MATS: Inference for potentially singular and heteroscedastic MANOVA

Sarah Friedrich *, Markus Pauly

Institut für Statistik, Universität Ulm, Germany

1. Motivation and introduction

In many experiments, notably in biology, ecology and psychology, several endpoints, potentially measured on different scales, are recorded per subject. As an example, we consider a data set on the 2016 presidential elections in the USA containing demographic data on counties from the US census. Suppose, for illustration purposes, that we wish to investigate whether the states differ with respect to some demographic factors. In addition to unequal empirical covariance matrices between groups, the analysis is complicated by their singularity.

The analysis of such multivariate data is typically based on classical MANOVA models assuming multivariate normality and/or homogeneity of the covariance matrices; see, e.g., [1,12,13,18,23,32,41]. These assumptions, however, are often not met in practice – as in the motivating example – and it is well known that the methods perform poorly when the data are heterogeneous [21,37]. Furthermore, the test statistic should be invariant under scale transformations of the components, since the endpoints may be measured on different scales. Thus, multivariate ANOVA-type test statistics (ATS) as, e.g., proposed in [4] and studied in [14], are only applicable if all endpoints are measured on the same scale, i.e., for repeated-measure designs.

Assuming non-singular covariance matrices and certain moment conditions, scale invariance is typically accomplished by resorting to Wald-type test statistics (WTS). However, inference procedures based thereon require (extremely) large sample

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

sizes to be accurate; see [21,34,38]. In particular, even the novel approaches of Konietschke et al. [21] and Smaga [34] tend to be liberal for skewed distributions. Moreover, their procedures cannot be used to analyze our motivating data example with possibly singular covariance matrices. Therefore, we follow a different approach by modifying the above mentioned ANOVA-type statistic (MATS). It is motivated by the modified Dempster statistic proposed by Srivastava and Kubokawa [36] for high-dimensional one-way MANOVA. This statistic is also invariant under changes in units of measurement for the null hypotheses considered. However, until now, it has only been developed for a homoscedastic one-way setting assuming non-singularity and a specific distributional structure that is motivated by multivariate normality.

It is the aim of the present paper to modify and extend the approach of Srivastava and Kubokawa [36] to factorial MANOVA designs, incorporating general heteroscedastic models. In particular, we only postulate the existence of the group-wise covariance matrices, which may even be singular. The small-sample behavior of our test statistic is enhanced through bootstrap techniques as in [21]. Thereby, the novel MATS procedure enables us to relax the usual MANOVA assumptions in several ways. While incorporating general heteroscedastic designs and allowing for potentially singular covariance matrices, we postulate their existence solely through a distributional assumption, i.e., only finite second moments are required. Moreover, we gain a procedure that is more robust against deviations from symmetry and homoscedasticity than the usual WTS approaches.

So far, only few approaches have been investigated which do not assume normality or equal covariance matrices (or both). Examples in the nonparametric framework include the permutation based nonparametric combination method [30,31] and the rank-based tests presented in [5,6] for split plot designs and in [2,26] for MANOVA designs. However, these methods are either not applicable for general MANOVA models or based on null hypotheses formulated in terms of distribution functions. In contrast we wish to derive inference procedures (tests and confidence regions) for contrasts of mean vectors. Here, beside all previously mentioned procedures, only methods for specific designs have been developed; see [10] for two-sample problems, [39,40] for robust but homoscedastic one-way MANOVA or [16] for a particular two-way MANOVA. To our knowledge, mean-based MANOVA procedures allowing for possibly singular covariance matrices have not been developed so far.

The paper is organized as follows. In Section 2 we describe the statistical model and hypotheses. Furthermore, we propose a new test statistic, which is applicable to singular covariance matrices and is invariant under scale transformations of the data. In Section 3, we present three different resampling approaches which are then used for the derivation of statistical tests as well as confidence regions and simultaneous confidence intervals for contrasts in Section 4. The different approaches are compared in a large simulation study (Section 5), where we analyze different factorial designs with a wide variety of distributions and covariance settings. The motivating data example is analyzed in Section 6 and we conclude with some final remarks and discussion in Section 7. All proofs are deferred to the Online Supplement, where we also provide further simulation results and the analysis of an additional data example.

2. Statistical model, hypotheses and test statistics

Throughout the paper, we will use the following notation. We denote by $\boldsymbol{I}_d$ the d-dimensional unit matrix and by $\boldsymbol{J}_d$ the $d \times d$ matrix of 1s, i.e., $\boldsymbol{J}_d = \mathbf{1}_d \mathbf{1}_d^\top$, where $\mathbf{1}_d = (1, \ldots, 1)^\top$ is the d-dimensional column vector of 1s. Furthermore, let $\boldsymbol{P}_d = \boldsymbol{I}_d - d^{-1}\boldsymbol{J}_d$ denote the d-dimensional centering matrix. By $\oplus$ and $\otimes$ we denote the direct sum and the Kronecker product, respectively.

In order to cover different factorial designs of interest, we consider the general model

$$\boldsymbol{X}_{ik} = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{ik}$$

for treatment group $i \in \{1, \ldots, a\}$ and individual $k \in \{1, \ldots, n_i\}$, on which we measure d-variate observations. Here $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{id})^\top \in \mathbb{R}^d$ for all $i \in \{1, \ldots, a\}$. A factorial structure can be incorporated by splitting up indices; see, e.g., [21]. For fixed $i \in \{1, \ldots, a\}$, the error terms $\boldsymbol{\epsilon}_{ik}$ are assumed to be independent and identically distributed d-dimensional random vectors, for which the following conditions hold:

(1) $\mathrm{E}(\boldsymbol{\epsilon}_{i1}) = 0$ for all $i \in \{1, \ldots, a\}$.
(2) $0 < \sigma_{is}^2 = \mathrm{var}(X_{iks}) < \infty$ for all $i \in \{1, \ldots, a\}$ and $s \in \{1, \ldots, d\}$.
(3) $\mathrm{cov}(\boldsymbol{\epsilon}_{i1}) = \boldsymbol{V}_i \geq 0$ for all $i \in \{1, \ldots, a\}$.

Thus, we only assume the existence of second moments. For convenience, we aggregate the individual vectors into $\boldsymbol{X} = (\boldsymbol{X}_{11}^\top, \ldots, \boldsymbol{X}_{an_a}^\top)^\top$ as well as $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_a^\top)^\top$. Denote by $N = n_1 + \cdots + n_a$ the total sample size. In order to derive asymptotic results in this framework, we will assume throughout the usual sample size condition, viz.

$$\forall_{i \in \{1,\ldots,a\}} \quad \lim_{N \to \infty} n_i/N = \kappa_i > 0.$$

An estimator for $\boldsymbol{\mu}$ is given by the vector of pooled group means $\overline{\boldsymbol{X}}_{i\cdot} = (\boldsymbol{X}_{i1} + \cdots + \boldsymbol{X}_{in_i})/n_i$ for all $i \in \{1, \ldots, a\}$, which we denote by $\overline{\boldsymbol{X}}_{\cdot} = (\overline{\boldsymbol{X}}_{1\cdot}^\top, \ldots, \overline{\boldsymbol{X}}_{a\cdot}^\top)^\top$. The covariance matrix of $\sqrt{N}\,\overline{\boldsymbol{X}}_{\cdot}$ is given by

$$\boldsymbol{\Sigma}_N = \mathrm{cov}(\sqrt{N}\,\overline{\boldsymbol{X}}_{\cdot}) = \mathrm{diag}\,(N\boldsymbol{V}_i/n_i : 1 \leq i \leq a)\,,$$

where the group-specific covariance matrices $\boldsymbol{V}_i$ are estimated by the empirical covariance matrices

$$\widehat{\boldsymbol{V}}_i = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_{i\cdot})^\top$$

resulting in $\widehat{\boldsymbol{\Sigma}}_N = \mathrm{diag}(N\widehat{\boldsymbol{V}}_i/n_i : 1 \leq i \leq a)$.

In this semi-parametric framework, hypotheses are formulated in terms of the mean vector as $\mathcal{H}_0 : \boldsymbol{H}\boldsymbol{\mu} = \boldsymbol{0}$, where $\boldsymbol{H}$ is a suitable contrast matrix, i.e., $\boldsymbol{H}\boldsymbol{1}_{ad} = \boldsymbol{0}$. Note that we can use the unique projection matrix $\boldsymbol{T} = \boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^+\boldsymbol{H}$, where $(\boldsymbol{H}\boldsymbol{H}^\top)^+$ denotes the Moore–Penrose inverse of $\boldsymbol{H}\boldsymbol{H}^\top$. One has $\boldsymbol{T} = \boldsymbol{T}^2$, $\boldsymbol{T} = \boldsymbol{T}^\top$, and $\boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0} \Leftrightarrow \boldsymbol{H}\boldsymbol{\mu} = \boldsymbol{0}$; see, e.g., [7].

A commonly used test statistic for multivariate data is the Wald-type statistic (WTS)

$$T_N = N\overline{\boldsymbol{X}}_{\cdot}^\top \boldsymbol{T}(\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}_N\boldsymbol{T})^+\boldsymbol{T}\overline{\boldsymbol{X}}_{\cdot}, \tag{1}$$

which requires the additional assumption

(3') $\boldsymbol{V}_i > 0$ for all $i \in \{1, \ldots, a\}$.

It is easy to show that under $\mathcal{H}_0 : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}$, the WTS has, as $N \to \infty$, a χ^2_f distribution with $f = \mathrm{rank}(\boldsymbol{T})$ degrees of freedom if the above conditions (1)–(3') hold. However, large sample sizes are necessary to maintain a pre-assigned level α using quantiles of the limiting χ^2 distribution.

Konietschke et al. [21] proposed different resampling procedures in order to improve the small-sample behavior of the WTS for multivariate data. Therein, a parametric bootstrap approach turned out to be the best when the underlying distributions are not too skewed and/or too heteroscedastic. In these cases, all considered procedures were more or less liberal. Moreover, assuming only (3) instead of (3') for the WTS would generally not lead to a χ^2_f limit distribution.

This is because of possible rank jumps between $\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}_N\boldsymbol{T}$, $\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}$ and $\boldsymbol{T}$. To see this, suppose that $\mathrm{rank}(\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}) = 2$, while $\mathrm{rank}(\boldsymbol{T}) = 4$; this corresponds to Scenario S5 in the simulation studies below. If additionally

$$\lim_{N\to\infty} \mathrm{rank}(\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}_N\boldsymbol{T}) = \mathrm{rank}(\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}) = 2,$$

we have that the WTS follows, asymptotically, a χ^2_g distribution under the null hypothesis, where $g = \mathrm{rank}(\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}) = 2$. The Wald-type test, however, compares T_N to the quantile of a χ^2_4 distribution. Thus, for a chosen significance level of $\alpha = 0.05$ this results in a true asymptotic ($N \to \infty$) type-I error of $\Pr(T_N > \chi^2_{4;0.95}) \approx 0.0087$, i.e., a strictly conservative behavior of the test. Here $\chi^2_{f;1-\alpha}$ denotes the $(1-\alpha)$-quantile of the χ^2_f distribution. Similarly, for $\alpha = 0.1$ and $\alpha = 0.01$ we obtain asymptotically inflated type-I error rates of 0.02 and 0.0013 (both again conservative), respectively. Moreover, the situation is even more complicated since $\lim_{N\to\infty} \mathrm{rank}(\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}_N\boldsymbol{T}) = \mathrm{rank}(\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T})$ is neither verifiable in practice nor holds in general.

We tackle this problem here. In addition to relaxing the assumption (3') on the unknown covariance matrices, we gain a procedure that is more robust to deviations from symmetry and homoscedasticity. To this end, we consider a different test statistic, namely a multivariate version of the ANOVA-type statistic (ATS) proposed by [4] for repeated measures designs, which we obtain by erasing the Moore–Penrose term from (1), viz. $\tilde{Q}_N = N\overline{\boldsymbol{X}}_{\cdot}^\top \boldsymbol{T}\overline{\boldsymbol{X}}_{\cdot}$. In the special two-sample case where we wish to test the null hypothesis $\mathcal{H}_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{0}$, this is equivalent to the test statistic proposed by [13].

The drawback of the ATS for multivariate data is that it is not invariant under scale transformations of the components, e.g., under change of units ($cm \mapsto m$ or $g \mapsto kg$) in one or more components. We demonstrate this problem in a real data analysis given in the Online Supplement, where we show that a simple unit change can completely alter the test decision of the ATS. Thus, we consider a slightly modified version of the ATS, which we denote as MATS:

$$Q_N = N\overline{\boldsymbol{X}}_{\cdot}^\top \boldsymbol{T}(\boldsymbol{T}\widehat{\boldsymbol{D}}_N\boldsymbol{T})^+\boldsymbol{T}\overline{\boldsymbol{X}}_{\cdot}. \tag{2}$$

Here, $\widehat{\boldsymbol{D}}_N = \mathrm{diag}(N/n_i \cdot \widehat{\sigma}^2_{is})$ for all $i \in \{1, \ldots, a\}$ and $s \in \{1, \ldots, d\}$, where $\widehat{\sigma}^2_{is}$ is the empirical variance of component s in group i. The MATS is invariant under scale transformations of the data for null hypotheses of the form $\boldsymbol{T} = \boldsymbol{M} \otimes \boldsymbol{I}_d$, i.e., hypotheses as formulated in Section 4 of [21]. See the Online Supplement for a more detailed discussion of this topic. A related test statistic has been proposed by Srivastava and Kubokawa [36] in the context of high-dimensional ($d \to \infty$) data for a special non-singular one-way MANOVA design. Here, we investigate in the classical multivariate case (with fixed d) how its finite-sample performance can be enhanced considerably. We start by analyzing its asymptotic limit behavior.

Theorem 1. *Under Conditions* (1)–(3) *and under* $\mathcal{H}_0 : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}$, *the test statistic* Q_N *in* (2) *has asymptotically, as* $N \to \infty$, *the same distribution as a weighted sum of* χ^2_1 *distributed random variables, where the weights* λ_{is} *are the eigenvalues of* $\boldsymbol{V} = \boldsymbol{T}(\boldsymbol{T}\boldsymbol{D}\boldsymbol{T})^+\boldsymbol{T}\boldsymbol{\Sigma}$ *for* $\boldsymbol{D} = \mathrm{diag}(\kappa_i^{-1}\sigma^2_{is})$ *and* $\boldsymbol{\Sigma} = \mathrm{diag}(\kappa_i^{-1}\boldsymbol{V}_i)$, *i.e.,*

$$Q_N = N\overline{\boldsymbol{X}}_{\cdot}^\top \boldsymbol{T}(\boldsymbol{T}\widehat{\boldsymbol{D}}_N\boldsymbol{T})^+\boldsymbol{T}\overline{\boldsymbol{X}}_{\cdot} \rightsquigarrow Z = \sum_{i=1}^{a}\sum_{s=1}^{d} \lambda_{is} Z_{is},$$

with $Z_{is} \overset{iid}{\sim} \chi^2_1$ *and* $\rightsquigarrow$ *denoting convergence in distribution.*

Thus, we obtain an asymptotic level α benchmark test $\varphi_N = \mathbf{1}(Q_N > c_{1-\alpha})$ for $\mathcal{H}_0 : \boldsymbol{T}\boldsymbol{\mu} = \mathbf{0}$, where $c_{1-\alpha}$ is the $(1-\alpha)$-quantile of Z. However, the distribution of Z depends on the unknown variances σ_{is}^2 with $i \in \{1, \ldots, a\}$ and $s \in \{1, \ldots, d\}$ so that φ_N is infeasible for most practical situations. For this reason we consider different bootstrap approaches in order to approximate the unknown limiting distribution and to derive adequate and asymptotically correct inference procedures based on Q_N in (2). This will be explained in detail in the next section. Apart from statistical test decisions discussed in Section 4.1, a central part of statistical analyses is the construction of confidence intervals, which allows for deeper insight into the variability and the magnitude of effects. This univariate concept can be generalized to multivariate endpoints by constructing multivariate confidence regions and simultaneous confidence intervals for contrasts $\boldsymbol{h}^\top\boldsymbol{\mu}$ for any contrast vector $\boldsymbol{h} \in \mathbb{R}^{ad}$ of interest. Details on the derivation of such confidence regions for $\boldsymbol{h}^\top\boldsymbol{\mu}$ are given in Section 4.2.

3. Bootstrap procedures

The first bootstrap procedure we consider is a parametric bootstrap approach as proposed by Konietschke et al. [21] for the WTS. The second is a wild bootstrap approach, which has already been successfully applied in the context of repeated measures or clustered data; see, e.g., [8,9,15]. The third procedure is a group-wise, nonparametric bootstrap approach. All of these bootstrap approaches are based on the test statistic Q_N in (2). Note that the procedures derived in the following can also be used for multiple testing problems, either by applying the closed testing principle [28,35] or in the context of simultaneous contrast tests [17,19].

3.1. A parametric bootstrap approach

This asymptotic model-based bootstrap approach has successfully been used in univariate one- and two-way factorial designs [22,43] and has recently been applied to Wald-type statistics for general MANOVA in [21,34] under the additional assumption that fourth moments are finite. The approach is as follows. Given the data, we generate a parametric bootstrap sample as

$$\forall_{i\in\{1,\ldots,a\}} \quad \boldsymbol{X}_{i1}^*, \ldots, \boldsymbol{X}_{in_i}^* \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \widehat{\boldsymbol{V}}_i).$$

The idea behind this method is to obtain an accurate finite-sample approximation by mimicking the covariance structure given in the observed data. This is achieved by calculating Q_N^* from the bootstrap variables $\boldsymbol{X}_{i1}^*, \ldots, \boldsymbol{X}_{in_i}^*$, i.e.,

$$Q_N^* = N(\overline{\boldsymbol{X}}_{\cdot}^*)^\top \boldsymbol{T}(\boldsymbol{T}\widehat{\boldsymbol{D}}_N^*\boldsymbol{T})^+\boldsymbol{T}\overline{\boldsymbol{X}}_{\cdot}^*. \tag{3}$$

We then obtain a parametric bootstrap test by comparing the original test statistic Q_N with the conditional $(1-\alpha)$-quantile $c_{1-\alpha}^*$ of its bootstrap version Q_N^*.

Theorem 2. *The conditional distribution of Q_N^* weakly approximates the null distribution of Q_N in probability for any parameter $\boldsymbol{\mu} \in \mathbb{R}^{ad}$ and $\boldsymbol{\mu}_0$ with $\boldsymbol{T}\boldsymbol{\mu}_0 = \mathbf{0}$, i.e.,*

$$\sup_{x\in\mathbb{R}} |\mathrm{Pr}_{\boldsymbol{\mu}}(Q_N^* \le x|\boldsymbol{X}) - \mathrm{Pr}_{\boldsymbol{\mu}_0}(Q_N \le x)| \overset{\mathrm{Pr}}{\to} 0,$$

where $\mathrm{Pr}_{\boldsymbol{\mu}}(Q_N \le x)$ and $\mathrm{Pr}_{\boldsymbol{\mu}}(Q_N^* \le x|\boldsymbol{X})$ denote the unconditional and conditional distribution of Q_N and Q_N^*, respectively, if $\boldsymbol{\mu}$ is the true underlying mean vector.

3.2. A wild bootstrap approach

Another resampling approach, which is based on multiplying the fixed data with random weights, is the so-called wild bootstrap procedure. To this end, let W_{ik} denote iid random variables, independent of $\boldsymbol{X}$, with $\mathrm{E}(W_{ik}) = 0$, $\mathrm{var}(W_{ik}) = 1$ and $\sup_{i,k}\mathrm{E}(W_{ik}^4) < \infty$. We obtain a bootstrap sample as

$$\forall_{i\in\{1,\ldots,a\}} \quad \forall_{k\in\{1,\ldots,n_i\}} \quad \boldsymbol{X}_{ik}^\star = W_{ik}(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_{i\cdot}).$$

Note that there are different choices for the random weights W_{ik}, e.g., Rademacher distributed random variables [11] or weights satisfying different moment conditions; see, e.g., [3,25,27,42]. The choice of the weights typically depends on the situation. In our simulation studies, we have investigated the performance of different weights such as Rademacher distributed as well as $\mathcal{N}(0, 1)$ distributed weights; see [24] for this specific choice. The results were quite similar and therefore Section 5 includes only those that pertain to standard normal weights.

Based on the bootstrap variables $\boldsymbol{X}_{ik}^\star$, we can compute $Q_N^\star$ in the same way as described for Q_N^* in (3) above. A wild bootstrap test is finally obtained by comparing Q_N to the conditional $(1-\alpha)$-quantile of its wild bootstrap version $Q_N^\star$.

Theorem 3. *The conditional distribution of $Q_N^\star$ weakly approximates the null distribution of Q_N in probability for any parameter $\boldsymbol{\mu} \in \mathbb{R}^{ad}$ and $\boldsymbol{\mu}_0$ with $\boldsymbol{T}\boldsymbol{\mu}_0 = \mathbf{0}$, i.e.,*

$$\sup_{x\in\mathbb{R}} |\mathrm{Pr}_{\boldsymbol{\mu}}(Q_N^\star \le x|\boldsymbol{X}) - \mathrm{Pr}_{\boldsymbol{\mu}_0}(Q_N \le x)| \overset{\mathrm{Pr}}{\to} 0.$$

3.3. A nonparametric bootstrap approach

The third approach we consider is the nonparametric bootstrap. Here, for each group $i \in \{1, \dots, a\}$, we randomly draw n_i independent selections $\boldsymbol{X}_{ik}^{\dagger}$ with replacement from the ith sample $\{\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{in_i}\}$. The bootstrap test statistic $Q_N^{\dagger}$ is then computed in the same way as described above, i.e., by recalculating Q_N with $\boldsymbol{X}_{ik}^{\dagger}$ for all $i \in \{1, \dots, a\}$ and $k \in \{1, \dots, n_i\}$. Finally, a nonparametric bootstrap test is obtained by comparing the original test statistic Q_N to the empirical $(1-\alpha)$-quantile of $Q_N^{\dagger}$. The asymptotic validity of this method is guaranteed by the following result.

Theorem 4. *The conditional distribution of $Q_N^{\dagger}$ weakly approximates the null distribution of Q_N in probability for any parameter $\boldsymbol{\mu} \in \mathbb{R}^{ad}$ and $\boldsymbol{\mu}_0$ with $\boldsymbol{T}\boldsymbol{\mu}_0 = \boldsymbol{0}$, i.e.,*

$$\sup_{x \in \mathbb{R}} |\mathrm{Pr}_{\boldsymbol{\mu}}(Q_N^{\dagger} \leq x | \boldsymbol{X}) - \mathrm{Pr}_{\boldsymbol{\mu}_0}(Q_N \leq x)| \xrightarrow{\mathrm{Pr}} 0.$$

4. Statistical applications

We now want to base statistical inference on the modified test statistic in (2) using the bootstrap approaches described above. A thorough statistical analysis should ideally consist of two parts. First, statistical tests give insight into significant effects of the different factors as well as possible interactions. We therefore consider important properties of statistical tests based on the bootstrap approaches in Section 4.1. Second, it is helpful to construct confidence regions for the unknown parameters of interest in order to gain a more detailed insight into the nature of the estimates. The derivation of such confidence regions is discussed in Section 4.2.

4.1. Statistical tests

In this section, we analyze the statistical properties of the bootstrap procedures described above. For ease of notation, we will only state the results for the parametric bootstrap procedure, i.e., we consider the test statistic Q_N^* based on $\boldsymbol{X}_{ik}^*$ throughout. Note, however, that the results are also valid for the wild and the nonparametric bootstrap procedure, i.e., the test statistics $Q_N^{\star}$ and $Q_N^{\dagger}$.

As mentioned above, a bootstrap test $\varphi^* = \mathbf{1}(Q_N > c_{1-\alpha}^*)$ is obtained by comparing the original test statistic Q_N to the $(1-\alpha)$-quantile $c_{1-\alpha}^*$ of its bootstrap version. In particular, p-values are numerically computed as follows:

(1) Given the data $\boldsymbol{X}$, compute the MATS Q_N for the null hypothesis of interest.
(2) Bootstrap the data with either of the bootstrap approaches described above and compute the corresponding test statistic $Q_N^{*,1}$.
(3) Repeat step (2) a large number of times, e.g., $B = 10{,}000$ times, and obtain values $Q_N^{*,1}, \dots, Q_N^{*,B}$.
(4) Calculate the p-value based on the empirical distribution of $Q_N^{*,1}, \dots, Q_N^{*,B}$ as

$$p\text{-value} = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}(Q_N \leq Q_N^{*,b}).$$

Theorems 2–4 imply that the corresponding tests asymptotically keep the pre-assigned level α under the null hypothesis and are consistent for any fixed alternative $\boldsymbol{T}\boldsymbol{\mu} \neq \boldsymbol{0}$, i.e., $E_{\boldsymbol{\mu}}(\varphi^*) \to \alpha\, \mathbf{1}(\boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}) + \mathbf{1}(\boldsymbol{T}\boldsymbol{\mu} \neq \boldsymbol{0})$. Moreover, for local alternatives $\mathcal{H}_1 : \boldsymbol{T}\boldsymbol{\mu} = N^{-1/2}\boldsymbol{T}\boldsymbol{\nu}$ with $\boldsymbol{\nu} \in \mathbb{R}^{ad}$, the bootstrap tests have the same asymptotic power as $\varphi_N = \mathbf{1}(Q_N > c_{1-\alpha})$, where $c_{1-\alpha}$ is the $(1-\alpha)$-quantile of Z given in Theorem 1. In particular, the asymptotic relative efficiency of the bootstrap tests compared to φ_N is 1 in this situation.

4.2. Confidence regions and confidence intervals for contrasts

In order to conduct a thorough statistical analysis, interpretation of the results should not be based on p-values alone. In addition, it is helpful to construct confidence regions for the unknown parameter. The concept of a confidence region is the same as that of a confidence interval in the univariate setting. We want to construct a multivariate region, which is likely to contain the true, but unknown parameter of interest.

The aim of this section is to derive multivariate confidence regions and simultaneous confidence intervals for contrasts $\boldsymbol{h}^{\top}\boldsymbol{\mu}$ for any contrast vector $\boldsymbol{h}$ of interest. Such contrasts include, e.g., the difference in means $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ in two-sample problems, Dunnett's many-to-one comparisons, Tukey's all-pairwise comparisons, and many more; see, e.g., [19] for specific examples. In this section, we will base the derivation of confidence regions on the bootstrap approximations given in Section 3, i.e., we will use one of the bootstrap quantiles. Again, we only formulate the results for $c_{1-\alpha}^*$.

For the derivation of a confidence region, first note that the results from Section 4.1 imply that the null hypothesis $\mathcal{H}_0 : \boldsymbol{H}\boldsymbol{\mu} = \boldsymbol{H}\boldsymbol{\mu}_0$ for a vector of contrasts $\boldsymbol{H}\boldsymbol{\mu}_0$, $\boldsymbol{H} = (\boldsymbol{h}_1 | \cdots | \boldsymbol{h}_q)^{\top} \in \mathbb{R}^{q \times ad}$, $\boldsymbol{\mu}_0 \in \mathbb{R}^{ad}$, is rejected at asymptotic level α,

if $N(\boldsymbol{H}\overline{\boldsymbol{X}}_{\cdot} - \boldsymbol{H}\boldsymbol{\mu}_0)^\top(\boldsymbol{H}\widehat{\boldsymbol{D}}_N\boldsymbol{H}^\top)^+(\boldsymbol{H}\overline{\boldsymbol{X}}_{\cdot} - \boldsymbol{H}\boldsymbol{\mu}_0)$ is larger than the bootstrap quantile $c^*_{1-\alpha}$. Thus, a confidence region for the vector of contrasts $\boldsymbol{H}\boldsymbol{\mu}$ is determined by the set of all $\boldsymbol{H}\boldsymbol{\mu}$ such that

$$N(\boldsymbol{H}\overline{\boldsymbol{X}}_{\cdot} - \boldsymbol{H}\boldsymbol{\mu})^\top(\boldsymbol{H}\widehat{\boldsymbol{D}}_N\boldsymbol{H}^\top)^+(\boldsymbol{H}\overline{\boldsymbol{X}}_{\cdot} - \boldsymbol{H}\boldsymbol{\mu}) \le c^*_{1-\alpha}.$$

A confidence ellipsoid is now obtained based on the eigenvalues $\widehat{\lambda}_s$ and eigenvectors $\widehat{\boldsymbol{e}}_s$ of $\boldsymbol{H}\widehat{\boldsymbol{D}}_N\boldsymbol{H}^\top$. As in [20], the direction and lengths of its axes are determined by going $(\widehat{\lambda}_s\, c^*_{1-\alpha}/N)^{1/2}$ units along the eigenvectors $\widehat{\boldsymbol{e}}_s$ of $\boldsymbol{H}\widehat{\boldsymbol{D}}_N\boldsymbol{H}^\top$. In other words, the axes of the ellipsoid are given, for all $s \in \{1, \ldots, d\}$, by

$$\boldsymbol{H}\overline{\boldsymbol{X}}_{\cdot} \pm (\widehat{\lambda}_s\, c^*_{1-\alpha}/N)^{1/2}\,\widehat{\boldsymbol{e}}_s. \tag{4}$$

Note that this approach is similar to the construction of confidence intervals in the univariate case, where we exploit the one-to-one relationship between CIs and tests. While we can compute (4) for arbitrary dimension d, we cannot display the joint confidence region graphically for $d \ge 4$. In the two-sample case with $d = 2$ endpoints, however, the ellipse can be plotted.

Beginning at the center $\boldsymbol{H}\overline{\boldsymbol{X}}_{\cdot}$ the axes of the ellipsoid are given by $\pm(\widehat{\lambda}_s\, c^*_{1-\alpha}/N)^{1/2} t\widehat{\boldsymbol{e}}_s$ for $s \in \{1, 2\}$. That is, the confidence ellipse extends $(\widehat{\lambda}_s \cdot c^*_{1-\alpha}/N)^{1/2}$ units along the estimated eigenvector $\widehat{\boldsymbol{e}}_s$ for $s \in \{1, 2\}$. Therefore, we get a graphical representation of the relation between the group-mean differences $\mu_{11} - \mu_{21}$ and $\mu_{12} - \mu_{22}$ of the first and second components; see Section 10 and Figure 3 in the Online Supplement for an example.

Concerning the derivation of multiple contrast tests and simultaneous confidence intervals for contrasts, we consider the family of hypotheses

$$\boldsymbol{\Omega} = \{\mathcal{H}_0 : \boldsymbol{h}_\ell^\top\boldsymbol{\mu} = \boldsymbol{0} \text{ with } \boldsymbol{h}_\ell \ne \boldsymbol{0},\, \ell = 1, \ldots, q\}.$$

As shown in Sections 2–3, a test statistic for testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{H}\boldsymbol{\mu} = \boldsymbol{0}$ is given by Q_N in (2). Consequently, working with a single contrast $\boldsymbol{h}_\ell$ as contrast matrix leads to the test statistic

$$Q_N^\ell = N(\boldsymbol{h}_\ell^\top\overline{\boldsymbol{X}}_{\cdot})^\top(\boldsymbol{h}_\ell^\top\widehat{\boldsymbol{D}}_N\boldsymbol{h}_\ell)^{-1}(\boldsymbol{h}_\ell^\top\overline{\boldsymbol{X}}_{\cdot}) = N\,\frac{(\sum_{i=1}^a\sum_{s=1}^d h_{\ell,is}\overline{X}_{i\cdot s})^2}{\sum_{i=1}^a\sum_{s=1}^d h_{\ell,is}^2\widehat{\sigma}_{is}^2}$$

for the null hypotheses $\mathcal{H}_0^\ell : \forall_{\ell\in\{1,\ldots,q\}}\ \boldsymbol{h}_\ell^\top\boldsymbol{\mu} = \boldsymbol{0}$. Here, $\boldsymbol{h}_\ell = (h_{\ell,11}, \ldots, h_{\ell,ad})^\top$. To obtain a single critical value with one of the bootstrap methods we may, e.g., consider the usual maximum or sum statistics. We exemplify the idea for the latter. Thus, let

$$S_N \equiv N(\boldsymbol{H}\overline{\boldsymbol{X}}_{\bullet})^\top \operatorname{diag}\{(\boldsymbol{h}_\ell^\top\widehat{\boldsymbol{D}}_N\boldsymbol{h}_\ell)^{-1} : \ell = 1, \ldots, q\}\ \boldsymbol{H}\overline{\boldsymbol{X}}_{\bullet} = \sum_{\ell=1}^q Q_N^\ell$$

and denote by $q^*_{1-\alpha}$ the conditional $(1-\alpha)$-quantile of its corresponding bootstrap version S_N^*. From the proofs of Theorems 2–4 given in the Online Supplement, it follows that, for any of the three bootstrap methods described in Section 3, $\widehat{\sigma}_{is}^*$ is a consistent estimate of σ_{is} for all $i \in \{1, \ldots, a\}$ and $s \in \{1, \ldots, d\}$, and that $\sqrt{N}\boldsymbol{H}\overline{\boldsymbol{X}}_{\bullet}^*$ asymptotically mimics the distribution of $\sqrt{N}\boldsymbol{H}(\overline{\boldsymbol{X}}_{\bullet} - \boldsymbol{\mu})$. Thus, the continuous mapping theorem implies $\Pr(S_N \le q^*_{1-\alpha}) \to 1-\alpha$ as $N \to \infty$ and therefore

$$\Pr\left(\bigcap_{\ell=1}^q\{Q_N^\ell \le q^*_{1-\alpha}\}\right) \le \Pr\left(\sum_{\ell=1}^q Q_N^\ell \le q^*_{1-\alpha}\right) \to 1-\alpha, \quad \text{as } N \to \infty.$$

This implies, that simultaneous $100 \times (1-\alpha)\%$ confidence intervals for contrasts $\boldsymbol{h}_1^\top\boldsymbol{\mu}, \ldots, \boldsymbol{h}_q^\top\boldsymbol{\mu}$ are given by

$$\boldsymbol{h}_\ell^\top\overline{\boldsymbol{X}}_{\cdot} \pm \sqrt{q^*_{1-\alpha}\cdot\boldsymbol{h}_\ell^\top\widehat{\boldsymbol{D}}_N\boldsymbol{h}_\ell/N}.$$

In the Online Supplement, we also explain that the bootstrap idea also works for the usual maximum statistic.

5. Simulations

The procedures described in Section 3 are valid for large sample sizes. In order to investigate their behavior for small samples, we have conducted various simulations. In the simulation studies, the behavior of the proposed approaches was compared to a parametric bootstrap approach for the WTS as in [21] since this turned out to perform better than other resampling versions of the WTS and Wilk's Λ. For comparison, we also included the asymptotic χ^2 approximation of the WTS. All simulations were conducted using `R` Version 3.3.1 [33] each with `nsim` = 5000 simulation and `nboot` = 5000 bootstrap runs. We investigated a one- and a two-factorial design.

5.1. One-way layout

For the one-way layout, data were generated as in [21]. We considered $a = 2$ treatment groups and $d \in \{4, 8\}$ endpoints as well as the following covariance settings:

$$\begin{aligned}
&\text{Setting 1:} \quad \boldsymbol{V}_1 = \boldsymbol{I}_d + 0.5(\boldsymbol{J}_d - \boldsymbol{I}_d) = \boldsymbol{V}_2,\\
&\text{Setting 2:} \quad \boldsymbol{V}_1 = ((0.6)^{|r-s|})_{r,s=1}^{d} = \boldsymbol{V}_2,\\
&\text{Setting 3:} \quad \boldsymbol{V}_1 = \boldsymbol{I}_d + 0.5(\boldsymbol{J}_d - \boldsymbol{I}_d) \text{ and } \boldsymbol{V}_2 = \boldsymbol{I}_d\, 3 + 0.5(\boldsymbol{J}_d - \boldsymbol{I}_d),\\
&\text{Setting 4:} \quad \boldsymbol{V}_1 = ((0.6)^{|r-s|})_{r,s=1}^{d} \text{ and } \boldsymbol{V}_2 = ((0.6)^{|r-s|})_{r,s=1}^{d} + \boldsymbol{I}_d\, 2.
\end{aligned}$$

Setting 1 represents a compound symmetry structure, while Setting 2 is an autoregressive covariance structure. Both settings 1 and 2 represent homoscedastic scenarios while Settings 3 and 4 display two scenarios with unequal covariance structures. Data were generated, for each $i \in \{1, \ldots, a\}$ and $k \in \{1, \ldots, n_i\}$, by

$$\boldsymbol{X}_{ik} = \boldsymbol{\mu}_i + \boldsymbol{V}_i^{1/2}\boldsymbol{\epsilon}_{ik},$$

where $\boldsymbol{V}_i^{1/2}$ denotes the square root of the matrix $\boldsymbol{V}_i$, i.e., $\boldsymbol{V}_i = \boldsymbol{V}_i^{1/2} \cdot \boldsymbol{V}_i^{1/2}$. The mean vectors $\boldsymbol{\mu}_i$ were set to $\mathbf{0}$ in both groups. The iid random errors $\boldsymbol{\epsilon}_{ik} = (\epsilon_{ik1}, \ldots, \epsilon_{ikd})^\top$ with mean $\mathrm{E}(\boldsymbol{\epsilon}_{ik}) = \mathbf{0}_d$ and $\mathrm{cov}(\boldsymbol{\epsilon}_{ik}) = \boldsymbol{I}_{d\times d}$ were generated by simulating independent standardized components

$$_{iks} = \frac{Y_{iks} - \mathrm{E}(Y_{iks})}{\sqrt{\mathrm{var}(Y_{iks})}}$$

for various distributions of Y_{iks}. In particular, we simulated normal, χ_3^2, lognormal, t_3 and double-exponential (or Laplace) distributed random variables. We investigated balanced as well as unbalanced designs with sample size vectors $\boldsymbol{n}^{(1)} = (10, 10)^\top$, $\boldsymbol{n}^{(2)} = (20, 20)^\top$, $\boldsymbol{n}^{(3)} = (10, 20)^\top$ and $\boldsymbol{n}^{(4)} = (20, 10)^\top$, respectively. A major criterion concerning the accuracy of the procedures is their behavior in situations where increasing variances (Settings 3–4 above) are combined with increasing sample sizes ($\boldsymbol{n}^{(3)}$, positive pairing) or decreasing sample sizes ($\boldsymbol{n}^{(4)}$, negative pairing).

In this setting, we tested the null hypothesis $\mathcal{H}_0^\mu : \{(\boldsymbol{P}_a \otimes \boldsymbol{I}_d)\boldsymbol{\mu} = \mathbf{0}\} = \{\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2\}$, i.e., no treatment effect. The resulting type-I error rates (nominal level $\alpha = 5\%$) for $d = 4$ and $d = 8$ endpoints are displayed in Table 1 (normal distribution) and Table 2 (χ_3^2 distribution), respectively. Further simulation results for lognormal, t_3 and Laplace distributed errors for the parametric bootstrap of WTS and MATS can be found in Tables 7–9 in the Online Supplement.

As already noticed by [21], the WTS with the χ^2 approximation is far too liberal, reaching type-I error rates of more than 50% in some scenarios (e.g., for $d = 8$ with negative pairing, i.e., covariance setting S3 and $\boldsymbol{n} = (20, 10)^\top$). Even in the scenarios with only $d = 4$ dimensions and $\boldsymbol{n} = (20, 20)^\top$, the error rates are around 9% instead of 5%. The parametric bootstrap of the WTS greatly improves this behavior for all situations. However, it still shows a rather liberal behavior with type-I error rates of around 10% in some situations, e.g., $d = 8$ dimensions with S3 or S4 and $\boldsymbol{n} = (20, 10)^\top$ in Tables 1 and 2.

The wild bootstrap of the MATS shows a rather liberal behavior across all scenarios and can therefore not be recommended in practice. In contrast, both the parametric and the nonparametric bootstrap of the MATS show a very accurate type-I error rate control. The nonparametric bootstrap is often slightly more conservative than the parametric bootstrap and thus works better in situations with negative pairing, especially for the χ_3^2 distribution, i.e., for S3 and S4 with $\boldsymbol{n} = (20, 10)^\top$ and $d = 4$ or $d = 8$ dimensions in Table 2. In most other scenarios, however, the parametric bootstrap yields slightly better results. The improvement of the parametric bootstrap MATS over WTS (PBS) and nonparametric bootstrap MATS is most pronounced for large d, i.e., in situations where d is close to $\min(n_1, n_2)$.

However, in situations with negative pairing and skewed distributions (see Table 2 as well as Table 7 in the Online Supplement), the parametric bootstrap MATS shows a slightly liberal behavior. For t_3 and Laplace distributed errors and negative pairing, in contrast, the parametric bootstrap MATS is slightly conservative, see Tables 8–9 in the Online Supplement, respectively.

Surprisingly, the resampling approaches based on the MATS improve with growing d in most settings, i.e., when the number of endpoints is closer to the sample size. The WTS approach, in contrast, gets worse in these scenarios. This might be an interesting approach for future research in high-dimensional settings such as in [29].

As a result, we find that the MATS with the parametric bootstrap approximation is the best procedure in most scenarios. Especially, it is less conservative than the nonparametric bootstrap approximation and less liberal than the WTS equipped with the parametric bootstrap approach over all simulation settings. Only in situations with negative pairing and skewed distributions, the new procedure shows a slightly liberal behavior.

5.1.1. Singular covariance matrix

In order to analyze the behavior of the discussed methods in designs involving singular covariance matrices, we considered the one-way layout described above with $a = 2$ groups and $d \in \{4, 8\}$ observations involving the following

Table 1
Type-I error rates in % (nominal level $\alpha = 5\%$) for the WTS with χ^2 approximation and parametric bootstrap (PBS) and the MATS with wild bootstrap (wild), parametric bootstrap (PBS) and nonparametric bootstrap (NPBS) in the one-way layout for the normal distribution.

d	Cov	$\boldsymbol{n}$	WTS (χ^2)	WTS (PBS)	MATS (wild)	MATS (PBS)	MATS (NPBS)
$d = 4$	S1	(10, 10)	15.2	4.5	6.9	5.2	4.4
		(10, 20)	14.5	5.9	6.9	5.1	4.5
		(20, 10)	14.0	5.6	7.2	5.3	4.8
		(20, 20)	9.5	5.3	6.0	5.1	5.1
	S2	(10, 10)	15.2	4.5	7.0	5.0	4.5
		(10, 20)	14.5	5.8	6.9	5.0	4.5
		(20, 10)	14.0	5.6	7.3	5.5	4.9
		(20, 20)	9.5	5.4	6.3	5.2	5.0
	S3	(10, 10)	18.3	5.5	7.3	4.8	3.6
		(10, 20)	10.9	4.7	6.4	4.8	4.4
		(20, 10)	21.4	6.6	7.8	4.8	3.4
		(20, 20)	11.2	5.7	6.3	5.1	4.6
	S4	(10, 10)	18.3	5.6	7.5	4.8	3.9
		(10, 20)	11.0	5.2	6.1	4.6	4.3
		(20, 10)	21.0	6.7	7.9	4.7	3.2
		(20, 20)	10.9	5.7	6.2	5.0	4.7
$d = 8$	S1	(10, 10)	38.6	4.7	7.7	5.1	4.3
		(10, 20)	31.0	6.2	6.9	5.0	4.2
		(20, 10)	32.1	6.1	6.6	4.6	4
		(20, 20)	17.0	4.9	5.8	4.8	4.8
	S2	(10, 10)	38.6	4.5	7.9	4.3	3.4
		(10, 20)	31.0	6.3	7.4	4.3	3.6
		(20, 10)	32.1	6.1	7.0	4.1	3.4
		(20, 20)	17.0	4.7	6.2	4.8	4.5
	S3	(10, 10)	50.1	6.6	7.9	4.2	2.8
		(10, 20)	21.8	4.1	6.3	4.4	4.1
		(20, 10)	55.0	10.3	8.5	3.6	2.2
		(20, 20)	21.9	5.4	6.1	4.0	3.6
	S4	(10, 10)	48.9	6.3	7.8	3.6	2.4
		(10, 20)	21.9	4.2	6.3	3.8	3.4
		(20, 10)	54.1	10.4	8.4	3.5	2.0
		(20, 20)	21.8	5.2	6.0	3.9	3.6

covariance settings (displayed for $d = 4$):

$$\text{Setting 5:} \quad \boldsymbol{V}_1 = \begin{pmatrix} 1 & 1/2 & 1 & 1 \\ 1/2 & 1 & 1/2 & 1/2 \\ 1 & 1/2 & 1 & 1 \\ 1 & 1/2 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{V}_2 = \boldsymbol{V}_1 + 0.5\boldsymbol{J}_d$$

$$\text{Setting 6:} \quad \boldsymbol{V}_1 = \begin{pmatrix} 1 & 0.6 & 0.36 & 0.18 \\ 0.6 & 1 & 0.6 & 0.3 \\ 0.36 & 0.6 & 1 & 0.5 \\ 0.18 & 0.3 & 0.5 & 0.25 \end{pmatrix}, \quad \boldsymbol{V}_2 = \boldsymbol{V}_1 + 0.5\boldsymbol{J}_d$$

$$\text{Setting 7:} \quad \boldsymbol{V}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 0.5 \end{pmatrix}, \quad \boldsymbol{V}_2 = \boldsymbol{V}_1 + 0.5\boldsymbol{J}_d$$

Setting 6 is based on an $AR(0.6)$ covariance matrix (see Setting 2 above), where the last row and column have been replaced by half the row/column before, respectively. Setting 7 is based on $\widetilde{\boldsymbol{V}}_1 = \text{diag}(2^{s/2})$ for all $s \in \{0, \ldots, d-1\}$, where the last row and column have been replaced by half the row/column before. We have considered the same sample size vectors as above.

The results are displayed in Tables 3–4. The parametric bootstrap of the MATS again yields the best results in almost all scenarios. The wild bootstrap, in contrast, is again rather liberal. For the χ^2 approximation of the WTS, the results are in concordance with the theoretical reflections mentioned in Section 2. Covariance setting S5 corresponds to the case, where the rank of $\boldsymbol{T}$ and $\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}$ differs and as calculated above, the χ^2 approximation becomes very conservative here. In Settings S6–S7, in contrast, there is no rank jump despite the singular covariance matrices and the χ^2 approximation shows its usual liberal behavior.

Table 2
Type-I error rates in % (nominal level $\alpha = 5\%$) for the WTS with χ^2 approximation and parametric bootstrap (PBS) and the MATS with wild bootstrap (wild), parametric bootstrap (PBS) and nonparametric bootstrap (NPBS) in the one-way layout for the χ^2_3-distribution.

d	Cov	$\boldsymbol{n}$	WTS (χ^2)	WTS (PBS)	MATS (wild)	MATS (PBS)	MATS (NPBS)
$d = 4$	S1	(10, 10)	15.3	4.0	7.1	4.8	3.4
		(10, 20)	13.9	5.5	7.3	5.6	4.6
		(20, 10)	14.6	5.7	7.7	5.9	4.6
		(20, 20)	8.9	4.7	6.3	5.5	5.0
	S2	(10, 10)	15.3	4.1	7.1	4.5	3.2
		(10, 20)	13.9	5.6	7.5	5.5	4.5
		(20, 10)	14.6	5.8	7.7	5.5	4.5
		(20, 20)	8.9	4.7	6.3	5.3	4.7
	S3	(10, 10)	20.6	7.1	9.5	6.1	3.8
		(10, 20)	11.2	4.8	6.9	4.8	3.7
		(20, 10)	26.2	10.9	12.3	8.9	5.6
		(20, 20)	12.8	6.6	7.6	6.0	4.7
	S4	(10, 10)	21.2	7.2	9.6	6.2	3.8
		(10, 20)	11.1	5.0	6.9	4.7	3.4
		(20, 10)	26.5	10.7	12.7	8.9	5.6
		(20, 20)	12.9	6.7	7.7	6.2	4.7
$d = 8$	S1	(10, 10)	39.3	3.8	7.7	4.9	3.4
		(10, 20)	32.3	5.5	7.6	5.9	4.7
		(20, 10)	33.4	6.3	7.2	5.1	4.2
		(20, 20)	16.9	4.5	5.9	4.9	4.6
	S2	(10, 10)	39.3	3.8	8.1	4.3	2.7
		(10, 20)	32.3	5.5	8.6	5.2	4.0
		(20, 10)	33.4	6.3	7.6	4.9	3.9
		(20, 20)	16.9	4.5	6.2	4.5	4.0
	S3	(10, 10)	53.1	6.8	10.2	5.5	3.1
		(10, 20)	23.4	4.8	6.8	4.6	3.5
		(20, 10)	59.9	13.9	13.7	8.1	4.6
		(20, 20)	24.8	6.9	7.9	5.7	4.1
	S4	(10, 10)	52.5	6.3	11.0	5.3	2.6
		(10, 20)	24.3	4.5	7.1	4.1	2.8
		(20, 10)	59.0	13.6	14.8	8.4	4.5
		(20, 20)	24.3	6.9	7.7	5.7	4.0

Since the rank of $\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}$ is not known in practice, the WTS should not be used for data with possibly singular covariance matrices. It turns out, however, that the parametric bootstrap of the WTS is relatively robust against singular covariance matrices. Its behavior is comparable to the scenarios above with non-singular covariance matrices. It is, however, rather liberal for $\boldsymbol{n} = (20, 10)^\top$, especially with the χ^2_3 distribution; see Table 4. This behavior is improved by the parametric bootstrap MATS, e.g., for $d = 8$ and S7, the WTS (PBS) leads to a type-I error of 9%, whereas the MATS (PBS) is at 5.1%. The nonparametric bootstrap, in contrast, sometimes leads to strictly conservative test decisions. This is especially apparent for $d = 8$ and covariance setting S7 in Tables 3–4.

5.2. Two-way layout

We have investigated the behavior of the methods in a setting with two crossed factors A and B, which is again adapted from [21]. In particular, we simulated 2×2 designs with covariance matrices similar to the one-way layout above. A detailed description of the simulation settings as well as the results for the main and interaction effects are deferred to the Online Supplement. Here we only summarize our findings. Since the total sample size N is larger in this scenario, the asymptotic results come into play and therefore all methods lead to more accurate results than in the one-way layout. Nevertheless we find a similar behavior as in the one-way layout. Again, the MATS and the WTS with the parametric bootstrap approach control the type-I error very accurately, whereas the nonparametric bootstrap approach leads to slightly more conservative results. Both the WTS with χ^2 approximation and the wild bootstrap MATS cannot be recommended due to their liberal behavior. In situations with negative pairing (covariance setting 10 and 11 with sample size vector $\boldsymbol{n}^{(3)}$), the parametric bootstrap MATS improves the slightly liberal behavior of the WTS; see e.g., Table 10 for the normal distribution, where the WTS (PBS) leads to a type-I error of 6.1% while the MATS (PBS) is at 4.9%.

5.3. Power

We have investigated the empirical power of the proposed methods to detect a fixed alternative in the simulation scenarios above. Data were simulated as described in Section 5.1 but now with $\boldsymbol{\mu}_1 = \boldsymbol{0}$ and $\boldsymbol{\mu}_2 = (\delta, \ldots, \delta)^\top$ for varying

Table 3
Type-I error rates in % (nominal level $\alpha = 5\%$) for the WTS with χ^2 approximation and parametric bootstrap (PBS) and the MATS with wild bootstrap (wild), parametric bootstrap (PBS) and nonparametric bootstrap (NPBS) in the one-way layout with singular covariance matrices for the normal distribution.

d	Cov	$\boldsymbol{n}$	WTS (χ^2)	WTS (PBS)	MATS (wild)	MATS (PBS)	MATS (NPBS)
$d = 4$	S5	(10, 10)	3.1	5.1	5.9	4.8	4.4
		(10, 20)	2.6	5.1	5.7	4.9	4.6
		(20, 10)	2.7	5.3	6.6	5.3	4.7
		(20, 20)	1.7	4.8	5.6	5.2	5.0
	S6	(10, 10)	16.7	5.3	7.1	5.0	4.6
		(10, 20)	12.4	5.1	6.0	4.7	4.3
		(20, 10)	17.1	6.1	7.1	5.3	4.6
		(20, 20)	10.2	5.5	6.0	5.2	5.1
	S7	(10, 10)	16.6	5.2	7.3	4.7	4.0
		(10, 20)	12.3	5.8	6.6	4.6	4.2
		(20, 10)	16.3	5.7	6.9	4.5	4.0
		(20, 20)	9.4	4.8	5.9	4.8	4.8
$d = 8$	S5	(10, 10)	2.8	4.5	6.2	5.0	4.7
		(10, 20)	2.3	4.9	5.5	4.9	4.7
		(20, 10)	2.6	4.4	5.6	4.7	4.3
		(20, 20)	1.5	4.6	5.5	4.9	4.8
	S6	(10, 10)	39.5	4.4	8.2	5.0	4.2
		(10, 20)	28.8	5.4	6.6	4.6	4.2
		(20, 10)	35.7	7.0	7.5	4.8	4.0
		(20, 20)	17.3	4.7	6.1	4.8	4.5
	S7	(10, 10)	38.8	4.2	7.4	4.0	2.9
		(10, 20)	27.4	5.2	6.6	3.6	3.0
		(20, 10)	36.3	6.3	7.4	3.8	3.2
		(20, 20)	17.3	5.1	5.9	4.2	4.0

Table 4
Type-I error rates in % (nominal level $\alpha = 5\%$) for the WTS with χ^2-approximation and parametric bootstrap (PBS) and the MATS with wild bootstrap (wild), parametric bootstrap (PBS) and nonparametric bootstrap (NPBS) in the one-way layout with singular covariance matrices for the χ^2_3 distribution.

d	Cov	$\boldsymbol{n}$	WTS (χ^2)	WTS (PBS)	MATS (wild)	MATS (PBS)	MATS (NPBS)
$d = 4$	S5	(10, 10)	2.7	4.2	6.7	5.4	4.5
		(10, 20)	1.9	4.8	5.9	4.9	4.5
		(20, 10)	3.5	6.5	7.3	5.9	5.2
		(20, 20)	1.7	4.9	6.2	5.7	5.5
	S6	(10, 10)	19.7	7.1	7.4	5.2	4.1
		(10, 20)	14.6	7.1	6.4	5.0	4.3
		(20, 10)	20.1	8.5	8.1	6.4	5.4
		(20, 20)	11.4	6.3	6.5	5.7	5.3
	S7	(10, 10)	19.4	7.0	7.1	4.1	3.1
		(10, 20)	14.5	6.7	6.4	4.2	3.4
		(20, 10)	20.3	8.7	8.3	6.1	4.5
		(20, 20)	11.7	6.4	6.1	5.1	4.5
$d = 8$	S5	(10, 10)	2.4	4.7	6.1	5.1	4.6
		(10, 20)	2.6	5.3	6.1	5.3	5.0
		(20, 10)	3.0	5.6	6.0	5.1	4.6
		(20, 20)	1.2	4.5	5.9	5.3	5.1
	S6	(10, 10)	43.1	5.4	8.2	5.1	3.9
		(10, 20)	30.7	6.6	7.3	5.2	4.2
		(20, 10)	39.2	8.7	8.3	5.6	4.5
		(20, 20)	19.3	5.6	6.8	5.1	4.7
	S7	(10, 10)	42.4	5.5	7.5	3.3	1.7
		(10, 20)	31.1	6.3	7.1	4.0	2.4
		(20, 10)	39.5	9.0	9.2	5.1	3.4
		(20, 20)	18.7	5.3	5.1	3.2	2.6

shifts $\delta \in \{0, 0.5, 1, 1.5, 2, 3\}$. Due to the liberality of the classical Wald-type test and the wild bootstrapped MATS, we only considered the WTS with parametric bootstrap as well as the parametric and nonparametric bootstrap of the MATS. The results for selected scenarios are displayed in Figs. 1–2. The plots show that both resampling versions of the MATS have a higher power for detecting the fixed alternative than the WTS. The parametric bootstrap of the MATS has a slightly higher power than the nonparametric bootstrap, a behavior that is more pronounced for the χ^2 distribution (Fig. 1). Moreover, the power analysis shows a clear advantage of applying the parametric bootstrap approach to the MATS over its application to the WTS. For example, in the scenario with normally distributed data, $d = 8$ dimensions, covariance setting S4 and

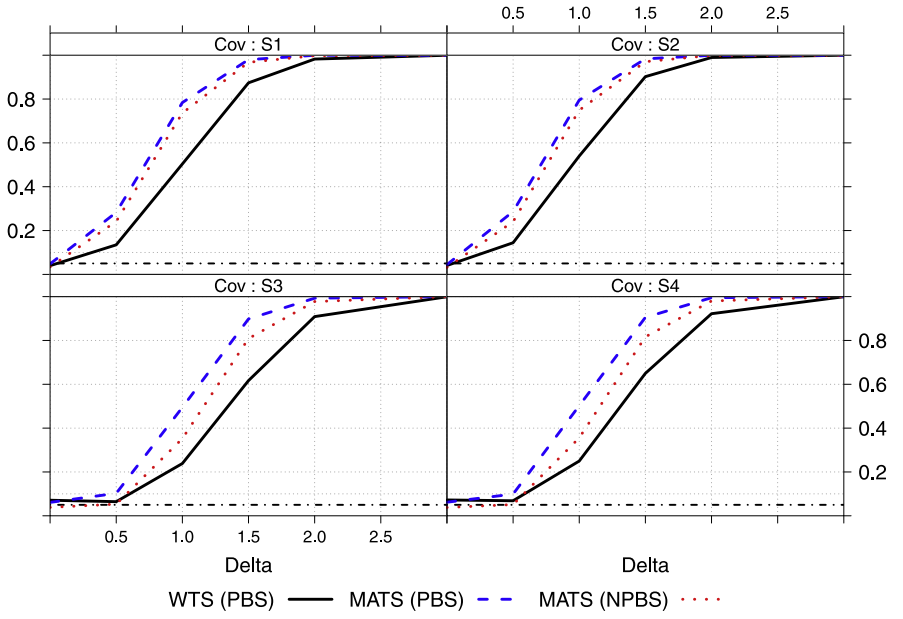

Fig. 1. Empirical power results for the WTS with parametric bootstrap as well as the MATS with parametric (PBS) and nonparametric (NPBS) bootstrap for χ^2_3 distributed data with $d = 4$ dimensions and sample sizes $\boldsymbol{n} = (10, 10)^\top$.

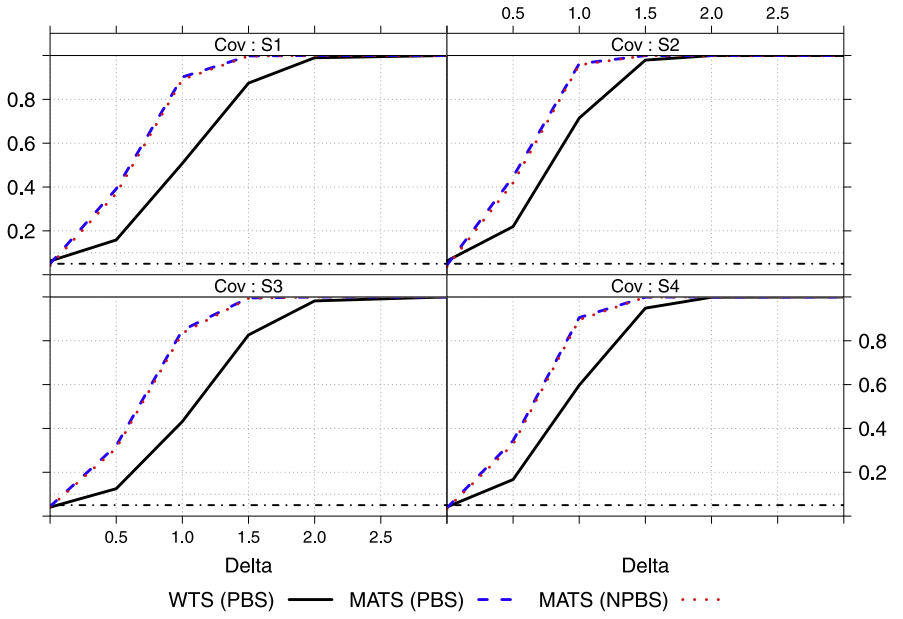

Fig. 2. Empirical power results for the WTS with parametric bootstrap as well as the MATS with parametric (PBS) and nonparametric (NPBS) bootstrap for normally distributed data with $d = 8$ dimensions and sample sizes $\boldsymbol{n} = (10, 20)^\top$.

$\boldsymbol{n} = (10, 20)^\top$ observations (Fig. 2), the parametric bootstrap MATS has twice as much power as its WTS version in case of $\delta = 0.5$ (34.4% as compared to 16.7%). Similar differences can also be observed in some of the other settings.

6. Application: analysis of the data example

As a data example, we consider seven demographic factors of US citizens in 43 states. Our aim is to investigate whether these factors differ between the states. The full data set 'county_facts.csv' is available from kaggle (www.kaggle.com/

Table 5
Descriptive statistics of the data example: Reported are the sample sizes and the 7-dimensional mean vectors for each of the 43 states.

State	n	PST045214	SEX255214	RHI125214	RHI225214	RHI325214	RHI425214	RHI525214
AK	29	25 404.55	45.73	52.51	1.92	31.89	5.68	0.55
AL	67	72 378.76	51.26	68.26	28.66	0.80	0.73	0.11
AR	75	39 551.59	50.54	80.41	16.13	0.89	0.78	0.10
AZ	15	448 765.60	49.63	78.99	2.33	14.76	1.45	0.20
CA	58	669 008.62	49.52	81.59	3.57	3.16	7.50	0.39
CO	64	83 685.41	48.05	92.62	1.81	2.06	1.30	0.12
FL	67	296 914.88	48.60	80.61	15.00	0.73	1.70	0.10
GA	159	63 505.30	50.31	68.19	28.36	0.49	1.30	0.12
IA	99	31 385.11	50.15	95.69	1.43	0.46	1.13	0.08
ID	44	37 146.91	49.26	94.36	0.58	2.03	0.82	0.16
IL	102	126 280.20	49.93	91.67	5.30	0.35	1.23	0.03
IN	92	71 704.95	50.20	94.42	2.85	0.36	0.99	0.04
KS	105	27 657.34	49.74	93.64	2.11	1.23	0.91	0.08
KY	120	36 778.81	50.25	93.87	3.86	0.29	0.57	0.06
LA	64	72 651.19	49.95	64.63	32.08	0.84	0.94	0.05
MD	24	249 016.96	50.88	73.14	20.58	0.44	3.41	0.10
ME	16	83 130.56	50.89	95.67	0.98	0.87	0.89	0.02
MI	83	119 396.11	49.67	91.18	4.05	1.69	1.03	0.03
MN	87	62 726.13	49.88	92.93	1.59	2.24	1.47	0.06
MO	115	52 726.86	50.11	93.08	3.71	0.62	0.74	0.12
MS	82	36 513.16	51.01	56.48	41.21	0.68	0.57	0.04
MT	56	18 278.20	49.16	88.81	0.41	8.00	0.48	0.05
NC	100	99 439.64	50.82	74.15	20.80	1.93	1.26	0.10
ND	53	13 952.49	48.69	90.26	0.79	6.79	0.59	0.04
NE	93	20 231.22	49.82	95.43	0.93	1.73	0.59	0.07
NJ	21	425 627.38	51.06	76.97	13.19	0.56	7.16	0.10
NM	33	63 199.15	49.37	86.02	1.83	8.70	1.11	0.14
NV	17	167 005.82	47.41	87.34	2.88	4.27	2.26	0.31
NY	62	318 487.53	50.22	87.43	6.98	0.70	2.90	0.06
OH	88	131 751.85	50.38	92.70	4.33	0.28	0.97	0.02
OK	77	50 364.30	49.84	78.41	3.78	11.18	0.86	0.14
OR	36	110 284.42	50.04	91.49	0.89	2.48	1.77	0.30
PA	67	190 853.87	50.03	91.99	4.86	0.27	1.46	0.03
SC	46	105 053.96	50.93	60.75	36.10	0.64	0.93	0.09
SD	66	12 926.89	49.48	82.51	0.78	14.02	0.63	0.04
TN	95	68 940.55	50.46	89.75	7.50	0.45	0.73	0.05
TX	254	106 129.76	49.20	89.24	6.82	1.17	1.16	0.08
UT	29	101 479.38	49.12	92.88	0.66	3.30	1.04	0.36
VA	134	62 136.49	50.25	75.67	18.84	0.51	2.09	0.07
WA	39	181 064.87	49.85	88.87	1.61	3.01	2.79	0.34
WI	72	79 966.17	49.64	92.55	1.71	2.96	1.31	0.04
WV	55	33 642.29	50.12	95.57	2.38	0.25	0.49	0.01
WY	23	25 397.96	49.04	94.04	1.17	2.13	0.83	0.10

joelwilson/2012-2016-presidential-elections). In order to have sufficient sample sizes for the analysis and to avoid a high-dimensional setting, we exclude all states with less than 15 counties. In particular, we removed Connecticut, Delaware, Hawaii, Massachusetts, New Hampshire, Rhode Island and Vermont.

We consider the following demographic factors: the population estimate for 2014 (PST045214), the percentage of female citizens in 2014 (SEX255214) as well as the percentage of white (RHI125214), black or African American (RHI225214), American Indian and Alaska native (RHI325214), Asian (RHI425214) and native Hawaiian and other Pacific islanders (RHI525214) citizens in 2014. This results in a one-way layout with $a = 43$ levels of the factor 'state' and $d = 7$ dimensions. The sample sizes and mean values for the different states can be found in Table 5. As an example, Fig. 3 displays boxplots for the percentage of white citizens across the different states.

We now want to analyze, whether there is a significant difference in the multivariate means for the different states. The null hypothesis of interest thus is $\mathcal{H}_0 : \{(\boldsymbol{P}_{43} \otimes \boldsymbol{I}_7)\boldsymbol{\mu} = \boldsymbol{0}\}$. Since the empirical covariance matrix is computationally singular in this example (reciprocal condition number 1.7e−16), we cannot apply the Wald-type test. Thus, we consider the parametric bootstrap approach of the MATS which yielded the best results in the simulation study. Computation of the MATS results in a value of $Q_N = 393.927$ and the parametric bootstrap routine with 1000 bootstrap runs gives a p-value smaller than 0.0001, implying that there is indeed a significant difference between the states with respect to the seven demographic measurements.

A confidence region for this effect can be constructed as described in Section 4. The analysis of this example, including the calculation of the confidence region, can be conducted using the R package `MANOVA.RM`.

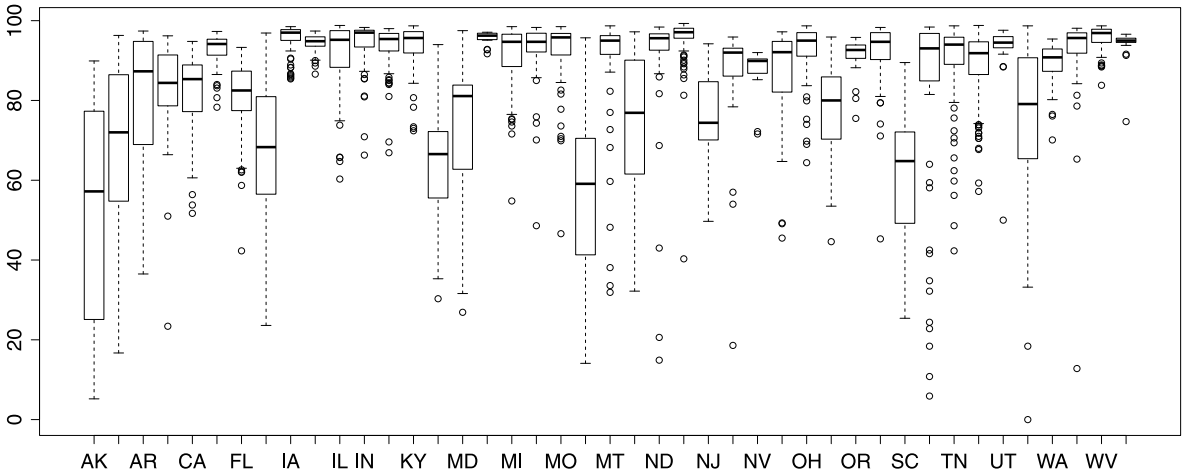

Fig. 3. Boxplots of the percentage of white citizens across the different states.

7. Conclusions and discussion

We have investigated a test statistic for multivariate data (MATS) which is based on a modified Dempster statistic. Contrary to classical MANOVA models, we incorporate general heteroscedastic designs and allow for singular covariance matrices while postulating their existence as solely distributional assumption. Moreover, our proposed MATS statistic is invariant under linear transformations of the response variables.

In order to improve the small-sample behavior of the test statistic, we have investigated different bootstrap approaches, namely a parametric bootstrap, a wild bootstrap and a nonparametric bootstrap procedure. We have rigorously proven that they lead to asymptotically exact and consistent tests and even analyzed their local power behavior.

In a large simulation study, the parametric bootstrap turned out to perform best in most scenarios, even with skewed data and heteroscedastic variances. Although the type-I error control is still not ideal in the latter case, the method performed advantageously over the parametric bootstrap of the WTS proposed in [21] and has the additional advantage of being applicable to situations with singular covariance matrices. In situations with skewed distributions, the parametric bootstrap of the MATS yielded more robust results than the WTS. The wild bootstrap approach, in contrast, turned out to be very liberal in all scenarios, while the nonparametric bootstrap was mostly slightly more conservative than the parametric bootstrap. Power simulations showed a clear advantage of the parametric bootstrap MATS compared to the WTS (PBS) as well as the nonparametric bootstrap. All in all, we therefore recommend the parametric bootstrap based on the MATS for practical applications in a multivariate setting.

Furthermore, we have constructed confidence regions and simultaneous confidence intervals for contrasts $\boldsymbol{h}^\top \boldsymbol{\mu}$ based on the bootstrap quantiles. These confidence regions provide an additional benefit for the analysis of multivariate data since they allow for more detailed insight into the nature of the estimates.

In order to facilitate application of the proposed methods, the parametric bootstrap test and the calculation of confidence regions are implemented in the `R` package `MANOVA.RM`.

Following the idea of [36] we plan to extend our concepts to the high-dimensional setting, i.e., where the sample size N may be less than the dimension d. This approach looks promising, since we have seen in the simulation study that the MATS with the parametric bootstrap approach exhibited an improved type-I error control with increasing d. However, the extension to high-dimensional data requires different techniques and will be part of future research.

Acknowledgments

The authors thank Dr. Jan Paul and Prof. Dr. Volker Rasche for providing the cardiology data example used in the Online Supplement as well as the Editor-in-Chief, the Associate Editor and two anonymous referees for their comments, which greatly improved this paper. This work was supported by the German Research Foundation, projects DFG PA 2409/3-1 and PA 2409/4-1.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2017.12.008.

References

[1] M.S. Bartlett, A note on tests of significance in multivariate analysis, Math. Proc. Cambridge Philos. Soc. 35 (1939) 180–185.
[2] A.C. Bathke, S.W. Harrar, L.V. Madden, How to compare small multivariate samples using nonparametric tests, Comput. Statist. Data Anal. 52 (2008) 4951–4965.
[3] J. Beyersmann, S.D. Termini, M. Pauly, Weak convergence of the wild bootstrap for the Aalen–Johansen estimator of the cumulative incidence function of a competing risk, Scand. J. Stat. 40 (2013) 387–402.
[4] E. Brunner, Asymptotic and approximate analysis of repeated measures designs under heteroscedasticity, in: Mathematical Statistics with Applications in Biometry, 2001.
[5] E. Brunner, F. Konietschke, M. Pauly, M.L. Puri, Rank-based procedures in factorial designs: Hypotheses about non-parametric treatment effects, J. R. Stat. Soc. Ser. B Stat. Methodol. 79 (2017) 1463–1485.
[6] E. Brunner, U. Munzel, M.L. Puri, Rank-score tests in factorial designs with repeated measures, J. Multivariate Anal. 70 (1999) 286–317.
[7] E. Brunner, M.L. Puri, Nonparametric methods in factorial designs, Statist. Papers 42 (2001) 1–52.
[8] A.C. Cameron, J.B. Gelbach, D.L. Miller, Bootstrap-based improvements for inference with clustered errors, Rev. Econ. Stat. 90 (2008) 414–427.
[9] A.C. Cameron, D.L. Miller, A practitioner's guide to cluster-robust inference, J. Hum. Resour. 50 (2015) 317–372.
[10] E. Chung, J.P. Romano, Multivariate and multiple permutation tests, J. Econometrics 193 (2016) 76–91.
[11] R. Davidson, E. Flachaire, The wild bootstrap, tamed at last, J. Econometrics 146 (2008) 162–169.
[12] A.P. Dempster, A high dimensional two sample significance test, Ann. Math. Statist. 29 (1958) 995–1010.
[13] A.P. Dempster, A significance test for the separation of two highly multivariate small samples, Biometrics 16 (1960) 41–50.
[14] S. Friedrich, E. Brunner, M. Pauly, Permuting longitudinal data in spite of the dependencies, J. Multivariate Anal. 153 (2017) 255–265.
[15] S. Friedrich, F. Konietschke, M. Pauly, A wild bootstrap approach for nonparametric repeated measurements, Comput. Statist. Data Anal. 113 (2017) 38–52.
[16] S.W. Harrar, A.C. Bathke, A modified two-factor multivariate analysis of variance: Asymptotics and small sample approximations, Ann. Inst. Statist. Math. 64 (2012) 135–165.
[17] M. Hasler, L.A. Hothorn, Multiple contrast tests in the presence of heteroscedasticity, Biometrical J. 50 (2008) 793–800.
[18] H. Hotelling, A generalized t-test and measure of multivariate dispersion, in: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, The Regents of the University of California, 1951.
[19] T. Hothorn, F. Bretz, P. Westfall, Simultaneous inference in general parametric models, Biometrical J. 50 (2008) 346–363.
[20] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, sixth ed., Prentice Hall, 2007.
[21] F. Konietschke, A. Bathke, S. Harrar, M. Pauly, Parametric and nonparametric bootstrap methods for general MANOVA, J. Multivariate Anal. 140 (2015) 291–301.
[22] K. Krishnamoorthy, F. Lu, A parametric bootstrap solution to the MANOVA under heteroscedasticity, J. Stat. Comput. Simul. 80 (2010) 873–887.
[23] D. Lawley, A generalization of Fisher's z test, Biometrika 30 (1938) 180–187.
[24] D. Lin, Non-parametric inference for cumulative incidence functions in competing risks studies, Stat. Med. 16 (1997) 901–910.
[25] R.Y. Liu, Bootstrap procedures under some non-iid models, Ann. Statist. 16 (1988) 1696–1708.
[26] C. Liu, A.C. Bathke, S.W. Harrar, A nonparametric version of Wilks' lambda: Asymptotic results and small sample approximations, Statist. Probab. Lett. 81 (2011) 1502–1506.
[27] E. Mammen, When Does Bootstrap Work? Asymptotic Results and Simulations, Springer Science & Business Media, 1993.
[28] R. Marcus, P. Eric, K.R. Gabriel, On closed testing procedures with special reference to ordered analysis of variance, Biometrika 63 (1976) 655–660.
[29] M. Pauly, D. Ellenberger, E. Brunner, Analysis of high-dimensional one group repeated measures designs, Statistics 49 (2015) 1243–1261.
[30] F. Pesarin, L. Salmaso, Permutation Tests for Complex Data: Theory, Applications and Software, Wiley, New York, 2010.
[31] F. Pesarin, L. Salmaso, A review and some new results on permutation testing for multivariate problems, Statist. Comput. 22 (2012) 639–646.
[32] K. Pillai, Some new test criteria in multivariate analysis, Ann. Math. Statist. 26 (1955) 117–121.
[33] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2016.
[34] Ł. Smaga, Bootstrap methods for multivariate hypothesis testing, Comm. Statist. Simulation Comput. (2017).
[35] E. Sonnemann, General solutions to multiple testing problems, Biometrical J. 50 (2008) 641–656.
[36] M.S. Srivastava, T. Kubokawa, Tests for multivariate analysis of variance in high dimension under non-normality, J. Multivariate Anal. 115 (2013) 204–216.
[37] G. Vallejo, M. Ato, Robust tests for multivariate factorial designs under heteroscedasticity, Behav. Res. Methods 44 (2012) 471–489.
[38] G. Vallejo, M. Fernández, P.E. Livacic-Rojas, Analysis of unbalanced factorial designs with heteroscedastic data, J. Stat. Comput. Simul. 80 (2010) 75–88.
[39] S. Van Aelst, G. Willems, Robust and efficient one-way MANOVA tests, J. Amer. Statist. Assoc. 106 (2011) 706–718.
[40] S. Van Aelst, G. Willems, Fast and robust bootstrap for multivariate inference: The R package FRB, J. Stat. Softw. 53 (2013) 1–32.
[41] S.S. Wilks, Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution, Ann. Math. Statist. 17 (1946) 257–281.
[42] C.-F.J. Wu, Jackknife, bootstrap and other resampling methods in regression analysis, Ann. Statist. 14 (1986) 1261–1295.
[43] L.-W. Xu, F.-Q. Yang, A. Abula, S. Qin, A parametric bootstrap approach for two-way ANOVA in presence of possible interactions with unequal variances, J. Multivariate Anal. 115 (2013) 172–180.