Statistical Methodology 30 (2016) 31–44

Contents lists available at ScienceDirect

Statistical Methodology

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

A distribution-free test of parallelism for two-sample repeated measurements

Mehrdad Vossoughi [a], S.M.T. Ayatollahi [b,*], Mina Towhidi [c], Seyyed Taghi Heydari [d]

[a] Department of Dental Public Health, School of Dentistry, Shiraz University of Medical Sciences, Shiraz, Iran

[b] Department of Biostatistics, Medical School, Shiraz University of Medical Sciences, Shiraz, Iran

[c] Department of Statistics, College of Sciences, Shiraz University, Shiraz, Iran

[d] Health Policy Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

ARTICLE INFO

Article history:
Received 31 January 2015
Received in revised form
9 October 2015
Accepted 18 December 2015
Available online 29 December 2015

Keywords:
Distribution-free
Interaction
Repeated measurements
Mixed model
Brownian bridge
Simulation

ABSTRACT

In this paper, we propose a new two-sample distribution-free procedure for testing group-by-time interaction effect in repeated measurements from a linear mixed model setting. The test statistic is based on the maximum difference of partial sums (MDPS) over time points between the two groups. Although the test has a biomedical focus, it can be applied in fields that the study is designed and monitored to be balanced and complete with equal sample sizes as would be generally done in a controlled experiment. The asymptotic null distribution of the test statistic was also derived based on the maxima of Brownian bridge under two different conditions. The simulations revealed that MDPS performed markedly better than the commonly used unstructured multivariate approach (UMA) to profile analysis. However, the empirical powers of MDPS test were convincingly close to those of the best-fitting linear mixed model (LMM).

* Corresponding author. Tel.: +98 71 32349330; fax: +98 71 32349330.
E-mail addresses: vossoughim@sums.ac.ir (M. Vossoughi), ayatolahim@sums.ac.ir (S.M.T. Ayatollahi), mtowhidi@susc.ac.ir (M. Towhidi), heidaryt@sums.ac.ir (S.T. Heydari).

http://dx.doi.org/10.1016/j.stamet.2015.12.001

1. Introduction

Randomized repeated measures design is one of the most widely used statistical designs in many areas of research. Testing the hypothesis of parallelism or no group $\times$ time interaction effect is of primary interest in such designs. This is due to the fact that inferences on the main effects, i.e. time (within-subject) and group (between-subject), may be misleading in the presence of a significant interaction effect. Moreover, the investigators are essentially interested in determining whether the patterns of change in response profiles are the same across the groups.

The analysis of repeated measurements must take into account the relationships among the measurements in the subject. The linear mixed model (LMM) to repeated measurements [15,38] has the main advantages over traditional approaches of modeling mean response and covariance structure simultaneously as well as accommodating incomplete and unbalanced data. However, as a main disadvantage, the performance of LMM is not robust to the misspecification of covariance structure for errors [19,26,37]. The other disadvantages are that the inferences may be inconsistent with respect to the different choice of number and distribution of random effects in the model [11,18,40], method of estimation [26,25], and type of test [8,13,32].

Because of the inconvenience encountered in using the LMM, investigators often tend to employ simpler approaches, especially when the circumstances are fully controlled to obtain a balanced and complete data set. For example, in a controlled clinical trial, it is often expected that the subjects have the same number of responses with no missing measurements at a common set of pre-specified time points. In this respect, the extensions of the unstructured multivariate and univariate split-plot ANOVA approaches to the repeated measurements are still the most widely used tools. Several authors studied the performance of tests used in profile analysis [2,23,34]. Boik [2] showed the superiority of unstructured multivariate approach (UMA) over its univariate counterpart for the analysis of repeated measurements.

In this paper, we have proposed a new procedure to test the parallelism of mean profiles in two-sample repeated measurements from a linear mixed model setting. It is a new distribution-free test for LMMs requiring complete and balanced data as well as equal sample sizes in the groups. However, it does not require any specification of covariance structure and assumption concerning the distribution of errors and random effects. In addition, the restrictions are generally intrinsic parts of controlled clinical trials and experimental studies designed with equal sizes and fully controlled to achieve balanced and complete data. The proposed statistic is based on the maximum difference of partial sums (MDPS) between the two groups where the partial sums at each time point are taken from individually centered measurements.

In Section 2, we first provide brief discussions of the LMM and UMA methodologies to the repeated measurements and then propose the new test statistic. Section 3 describes the structure of Monte Carlo simulation studies for evaluating the performance of methods. In Section 4, we report the simulation results about the empirical type I error rate and power of MDPS test and other competitors. This section also reports the results for evaluating a real data example. Finally, we provide discussion and conclusion in Sections 5 and 6, respectively.

We have used free statistical package R to perform the analyses and simulations demonstrated in this paper.

2. Methods

2.1. Linear mixed model (LMM)

The general form of LMM [15] can be written as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \tag{1}$$

where $\boldsymbol{Y_i} = (Y_{i1}, \dots, Y_{im_i})^T$ is the m_i-dimensional vector of response variables for subject i, $i = 1, \dots, N$, N is the total number of subjects, $\boldsymbol{\beta}$ is a p-dimensional vector containing fixed effects, $\boldsymbol{b}_i$ is a q-dimensional vector containing random effects, $\boldsymbol{X}_i^{\mathrm{T}}$ is an $m_i \times p$ covariate matrix for the fixed