be modified to avoid a zero coverage probability when *P* goes to 0 or 1 (Brown, Cai and Dasgupta, 2001).

3. ICT usage in Swedish enterprises

The ICT usage in Swedish enterprises is a survey that aims to investigate the usage of information technology in Swedish enterprises. More specifically questions about access to computers and networks, usage of IT systems and internet and e commerce are asked. The survey was from the beginning a joint Nordic project that started in 1999. In 2001 Eurostat developed the model in order to make a consistent survey in Europe. From 2006 and on the survey, ICT usage in enterprises, is regulated by the European Union and the survey are now conducted on behalf of Eurostat (ICT usage in enterprises, 2013). The survey has been conducted yearly since 2001.

The ICT usage in enterprises is stratified random sample with optimal allocation. The sampling frame is stratified after industry and size of enterprise, measured as number of employees. For enterprises with more than 250 employees all units are sampled. All units have also been sampled if a stratum consists of 7 units or less. At least 7 units are sampled if there are more than 7 units in the stratum. For the rest of the enterprises with 10-249 employees a stratified random sample with Neyman allocation is conducted. Both number of enterprises in the stratum and turnover is used for allocation. This will result in two different alternative sample sizes. The alternative that gives the largest sample size is used (ICT usage in enterprises, 2013).

The data for the survey is collected through both internet-based and mailed questionnaires (ICT usage in enterprises, 2013). In 2103 the unit non-response was about 19%. The unit non-response rates for type of industry are about 18 percent for most classes. Accommodation and food service activities stand out with the highest non-response rate at 33 percent. Enterprises with 10-49 tend to be less inclined to answer than large enterprises, with non-response rates of 22 and 15 percent respectively (ICT usage in enterprises, 2013).

The total population in the survey, 2013, consists of 34465 enterprises and the total sample size is 3419. The observations are distributed across 212 stratums. As already mentioned the estimated proportions are very large for nearly all industry classes and sizes. Most of them are equal to one, roughly about 70 percent. The remaining proportions are usually above 0.9.

Further the number of sampled units in each stratum varies a lot. This is also true for the strata sizes. The smallest stratum size in the survey is in fact one and there are several strata that consist only of a few units. Since all units are sampled when the total strata size is seven or less and likewise when an enterprise has more than 250 employees the result are many total selections. The amount of strata where all units are sampled will decrease some due to non-response but the quantity of total selections is still large. About 80 percent of the strata have sample sizes relative to stratum size around 10 percent and above and in 20 percent of the strata the sampling fraction is below 10 percent.

Ignoring the finite population and assume that the data follow a binomial distribution in every strata may not be reasonable in this case. Assuming an infinite population would have made the calculation of a confidence interval for this data slightly less complicated. Ignoring the finite population will probably result in wider confidence intervals and hence more uncertainty about the true proportion. The data cannot be presented in full here due to confidentiality reasons.

3.2 Results, ICT usage in Swedish enterprises 2013

Some of the findings related to question of interest, will be summarized in short in the following section. Only a part of the results from the 2103 survey will be reported here. Many questions in the survey will result in quite large proportions but only the question of usage of computers will be addressed here. See, SCB, ICT usage in Swedish enterprises 2013, for more information of the survey 2013.

The proportions of enterprises that use computers have since the beginning of the survey been high. It has also been more or less been constant over time. The smaller enterprises, 1 to 9 employees, show a similar pattern. These enterprises were investigated for the first time in 2008, (ICT usage in Swedish enterprises 2013).

The results are reported broken down after industry and size of enterprise. The figures in table 1 display the proportion of enterprises by industry for 10 employees or more that use computers. The proportions of computer usage for the different industry classes are ranging from 94-100 percent. Accommodation and food service activities have the lowest usage and Electricity and waste management and the It sector have the highest usage with 100 percent.

Table 1. Use computers, share of enterprises by industry, 2013, for 10 employees or more.

	Proportion %	Confidence interval
Total	98	± 1
Manufacturing	99	± 1
Electricity and waste management	100	± 0
Construction	97	± 3
Wholesale and retail trade; repair of motor vehicles	99	± 1

and motorcycles		
Transportation and storage	99	± 2
Accommodation and food service activities	94	± 4
Information and communication	99	± 0
Financial and insurance activities	98	± 1
Real estate activities	97	± 3
Other service activities	98	± 1
It-sector	100	± 0

Source: Statistic Sweden, ICT usage in enterprises 2013

In table 2 the proportions of computer usage divided after size of enterprises, measured by number of employees, shows that enterprises with a small number of employees to a less extent use computers than large enterprises. The proportions are ranging from 90-99.

Table 2. Use computers, share of enterprises by size, 2013, 1 employee or more.

Number of employees	Proportion %	Confidence interval
Total, 1-9 employees	91	± 2
1-4 employees	90	± 3
5-9 employees	96	± 4
10-49 employees	98	± 1
50-249 employees	99	± 1
250 or more employees	99	± 1

Source: Statistics Sweden, ICT usage in enterprises 2013

3.3 Stratified random sampling

The ICT survey is a stratified random sample. A simple random sample without replacement is taken in each stratum. The notations for a stratified random sample are as follows.

The population proportion is given by,

$$p_{str} = \frac{1}{N}\sum_{h=1}^{H}\sum_{j=1}^{N_h} y_{hj} \qquad (3.1)$$

(Lohr, 2010)

Where y_{hj} = 1 if enterprise j use computers and 0 otherwise. *N* is the total population and N_h is the number of enterprises in stratum *h.*

The estimate of the proportion is given by,

$$\hat{p}_{str} = \sum_{h=1}^{H} \frac{N_h}{N} \hat{p}_h \qquad (3.2)$$

The aggregated proportions are a weighted average of the sample stratum averages and are multiplied by N_h/N (Lohr,2010), where $\hat{p}_h = \bar{y}_h$ is the estimated proportion in stratum h.

The variance in the population is estimated according to,

$$\hat{V}(\hat{p}) = \frac{1}{N^2} \sum_{h=1}^{H} \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) \hat{S}_h^2 \qquad (3.3)$$

and the standard error, $\hat{S}_h^2$ is given by,

$$\hat{S}_h^2 = \frac{N_h}{N_h - 1} \hat{p}_h (1 - \hat{p}_h) \qquad (3.4)$$

(ICT usage in Swedish enterprises, 2013)

If the number of strata is large enough or the sample sizes within each stratum is sufficiently large (Lohr, 2010), a 95 % Wald confidence interval for the population proportion, $\hat{p}$, is given by,

$$\hat{p}_{str} \pm z_{\alpha/2} \sqrt{\hat{V}(\hat{p})} \qquad (3.5)$$

4. Confidence intervals for a sum of weighted proportions

The methods presented in section two can be used to calculate a confidence interval for a proportion in a single stratum. For a sum of proportions obtained from a stratified random sample some adjustment of the methods is needed, not obvious how though. Quite a lot of attention has been given to the difference of proportions from independent binomial distributions and there are known procedures for these situations; see Newcombe (1998b) and Agresti and Caffo (2000). However confidence intervals for aggregated proportions seem to be much less investigated. Some of the previous efforts in this area are presented in the following section.

Decrouez and Robinson (2012) evaluate confidence intervals for the sum of two weighted proportions based on inverting the Wald, score and the log likelihood ratio test. The score test based intervals performed best. Their proposed score interval relies on some heavy calculations though and the interval must be obtained numerically. Not an approach applicable for this problem.

Hamada, Mitchell and Necker (2014) compare three different methods for constructing an upper confidence bound for small weighted proportions in a stratified random sample. Assuming that the number of events follow a hypergeometric distribution in each stratum they evaluate a Clopper-Pearson type interval based interval, a standard Wald interval and a Bayesan approach. The Clopper-Pearson based on the hypergemetric distribution generated very conservative intervals since an adjustment for simultaneous intervals was needed. Their analysis is based on only three stratums and this method yielded intervals with 100 % coverage. The Wald interval and the Bayesan approach resulted in shorter intervals with coverage around 90 and 95 % respectively. For the Wald interval zero stratum variances was replaced by 0.1 to increase the length of the interval.

The method proposed by Yan and Su (2010) for a sum of weighted independent proportions is perhaps the most interesting solution to the problem at hand. They propose a stratified score confidence interval. Instead of a single parameter we now have multiple independent parameters.

The stratified proportion is now defined as,

$$p_{str} = \sum_{h=1}^{H} w_h p_h, \qquad (4.1)$$

where w_h are the stratum weights and $h = 1,2,3 \ldots H$ are the strata.

The estimate of the stratified proportion is,

$$\hat{p}_{str} = \sum_{h=1}^{H} w_h \hat{p}_h \qquad (4.2)$$

And finally the variance of the point estimate is given by

$$\sum_{h=1}^{H} w_h^2 \hat{p}_h \, (1 - \hat{p})/n_h \qquad (4.3)$$

The score interval for the stratified proportion is quite similar to the score interval for a single proportion and can be solved in a similar fashion. The derivation of the score interval involves a solution of a quadratic expression as shown in section 2. However the expression for the variance of the point estimate cannot be expressed as a quadratic form of $\sum_{h=1}^{H} w_h p_h$ unless all the proportions are equal. Recall that the expression for the standard error in the denominator in the score test is the null standard error, $\sqrt{\sum_{h=1}^{H} \frac{w_h^2}{n_h} \, p(1-p)}$. If the proportions are equal the roots of equation (4.4) gives the confidence bounds for p (Yan and Su ,2010).

$$\left(\sum_{h=1}^{H} w_h \, \hat{p}_h - p\right)^2 = \; z_{1-\alpha/2}^2 \sum_{h=1}^{H} \frac{w_h^2}{n_h} \, p(1-p) \qquad (4.4)$$

The roots of equation (4.4) are given by,

$$(L) = \frac{\sum_{h=1}^{H} w_h \hat{p}_h + \lambda Z_{\alpha/2}^2}{1+\lambda Z_{\alpha/2}^2} - \frac{Z_{1-\alpha/2}\sqrt{\lambda \sum_{h=1}^{H} w_h \hat{p}_h (1 - \sum_{h=1}^{H} w_h \hat{p}_h) + \frac{\lambda^2}{4} Z_{\alpha/2}^2}}{1+\lambda Z_{\alpha/2}^2} \qquad (4.5)$$

$$(U) = \frac{\sum_{h=1}^{H} w_h \hat{p}_h + \lambda Z_{\alpha/2}^2}{1+\lambda Z_{\alpha/2}^2} + \frac{Z_{1-\alpha/2}\sqrt{\lambda \sum_{h=1}^{H} w_h \hat{p}_h (1-\sum_{h=1}^{H} w_h \hat{p}_h) + \frac{\lambda^2}{4} Z_{\alpha/2}^2}}{1+\lambda Z_{\alpha/2}^2} \quad (4.6)$$

(Yan and Su, 2010)

Where $\lambda = \sum_{h=1}^{H} \frac{w_h^2}{n_h}$ and $z_{1-\alpha/2}$ is the $1-\alpha/2$ percentile of a standard normal distribution.

The weights for the proportions in the different strata are as previously *Nh/N*, where *Nh* is the stratum size and *N* is the total population size. Yan and Su (2010) use different kind of weights based on variances but state that the stratified score interval are not dependent on any specific weights. A minimum requirement is that the weights are nonnegative and that they sum to 1 (Yan and Su, 2010). Since we in this case have a finite population and the stratum sizes are known the weights, N_h/N, is a natural choice and more in line with the survey. Using other weights the estimated proportions will be slightly different and hence a confidence interval would not be constructed around the same point estimate as in the survey. No other weights than *Nh/N* will be used in the analysis.

A drawback of this method is naturally that we have to assume that the true proportions are in fact equal, $p_1 = p_2, \ldots, p_H$. The assumption of equal proportions may not be feasible in many situations. For this particular data this assumption is not unrealistic. A common situation is several estimates equal to 1 and one or a few estimates above 0.9, often closer to 1 than 0.9. For some of the combinations of strata all the estimates are equal to 1.

Presumably we can proceed in a similar fashion to find a score interval for a stratified proportion obtained from a finite sample by replacing the variance of the standard error for the sample parameter with the expression of the variance for the population parameter.

By the central limit theorem,

$$\frac{\sum_{h=1}^{k} w_h \hat{p}_h - p}{\sqrt{\left(\frac{N_h}{N}\right)^2 \left(\frac{1}{n_h}\right)\left(1-\frac{n_h}{N_h}\right)\left(\frac{N_h}{N_h-1}\right) p(1-p)}} \to n(0,1) \quad (4.7)$$

And the roots of equation (4.8) give the lower and upper bound of the score interval for the stratified proportion in a finite sample.

$$\left(\sum_{h=1}^{H} w_h \hat{p}_h - p\right)^2 = z_{\alpha/2}^2 \delta p(1-p) \quad (4.8)$$

The roots are explicitly given by,

$$(L) = \frac{\sum_{h=1}^{H} w_h \hat{p}_h + \delta Z_{\alpha/2}^2}{1+\delta Z_{\alpha/2}^2} - \frac{Z_{1-\alpha/2}\sqrt{\delta \sum_{h=1}^{H} w_h \hat{p}_h (1-\sum_{h=1}^{H} w_h \hat{p}_h) + \frac{\delta^2}{4} Z_{\alpha/2}^2}}{1+\delta Z_{\alpha/2}^2} \quad (4.9)$$

$$(U) = \frac{\sum_{h=1}^{H} w_h \hat{p}_h + \delta Z_{\alpha/2}^2}{1+\delta Z_{\alpha/2}^2} + \frac{Z_{1-\alpha/2}\sqrt{\delta \sum_{h=1}^{H} w_h \hat{p}_h (1-\sum_{h=1}^{H} w_h \hat{p}_h) + \frac{\delta^2}{4} Z_{\alpha/2}^2}}{1+\delta Z_{\alpha/2}^2} \quad (4.10)$$

Where $\delta = \sum_{h=1}^{H} \left(\frac{N_h}{N}\right)^2 \left(\frac{1}{n_h}\right)\left(1 - \frac{n_h}{N_h}\right)\left(\frac{N_h}{N_h - 1}\right)$, and $w_h = \sum_{h=1}^{H} \frac{N_h}{N}$

Yan and Su (2010) do also propose a stratified score interval when the proportions cannot be assumed to be equal. A confidence interval for the overall proportion is obtained by first creating individual intervals around the proportions in each stratum. The stratified interval is then formed by the weighted sum of the intervals.

The lower and upper confidence limits for *p* are given by

$$L = \sum_{h=1}^{H} w_h \left(\frac{\hat{p}_h + z_y^2/2n_h}{1+z_y^2/n_h} - \frac{z_y}{1+z_y^2/n_h} \sqrt{\frac{\hat{p}_h(1-\hat{p}_h)}{n_h} + \frac{z_y^2}{4n_h^2}} \right) \quad (4.11)$$

$$U = \sum_{h=1}^{H} w_h \left(\frac{\hat{p}_h + z_y^2/2n_h}{1+z_y^2/n_h} + \frac{z_y}{1+z_y^2/n_h} \sqrt{\frac{\hat{p}_h(1-\hat{p}_h)}{n_h} + \frac{z_y^2}{4n_h^2}} \right) \quad (4.12)$$

In order to get an overall confidence level of 1-α, the individual intervals must be constructed with an appropriate confidence level. Yan and Su (2010) provide a scheme for obtaining both weights and the adjusted confidence level *y* for constructing the individual confidence intervals. The method could unfortunately not be implemented with the weights, *Nh/N*, used in this analysis and the method will not be under consideration. It is worth mentioning though because the authors present a nice solution to the problem when the proportions are unequal, at least for the situation with an infinite population. The proposed score interval performed satisfactory and the coverage rates where close to the nominal 95% level (Yan and Su, 2010).

4.2 Simulations

To assess how the stratified score interval with a finite population correction works some simulations will be carried out. Evaluating the stratified score interval only based on the data from the ICT usage in Swedish enterprises might not give a clear picture of the performance in general. The coverage and mean length are based on random intervals. Assuming we can´t ignore the finite population, 50 000 random samples are generated from the hypergeometric distribution with parameters *(x;N,n,p).* Where *N* is the population size, *n* is the sample size and *p* is the proportion of success. The population proportion is assumed to be known in this scenario. The sample sizes and stratum sizes resembles that of one from the actual survey. The smallest sample sizes relative to stratum size have deliberately been left out, to see how the stratified score performs under more normal conditions. The samples sizes equals *19, 36, 15 and 18* and the strata sizes equals *312, 148, 74 and 40.* The proportions are assumed to be equal and the score interval is based on equation (4.9 and 4.10). The Wald interval is based on equation (3.5) as in the survey. Only large proportions above 0.9 are considered in the following simulations and the stratum weights are *Nh/N*.

Table 3. Stratified score confidence interval, mean length and coverage based on 4 strata with nh=19, 36, 15 and 18 and Nh=312, 148, 74 and 40.

p_1 , p_2 , p_3, p_4	$\hat{p}_{str}$	C.I. 95%	Mean length	Coverage (%)
1, 1, 1, 1,	1			
Wald		(1, 1)	.	.
Score		(0.9383, 1)	0.06166	100
0.98,0.98,0.98,0.98	0.98			
Wald		(0.9441 ,1.0159)	0.0511	63.94
Score		(0.9048, 0.9961)	0.0877	95.23
0.96,0.96,0.96,0.96	0.96			
Wald		(0.9098, 1.0102)	0.0838	72.43
$Score_1$		(0.8753, 0.9880)	0.1096	95.90
0.94,0.94,0.94,0.94	0.94			
Wald		(0.8791, 1.0010)	0.1093	76.92
$Score_1$		(0.8480, 0.9778)	0.1261	96.10
0.92,0.92,0.92,0.92	0.92			
Wald		(0.8505, 0.9895)	0.1281	83.64
$Score_1$		(0.8220, 0.9663)	0.1417	96.01
0.90,0.90,0.90,0.90	0.90			
Wald		(0.8231 ,0.9769)	0.1431	87.19
$Score_1$		(0.7969, 0.9538)	0.1536	95.60

The simulations indicate that the stratified score interval works well when the proportions are equal. The coverage rates are around 95 and 96 percent and close to the nominal level. The Wald interval shows the expected pattern of decreasing coverage rates as the proportions get larger. The coverage rates are ranging from roughly 64 to 87 %. The Wald interval is also shorter than the score interval, which of course affect the coverage negatively. Increasing the number of stratums to 8 and maintaining the same sample and stratum sizes increase the coverage rates for the Wald interval now ranging from 0.6883 when the proportions are 0.98 and 0.9087 when the proportions are 0.9. The coverage rates for the score interval for the same setting are still about the nominal level, 0.95-0.96 percent.

In table 5 the proportions are slightly different. The sample sizes and strata sizes are the same as before. The stratified score interval assumes that the proportions are equal. The coverage rates for the score interval are consistently higher than the nominal 95 percent level. When the proportions are unequal the score interval gets more conservative, at least in this case. The score method will assume that the proportions are all equal to the estimate of the aggregated proportion. It cannot take the variance of the different proportions into account. The Wald interval has poor coverage also in this situation and produce narrower intervals compared to the score interval. The results here cannot be generalized but gives a picture of how the score interval behaves when the proportions are unequal. The coverage rates will depend on sample sizes, the total population and weights.

Table 4. Stratified score confidence interval, mean length and coverage based on 4 strata with nh=19, 36, 15 and 18 and Nh=312, 148, 74 and 40.

p_1, p_2, p_3, p_4	$\hat{p}_{str}$	C.I. 95 %	Mean length	Coverage %
0.96,0.97,0.98,0.99	0.967			
Wald		(0.9183, 1.0162)	0.0771	0.6219
Score		(0.8857, 0.9912)	0.0983	0.9611
0.99,.98,0.97,0.96	0.983			
Wald		(0.9547 ,1.0109)	0.0413	0.7706
Score		(0.9090 ,0.9969)	0.0860	0.9707
0.96,0,95,0.94,0.93	0.953			
Wald		(0.9013 , 1.004)	0.0864	0.7352
Score		(0.8652 ,0.9845)	0.1145	0.9736
0.99,0.96,0.93,0.90	0.968			
Wald		(0.9360, 1.001)	0.0546	0.8508
Score		(0.8872, 0.9916)	0.1027	0.9872
0.99,0.98,0.91,0.90	0.971			
Wald		(0.9393 ,1.0024)	0.0529	0.8545
Score		(0.8909 ,0.9927)	0.1006	0.9823

5. Application on ICT usage in Swedish enterprises

The suggested score interval is applied on the ICT usage in Swedish enterprises 2013. The question of interest is the shares of enterprises that use computers. The score interval is compared with the standard Wald interval that is the currently used method for interval estimation in the survey. The results are based on a limited collection of the aggregated proportions reported in the survey, this due to space considerations. Table 5 and 6 lists the estimated stratified proportion, the 95 % confidence interval and the length for the Wald and score interval. The data is stratified after industry and size and a total of 19 groups or aggregated measures are presented. The number between the brackets refers to the number of strata that the groups consist of.

Table 5. Wald and stratified score confidence intervals. Group, number of strata, estimated proportion and length of intervals.

Group (# of strata)	$\hat{p}_{str}$	C.I. 95 %	Length
1 (7)	1		
Wald		(1,1)	.
Score		(0.9754, 1)	0.0246
2 (7)	1		
Wald		(1,1)	.
Score		(0.9784, 1)	0.0216
3 (7)	1		
Wald		(1,1)	.
Score		(0.9314, 1)	0.0686
4 (7)	1		
Wald		(1,1)	.
Score		(0.9755, 1)	0.0245
5 (11)	1		
Wald		(1,1	.
Score		(0.9615, 1)	0.0385
6 (5)	1		
Wald		(1,1)	.
Score		(0.7436, 1)	0.2564
7 (12)	1		
Wald		(1,1)	.
Score		(0.9204, 1)	0.0796
8 (6)	1		
Wald		(1,1)	.
Score2		(0.9646, 1)	0.0354

In table 5 the estimated proportions in each strata is equal to 1. In this case the Wald interval produces zero width intervals. The length of the intervals is in most cases around 0.2 and 0.3. Group 3 and 7 have slightly wider intervals than the majority with a length of 0.0686 and 0.0796 respectively. The confidence interval for group 5 is considerably wider than the rest.

Table 6 displays the situation when not all proportions are equal. The estimated stratified proportions vary between roughly 0.92 and 0.99. In most groups the majority of stratums have proportions equal to one. There are some exceptions though. Group 12 and 15 have a considerably larger share of proportions other than one compared to the other groups.

The Wald intervals appear to have shorter length than the score intervals in general. In most cases the difference is small though. For group 9, 13 and 19, the length of the Wald interval is shorter than for the score interval. The longest intervals are found in group 9 and 12. In some instances the upper bound of the Wald interval exceeds one.

Table 6. Wald and stratified score confidence intervals. Group, number of strata, estimated proportion and length of intervals.

Group (# of strata)	$\hat{p}_{str}$	CI, 95 %	Length
9 (7)	0.9410		
Wald		(0.8373, 1.0443)	0.2070
Score		(0.7937, 0.9850)	0.1913
10 (19)	0. 9900		
Wald		(0.979 , 1.0040)	0.0250
Score		(0.9661, 0.9989)	0.0328
11 (7)	0.9740		
Wald		(0.9466, 1.0020)	0.0554
Score		(0.9331, 0.9903)	0.0572
12 (7)	0.9476		
Wald		(0.8845, 1.0107)	0.1262
Score		(0.8470, 0.9834)	0.1364
13 (33)	0.9587		
Wald		(0.9264, 0.9912)	0.0648
Score		(0.9183, 0.9796)	0.0613
14 (13)	0.9813		
Wald		(0.9632, 0.9994)	0.0362
Score		(0.9275, 0.9954)	0.0679
15 (7)	0.9529		
Wald		(0.9174, 0.9884)	0.0710
Score		(0.9042, 0.9778)	0.0736
16 (14)	0.9810		
Wald		(0.9608, 1.0012)	0.0404
Score		(0.9485, 0.9935)	0.0450
17 (28)	0.9819		
Wald		(0.9687 ,0.9952)	0.0265
Score		(0.9633 ,0.9912)	0.0279
18 (31)	0.9921		
Wald		(0.9864 ,0.9978)	0.0114
Score		(0.9658 ,0.9982)	0.0324
19 (14)	0.9242		
Wald		(0.8812, 0.9673)	0.0861
Score		(0.8756. 0.9548)	0.0792

6. Conclusions

The goal of this study was to find a procedure for interval estimation for a weighted sum of large proportions that performed better than the standard Wald interval. To this end a modified version of a stratified score interval proposed by Yan and Su (2010), which account for the finite population was tested.

Assuming the true proportions are equal the stratified score interval performs better than the standard Wald interval in terms of coverage. Based on the simulations, for proportions above 0.9, the stratified score interval has coverage probabilities close to the nominal 95 percent level when the true proportions are equal. Letting the true proportions vary slightly increase the coverage rates somewhat for the stratified score interval to reach above 95 percent nominal level. The standard Wald interval performs much worse in terms of coverage for large proportions. Further the stratified score interval does not have the inherent problems of the Wald interval, that is intervals stretching outside the limit of [0, 1] and zero width intervals when the proportions are exactly zero and one.

The conclusions are, under the assumption that the true proportions are equal, that the stratified score interval in general performs better than the standard Wald interval for large proportions. For a question regarding computer usage in Swedish enterprises with proportions reaching a hundred percent in most strata the method can be a good alternative.

Only the question of computer usage in Swedish enterprises has been addressed here. The survey consists of many other measures that do not render the same extreme values. The assumption of equal proportions is obviously not reasonable in many cases. It is difficult to say how different the proportions can be in order for the proposed method to work properly and if this can be measured and evaluated. Yan and Su (2010) do present an alternative method when proportions are unequal. This method could not be implemented due to the weighting scheme used here. A natural development of this method would be to find a stratified score interval with weights depending on relative stratum size that can be used when the proportions are unequal.

References

Agresti, A. and Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. The American Statistician. 52 119-126.

Agresti, A. and Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions, Result from Adding Two Successes and Two Failures. The American Statistician. 54, 280-288.

Brown ,L.T. Cai,T. T. DasGupta, A. (2001). Interval estimation for a binomial proportion, Statistical Science, 16, 101-117

Casella, G. (1986). Refining binomial confidence intervals, Canadian Journal of Statistics, 14, 113-129.

Casella, G. Berger, R.L. (2002).Statistical Inference, 2nd ed. Pacific Grove, Duxbury Press.

Decrouez, G, Robinson, A.P (2012). Confidence intervals for the weighted sum of two independent binomial proportions, Australian and New Zealand Journal of statistics, 54, 281-299.

Feng, C. (2006).On confidence intervals for proportions with focus on the U.S. National Health and Nutrition Examination Surveys. master thesis, Simon Fraiser University.

Hamada, M.S, Mitchell, B.L, Necker C.T. (2015) On uncertainty of a proportion from a stratified random sample of a small population, Journal of Applied Statistics, 42, 828-833.

Liu, Y.K. and Kott, P.S. (2007). Evaluating alternative one-sided coverage intervals for an extreme binomial proportion. Statistics of income, SOI paper series, National Agricultural Statistics Service.

Lohr, S.L. (2010).Sampling.Design and Analysis, 2nd ed, Boston, Brooks and Cole.

Newcombe, R. G. (1998a). Two-sided confidence intervals for the single proportion; comparison of seven methods. Statistics in Medicine 17 857-872.

Necombe, R.G. (1998b). Interval estimation for the difference between independent proportions: comparison of eleven methods. Statistics in Medicine. 17, 873–890.

Reiczigel, J, (2003). Confidence intervals for the binomial parameter: some new considerations, Statistics in medicine, 22, 611-621.

R Development Core Team.The R project for statistical computing. www.R-project.org.

Statistics Sweden, ICT usage in enterprises (2013), (SCB, Företagens användning av it 2013.) www.scb.se

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association, 22, 209-212.

Yan, X. Su, G (2010). Stratified Wilson and Newcombe confidence intervals for multiple binomial proportions, Statistics in Biopharmaceutical Research, 2, 329-335.

Appendix

R-code

```
#simulations
ph<-c()
nh<-c()
Nh<-c()

#Score interval

w<-Nh/sum(Nh)
pobs<-sum(w*ph)
Nhh<-Nh/(Nh-1)
Nhh[which(Nhh==Inf)] <- 0
m<- sum( (w^2) *(1/nh)*(1-nh/Nh)*Nhh )
k<-20000
CI<-matrix(NA,k,2)
alpha<-0.05
z<-qnorm(1-alpha/2)

#loop
for (i in 1:k){
x<-c(rhyper(4,Nh*ph,Nh-Nh*ph,nh))
p<-x/nh
#CI´s
A<- sum(w*p) + (m*z^2)/2
B<- (1 + m*z^2)
C<- m*sum(w*p)*(1-sum(w*p))
D<- (m^2*z^2)/4
L<- (A / B) - (z*sqrt(C + D))/B
U<- (A / B) + (z*sqrt(C + D))/B
CI[i,]<-cbind(L,U)
}

CIcov<-cbind( CI[,1] <= pobs & CI[,2] >= pobs )
Covprob<-mean(CIcov);Covprob
mean(CI[,1]);mean(CI[,2])
mean(CI[,2])-mean(CI[,1])

#Wald interval

w<-Nh/sum(Nh)
Pobs<-sum(w*ph)
alpha<-0.05
z<-qnorm(1-alpha/2)
Nhh<-Nh/(Nh-1)
Nhh[which(Nhh==Inf)] <- 0
k<-20000
CI<-matrix(NA,k,2)

#loop
for (i in 1:k){
x<-c(rhyper(14,Nh*ph,Nh-Nh*ph,nh))
```

```
p<-x/nh

phat<-sum(w*p)
Vp<-sum((w^2)*(1/nh)*(1-nh/Nh)*Nhh*(p*(1-p)))
L<-phat-z*sqrt(Vp)
U<-phat+z*sqrt(Vp)
CI[i,]<-cbind(L,U)
}

CIcov<-cbind( CI[,1] <= Pobs & CI[,2] >= Pobs )
Covprob<-mean(CIcov);Covprob
mean(CI[,1]);mean(CI[,2])
mean(CI[,2])-mean(CI[,1])


#Confidence intervals
ph<-c()
nh<-c()
Nh<-c()

#Score interval
w<-Nh/sum(Nh)
pobs<-sum(w*ph)
Nhh<-Nh/(Nh-1)
Nhh[which(Nhh==Inf)] <- 0
m<- sum( (w^2) *(1/nh)*(1-nh/Nh)*Nhh )
alpha<-0.05
z<-qnorm(1-alpha/2)
#CI´s
A<- sum(w*ph) + (m*z^2)/2
B<- (1 + m*z^2)
C<- m*sum(w*ph)*(1-sum(w*ph))
D<- (m^2*z^2)/4
L<- (A / B) - (z*sqrt(C + D))/B
U<- (A / B) + (z*sqrt(C + D))/B

CI<-cbind(L,U);CI

#Wald  confidence interval

w<-Nh/sum(Nh)
Pobs<-sum(w*ph)
alpha<-0.05
z<-qnorm(1-alpha/2)
Nhh<-Nh/(Nh-1)
Nhh[which(Nhh==Inf)] <- 0

phat<-sum(w*ph)
Vp<-sum((w^2)*(1/nh)*(1-nh/Nh)*Nhh*(ph*(1-ph)))
L<-phat-z*sqrt(Vp)
U<-phat+z*sqrt(Vp)
CI<-cbind(L,U);CI
```