Journal of Multivariate Analysis 153 (2017) 255–265

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Permuting longitudinal data in spite of the dependencies

Sarah Friedrich [a,*], Edgar Brunner [b], Markus Pauly [a]

[a] Ulm University, Institute of Statistics, Germany

[b] University Medical Center Göttingen, Institute of Medical Statistics, Germany

ARTICLE INFO

Article history:
Received 22 June 2016
Available online 25 October 2016

AMS subject classifications:
primary 62G10
secondary 62G09
62H15
62P10

Keywords:
Permutation tests
Longitudinal data
Quadratic forms
Repeated measures

ABSTRACT

For general repeated measures designs the Wald-type statistic (WTS) is an asymptotically valid procedure allowing for unequal covariance matrices and possibly non-normal multivariate observations. The drawback of this procedure is its poor performance for small to moderate samples, i.e., decisions based on the WTS may become quite liberal. It is the aim of the present paper to improve the small-sample behavior of the WTS by means of a novel permutation procedure. In particular, it is shown that a permutation version of the WTS inherits its good large-sample properties while yielding a very accurate finite-sample control of the type-I error as shown in extensive simulations. Moreover, the new permutation method is motivated by a practical data set of a split plot design with a factorial structure on the repeated measures.

1. Motivation and introduction

In many experiments in the life, social or psychological sciences the experimental units (e.g., subjects) are repeatedly observed at different occasions (e.g., at different time points) or under different treatment conditions. This leads to certain dependencies between observations from the same unit and results in a more complicated statistical analysis of such studies. In the context of experimental designs, the repeated measures are considered as levels of the *sub-plot* factor. If several groups are observed, these are considered as levels of the *whole-plot* factor.

Typical questions in repeated measures and profile analysis concern the investigation of a group effect, a non-constant effect of time or different time profiles in the groups; see, e.g., the monographs of Davis [14, Section 4.3] or Johnson and Wichern [25, Section 6.8]. Classical repeated measures models, where hypotheses are tested with Hotelling's T^2 [19] or Wilks's Λ [45], assume normally distributed observation vectors and a common covariance matrix for all groups; see e.g., the monograph of Davis [14]. In medical and biological research, however, the assumptions of equal covariance matrices and multivariate normally distributed outcomes are often not met and a violation of them may inflate the type-I error rates; see the comments in Xu and Cui [46], Suo et al. [40] or Konietschke et al. [28].

Therefore, other procedures have been developed for repeated measures which are based on certain approximation techniques [1,7–10,17,18,21,26,27,30,35,41,44]. However, these papers mainly assume the multivariate normal distribution and only discuss methods for specific models which are also asymptotically only approximations, i.e., they do not even lead to asymptotic exact tests. Another possibility is to apply a specific mixed model in the GEE context, see, e.g., the text books by Verbeke and Molenberghs [42,43]. These methods require that the data stem from a specific exponential family. An

* Corresponding author.
E-mail address: sarah.friedrich@uni-ulm.de (S. Friedrich).

http://dx.doi.org/10.1016/j.jmva.2016.10.004
0047-259X/

Table 1
Means and empirical standard deviations of oxygen consumption of leukocytes in the presence and absence of inactivated staphylococci.

O_2-Consumption [$\mu\ell$]							
		Staphylococci					
		With			Without		
		Time (min)			Time (min)		
		6	12	18	6	12	18
Placebo	Mean	1.618	2.434	3.527	1.322	2.430	3.425
($n = 12$)	Sd	0.157	0.303	0.285	0.193	0.263	0.339
Verum	Mean	1.656	2.799	4.029	1.394	2.57	3.677
($n = 12$)	Sd	0.207	0.336	0.256	0.218	0.242	0.340

exception is given by the multivariate Wald-type test statistic (WTS), which is asymptotically exact. However, it is well known that it requires large sample sizes to keep the pre-assigned type-I error level; see, e.g., [6,28,34].

To improve the small-sample behavior of the WTS in a MANOVA setting, Konietschke et al. [28] proposed different bootstrap techniques. Another possibility would be to apply permutation procedures. It is well known that permutation tests are finitely exact under the assumption of exchangeability; see, e.g., [5,31,36] or [37–39] as well as [2,3,12] for examples. In most of these examples, however, permutation tests are only applied in situations where the null distribution is invariant under the corresponding randomization group.

A modified permutation procedure may also be applied in situations where this invariance does not hold; see, e.g., [11,23,24,33,34]. The main idea in these papers is to apply a studentized test statistic and to use its permutation distribution (based on permuting the pooled sample) for calculating critical values. This leads to particularly good finite-sample properties even in case of general factorial designs with fixed factors [34]. It is the aim of the present paper to extend the concept of permuting all data to the context of longitudinal data in general (not necessarily normal and homoscedastic) split plot designs. Applied to the WTS this generalizes the results of Pauly et al. [34] and leads to astonishingly accurate results despite the dependencies in repeated measurements data.

The methodology derived in the present paper is motivated by the following data example on the O_2 consumption of leukocytes. To examine the breathability of leukocytes, an experiment with 44 HSD-rats was conducted. A group of 22 rats was treated with a placebo, while the other 22 rats were treated with a substance supposed to enhance the humoral immunity. 18 h prior to the opening of the abdominal cavity, all animals received 2.4 g sodium-caseinate for the production of a peritoneal exudate rich on leukocytes. In order to obtain a sufficient amount of material the peritoneal liquid of 3–4 animals was mixed and the leukocytes therein were rehashed in an experimental batch. One half of the experimental batch was mixed with inactivated staphylococci in a ratio of 100:1, the other half remained untreated and served as a control. Then, the oxygen consumption of the leukocytes was measured with a polarographic electrode after 6, 12 and 18 min, respectively. For each group separately, 12 experimental batches were carried out. Some descriptive statistics of the experimental batches in both treatment groups are listed in Table 1.

Questions of interest in this example concern the effect of the whole-plot factor 'treatment', the effect of the sub-plot factors 'staphylococci' and 'time' as well as interactions between these effects. We note that the empirical 6×6 covariance matrices of the two groups appear to be quite different (see the supplement (see Appendix A) for details). This also motivates the inclusion of unequal covariance matrices in our model. For such experimental designs, procedures are derived in this paper that lead to good small-sample control of the type-I error while being asymptotically exact.

The paper is organized as follows. The underlying statistical model is described in Section 2, where we also introduce the Wald-type (WTS) as well as the ANOVA-type statistic (ATS) and state their asymptotic behavior. In Section 3, we describe the novel permutation procedure used to improve the small sample behavior of the WTS. Afterwards, we present the results of extensive simulation studies in Section 4, analyzing the behavior of the permuted test statistic in different simulation designs with certain competitors. Additional simulation results have also been run for several other resampling schemes. They did not show a better performance than the permutation procedure and are only reported in the supplementary material, where also various power simulations can be found. The motivating data example is analyzed in detail in Section 5. The paper closes with a brief discussion of our results in Section 6. All proofs are given in the supplementary material (see Appendix A).

2. Statistical model, hypotheses and statistics

2.1. Statistical model and hypotheses

To establish the general model, let

$$\boldsymbol{Y}_{ik} = (Y_{ik1}, \ldots, Y_{ikt_i})^\top, \quad i = 1, \ldots, a;\ k = 1, \ldots, n_i \tag{2.1}$$

denote independent random vectors with distribution F_i and expectation $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{it_i})^\top = \mathrm{E}(\boldsymbol{Y}_{i1})$ in treatment group i. The underlying dependency structure is regulated by pairwise correlations. In particular, we do not assume *any special*