Computational Statistics and Data Analysis 85 (2015) 67–83

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Sample size methods for constructing confidence intervals for the intra-class correlation coefficient

Kevin K. Dobbin [a,*], Alexei C. Ionan [b]

[a] Department of Epid. and Biostat., University of Georgia, Athens GA, Georgia
[b] Department of Statistics, University of Georgia, Athens GA, Georgia

HIGHLIGHTS

- Sample size methods and programs for reliability studies are developed.
- The two-way balanced analysis of variance model without interaction is the focus.
- The sample size guarantees a user-specified mean confidence interval width.
- Modified large sample and generalized confidence interval methods are used.
- Novel computational algorithms are developed, studied, and implemented.

ARTICLE INFO

Article history:
Received 25 February 2014
Received in revised form 13 November 2014
Accepted 14 November 2014
Available online 9 December 2014

Keywords:
Sample size
Intraclass correlation coefficient
Variance components
Generalized confidence intervals (GCI)
Modified large sample (MLS)
Confidence intervals
Control variates
Rao-Blackwellization

ABSTRACT

The intraclass correlation coefficient (ICC) in a two-way analysis of variance is a ratio involving three variance components. Two recently developed methods for constructing confidence intervals (CI's) for the ICC are the Generalized Confidence Interval (GCI) and Modified Large Sample (MLS) methods. The resulting intervals have been shown to maintain nominal coverage. But methods for determining sample size for GCI and MLS intervals are lacking. Sample size methods that guarantee control of the mean width for GCI and MLS intervals are developed. In the process, two variance reduction methods are employed, called dependent conditioning and inverse Rao-Blackwellization. Asymptotic results provide lower bounds for mean CI widths, and show that MLS and GCI widths are asymptotically equivalent. Simulation studies are used to investigate the new methods. A real data example is used and application issues discussed. The new methods are shown to result in adequate sample size estimates, the asymptotic estimates are accurate, and the variance reduction techniques are effective. A sample size program is developed.[1] Future extensions of these results are discussed.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The two-way crossed analysis of variance layout is commonly encountered in psychometry (McGraw and Wong, 1996), radiology (Belge et al., 2006; Hing et al., 2007), inter-rater reliability studies (Potempa et al., 1995; Eliasziw et al., 1994), assay reproducibility studies (McShane et al., 2000; Dobbin et al., 2005), and gauge repeatability and reproducibility studies (Burdick et al., 2005). The intra-class correlation coefficient (denoted ICC_b below) represents agreement between

* Corresponding author.
E-mail address: dobbinke@uga.edu (K.K. Dobbin).

[1] R program can be downloaded at http://dobbinuga.com.

http://dx.doi.org/10.1016/j.csda.2014.11.010
0167-9473/© 2014 Elsevier B.V. All rights reserved.

laboratories, raters, or instruments in this layout, depending on the context. The ICC_b is a function of three variance components. A common goal of inter-rater reliability studies is to construct a confidence interval for the ICC_b. But traditional interval methods based on maximum likelihood asymptotics perform poorly (for examples of this poor performance, see supplement (Appendix B) and (Cappelleri and Ting, 2003)). As a result, both Modified Large Sample (MLS) interval (Cappelleri and Ting, 2003) and Generalized Confidence Interval (GCI) (Weerahandi, 1993; Burdick et al., 2005) methods were developed. MLS and GCI provide nominal coverage for the ICC_b (Cappelleri and Ting, 2003), (Ionan et al., 2014). But, no methods currently exist for determining the sample size required to construct GCI or MLS intervals in this layout.

The GCI and MLS intervals are constructed using exact statistical methods (Weerahandi, 1995). Sample sizes for confidence intervals based on exact methods have been developed for some important contexts (Bonett, 2002; Zou, 2012), but not the one addressed in this paper. Bonett (2002) developed a sample size method for the intraclass correlation coefficient for settings in which the *ICC* is a function of two variance components. None of the models in Bonett (2002) match the setting addressed in this paper. Also, Bonett's approach cannot be adopted to this setting because it is based on an asymptotic approximation to the variance of the ICC estimate. But no adequate approximation to this variance is available for the setting studied in this paper. Moreover, the MLS and GCI methods, since they do not use such a variance estimate in their construction (that is, they are based on exact calculations), may have widths unrelated to this variance. Nor can the MLS or GCI widths be expressed algebraically as simple functions of the sample size. Turning to the work of Zou (2012), he addressed the important setting of a one-way layout; the approach used an estimate of the width of the Wald interval for the sample size calculations. Zou also interestingly controlled the probability that the width is below a bound, rather than the expected width. Zou's Wald interval approximation approach is not well suited to the context addressed in this paper because the ICC_b estimator in the two-way layout is often highly skewed, the asymptotic convergence in this context is slow, and Wald intervals do not perform well in this setting.

The ICC_b in a two-way analysis of variance layout without interaction was discussed in McGraw and Wong (1996) and studied in Saito et al. (2006). McGraw and Wong (1996) did not give sample size formulas; they provided formulas for the lower and upper bounds of a confidence interval (their Table 7, Case 2A); in that context, the ICC_b is called the "degree of absolute agreement among measurements", the "criterion-referenced reliability", or the "Type I ICC". Importantly, the formulas from Table 7 in McGraw and Wong for the specific setting ICC(A, 1) and case 2A, which were taken from Shrout and Fleiss (1979) and developed by Haggard (1958) in 1958, only work well when the degrees of freedom for each mean square is large (Montgomery, 2013, p. 600; Cappelleri and Ting, 2003; Ionan et al., 2014). The MLS and GCI methods studied in this paper, on the other hand, work well even if the degrees of freedom for one or more mean squares are small; thus the MLS and GCI are more robust and appropriate to use for sample size estimation. Saito et al. (2006) studied different experimental designs for this model when $r_0 = 1$. To compare the efficiency of different designs, they used an approximation to $\mathrm{Var}(\mathrm{Ln}(ICC_b))$ based on the delta method (that is, a Taylor series expansion). But since the GCI and MLS methods are based on exact calculations, their widths may not be related to the efficiencies (or standard errors) of the point estimates. At one point in their discussion, Saito et al. (2006) did use exact methods to construct the intervals and compare widths (their Table III), but strictly in a limited context of comparing optimal allocation patterns, not sample size estimation. Since the primary objective of the 2006 paper was to compare designs, they did not present an explicit sample size method as we do here, or a set of tools for determining sample size for specific settings. In our study, we also allow $r_0 \geq 1$ because this is a reasonable design in some cases and is not uncommon in biomarker studies (MAQC Consortium, 2006; Dobbin et al., 2005; Polley et al., 2013; McShane et al., 2000).

The sample size estimates produced by our method ensure that the mean confidence interval width is below a user-specified target width. Ideally, one may wish to control the actual width of an interval to be below a target, rather than the mean width. But the actual width of both the GCI and MLS intervals are functions of the observed mean squares; since the observed mean squares are unknown at the study design phase, it is not feasible to control the actual interval width itself. One approach might be to "plug in" estimates of the values of these variance components that will be observed when the experiment is run. But such an approximation approach seems suboptimal since it leaves open the question of the sensitivity of the approximation to mis-specification of the variance component values that will be observed. Controlling the expected widths as we do in this paper averages over the observed values of mean squares. This averaging requires integrating the unobserved mean squares out of the width formulas. For both the MLS and GCI, this results in complex multiple integrations without clean analytic solutions. A variety of statistical computation methods are employed to make the integrations computationally tractable.

Statistical computation methods are reviewed in Robert and Casella (2013). As shown below, classic Monte Carlo integration is inadequate for sample size determination because of the computational demands. For the GCI sample size method, importance sampling (Ripley, 2006) was investigated but produced relatively minor computational gains. The method of Rao-Blackwellization (Robert and Casella, 2013), which involves conditioning on variables so as to reduce the variance, produced substantial variance reduction for estimates of the cumulative distribution function. We modified this method and call the resulting procedure inverse Rao-Blackwellization. Control variates (Rothery, 1982) are another widely used variance reduction technique. They depend on identification of an effective control variate. By investigation we identified a control variate for the ICC_b. For the MLS procedure, we find that the ICC_b width – which is initially a function of three variance components – can be rewritten as a function of two dependent F distributions. Subsequently, using mathematical manipulations we develop a method which we call "dependent conditioning". This method is closely related to two-stage Gibbs sampling (e.g., Robert and Casella, 2013). Gibbs sampling requires assessment of the convergence

Download English Version:

https://daneshyari.com/en/article/417434

Download Persian Version:

https://daneshyari.com/article/417434

Daneshyari.com