Variance Estimation

Mark Baillie, James Brown, Alan Taylor and Owen Abbott

1. Introduction

A basic requirement of any estimate is a measure of its precision or uncertainty. This is typically provided by estimates of variance and associated confidence intervals. Derivation of computationally and statistically efficient[1] estimates of the variance of the coverage adjustment estimates is therefore crucial for the evaluation of these estimates. Variance estimates provide information on their level of accuracy and 95% confidence intervals will allow them to be appropriately compared to external population estimates.

The paper outlines the work undertaken to explore and evaluate an appropriate variance estimation methodology for use in the 2011 Census. As in 2001, these methods will provide the basis for obtaining confidence intervals for the primary estimates obtained from the coverage assessment methodology: Local Authority estimates by age-sex groups. However, the estimation process does not in general provide direct LA level estimates, as in most cases we group LAs into Estimation Areas to ensure sample sizes are sufficient for ratio estimation. Therefore in the first instance we discuss variance estimation at the Estimation Area level. The paper provides a comparison of the variance estimation methodology used in 2001 alongside an implementation of the bootstrap re-sampling method. The research presented provides the foundation for developing a Local Authority level variance estimation methodology.

2. Background

2.1 Variance Estimators

An important measure of a "good" variance estimator is consistency. This means that for repeated samples the estimator converges to the 'true' value of the parameter of interest (Wolter, 2007). Equally, confidence intervals derived from the variance estimation methodology should have good coverage probabilities. For example, for a 95% interval, over repeated samples, 95% of the estimated intervals will contain the true value of the estimate. These two characteristics may sometime conflict in that the variance estimator with the optimal coverage characteristics may not be consistent, and vice versa. The chosen estimator may therefore need to be based on a compromise between these characteristics. The choice of estimator should also include consideration of its ease of calculation. This is particularly important for more recently developed estimators that use computationally intensive methods which can have a high time overhead. If two estimators have similar properties the simpler method should be chosen to ensure the wider user community will have a clearer understanding and acceptance of the methodology.

For variance estimation in the 2011 Census a number of approaches can be considered and have been investigated: (i) The jack-knife approach, similar to 2001 approach, adjusted for the revised Census Coverage Study (CCS) sample design; and (ii) the bootstrap method. The jack-knife and bootstrap are described below.

2.1.1 The Jack-knife

In 2001 the jack-knife approach was applied, stratified by hard-to-count group within Estimation Area. Symmetric 95% Confidence intervals were then derived using the standard errors estimated

[1] Computational efficiency refers to the time and volume of computing resource required to produce a variance estimate. Statistical efficiency refers to whether the variance estimate of the parameter of interest is estimated in the best possible way with respect to specified criteria such as confidence interval width and coverage.

using this approach based on the assumption of a normally distributed estimator. An important point to note about these intervals is that they are not bounded by the Census count and therefore some intervals went below this theoretical lower bound. This theoretical lower bound does assume that there are negligible levels of over-count in the Census counts and this will not be a lower bound in circumstances with high levels of over-count.

2.1.2 The Bootstrap

The bootstrap method (Efron & Tibshirani, 1993; Wolter, 2007) has been developed to estimate the variance and confidence intervals of complex estimators, often derived from multistage survey samples. An advantage of the bootstrap method is that, at least in summary, it provides an easily understood methodology for the non-technical user. The method provides empirical variance estimates as well as empirical confidence intervals based on the empirical distribution of estimates over repeated sample replicates. These confidence intervals are based on an empirically derived sampling distribution and therefore do not make an assumption of normality. They are therefore potentially non-symmetric and so can provide appropriately bounded confidence intervals. However, to be able to adequately define the bounds for these intervals the replication has to be relatively large to ensue enough of the replicates provide an appropriate level of information to define the confidence interval bounds. For example, for 1000 replicates one would expect only 25 of these to be above and 25 to be below the 2.5% cut point on the empirical distribution that defines the 95% confidence bounds.

As Efron and Tibshirani (1993) note the distribution of $\hat{\theta}$ percentile confidence intervals can be used to provide superior nominal coverage levels rates for non-normal estimators. Both the jack-knife and bootstrap methods are particularly useful for complex estimators where other standard methods, such as Taylor series linearization require complicated derivations to produce a variance estimate. An additional advantage of the bootstrap method for our situation it that it may be possible to build an implementation that reruns the full estimation methodology repeatedly. This will allow for inclusion of additional adjustments, such as those for dependence or over-count, as integral parts of the variance estimation. This may provide advantages over the other approaches as they would have to estimate each component of variance separately and add them together to provide a final variance estimator. This is basically the approach used for variance estimation at the LA level in 2001 where an Estimation Area variance estimate was combined with a residual LA variance estimate.

To be able to define the variance and confidence intervals for totals the full variance-covariance matrix of the age-sex specific groups is required as the total is defined by the sum of the age-sex estimates. In this paper, each method has been extended to provide these.

2.2 Previous Research

Initial research comparing a bootstrap and jack-knife approach concluded that the bootstrap showed promise and has the potential to easily incorporate estimation at the LA level, as well as national adjustments such as those that are likely to be made for dependence and over-count (Baillie *et al* 2010). However, there were three main issues that required further consideration:

1. The jack-knife estimator is, in general, tracking the empirical variance closely. This confirmed simulation results from 2001 (Brown, 2000). The initial bootstrap estimator is, in general, over-estimating compared to the empirical variance. This then results in over-stating the relative standard error of estimates.

2. The resulting 95% confidence intervals from the jack-knife tend to have coverage around or just under 95%, again confirming results from 2001. However, the over-estimation of the

variance using the bootstrap feeds through to 95% confidence intervals with coverage consistently in excess of 95%.

3. The bootstrap approach also allows for the construction of empirical confidence intervals based on the bootstrap distribution. These have even higher coverage and there is the issue of correcting bias in the bootstrap distribution as well as the acceptability of non-symmetric confidence intervals.

This paper outlines the research undertaken to resolve these issues, and outlines the remaining work.

3. Over-Estimation with the Bootstrap

To implement the bootstrap we first create a pseudo-population by re-sampling with replacement from the CCS sampled areas. In the previous work by Baillie *et al* (2010) we sampled output areas (OAs) with replacement and then sampled from the selected postcodes with replacement to build the within OA populations. From this pseudo-population we then repeatedly applied the CCS sample design to generate the set of bootstrap estimates, and then the variance estimate and confidence interval. However, this second stage is not required because the dual-system estimate produced for each sampled OA encapsulates the within PSU variability. Using the variation between estimates at the OA level therefore captures both the between and within variability, so adding this extra variation by re-sampling within the OAs will tend to over-estimate the variance (Wolter, 2007). The same argument is used with the ultimate-cluster variance estimator; this just considers variation between the sampled PSUs as variation between estimates at the PSU level captures both the variation within the PSU feeding in to each PSU estimate as well as true variation between the PSUs.

As in the previous work, a simulation study is used to evaluate the performance of the variance estimators. Results from creating the pseudo-population by just re-sampling OAs and fixing the within OA sample of postcodes results in the bootstrap variance estimator tracking the empirical variance even more closely than the jack-knife approach. This can be seen in Table 1 and Figure 1, which compare the bias of the original 2-stage bootstrap with the revised bootstrap and the jack-knife.

Table 1. Variance estimates for 5 EAs comparing the jack-knife with the single and two-stage bootstrap methods.

EA	*METHOD*	*VARIANCE ESTIMATE*
KK	SIMULATION	30,433,847
	JACK-KNIFE	33,012,073
	BOOTSTRAP 95%	29,643,367
	2-STAGE BOOTSTRAP 95%	51,542,906
KO	SIMULATION	19,683,414
	JACK-KNIFE	21,418,196
	BOOTSTRAP 95%	19,468,872
	2-STAGE BOOTSTRAP 95%	34,301,484
LB	SIMULATION	167,990,994
	JACK-KNIFE	184,929,344
	BOOTSTRAP 95%	176,120,843
	2-STAGE BOOTSTRAP 95%	300,476,386
SH	SIMULATION	16,837,951
	JACK-KNIFE	16,916,600
	BOOTSTRAP 95%	17,452,585
	2-STAGE BOOTSTRAP 95%	32,322,299
NA	SIMULATION	15,190,457
	JACK-KNIFE	18,267,049
	BOOTSTRAP 95%	19,716,529
	2-STAGE BOOTSTRAP 95%	35,434,868

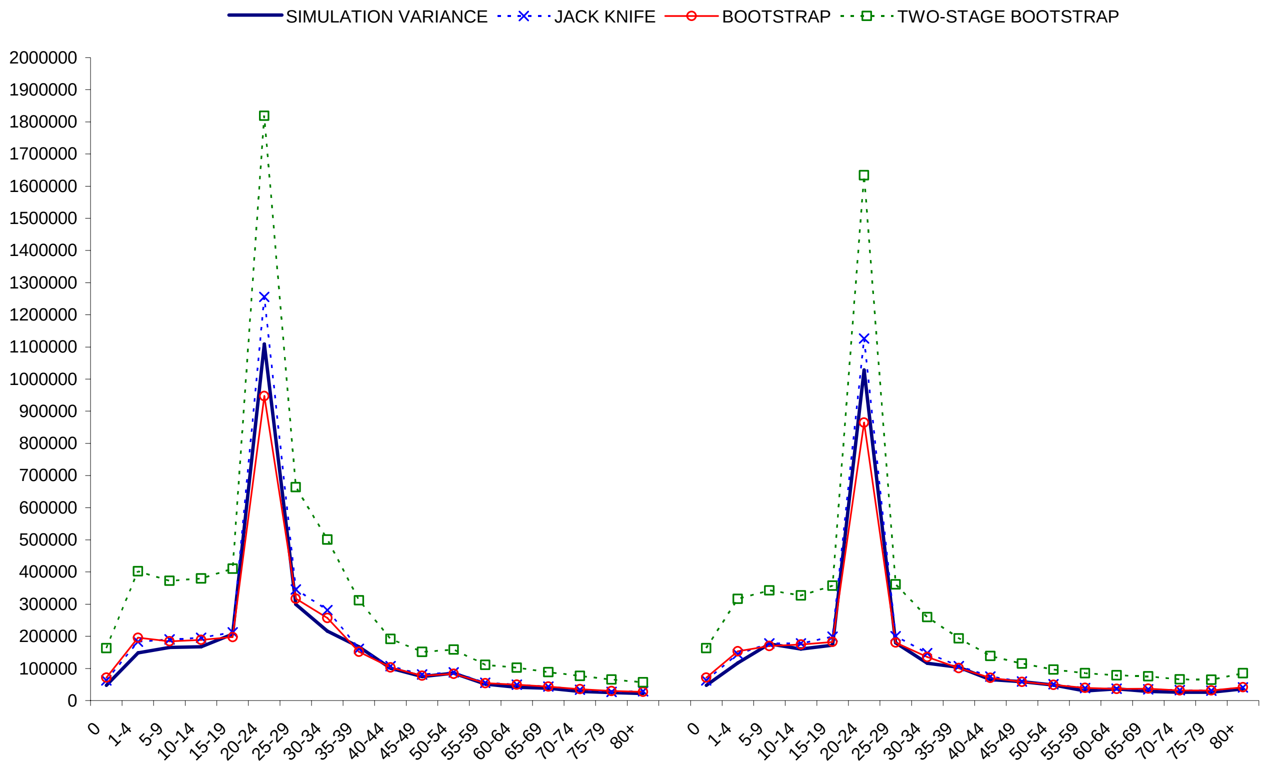


Figure 1. Summary plot of the variance estimates for KK. The simulation variance ('truth'), jack-knife and bootstrap mean estimates by age sex group are displayed. For comparison, the previous two-stage Bootstrap is also presented.

Table 2 and Figure 2 evaluate the coverage of confidence intervals derived from the alternative variance estimators.

Table 2. Coverage of the jack-knife and bootstrap. A number of bootstrap confidence intervals were implemented and compared.

EA	*METHOD*	*CI COVERAGE*
KK	SIMULATION	
	JACK-KNIFE	95
	BOOTSTRAP 95%	94.75
	BS BC	95.75
	BS EMP	94
	BS Approx T	95
	BCa	94.25
KO	SIMULATION	
	JACK-KNIFE	93
	BOOTSTRAP 95%	91.75
	BS BC	92.5
	BS EMP	91.75
	BS Approx T	92.25
	BCa	91.75
LB	SIMULATION	
	JACK-KNIFE	92.75
	BOOTSTRAP 95%	92.75
	BS BC	93.75
	BS EMP	93.5
	BS Approx T	92.75
	BCa	92.25
SH	SIMULATION	
	JACK-KNIFE	90.25
	BOOTSTRAP 95%	91
	BS BC	92.25
	BS EMP	92.75
	BS Approx T	92
	BCa	91.75
NA	SIMULATION	
	JACK-KNIFE	94
	BOOTSTRAP 95%	94.75
	BS BC	96
	BS EMP	95.75
	BS Approx T	96
	BCa	95.75

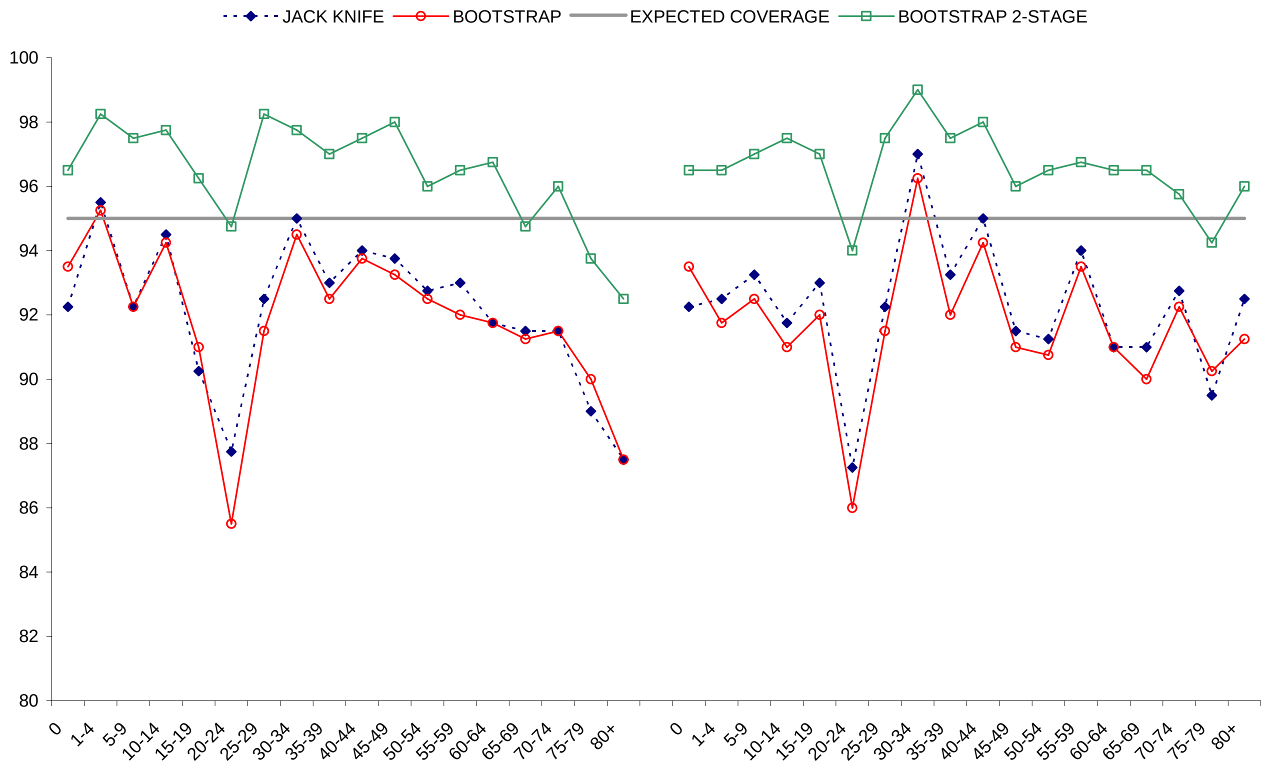


Figure 2. Coverage of the simulation population truth using jack-knife and bootstrap 95% CI. Summarised by age sex group.

4. Constructing 95% Confidence Intervals

The slight reduction in the bootstrap variance estimator now leads to standard +/-1.96 based 95% confidence intervals (i.e. *Z* is approximately distributed to the Normal distribution with zero mean and unit variance) that have similar coverage to the jack-knife. In other words, we can now recover similar performance to the 2001 approach with the advantage of bootstrapping for things like LA level estimates. However, like the jack-knife approach, the coverage of the 95% confidence intervals tend to be just below the 95% level.

The above assumption is valid as the sample size tends to infinity but is only an approximation when the sample size is small and/or finite. A simple additional correction is to recognise that we should be using the t-distribution rather than a standard normal e.g.

$$Z = \frac{\hat{\theta} - \theta}{\delta / \sqrt{n}} \sim \mathrm{t}_{(n-1)}$$

In other words, *Z* is approximately distributed to the t-distribution with n-1 degrees of freedom. Work by (DiCiccio and Efron, 1996) look at using a simulation within the bootstrap to get the correct t-distribution. A second bootstrap simulation is required to estimate the standard error for each of the parent bootstrap re-samples. For small finite samples results indicate this is a better approximation than using the standard normal distribution.

We explored this method using a t-distribution with n-1 degrees of freedom, where n is the number of OAs sampled within the estimation area. For a sample size of 45 OAs, this would be around 2.01

for the confidence interval rather than the standard 1.96. The adjustment is small but the loss of coverage is also small. This adjustment can also be applied to the jack-knife confidence intervals.

Using the empirical distribution to produce the confidence interval is an attraction of the bootstrap. Related research indicates that both 95% Normal approximation and Bootstrap-t intervals have good theoretical coverage but can tend to be erratic in practice. Additionally, as discussed above, often no statistical formula is available to estimate the standard error for the parameter of interest therefore a double bootstrap routine is required increasing processing time exponentially.

Percentile or empirical confidence intervals are simpler to use and are assumed to be more robust but comparisons with exact intervals highlight that these intervals can provide poor coverage in certain cases. There are two main reasons why: (i) bootstrap estimates are biased with respect to the original estimate, and (ii) the standard error varies with the value of the estimate (e.g. for each bootstrap resample we will have a different standard error estimate). In small samples the percentile method, which simply uses the alpha/2 and 1-alpha/2 percentiles of the bootstrap distribution to define the interval, performs well with an unbiased estimator. However, coverage can be affected when applied to a biased estimator.

The Bias Corrected and Accelerated (BCa) interval (Efron, 1987) was designed to address both issues. The Bias component is estimated as a median bias; a systematic under or overestimation of the original sample estimate $\widehat{\theta}$ when calculated through with each of the B bootstrap re-samples e.g $\widehat{\theta}_b^{\,*}$. Therefore the bias component is estimated by:

$$\widehat{z}_0 = \Phi^{-1}\left(\frac{\#\{\widehat{\theta}_b^{\,*} < \widehat{\theta}\}}{B}\right)$$

This is essentially the proportion of bootstrap resamples whose estimate of the parameter of interest is less than the full sample estimate of that statistic. This proportion is entered into the inverse of the standard normal distribution e.g. $\Phi^{-1}(0.975) = 1.96$.

The acceleration or skew component $\hat{a}$ is an estimate of the rate of change of the standard error of the parameter of interest with respect to the true value of the statistic. This provides an estimate of the influence of each sample observation as well as how the empirical bootstrap distribution skews from an idealised normal distribution.

To estimate the acceleration value, a jack-knife simulation can be used, removing one sample point in turn.

$$\widehat{a} = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{.} - \hat{\theta}_i\right)^3}{6\left\{\sum_{i=1}^{n}\left(\hat{\theta}_{.} - \hat{\theta}_i\right)^2\right\}^{\frac{3}{2}}}$$

$\hat{\theta}_i$ is the jack-knife estimator of the parameter of interest with the i-th observation removed and $\hat{\theta}_{.}$ is the mean of the jack-knife samples for a single bootstrap sample b.

Given an estimate for both quantities a BCa confidence interval is constructed as follows:

I. For a 95% CI set $\alpha = 0.025$

II. Calculate:

a. $\alpha_1 = \Phi\left(\hat{z}_0 + \dfrac{\hat{z}_0 + \Phi^{-1}(\alpha)}{1 - \hat{a}(\hat{z}_0 + \Phi^{-1}(\alpha))}\right)$

b. $\alpha_2 = \Phi\left(\hat{z}_0 + \dfrac{\hat{z}_0 + \Phi^{-1}(1-\alpha)}{1 - \hat{a}(\hat{z}_0 + \Phi^{-1}(1-\alpha))}\right)$

III. Calculate:

a. $N_1 = B \times \alpha_1$

b. $N_2 = B \times \alpha_2$

IV. Take the N_1 and N_2 ordered bootstrap sample estimates as the lower and upper confidence limits.

This approach has been implemented in a set of Estimation Areas and is now leading to good coverage for the empirical confidence intervals (see Appendix for results for some estimation areas). Based on the findings of this research we will implement BCa confidence intervals derived from the application of the bootstrap. This allows us to retain flexibility in how the confidence bounds are calculated. Our default strategy will be to implement the asymmetric confidence bounds described above.

5. Publishing Confidence Intervals

One advantage of the bootstrap approach is that, based on the empirical distribution, the confidence bounds do not have to be assumed to be symmetric, as was assumed in the 2001 method. Therefore more appropriate asymmetric confidence bounds can be derived as they appropriately reflect the confidence associated with the estimates given they will be bounded by the Census count. For areas with moderate to high levels of undercount these confidence bounds will be approximately symmetric. They will only be asymmetric when the estimated undercount rate is small. We will use these asymmetric confidence intervals for the Census Quality Assurance process and for all census outputs. Information notes will be provided with these confidence bounds to ensure their appropriate use and interpretation.

However, many users may not be used to seeing non-symmetric confidence bounds and therefore we need to ensure that the associated notes deal with any concerns or potential misunderstanding. We will look at best practice to try to achieve this. We may also wish to test our notes with some users in advance of the first release publication, and prepare briefing material for those who will be presenting the results/answering questions.

6. Remaining work

Given the revised bootstrap approach is performing consistently, performance at the LA level needs to be considered. Work is ongoing to simulate and implement this.

7. References

Brown, J. J. (2000) Design of a census coverage survey and its use in the estimation and adjustment of census underenumeration. *University of Southampton*, unpublished PhD thesis.

Baillie, M., Brown, J., Taylor, A. and Abbott, O. (2010) An evaluation of Bootstrapping for Variance Estimation. Internal Report. Available on request.

DiCiccio, T. J. and Efron, B (1996). Bootstrap Confidence Intervals. Statistical Science 11(3), pages 189-228.

Efron, B. (1987). Better Bootstrap Confidence Intervals. Journal of the American Statistical Association (Journal of the American Statistical Association, Vol. 82, No. 397) 82 (397): 171–185. doi:10.2307/2289144. JSTOR 2289144. http://jstor.org/stable/2289144.

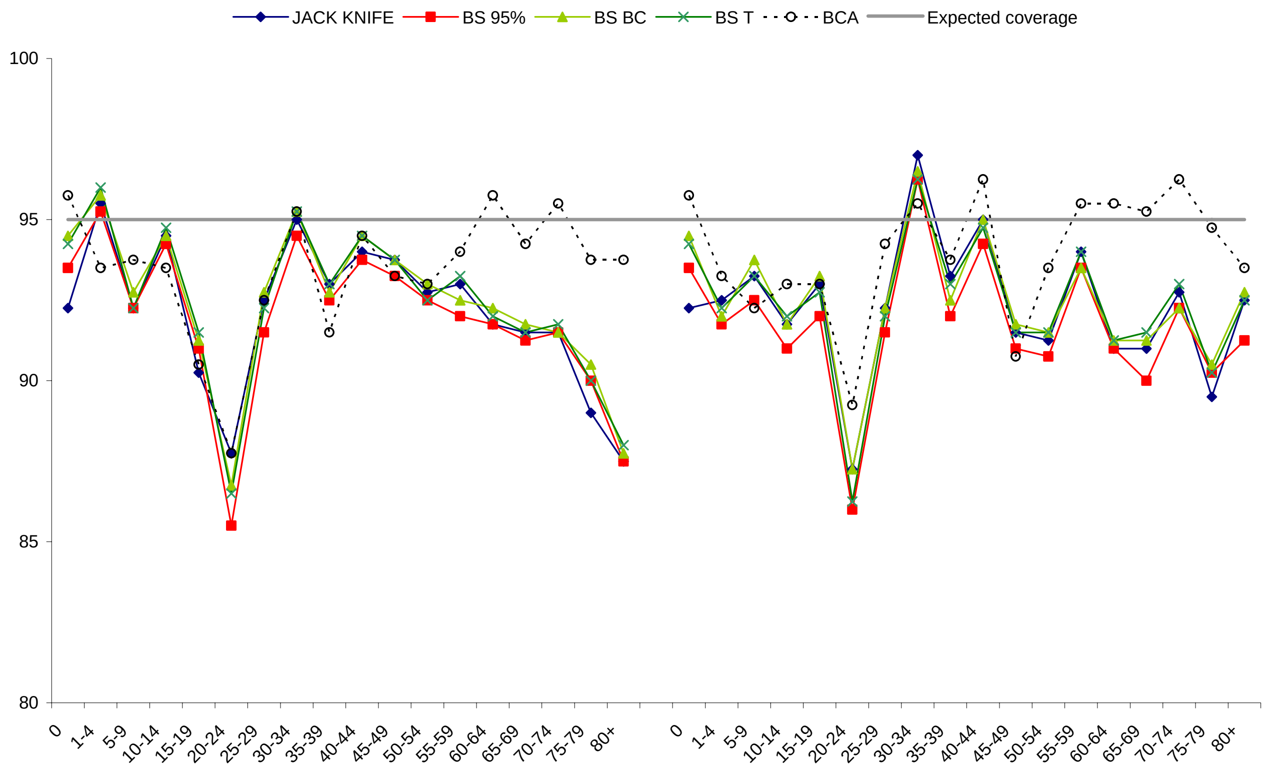
Efron, B. & Tibshirani. R.J. (1993). An Introduction to the Bootstrap. Chapman and Hall: New York.

ONC (2000). One Number Census Steering Committee Report SC 00/16. Available at http://www.statistics.gov.uk/census2001/pdfs/sc0016.pdf

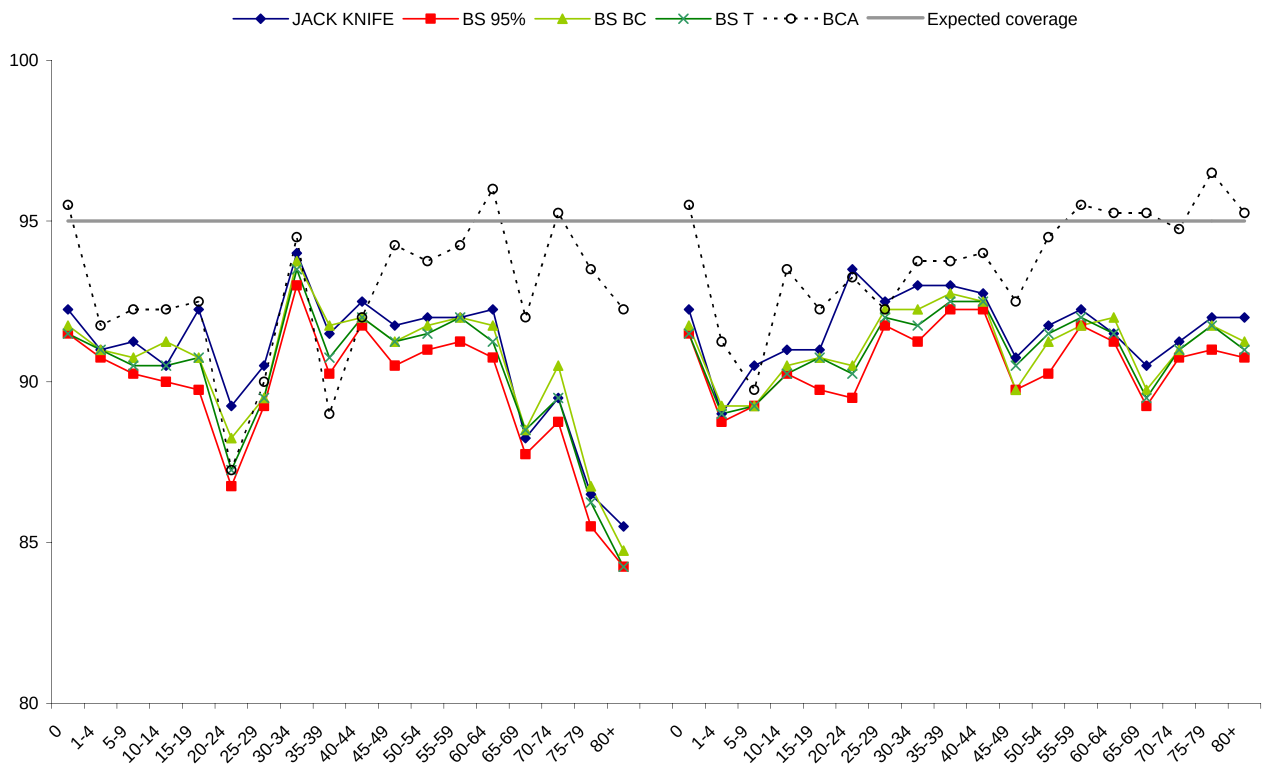
Wolter, K. M (2007). Introduction to Variance Estimation. 2nd Edition. New York: Springer.

Appendix

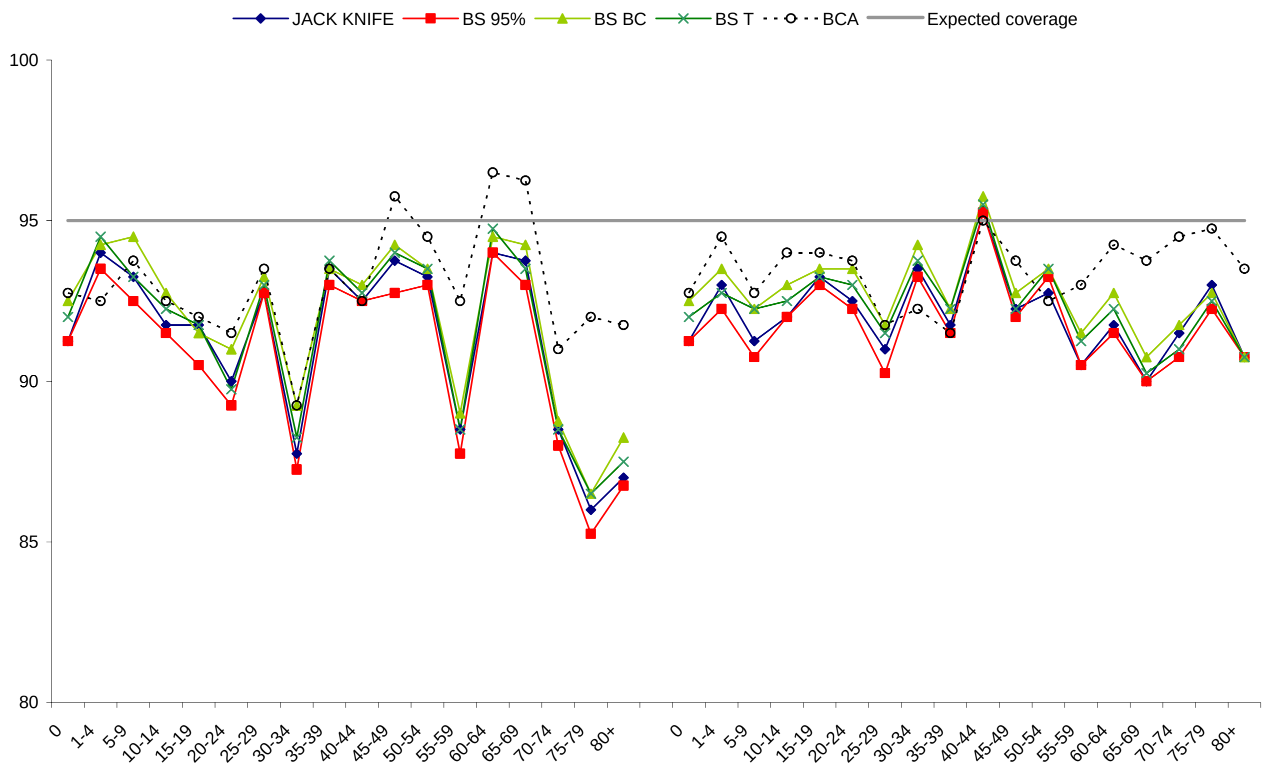
KK



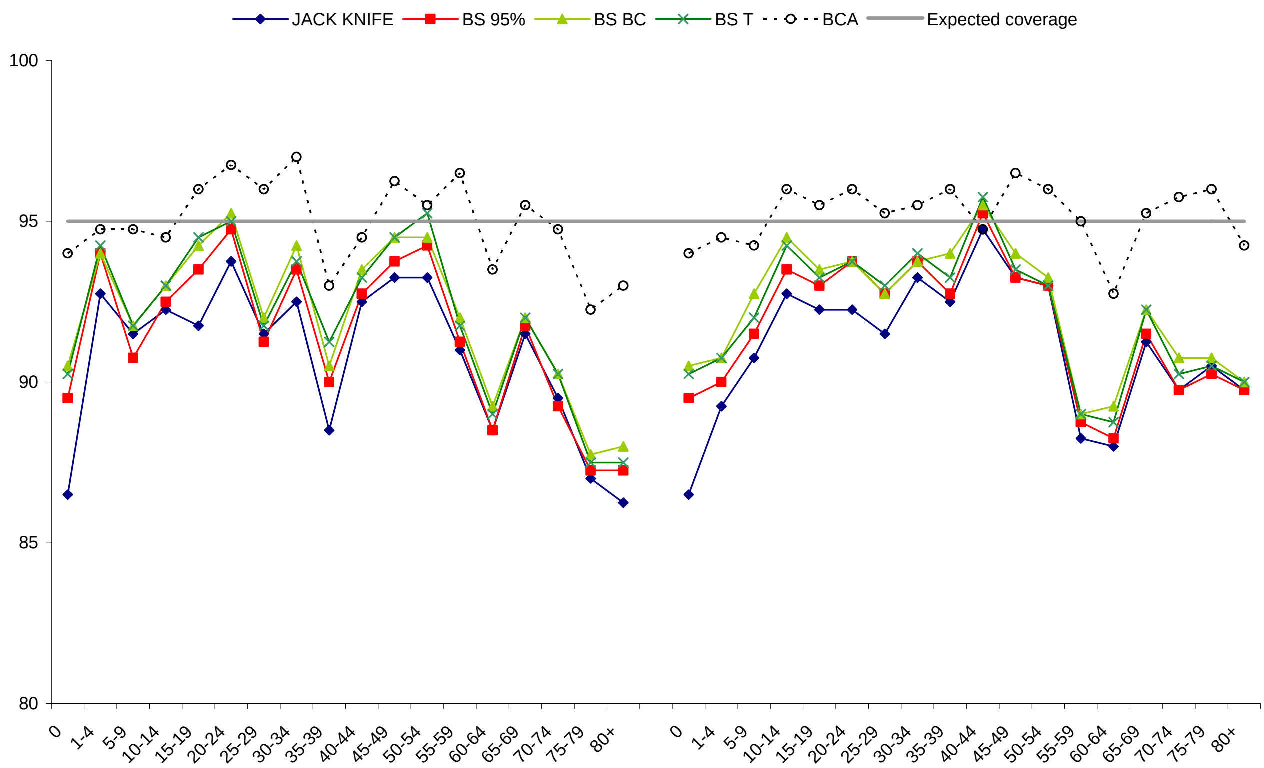
KO



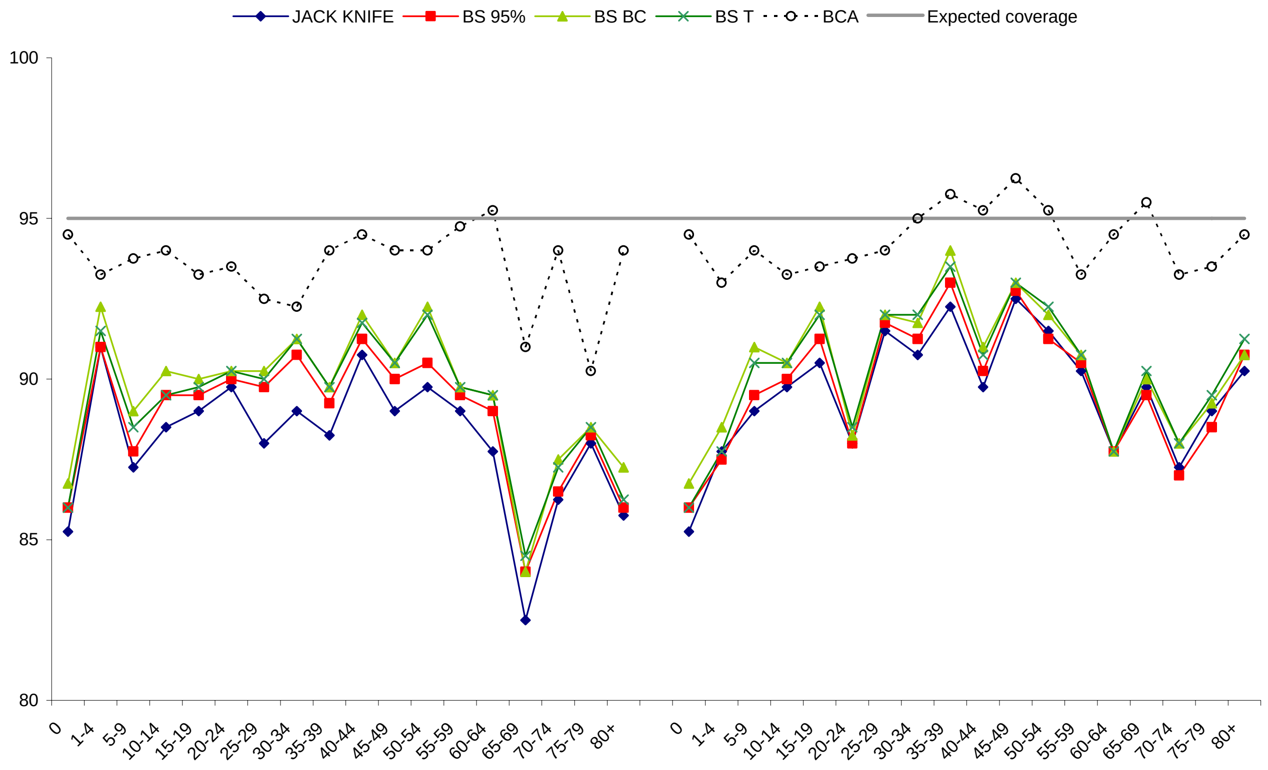
LB


JACK KNIFE
BS 95%
BS BC
BS T
BCA
Expected coverage
100
95
90
85
80
0
1-4
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80+

NA


JACK KNIFE
BS 95%
BS BC
BS T
BCA
Expected coverage
100
95
90
85
80
0
1-4
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80+

SH



JACK KNIFE
BS 95%
BS BC
BS T
BCA
Expected coverage
100
95
90
85
80
0
1-4
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80+
0
1-4
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80+