Statistics and Probability Letters 208 (2024) 110038

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Choice of the hypothesis matrix for using the Wald-type-statistic

Paavo Sattler[a,*], Georg Zimmermann[b,c]

[a] Institute for Mathematical Statistics and Industrial Applications, Faculty of Statistics, Technical University of Dortmund, Joseph-von-Fraunhofer-Strasse 2-4, 44227 Dortmund, Germany
[b] Team Biostatistics and Big Medical Data, IDA Lab Salzburg, Paracelsus Medical University Salzburg, Strubergasse 16 5020 Salzburg, Austria
[c] Department of Artificial Intelligence and Human Interfaces, Paris Lodron University of Salzburg, Hellbrunner Strasse 34 5020 Salzburg, Austria

A R T I C L E I N F O

Keywords:
Multivariate data
Hypothesis matrix

A B S T R A C T

A widely used formulation for null hypotheses in the analysis of multivariate d-dimensional data is $\mathcal{H}_0 : \boldsymbol{H}\boldsymbol{\theta} = \boldsymbol{y}$ with $\boldsymbol{H} \in \mathbb{R}^{m\times d}$, $\boldsymbol{\theta} \in \mathbb{R}^d$ and $\boldsymbol{y} \in \mathbb{R}^m$, where $m \leq d$. Here the unknown parameter vector $\boldsymbol{\theta}$ can, for example, be the expectation vector $\boldsymbol{\mu}$, a vector $\boldsymbol{\beta}$ containing regression coefficients or a quantile vector $\boldsymbol{q}$. Also, the vector of nonparametric relative effects $\boldsymbol{p}$ or an upper triangular vectorized covariance matrix $\boldsymbol{v}$ are useful choices. However, even without multiplying the hypothesis with a scalar $\gamma \neq 0$, there is a multitude of possibilities to formulate the same null hypothesis with different hypothesis matrices $\boldsymbol{H}$ and corresponding vectors $\boldsymbol{y}$. Although it is a well-known fact that in case of $\boldsymbol{y} = \boldsymbol{0}$ there exists a unique projection matrix $\boldsymbol{P}$ with $\boldsymbol{H}\boldsymbol{\theta} = \boldsymbol{0} \Leftrightarrow \boldsymbol{P}\boldsymbol{\theta} = \boldsymbol{0}$, for $\boldsymbol{y} \neq \boldsymbol{0}$ such a projection matrix does not necessarily exist. Moreover, such hypotheses are often investigated using a quadratic form as the test statistic and the corresponding projection matrices frequently contain zero rows; so, they are not even efficient from a computational point of view. In this manuscript, we show that for the Wald-type-statistic (WTS), which is one of the most frequently used quadratic forms, the choice of the concrete hypothesis matrix does not affect the test decision. Moreover, some simulations are conducted to investigate the possible influence of the hypothesis matrix on the computation time.

Formulating an appropriate null hypothesis is an important aspect of statistical data analysis. For multivariate analysis of an unknown d-dimensional parameter vector $\boldsymbol{\theta}$, a commonly used hypothesis is given as $\mathcal{H}_0 : \boldsymbol{H}\boldsymbol{\theta} = \boldsymbol{y}$. Here $\boldsymbol{H} \in \mathbb{R}^{m\times d}$ denotes the so-called hypothesis matrix, and $\boldsymbol{y} \in \mathbb{R}^m$, where $m \leq d$. Thereby, the parameter vector $\boldsymbol{\theta}$ depends on the model and the underlying research question of interest. For example, $\boldsymbol{\theta}$ may consist of expectation vectors of one or more groups of observations in a one- or multi-way analysis of variance (*e.g.*, Brunner et al. (2016) and Friedrich et al. (2017)). In Friedrich et al. (2016) and Rubarth et al. (2022), the nonparametric relative effect $\boldsymbol{p}$, which is an important effect measure in nonparametric statistics, is used. Furthermore, also vectors of quantiles $\boldsymbol{q}$ or vectorized covariance matrices $\boldsymbol{v}$ might be considered (*e.g.,* (Ditzhaus et al., 2021b) and Sattler et al. (2022)). In order to demonstrate that these frequently encountered settings are within the scope of the above-mentioned general null hypothesis formulation, we consider the following three examples in more detail now:

(A): Let $X_{i1}, \ldots, X_{in_i}$, $i \in \{1, 2, 3\}$ denote three independent samples of i.i.d. random variables with expectations $\mu_i := \mathbb{E}(\boldsymbol{X}_{i1})$. In such a classical one-way analysis-of-variance setting with three groups, it is usually of interest to compare the expectations μ_1, μ_2, μ_3, that is, to consider the null hypothesis $\mu_1 = \mu_2 = \mu_3$. Now, stacking together the three expectations into one vector, $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^\top$, this null hypothesis could be formulated with $\boldsymbol{H}\boldsymbol{\mu} = \boldsymbol{0}$, where each of the following three matrices might be plugged in for $\boldsymbol{H}$

* Corresponding author.
E-mail address: paavo.sattler@tu-dortmund.de (P. Sattler).

https://doi.org/10.1016/j.spl.2024.110038
Received 11 April 2023; Received in revised form 1 August 2023; Accepted 9 January 2024
Available online 13 January 2024

(actually, there are many more possible choices):

$$\boldsymbol{H}_1 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{pmatrix}, \quad \boldsymbol{H}_2 = 1/3 \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}, \quad \boldsymbol{H}_3 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

(B): Consider the same situation as above, yet assuming that the random variables are ordinal, with possibly different distribution functions. Consequently, expectations cannot be used any more. Instead, for example, an important effect measure could be the so-called pairwise relative effect, which is given through $p_{ij} := P(X_{i1} < X_{j1}) + 1/2P(X_{i1} = X_{j1})$ for $i, j \in \{1, 2, 3\}$. One of the frequently used null hypotheses here is $p_{12} = p_{23} = p_{31} = 1/2$, which means that no group tends to have smaller or bigger values than the other one. Possible hypothesis matrices for the parameter vector $\boldsymbol{p} = (p_{12}, p_{23}, p_{31})^\top$ would be

$$\boldsymbol{H}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{H}_2 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

These correspond to the formulations $\boldsymbol{H}_1 \boldsymbol{p} = (1/2, 1/2, 1/2)^\top$ and $\boldsymbol{H}_2 \boldsymbol{p} = (0, 0, 1/2)^\top$, respectively.

(C): Now, we consider a 2-dimensional random vector $\boldsymbol{X}$ with covariance matrix $\boldsymbol{V}$. If we are interested in testing whether this matrix is the 2-dimensional identity, this could be done by means of the vectorized covariance matrix $\boldsymbol{v} := (v_{11}, v_{12}, v_{22})^\top$. Thereby, v_{rs} denotes the element in row r and column s of $\boldsymbol{V}$. Suitable hypothesis matrices would be

$$\boldsymbol{H}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{H}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$$

and corresponding vectors $\boldsymbol{y}_1 = (1, 0, 1)^\top$ and $\boldsymbol{y}_2 = (1, 0, 0)^\top$, which enables us to formulate the null hypothesis (*i.e.,* that the covariance matrix is equal to the 2-dimensional identity matrix) as $\boldsymbol{H}_1 \boldsymbol{v} = \boldsymbol{y}_1$ or $\boldsymbol{H}_2 \boldsymbol{v} = \boldsymbol{y}_2$, respectively.

These few examples show the wide range of possible parameters and corresponding hypothesis formulations within this class of null hypotheses. However, we have also observed that each of the above-mentioned null hypotheses can be formulated with a multitude of different hypothesis matrices. These matrices might have different numbers of rows or different numbers of non-zero elements; moreover, some of these matrices are idempotent and symmetric, while others are neither. Thus, the question arises which hypothesis matrix to choose, and, importantly, whether the choice of the hypothesis matrix has an impact on the results of the subsequent statistical analysis (i.e., hypothesis tests). For $\boldsymbol{y} = \boldsymbol{0}$ (*i.e.,* $\boldsymbol{H\theta} = \boldsymbol{0}$), there is a general consensus to use the projection matrix given through $\boldsymbol{P} = \boldsymbol{H}^\top (\boldsymbol{H}\boldsymbol{H}^\top)^+ \boldsymbol{H}$, where $\boldsymbol{A}^+$ denotes the Moore–Penrose-Inverse of a matrix $\boldsymbol{A}$. An important advantage of this matrix is its uniqueness. So if $\boldsymbol{H}_1 \boldsymbol{\theta} = \boldsymbol{0}$ and $\boldsymbol{H}_2 \boldsymbol{\theta} = \boldsymbol{0}$ are two ways to express the same hypothesis, this hypothesis can also be formulated by $\boldsymbol{P\theta} = \boldsymbol{H}_1^\top (\boldsymbol{H}_1 \boldsymbol{H}_1^\top)^+ \boldsymbol{H}_1 \boldsymbol{\theta} = \boldsymbol{H}_2^\top (\boldsymbol{H}_2 \boldsymbol{H}_2^\top)^+ \boldsymbol{H}_2 \boldsymbol{\theta} = \boldsymbol{0}$ (see, e.g. Brunner and M.L. (2001)). Thus, the appropriate hypothesis matrix choice is straightforward. Unfortunately, for $\boldsymbol{y} \neq \boldsymbol{0}$, the solution sets of $\boldsymbol{H\theta} = \boldsymbol{y}$ and $\boldsymbol{P\theta} = \boldsymbol{y}$ are not the same. Possibly, by using an appropriate vector $\widetilde{\boldsymbol{y}}$, the equivalence of the hypotheses $\boldsymbol{H\theta} = \boldsymbol{y}$ and $\boldsymbol{P\theta} = \widetilde{\boldsymbol{y}}$ might be obtained, but no universally applicable technique for constructing such $\widetilde{\boldsymbol{y}}$ exists. Another solution could be shifting the original data to reformulate the hypothesis with $\boldsymbol{y} = \boldsymbol{0}$. However, this might only work for some parameters $\boldsymbol{\theta}$, such as expectations or quantiles, yet not for others like the vectorized covariance matrix or relative effects, because the latter ones are invariant under such shifts. Hence, this approach to solving the problem is not universally applicable either. Moreover, even for $\boldsymbol{y} = \boldsymbol{0}$, the unique projection matrix $\boldsymbol{P}$ is always quadratic and often seems inefficient and unwieldy. This can be seen in example (A). The hypothesis matrix $\boldsymbol{H}_2$ is the unique projection matrix with no zero entries, rational numbers as entries and three rows. So in comparison with both other matrices, it is less attractive. For both reasons, we will show that the value of a popular quadratic form, the so-called Wald-type-statistic (WTS), does not depend on the chosen hypothesis matrix.

Main result

Let $\boldsymbol{X}$ be the available data set, which can be from one or more groups, and N the total sample size. Then, the above mentioned WTS is based on an appropriate vector valued statistic $\boldsymbol{T}(\boldsymbol{X})$, given through

$$WTS(\boldsymbol{H}, \boldsymbol{y}) = N \cdot (\boldsymbol{H}\boldsymbol{T}(\boldsymbol{X}) - \boldsymbol{y})^\top (\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^\top)^+ (\boldsymbol{H}\boldsymbol{T}(\boldsymbol{X}) - \boldsymbol{y})$$

with $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{T}(\boldsymbol{X})) \geq 0$, see for example (Munzel and E., 2000). This vector $\boldsymbol{T}(\boldsymbol{X})$ usually converges to a multivariate normal distribution and, for example, might consist of one or more weighted mean vectors. In the case of asymptotic normality of $\boldsymbol{T}(\boldsymbol{X})$, the limit distribution of the WTS is pivot (see, e.g. Rao and S. (1971)), which makes it attractive for permutation approaches and similar techniques, see for example (Pauly et al., 2015) and Ditzhaus et al. (2021a). Now we investigate the conjunction between hypothesis matrices and the corresponding projection matrices.

Theorem 1. *Let $\boldsymbol{H}_1 \boldsymbol{x} = \boldsymbol{y}_1$ and $\boldsymbol{H}_2 \boldsymbol{x} = \boldsymbol{y}_2$ be two systems of linear equations with the same non-trivial solution set, while $\boldsymbol{H}_1 \in \mathbb{R}^{m_1 \times d}$, $\boldsymbol{H}_2 \in \mathbb{R}^{m_2 \times d}$, $\boldsymbol{y}_1 \in \mathbb{R}^{m_1}$, $\boldsymbol{y}_2 \in \mathbb{R}^{m_2}$ and $m_1, m_2 \leq d$. Then, it follows that $\boldsymbol{P}_1 = \boldsymbol{P}_2$ holds.*

This result is already known for $\boldsymbol{y} = \boldsymbol{0}$ and quadratic matrices, where even an "if and only if" relation exists. But, this substantially more general result implies that the corresponding Wald-type statistics are the same, too.

Table 1
Average computation time in seconds of test statistics based on different hypothesis matrices with different dimensions.

	Calculation time in seconds					
d	5	10	20	50	100	200
$WTS(\boldsymbol{H}_1^A, \boldsymbol{0}_{2d})$	0.928	1.452	117.934	1976.501	1616.691	6724.439
$WTS(\boldsymbol{H}_2^A, 0)$	0.737	0.853	0.862	22.373	29.275	44.880
d	15	55	120	210	325	465
$WTS(\boldsymbol{H}_1^B, \gamma \cdot \boldsymbol{1}_d)$	0.818	1.031	83.394	323.360	1370.995	9339.278
$WTS(\boldsymbol{H}_2^B, \gamma)$	0.718	0.742	0.777	23.281	26.594	59.630

Corollary 1. *Let $\boldsymbol{H}_1\boldsymbol{x} = \boldsymbol{y}_1$ and $\boldsymbol{H}_2\boldsymbol{x} = \boldsymbol{y}_2$ be two systems of linear equations with the same non-trivial solution set using $\boldsymbol{H}_1 \in \mathbb{R}^{m_1 \times d}, \boldsymbol{H}_2 \in \mathbb{R}^{m_2 \times d}, \boldsymbol{y}_1 \in \mathbb{R}^{m_1}, \boldsymbol{y}_2 \in \mathbb{R}^{m_2}$ and $m_1, m_2 \le d$. Then $\boldsymbol{H}_1\boldsymbol{\theta} = \boldsymbol{y}_1$ and $\boldsymbol{H}_2\boldsymbol{\theta} = \boldsymbol{y}_2$ are two ways to express the same null hypothesis, and for $\boldsymbol{\Sigma} > 0$ the values of the two corresponding Wald-type-statistics are the same.*

This result thus ensures that the choice of the hypothesis matrix does not influence the test result as long as the chosen matrix appropriately reflects the hypothesis under consideration.

Remark 2.

- Under similar conditions, the result also holds for the so-called modified ANOVA-type-statistic (MATS), which was introduced in Friedrich and M. (2017) and has a similar construction as the WTS (see more details in the Appendix).
- This result does not mean that for $\boldsymbol{H}\boldsymbol{\theta} = \boldsymbol{y}$ an equivalent null hypothesis can be formulated by using $\boldsymbol{P}$, which is only the case for $\boldsymbol{y} = \boldsymbol{0}$.
- For any hypothesis $\boldsymbol{H}\boldsymbol{\theta} = \boldsymbol{y}$, alternative systems of linear equations can be developed by using elementary row operations and removing rows only containing zeros.

Since the chosen hypothesis matrix does not affect the WTS, using a simple and numerically efficient matrix is sensible. Even for $\boldsymbol{y} = \boldsymbol{0}$, the unique projection matrix $\boldsymbol{P}$ is frequently not the best choice as it is often relatively expensive from a numerical point of view, and somewhat complicated with respect to interpretability. From a computational view, choosing a matrix $\boldsymbol{H}$ with $m = \operatorname{rank}(\boldsymbol{H})$ and, therefore, a minimal number of rows is favourable. Depending on the hypothesis, this can result in an essential reduction of computation time, which is investigated in the following section.

Simulation

In this section, we will investigate the influence of the chosen hypothesis matrix on the computation time of the WTS. Here we focus on the following two settings:

A) Consider two independent groups with observations $\boldsymbol{X}_{ik}$, $i \in \{1,2\}$, $k \in \{1,\dots,n_i\}$ that are identically distributed within these groups, and $\mathbb{E}(\boldsymbol{X}_{ij}) = \boldsymbol{\mu}_i \in \mathbb{R}^d$, $\operatorname{Cov}(\boldsymbol{X}_{ij}) = \boldsymbol{V}_i > 0$. For example, consider the case of repeated measures, where the components of $\boldsymbol{\mu}_i$ represent the time-point-specific expectations in group i. In this context, a frequently used null hypothesis is the hypothesis of no group effect for $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$, given through $\overline{\mu}_{1\cdot} = \overline{\mu}_{2\cdot}$. In this setting, the projection matrix can be written as $\boldsymbol{P}_2 = \boldsymbol{I}_2 - \boldsymbol{J}_2/2$, where $\boldsymbol{I}_2$ denotes the identity matrix, and $\boldsymbol{J}_2$ is a 2×2 matrix containing only 1's. Then, the hypothesis of no group effect can be formulated through the hypothesis matrix $\boldsymbol{H}_1^A = (\boldsymbol{P}_2 \otimes \boldsymbol{J}_d)$ by $\boldsymbol{H}_1^A\boldsymbol{\mu} = \boldsymbol{0}_{2d}$. Another way to formulate the same hypothesis would be $\boldsymbol{H}_2^A = (1,\dots,1,-1,\dots,-1) \in \mathbb{R}^{1\times 2d}$ and $\boldsymbol{H}_2^A\boldsymbol{\mu} = 0$.

B) Consider a positive definite symmetric matrix $\boldsymbol{V} \in \mathbb{R}^{p\times p}$ as covariance matrix of observations $\boldsymbol{X}_1,\dots,\boldsymbol{X}_n$. The hypothesis $\mathcal{H}_0 : \operatorname{tr}(\boldsymbol{V}) = \gamma$ can be investigated by using the upper triangular vectorization $\operatorname{vech}(\boldsymbol{V}) = (v_{11},\dots,v_{1p},v_{22},\dots,v_{2p},\dots,v_{pp})^\top \in \mathbb{R}^d$, $d = p(p+1)/2$ as it was introduced in Sattler et al. (2022). Thereby we use the vector $\boldsymbol{h}_p := (1,\boldsymbol{0}_{p-1}^\top,1,\boldsymbol{0}_{p-2}^\top,\dots,1,0,1)^\top \in \mathbb{R}^d$ to define two hypothesis matrices $\boldsymbol{H}_1^B = \boldsymbol{h}_p \cdot \boldsymbol{h}_p^\top$ or $\boldsymbol{H}_2^B = \boldsymbol{h}_p^\top \in \mathbb{R}^{1\times d}$. Then the hypothesis can be formulate as $\boldsymbol{H}_1^B \operatorname{vech}(\boldsymbol{V}) = \gamma \cdot \boldsymbol{1}_d$ and $\boldsymbol{H}_2^B \operatorname{vech}(\boldsymbol{V}) = \gamma$.

Since we want to focus on the calculation of the WTS and not on the vector T(X), nor on the estimation of the covariance matrix of T(X) we will not test the above hypotheses but use it as motivation. To get a pure result for the quadratic form, we generate one random vector $\boldsymbol{X}_{A)} \sim \mathcal{N}_{2d}(\boldsymbol{0}_{2d}, \boldsymbol{V}_{2d})$ resp. $\boldsymbol{X}_{B)} \sim \mathcal{N}_d(\boldsymbol{1}_d, \boldsymbol{V}_d)$. Here $\boldsymbol{V}_d$ denotes a d-dimensional compound symmetry matrix given through $\boldsymbol{V}_d = \boldsymbol{I}_d + \boldsymbol{J}_d$. With $\boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{X}$ for each setting, the computation time of the WTS using both respective hypothesis matrices is calculated for different dimensions $d_{A)} = (5,10,20,50,100,200)$ resp. $d_{B)} = (15,55,120,210,325,465)$. These dimensions $d_{B)}$ are based on the dimension of the observation vector $p = (5,10,15,20,25,30)$ and the clear connection between p and d through $d = p(p+1)/2$. Since the WTS is attractive for bootstrap and especially permutation approaches, we measure the time that is needed for 5.000 calculations of the quadratic form, which is a commonly used number of bootstrap steps resp. permutations. The computations were run by means of the R-computing environment version 3.6.1 (R Core Team, 2019) on an Intel Xeon E5430 quad-core CPU running at 2.66 GHz using 16 GB DDR2 memory on a Debian GNU Linux 7.8. The required time in seconds is displayed in Table 1.

The average computation time for both settings and hypothesis matrices can be seen in Table 1. For larger dimension $d \geq 20$ resp. $d \geq 120$ (which means $p \geq 15$), using the smaller hypothesis matrix saves over 99% of the calculation time, which makes the difference between minutes and multiple hours. For smaller dimensions, the savings in terms of computation time are pronounced too, yet somewhat less relevant with respect to practical applications, since the differences are only fractions of seconds.

The results for settings A and B are comparable because the relations between the number of rows of the quadratic matrix and of the smaller matrix are similar. Of course, the time-saving effect increases with the increasing number of removed rows.

However, this simulation demonstrates only one aspect of how a smaller hypothesis matrix can reduce the required calculation time. It is widely known that calculating the empirical covariance matrix is a substantial part of the required computations, especially for higher dimensions. Depending on the setting, it could simplify the calculation if one multiplies the underlying data with the hypothesis matrix first and afterwards estimates the empirical covariance matrix of these transformed observations. Similar approaches exist for bootstrap techniques, where for example, it is often more efficient to generate $\boldsymbol{Y} \sim \mathcal{N}_m(\boldsymbol{0}_m, \boldsymbol{H}^\top \boldsymbol{V} \boldsymbol{H})$, compared to $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{0}_d, \boldsymbol{V})$ and to multiply them with the hypothesis matrix afterwards. Therefore, choosing a smaller hypothesis matrix for the WTS has a huge potential, particularly for large dimensions and resampling approaches.

Other quadratic forms

Unfortunately, our result only holds for the WTS and the MATS, but not for the ATS, which is a widely used quadratic form-based test statistic (see, e.g. Brunner (2001)). The ATS is given through

$$ATS(\boldsymbol{H}, \boldsymbol{y}) = N \cdot (\boldsymbol{H}\boldsymbol{T}(\boldsymbol{X}) - \boldsymbol{y})^\top (\boldsymbol{H}\boldsymbol{T}(\boldsymbol{X}) - \boldsymbol{y})$$

resp.

$$ATS(\boldsymbol{H}, \boldsymbol{y})_s = N \cdot (\boldsymbol{H}\boldsymbol{T}(\boldsymbol{X}) - \boldsymbol{y})^\top (\boldsymbol{H}\boldsymbol{T}(\boldsymbol{X}) - \boldsymbol{y}) / \operatorname{tr}(\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^\top),$$

where the second version might be considered as a standardized ATS. Here the results strongly depend on the chosen hypothesis matrix, as can already be seen from simple examples. Nevertheless, no general statement regarding the choice of the hypothesis matrix exists for $\boldsymbol{y} \neq \boldsymbol{0}$. In addition to uniqueness, which would be the most important criterion for such a matrix, also other properties would be meaningful. In particular, this matrix should have a minimal number of rows and many zero entries to be computationally efficient. A matrix fulfilling this would be the reduced-row-echelon-form of a matrix, which is known from the Gauß-elimination algorithm. Thus, a hypothesis could be defined as the solution of a linear equation system, and afterwards, the reduced-row-echelon-form would be calculated and used for calculating the test statistic. Although there are different ways to define the linear equation system, for each solution set and the corresponding null hypothesis, this procedure would result in one and the same hypothesis matrix, eventually leading to a unique test statistic. Other meaningful properties of a hypothesis matrix, like symmetry or idempotence, would be desirable but usually are in conflict with computational advantages.

On the other hand, for $\boldsymbol{y} = \boldsymbol{0}$, the "default choice" $\boldsymbol{P}$ is unique but possibly contains multiple zero rows or rows, which are linear combinations of other rows. If we reconsider example (C) and the structure of the covariance matrix, simple examples for both cases exist. Considering a two-dimensional random vector, the unique projection matrices for testing diagonality resp. sphericity (i.e., diagonality with equal diagonal elements) would be

$$\boldsymbol{H} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{resp.} \quad \boldsymbol{H} = \frac{1}{2} \cdot \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

Both, the zero rows and the linear dependencies within the hypothesis matrices make them dispensable and time-consuming. Fortunately, the removal of zero rows does not affect the value of the ATS. Also, for linear dependencies, it is possible to remove multiple rows without affecting the value of the test statistic. Without loss of generality, let $\boldsymbol{H}_{\bullet \ell_1}, \dots, \boldsymbol{H}_{\bullet \ell_k}$ be pairwise linearly dependent rows with $\boldsymbol{H}_{\bullet \ell_2} = \beta_2 \cdot \boldsymbol{H}_{\bullet \ell_1}, \dots, \boldsymbol{H}_{\bullet \ell_k} = \beta_k \cdot \boldsymbol{H}_{\bullet \ell_1}$ and $\beta_2, \dots, \beta_k \in \mathbb{R} \setminus \{0\}$. Then we could remove $\boldsymbol{H}_{\bullet \ell_2}, \dots, \boldsymbol{H}_{\bullet \ell_k}$, if we replace $\boldsymbol{H}_{\bullet \ell_1}$ by $\sqrt{1 + \beta_2^2 + \dots + \beta_k^2}\, \boldsymbol{H}_{\bullet \ell_1}$, without influencing the value of the test statistic. So, their first row remains but is multiplied by the square root of its summed squared frequencies. It is clear that this also works if various rows have multiples, and it could be adapted for $\boldsymbol{y} \neq \boldsymbol{0}$. This result is summarized in the following corollary:

Corollary 3. *Let $\boldsymbol{H}\boldsymbol{\theta} = \boldsymbol{y}$ be an arbitrary hypothesis with hypothesis matrix $\boldsymbol{H} \in \mathbb{R}^{m \times d}$ and let $\mathcal{M}_0(\boldsymbol{H})$ denote the set containing the indices of zero rows of $\boldsymbol{H}$. Then there are $\ell \leq m$ pairwise linear independent rows of the matrix $\boldsymbol{H}$, which allows for dividing $\{1, \dots, m\} \setminus \mathcal{M}_0(\boldsymbol{H})$ into ℓ disjoint index sets $\mathcal{M}_1, \dots \mathcal{M}_\ell$. For these index sets, $\min(\mathcal{M}_i)$ denotes the smallest element of a set and $|\mathcal{M}_i|$ its cardinality. For $k = 1, \dots, \ell$ let $m_{k,1}, \dots, m_{k,|\mathcal{M}_k|}$ be the elements of $\mathcal{M}_k$. Hence, for $k = 1, \dots, \ell$, $j = 1, \dots, |\mathcal{M}_k|$ there is a $\beta_{k,j} \in \mathbb{R} \setminus \{0\}$ with $\boldsymbol{H}_{\bullet m_{k,j}} = \beta_{k,j} \cdot \boldsymbol{H}_{\bullet \min(\mathcal{M}_k)}$. Then we can express the same hypothesis through $\widetilde{\boldsymbol{H}}\boldsymbol{\theta} = \widetilde{\boldsymbol{y}}$ using*

$$\widetilde{\boldsymbol{H}} = \begin{pmatrix} w_1 \cdot \boldsymbol{H}_{\bullet \min(\mathcal{M}_1)} \\ w_2 \cdot \boldsymbol{H}_{\bullet \min(\mathcal{M}_2)} \\ \vdots \\ w_\ell \cdot \boldsymbol{H}_{\bullet \min(\mathcal{M}_\ell)} \end{pmatrix} \quad \text{and} \quad \widetilde{\boldsymbol{y}} = \begin{pmatrix} w_1 \cdot y_{\min(\mathcal{M}_1)} \\ w_2 \cdot y_{\min(\mathcal{M}_2)} \\ \vdots \\ w_\ell \cdot y_{\min(\mathcal{M}_\ell)} \end{pmatrix},$$

with $w_k = \sqrt{1 + \beta_{k,2}^2 + \dots + \beta_{k,|\mathcal{M}_k|}^2}$. Moreover, the equalities $ATS(\boldsymbol{H}, \boldsymbol{y}) = ATS(\widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{y}})$ and $ATS_s(\boldsymbol{H}, \boldsymbol{y}) = ATS_s(\widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{y}})$ hold.

Both adaptations make the remaining matrix more efficient but neither idempotent nor symmetric. However, if $\min(\mathcal{M}_1) < \min(\mathcal{M}_2) < \dots < \min(\mathcal{M}_\ell)$, starting with an unique projection matrix leads to an adapted version which is also unique.

Conclusion

The choice of the hypothesis matrix for the corresponding research question and the null hypothesis $\mathcal{H}_0 : \boldsymbol{H}\boldsymbol{\theta} = \boldsymbol{y}$ for $\boldsymbol{y} \neq \boldsymbol{0}$ are two topics that until now have received little attention in the literature. At the same time, the chosen matrix can potentially have a substantial impact on the test result and the required computation time. For the Wald-type statistic, our result allows for choosing an easily interpretable and computationally efficient hypothesis matrix without any risk of affecting the test result.

This is impossible for other quadratic form-based test statistics such as the ATS. Therefore we introduced an approach which leads to an unique and computationally advantageous hypothesis matrix. Especially for $\boldsymbol{y} \neq \boldsymbol{0}$, where no statements regarding the choice of an appropriate hypothesis matrix exist, this might be considered as an appropriate solution.

Even for $\boldsymbol{y} = \boldsymbol{0}$, several difficulties have been discussed and simple ways have been proposed to make the hypothesis matrix computationally more efficient. All in all, the choice of the hypothesis matrix is an important but neglected topic. So for the future, a general guideline would be desirable, which extends our approaches towards dealing with further quadratic forms.

Data availability

No data was used for the research described in the article.

Acknowledgements

The work of Paavo Sattler was funded by Deutsche Forschungsgemeinschaft Grant/Award Number DFG PA 2409/3-2. Georg Zimmermann gratefully acknowledges the support of the WISS 2025 project 'IDA-Lab Salzburg' (20204-WISS/225/197-2019 and 20102-F1901166-KZP).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2024.110038.

References

Brunner, E., 2001. Asymptotic and approximate analysis of repeated measures designs under heteroscedasticity. Math. Stat. Appl. Biom. 381.

Brunner, E., Konietschke, F., Pauly, M., M., Puri, 2016. Rank-based procedures in factorial designs: Hypotheses about nonparametric treatment effects. J. R. Stat. Soc. Ser. B Stat. Methodol. 79, 1463–1485. http://dx.doi.org/10.1111/rssb.12222.

Brunner, E., M.L., Puri, 2001. Nonparametric methods in factorial designs. Stat. Pap. 42 (1), 1–52.

Ditzhaus, M., Dobler, D., M., Pauly, 2021a. Inferring median survival differences in general factorial designs via permutation tests. Stat. Methods Med. Res. 30 (3), 875–891.

Ditzhaus, M., Fried, R., M., Pauly, 2021b. QANOVA: quantile-based permutation methods for general factorial designs. TEST 30, 960–979.

Friedrich, S., Brunner, E., M., Pauly, 2017. Permuting longitudinal data in spite of the dependencies. J. Multivariate Anal. 153, 255–265.

Friedrich, S., Konietschke, F., M., Pauly, 2016. A wild bootstrap approach for nonparametric repeated measurements. Comput. Statist. Data Anal. 113, 38–52. http://dx.doi.org/10.1016/j.csda.2016.06.016.

Friedrich, S., M., Pauly, 2017. MATS: Inference for potentially singular and heteroscedastic MANOVA. J. Multivariate Anal. 165, 166–179. http://dx.doi.org/10.1016/j.jmva.2017.12.008.

Munzel, U., E., Brunner, 2000. Nonparametric methods in multivariate factorial designs. J. Statist. Plann. Inference 88, 117–132. http://dx.doi.org/10.1016/S0378-3758(99)00212-8.

Pauly, M., Brunner, E., F., Konietschke, 2015. Asymptotic permutation tests in general factorial designs. J. R. Stat. Soc. Ser. B Stat. Methodol. 77 (2), 461–473.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

Rao, C., S., Mitra, 1971. Generalized Inverse of Matrices and Its Applications. Wiley, New York.

Rubarth, K., Sattler, P., Zimmermann, H.G., F., Konietschke, 2022. Estimation and testing of wilcoxon-mann-whitney effects in factorial clustered data designs. Symmetry 14 (2), http://dx.doi.org/10.3390/sym14020244, URL https://www.mdpi.com/2073-8994/14/2/244.

Sattler, P., Bathke, A.C., M., Pauly, 2022. Testing hypotheses about covariance matrices in general MANOVA designs. J. Statist. Plann. Inference 219, 134–146. http://dx.doi.org/10.1016/j.jspi.2021.12.001.