COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 109 (2013) 293–304

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Estimation of the concordance correlation coefficient for repeated measures using SAS and R

Josep L. Carrasco[a,*], *Brenda R. Phillips*[b], *Josep Puig-Martinez*[a], *Tonya S. King*[b], *Vernon M. Chinchilli*[b]

[a] *Biostatistics Unit, Public Health Department, University of Barcelona, Barcelona, Spain*
[b] *Department of Public Health Sciences, College of Medicine, Pennsylvania State University, Hershey, USA*

ARTICLE INFO

Article history:
Received 29 December 2011
Received in revised form
4 September 2012
Accepted 4 September 2012

Keywords:
Concordance correlation coefficient
Linear mixed models
Repeated measures
U-statistics
Variance components

ABSTRACT

The concordance correlation coefficient is one of the most common approaches used to assess agreement among different observers or instruments when the outcome of interest is a continuous variable. A SAS macro and R package are provided here to estimate the concordance correlation coefficient (CCC) where the design of the data involves repeated measurements by subject and observer. The CCC is estimated using U-statistics (UST) and variance components (VC) approaches. Confidence intervals and standard errors are reported along with the point estimate of the CCC. In the case of the VC approach, the linear mixed model output and variance components estimates are also provided. The performance of each function is shown by means of some examples with real data sets.

1. Introduction

Measuring the amount of agreement between two observers or measurement methods is a common research goal. Among the methods to assess agreement when data is continuous, the concordance correlation coefficient (CCC) introduced by [1] has risen as one of the most used, studied and implemented approaches, see for example [2–5]. Briefly, the CCC is a standardized coefficient that takes values between −1 and 1, where −1 means perfect disagreement, 0 translates to an independence situation (all the readings are at random), and 1 indicates perfect agreement. Lin [1] expressed the CCC as a function of the means, variances and covariances of the bivariate distribution of two observers. Carrasco and Jover [4] demonstrated the equivalence between the CCC and the appropriate intraclass correlation coefficient (ICC). Thus the CCC is expressed in terms of the variance components of a linear mixed model. This approach readily allows the CCC to be defined for more than two observers and non-structured repeated measures to be constructed as replicates. However, this approach is not adequate for data from a longitudinal design where the repeated measurements (e.g. visits) are ordered within each subject.

King et al. [6] introduced the CCC for repeated measures to assess agreement between two observers for both non-longitudinal and longitudinal designs. Subsequently, Carrasco et al. [7] developed the CCC for longitudinal repeated measurements as the appropriate specification of the intraclass correlation coefficient from a variance components linear mixed model.

In this paper, we introduce a SAS macro and an R package to estimate the CCC for repeated measures using the methods proposed in King et al. [6] and Carrasco et al. [7]. The function

* *Corresponding author*. Tel.: +34 667231387.
E-mail address: jlcarrasco@ub.edu (J.L. Carrasco).
0169-2607/$ – see front matter
http://dx.doi.org/10.1016/j.cmpb.2012.09.002

gives standard errors and confidence intervals as well as the necessary information to make inferences about the CCC.

The CCC for repeated measures and other statistical details are included in Section 2. The details of the SAS macro and R package are provided in Section 3. The performance of each approach is shown in Section 4 through three examples of real data. Finally, Section 5 includes a brief summary.

2. The CCC for repeated measures

2.1. Definition

Lin [1] defined the CCC for two observers based on their variances, covariance and means as

$$\rho_{\mathrm{CCC}} = \frac{2\sigma_{12}}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2}$$

where σ_{12}, σ_1^2, σ_2^2, μ_1 and μ_2 are the covariance, variances and means of observers.

King et al. [6] extended the concordance correlation coefficient for measuring agreement between two observers with p repeated measures on n subjects.

Let $\mathbf{Y}_1$ and $\mathbf{Y}_2$ $n \times p$ matrices correspond to the measurements of first and second observers respectively. Let $\mathbf{D}$ be an arbitrary weight matrix with dimension $p \times p$. The CCC for repeated measures is defined as

$$\rho_{\mathrm{CCC,RM}} = 1 - \frac{E(\mathbf{Y}_1 - \mathbf{Y}_2)'\mathbf{D}(\mathbf{Y}_1 - \mathbf{Y}_2)}{E_I(\mathbf{Y}_1 - \mathbf{Y}_2)'\mathbf{D}(\mathbf{Y}_1 - \mathbf{Y}_2)} = \frac{\sum_{j=1}^{p}\sum_{k=1}^{p} d_{jk}(\sigma_{12_{jk}} + \sigma_{21_{jk}})}{\sum_{j=1}^{p}\sum_{k=1}^{p} d_{jk}(\sigma_{11_{jk}} + \sigma_{22_{jk}}) + \sum_{j=1}^{p}\sum_{k=1}^{p} d_{jk}(\mu_{1_j} + \mu_{2_j})(\mu_{1_k} + \mu_{2_k})}$$

where $E_I[]$ is the expectation, assuming independence between the observers; $\sigma_{12_{jk}}$ is the covariance between the measurements of observer 1 at time j and observer 2 at time k; $\sigma_{21_{jk}}$ is the covariance between the measurements of observer 2 at time j and observer 1 at time k; $\sigma_{11_{jk}}$ is the covariance between the measurements of observer 1 at time j and observer 1 at time k; $\sigma_{22_{jk}}$ is the covariance between the measurements of observer 2 at time j and observer 2 at time k; μ_{1_j} and μ_{2_j} are the means of the observers 1 and 2 at time j.

Carrasco and Jover [4] otherwise proposed estimating the CCC using the appropriate intraclass correlation coefficient (ICC) from the following variance components model:

$$Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

where Y_{ij} is the measurement taken by observer j ($j = 1, \ldots, k$) on subject i; μ is the overall mean over subjects and observers; α_i is the subject random effect assumed to be distributed as $\alpha_i \sim N(0, \sigma_\alpha^2)$; β_j is the mean deviation of observer j from the overall mean; and e_{ij} is the random error assumed to be distributed as $e_{ij} \sim N(0, \sigma_e^2)$. All the effects of the model are assumed to be independent. Under this model, the CCC can be expressed by the variance components as

$$\rho_{\mathrm{CCC}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_e^2}$$

where $\sigma_\beta^2 = 1/(k-1)\sum_{j=1}^{k}\beta_j^2$ accounts for systematic differences between observers.

Suppose now that p replicates are taken by subject and observer. The former linear mixed model may be conveniently modified to account for these replicates as

$$Y_{ijt} = \mu + \alpha_i + \beta_j + e_{ijt}$$

with $t = 1, \ldots, p$.

The expression of the CCC remains the same as above. The difference stems from the estimation process, as the variance components will be more efficiently estimated and the variability of the CCC estimate will be smaller in the presence of repeated measurements [4].

Carrasco et al. [7] modified the CCC to afford for longitudinal repeated measures. Following with the assumption that a continuous variable is measured on n subjects by k observers over p times, the variance components model that accounts for all sources of variation is

$$Y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \alpha\beta_{ij} + \alpha\gamma_{it} + \beta\gamma_{jt} + e_{ijt}$$

where μ is the overall mean; α_i is the random subject effect ($i = 1, \ldots, n$) assumed to be distributed as $\alpha_i \sim N(0, \sigma_\alpha^2)$; β_j is the mean deviation of observer j from the overall mean; γ_t is the fixed time effect ($t = 1, \ldots, p$); $\alpha\beta_{ij}$ is the random subject–observer interaction effect assumed to be distributed as $\alpha\beta_{ij} \sim N(0, \sigma_{\alpha\beta}^2)$; $\alpha\gamma_{it}$ is the random subject–time interaction effect assumed to be distributed as $\alpha\gamma_{it} \sim N(0, \sigma_{\alpha\gamma}^2)$; $\beta\gamma_{jt}$ is the fixed observer–time interaction effect and e_{ijt} is the random error effect assumed to be distributed as $\boldsymbol{e}_i \sim MVN(0, \sigma_e^2\mathbf{R}_i)$, where $\boldsymbol{e}_i$ is the vector of residuals of each subject. All the effects of the model are assumed to be independent.

$\mathbf{R}_i$ is a $p \times p$ matrix that accounts for the longitudinal correlation between residuals whose elements are defined as $cov(e_{ijt}, e_{ijs}) = \sigma_e^2\rho(\tau, |t-s|)$. The autocorrelation function $\rho()$ takes values that are between 0 and 1, and depends on the vector of parameters τ and the distance over time between residuals. It is assumed that $\rho(0) = 1$ and $\rho(\infty) = 1$. The appropriate expression of the ICC for measuring agreement between observers is

$$\rho_{\mathrm{CCC}} = \frac{\sigma_\alpha^2 + \sigma_{\alpha\gamma}^2}{\sigma_\alpha^2 + \sigma_{\alpha\gamma}^2 + \sigma_{\alpha\beta}^2 + \sigma_{\beta\gamma}^2 + \sigma_e^2}$$

It must be stated that this coefficient concurs with the CCC for repeated measures defined above when $\mathbf{D}$ is an identity matrix. If $\mathbf{D}$ is a diagonal matrix, the expression of the CCC is

$$\rho_{\mathrm{CCC}} = \frac{(\sigma_\alpha^2 + \sigma_{\alpha\gamma}^2)\sum_{j=1}^{p} d_{jj}}{(\sigma_\alpha^2 + \sigma_{\alpha\gamma}^2 + \sigma_{\alpha\beta}^2 + \sigma_e^2)\sum_{j=1}^{p} d_{ij} + (1/pk(k-1))\sum_{t=1}^{p}\sum_{j=1}^{k-1}\sum_{l=j+1}^{k} d_{tt}(\mu_{jt} - \mu_{lt})}$$

2.2. Estimation

King et al. [6] proposed estimating the CCC via U-statistics. Thus, with a moderately large sample size, the distributions of the estimators are asymptotically normal, and consistent