Variance Estimation in NMSSA

1. Introduction

This appendix describes the standard procedures undertaken to calculate design effects in NMSSA on an annual basis.

Design effects

A design effect is the ratio of the variance of an estimate calculated for a complex sample design compared to the variance calculated as if the sample was a simple random sample.

$$d = \frac{Var(\theta)_{Complex}}{Var(\theta)_{SRS}}$$

Design effects are calculated for all key groups and subgroups in NMSSA each year. Calculations are generally restricted to assessment data where the whole NMSSA sample has been involved in the assessment.

Effective sample size

The design effect tells us the extent of the loss of efficiency in variance estimation caused by the complex sample design. This loss of efficiency can be couched in terms of an *effective sample size.* In a simple random sample (SRS) the sample size influences the precision (efficiency) with which estimates can be calculated. A decrease in the sample size leads to a decrease in efficiency, and subsequently an increase in the variance of an estimate. Using the design effect we can calculate the effective sample size, the size of a SRS that would give us the same efficiency as our complex sample.

$$n_{eff} = \frac{n_{complex}}{d(\hat{\theta})}$$

where n_{eff} = the effective sample size

$n_{complex}$ = the sample size selected under the complex design

d = design effect

θ = the estimate in question

2. Variance estimation for complex survey data

The NMSSA sample is a stratified cluster sample, with a new sample being selected each year. Schools are the primary sampling unit and are stratified implicitly by region, decile and size. One hundred schools at each of Year levels 4 and 8 are selected. Within selected schools up to 25 students are systematically (randomly) selected rendering an (approximately) equal probability sample of students representing the New Zealand student population.

For reporting purposes key variables are year level, decile, gender and ethnicity.

Incorporating sample weights

Each year an investigation is carried out to confirm that it is **not** necessary to use sample weights in analysis. The current NMSSA sampling method ensures that the achieved sample represents the NZ student population accurately, and it is unlikely that sample weights will be needed unless the sampling method changes. In the event that sample weights are deemed necessary, they can be readily incorporated into the variance estimation routines.

Post-stratification and collapsing rules

The NMSSA sample is post-stratified by quintile, gender and ethnic group.

Ethnicity grouping: Throughout general analysis and reporting NMSSA allows for individuals to be assigned to multiple ethnicities. In the current context, however, allowing for multiple ethnicities results in many, very small post-strata. Approximately 10 percent of students at Year 4 and at Year 8 are reported as belonging to multiple ethnicities. For the purposes of variance estimation NMSSA uses a 'prioritised' approach to ethnicity. See the stratum collapsing rules below.

The Year 4 and Year 8 samples are treated separately. Small post-strata (less than 15 members) post-strata have to be collapsed[1]. The following collapsing rules are applied, in order, to small post-strata. After each collapsing step, strata are re-assigned and stratum size re-calculated. If there are remaining small strata, the next collapsing step is applied.

1. Remove 'other' classification from students who already belong to any of NZE[2], Māori, Pacific, or Asian.
2. Small strata containing dual ethnicities are collapsed into prioritised ethnicity groups: Māori → Pacific → Asian → NZE.
   Example: A small stratum specified by quintile 3-Female-Māori/Asian would be subsumed into the Quintile 3-Female-Māori stratum.
3. Collapse remaining small ethnicity strata into the appropriate gender group.
   Example: A small stratum identified by quintile 4-Male-Pacific would be subsumed into a quintile 4-Male stratum.
4. Remaining small strata are collapsed into the appropriate quintile stratum.
   Example: A small stratum identified by quintile 1-Female would be subsumed into a quintile 1 stratum.
5. Finally any small strata left make up a 'mop-up' stratum, with no specific quintile, gender or ethnic identification.

[1] For the purposes of variance estimation, Heeringa, West, & Berglund (2010 p.43) suggest that collapsing post-strata so that each contains a minimum of 15-25 members is advisable.

[2] New Zealand European

3. Choosing a variance estimation method for NMSSA

In previous years NMSSA has carried out analyses using a) Jackknife and b) Taylor series linearisation[3] methods for variance estimation, and compared results. These two methods render almost identical results for the NMSSA sample design.

With the introduction of plausible values methodology in NMSSA 2015 to estimate population statistics, it has become practical to use the Taylor series linearisation method for variance estimation in preference to the Jackknife method. Analysis with plausible values involves repeating every analysis multiple times – one for each set of plausible values generated. The Jackknife is a time-consuming, computer-intensive estimation method, whereas the Taylor series approximations can be calculated comparatively quickly.

4. Results and recommendations

In NMSSA design effects generally vary between about 1.0 and 2.5. Even with the larger design effects the confidence intervals do not increase in width very much. A general increase in width of less than 1.0 NMSSA scale score point is usually observed.

It is recommended that, for ease of calculation and to absorb most of the variance bias caused by the NMSSA complex sample design, a factor or multiplier of **0.7** should be used to reduce the sample size in standard error calculations. This assumes a design effect of 1.43, which is close to most design effects calculated.

The factor of 0.7 used to calculate an effective sample size is checked each year, employing the standard procedures set out in this paper. Unless it appears that a very different factor should be used, a standard 0.7 is recommended. While the sample selection methods remain the same, this is unlikely to change. See the example on the following page.

Example: Calculate the standard error of a NMSSA mean

m_x = estimated mean of variable x

Under a simple random sample we would use

s_m = standard error of the mean = $\frac{s}{\sqrt{n}}$

Applying the recommended factor to account for a complex sample design we use

s_m* = standard error* of the mean = $\frac{s}{\sqrt{n*0.7}}$

5. References

Heeringa, S. G., West, B. T., Berglund, P. A. (2010). *Applied survey data analysis*. Taylor and Francis Group, LLC.

Lumley, T. (2004). Analysis of complex survey samples, *Journal of statistical software 9*(1), pp. 1-19.

Software

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria. URL http://www.R-project.org/.

Lumley, T. (2014), *survey: Analysis of complex survey samples*. R package version 3.30. https://rdrr.io/rforge/survey/

[3] Taylor series approximations of complex sample variances for sample estimates of means and proportions have been widely used since the 1950s (Heeringa et al., 2010). It is not a replication method like the Jackknife and the bootstrap, but uses Taylor series approximations to estimate variances. When the sample is reasonably standard the TSL method generally offers similar results to the Jackknife.