Statistics and Probability Letters 240 (2027) 110924

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

A MANOVA statistic for increasing number of samples☆

Rauf Ahmad [a,*], Ejaz Ahmed [b]

[a] Department of Statistics, Uppsala University, Sweden
[b] Department of Mathematics and Statistics, Brock University, Canada

ARTICLE INFO

Keywords:
MANOVA model
Wilks' statistic
Testing hypothesis
Large factor levels

ABSTRACT

A statistic to test MANOVA hypothesis is proposed for increasing number of independent samples. The statistic, defined as linear combination of U-statistics, is robust to normality and homoscedasticity assumptions, and can be used for balanced or unbalanced models. Whereas the focus is on the number of samples to be large, the limit also incorporates large sample sizes, where the dimension is kept fixed, although may exceed the number of samples as well as the number of replications per sample. Accuracy of the tests statistic is shown through simulations.

1. Introduction

Consider the one-way MANOVA model (Mardia et al., 2024)

$$\mathbf{y}_{ik} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ik}, \tag{1}$$

where $\mathbf{y}_{ik} \in \mathbb{R}^p$ and $\boldsymbol{\epsilon}_{ik} \in \mathbb{R}^p$ are, respectively, the response and error vectors on kth unit in ith sample, $\boldsymbol{\mu} \in \mathbb{R}^p$ is the overall average, $\boldsymbol{\alpha}_i \in \mathbb{R}^p$ denotes the effect of ith population, $k \in \{1, \ldots, r_i\}$, $i \in \{1, \ldots, g\}$, with r_i as number of independent replications under each of g independent samples; further notations follow in Section 2. For Model (1), we are mostly interested to test $\mathrm{H}_0 : \boldsymbol{\alpha}_i = \mathbf{0}\ \forall\ i$.

Most modern modifications of multivariate theory, including multivariate general linear model which MANOVA is part of, concerns high-dimensional extensions, i.e., for $p \gg r_i$; for a list of references, see e.g., Ahmad (2019, 2026). At times, however, multivariate data are observed on a large number of independent samples, g. Gene expressions on variants of cancer are recorded on groups of individuals; in microbiome studies, a large variety of species are often observed; anthropologists study a large number of populations to study forms of evolution, etc.

Often, in such studies, the dimension of the observed vector is moderate but the number of samples, g, grows. In fact, the same problem may occur in univariate cases, to be addressed with ANOVA model. In these models, the number of parameters increases with the number of samples.

The primary aim of this article is to discuss tests for Model (1) for large g. To keep this aim effectively focused, we keep p fixed, but we allow it to be large, even exceed r_i and also g. Further, although the test statistic is mainly defined for (1), it is formulated in such a general form that it can be conveniently used for high-order MANOVA models, or likewise, for univariate case, for ANOVA models. The test statistic is constructed as a linear combination of U-statistics, and its properties are discussed for $g \to \infty$ and $r_i \to \infty$, where $p \lesssim r_i$ and $p \lesssim g$, while keeping p fixed.

Multivariate analysis of variance (MANOVA) is one of the most fundamental techniques in multivariate analysis. It extends univariate ANOVA model for multiple response variables, and is one of the two partitions of multivariate general linear model, the other being multivariate regression. Under certain conditions, mainly independence, normality, and homoscedasticity across g

☆ This article is part of a Special issue entitled: 'STAPRO_High-Dimensional Data Analytics' published in Statistics and Probability Letters.

* Corresponding author.
E-mail address: rauf.ahmad@statistik.uu.se (R. Ahmad).

https://doi.org/10.1016/j.spl.2026.110924

Available online 4 August 2026
0167-7152/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

populations, the classical multivariate inference offers many statistics to test H_0, with Wilks' likelihood ratio statistic, denoted Λ, being the most popular (Anderson, 2003; Mardia et al., 2024).

Although most modern modifications of classical inference have addressed questions related to high-dimensional data ($p \gg r_i$), there have been a few attempts to consider the testing problem for increasing g, including for ANOVA model. A detailed study of ANOVA F statistic in this direction, for one- and multi-way models, is given in Bathke (2002), with a more recent investigations in Park and Park (2012) and Jiménez-Gamero and Sillero-Denamiel (2025). For some other, particularly (or including) nonparametric modifications, we refer to Boos and Brownie (1995), Wang and Akritas (2004, 2006) and the references cited in Bathke (2002); see also Boos and Stefanski (2013, ch.12). One- and multi-way MANOVA models, including nonparametric variants, are considered in Bathke and Harrar (2008), Harrar and Bathke (2008) and Gupta et al. (2008).

2. Test statistics and their properties

Consider Model (1). Denote $\mathbf{Y}_i = (\mathbf{y}_{i1}^T, \ldots, \mathbf{y}_{ir_i}^T)^T \in \mathbb{R}^{r_i \times p}$, $\boldsymbol{\Xi}_i = (\boldsymbol{\epsilon}_{i1}^T, \ldots, \boldsymbol{\epsilon}_{ir_i}^T)^T \in \mathbb{R}^{r_i \times p}$ as the response and error matrices for ith independent population, $i \in \{1, \ldots, g\}$. We denote $\sum_{i=1}^g r_i = n$. This formulation also helps write Model (1) as multivariate general linear model (MGLM),

$$\mathbf{Y} = \mathbf{XB} + \boldsymbol{\Xi} \tag{2}$$

where $\mathbf{Y} = (\mathbf{Y}_1^T, \ldots, \mathbf{Y}_g^T)^T \in \mathbb{R}^{n \times p}$, $\boldsymbol{\Xi} = (\boldsymbol{\Xi}_1^T, \ldots, \boldsymbol{\Xi}_g^T)^T \in \mathbb{R}^{n \times p}$, with $\mathbf{B} \in \mathbb{R}^{q \times p}$ as matrix of unknown parameters, and $\mathbf{X} \in \mathbb{R}^{n \times q}$ as design matrix. Assuming same $\mathbf{X}$ across all p responses in each i, it extends both regression and ANOVA components of the univariate general linear model.

For $r(\mathbf{X}) = p$, $\widehat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{XY} \in \mathbb{R}^{q \times p}$ is the unique least-squares estimator of $\mathbf{B}$, whose optimality, as best linear unbiased estimator, holds by the Gauss–Markov theorem; in fact, under normality, $\widehat{\mathbf{B}}$ is also the likelihood estimator, and optimal in the class of all unbiased estimators.

MANOVA models, such as (1), belong to (2) when $r(\mathbf{X}) < q$ so that a unique estimator of $\mathbf{B}$ is not possible. Under certain constraints, such as $\sum_{i=1}^g r_i\boldsymbol{\alpha}_i = \mathbf{0}$, estimability and testability of certain linear combinations of $\mathbf{B}$ can be ensured. To see it precisely for Model (1), note that, in terms of Model (2), we can write $\mathbf{X} = (\mathbf{1}_n, \widetilde{\mathbf{X}}) \in \mathbb{R}^{n \times p}$, with $\widetilde{\mathbf{X}} = \oplus_{i=1}^g \mathbf{1}_{r_i} \in \mathbb{R}^{n \times (p-1)}$, $p = g + 1$, $\mathbf{1}$ is a vector of 1s of appropriate dimension, and $\oplus$ is the direct sum. As the columns of $\widetilde{\mathbf{X}}$ sum to $\mathbf{1}_n$, the column space of $\mathbf{X}$, $C(\mathbf{X})$, is spanned by a subspace of dimension $p - 1 = g$. The constraint $\sum_{i=1}^g r_i\boldsymbol{\alpha}_i = \mathbf{0}$ ensures the estimability of $\widetilde{\mathbf{B}} = (\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T, \ldots, \boldsymbol{\mu}_g^T)^T$ for the reparametrized model

$$\mathbf{y}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{ij}, \tag{3}$$

with $\boldsymbol{\mu}_i = \boldsymbol{\mu} + \boldsymbol{\alpha}_i \ \forall\ i$. It implies an equivalent formulation of the hypothesis of main interest, as

$$\mathrm{H}_0 : \boldsymbol{\alpha}_i = \mathbf{0} \ \Leftrightarrow \ \mathrm{H}_0 : \boldsymbol{\mu}_i = \boldsymbol{\mu}\ \forall\ i, \quad \text{vs} \quad \mathrm{H}_1 : \text{Not } \mathrm{H}_0\,. \tag{4}$$

Assuming $\boldsymbol{\epsilon}_{ij} \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})\ \forall\ i, j$, Wilks' Λ is often used with a Chi-square approximation for moderately large r_i. All classical statistics to test H_0, including Λ, are based on the hypothesis and error matrices, $\mathbf{H} = \sum_{i=1}^g r_i(\overline{\mathbf{Y}}_i - \overline{\mathbf{Y}})(\overline{\mathbf{Y}}_i - \overline{\mathbf{Y}})^T$ and $\mathbf{E} = \sum_{i=1}^g \sum_{j=1}^{r_i}(\mathbf{Y}_{ij} - \overline{\mathbf{Y}}_i)(\mathbf{Y}_{ij} - \overline{\mathbf{Y}}_i)^T$, often expressed as functions of eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$, where $\overline{\mathbf{Y}}_i = \sum_{j=1}^{r_i} \mathbf{Y}_{ij}/r_i$ and $\overline{\mathbf{Y}} = \sum_{i=1}^g \sum_{j=1}^{r_i} \mathbf{Y}_{ij}/n$ are unbiased estimators of $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}$, respectively (Johnson and Wichern, 2007). We can also write $\mathbf{E} = \mathbf{Y}^T(\mathbf{I}_n - \mathbf{P})\mathbf{Y}$, with $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X} : \mathbb{R}^{n \times q} \to \mathbb{R}^{n \times n}$, as the projection matrix. For a brief evaluation of Wilks' Λ for increasing number of samples, see Ahmed and Ahmad (2023). The validity of classical inference hinges on the set of assumptions, $\boldsymbol{\epsilon}_{ij} \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})\ \forall\ i, j$, including normality and homoscedasticity, with $\boldsymbol{\Sigma}$ estimated as $\widehat{\boldsymbol{\Sigma}} = \mathbf{E}/(n - g)$.

The proposed statistic holds under a general multivariate model, including normality, and dispenses with homoscedasticity assumption. In fact, a variant of the statistic was originally introduced for high-dimensional ($p \gg r_i$) case, again relaxing standard assumptions (Ahmad, 2019). We reformulate it to test H_0 for increasing g but now for fixed dimension. Let $\mathbf{y}_{ik}$ in (1) be defined on a probability space $(\mathcal{X}, \mathcal{A}, \mathcal{P}_\theta)$, with $\mathcal{P}_\theta$ indexed with parameter $\theta \in \Theta$, where Θ is the parameter space. For $\mathbf{Y}_i$ etc. defined around Eq. (2), unbiased estimators of $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ follow as

$$\overline{\mathbf{Y}}_i = \mathbf{Y}_i^T\mathbf{1}_{r_i}/r_i, \qquad \widehat{\boldsymbol{\Sigma}}_i = \mathbf{Y}_i^T\mathbf{C}_{r_i}\mathbf{Y}_i/(r_i - 1), \tag{5}$$

with $\mathbf{C}_{r_i} = \mathbf{I}_{r_i} - \mathbf{J}_{r_i}/r_i$ as the centering matrix, where $\mathbf{I}_{r_i}$ is the identity matrix and $\mathbf{J}_{r_i} = \mathbf{1}_{r_i}\mathbf{1}_{r_i}^T$, $\mathbf{1}$ being a vector of all 1s. Denote $\boldsymbol{\Delta} = (\boldsymbol{\delta}_{12}^T, \ldots, \boldsymbol{\delta}_{g-1,g}^T)^T \in \mathbb{R}^G$, with $\boldsymbol{\delta}_{ii'} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_{i'}$, $i, i' \in \{1, \ldots, g\}$, $i < i'$, and $G = g(g-1)/2$. Then, for H_0 in (4), $\mathrm{H}_0 \Leftrightarrow \{\mathrm{H}_{01,ii'} : \boldsymbol{\delta}_{ii'} = \mathbf{0}\ \forall\ i < i'\ i, i' \in \{1, \ldots, g\}\}$; essentially, all equivalent formulations imply $\boldsymbol{\Delta} = \mathbf{0}$ or $\widetilde{\mathbf{B}} = \mathbf{1}_g \otimes \boldsymbol{\mu}^T$, with $\boldsymbol{\mu}$ as in (4) and $\otimes$ is the Kronecker product. For convenience, we shall use the following notation in the sequel

$$\sum_{\pi_<(i,i')}^{g}\sum^{g} := \sum_{\substack{i=1 \\ i<i'}}^{g}\sum_{i'=1}^{g}, \quad \sum_{\pi_<(k,r)}^{r_i}\sum^{r_i} := \sum_{\substack{k=1 \\ k<r}}^{r_i}\sum_{r=1}^{r_i},$$

which may extend for multiple indices, and change to $\pi_{\neq}$ for unequal indices. Now, let $\mathrm{T}_{ii'}$ denote the statistic corresponding to $\mathrm{H}_{01,ii'} : \boldsymbol{\delta}_{ii'} = \mathbf{0}$, with $\boldsymbol{\delta}_{ii'} \in \boldsymbol{\Delta}$. We can then test H_0 using

$$\mathrm{T}_{1\mathrm{M}} = \frac{1}{\binom{g}{2}} \sum_{\pi_<(i,i')}^{g}\sum^{g} \mathrm{T}_{ii'}\,. \tag{6}$$

Although Eq. (6) is specifically defined to test H_0 for (1), hence the subscript 1M, the formulation is kept general to adapt the statistic to other cases, by appropriately defining $T_{ii'}$. For ex., for a two-way MANOVA model, g may refer to the cardinality of the set of independent samples under any factor, or of their interaction. In our context, $T_{ii'}$ is an unbiased estimator of $\|\boldsymbol{\delta}_{ii'}\|^2 = \|\boldsymbol{\mu}_i\|^2 + \|\boldsymbol{\mu}_{i'}\|^2 - 2\langle \boldsymbol{\mu}_i, \boldsymbol{\mu}_{i'}\rangle$, where $\langle\cdot,\cdot\rangle : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ denotes the Euclidean inner product, so that $\|\cdot\|^2 : \mathbb{R}^p \to \mathbb{R}$ is the vector norm. Consider

Assumption 1. The error vectors, $\boldsymbol{\epsilon}_{ik}$, are independent with $\mathrm{E}(\boldsymbol{\epsilon}_{ik}) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{\epsilon}_{ik}) = \boldsymbol{\Sigma}_i \in \mathbb{R}^{p\times p}$, $\boldsymbol{\Sigma}_i > 0$, $\forall\ k \in \{1,\dots,r_i\}$, $i \in \{1,\dots,g\}$.

By Assumption 1, $\mathrm{E}(\mathbf{y}_{ik}) = \boldsymbol{\mu}_i = \boldsymbol{\mu} + \boldsymbol{\alpha}_i$, $\mathrm{Cov}(\mathbf{y}_{ik}) = \boldsymbol{\Sigma}_i$, $\forall\ k \in \{1,\dots,r_i\}$. Further, by independence, $\mathrm{E}\langle \bar{\mathbf{y}}_i, \bar{\mathbf{y}}_{i'}\rangle = \langle \boldsymbol{\mu}_i, \boldsymbol{\mu}_{i'}\rangle$. Since $\mathrm{E}\,\|\bar{\mathbf{y}}_i\|^2 = \|\boldsymbol{\mu}_i\|^2 + \|\boldsymbol{\Lambda}_i\|^2/r_i$, it gives $\mathrm{E}(\|\bar{\mathbf{y}}_i\|^2 - \|\widehat{\boldsymbol{\Lambda}}_i\|^2/r_i) = \|\boldsymbol{\mu}_i\|^2$, where $\boldsymbol{\Lambda}_i = \boldsymbol{\Sigma}_i^{1/2}$, $\widehat{\boldsymbol{\Lambda}}_i = \widehat{\boldsymbol{\Sigma}}_i^{1/2}$, $\{1,\dots,g\}$, with $\widehat{\boldsymbol{\Sigma}}_i$ as given in (5). A simple expansion shows that $\|\bar{\mathbf{y}}_i\|^2 - \|\widehat{\boldsymbol{\Lambda}}_i\|^2/r_i$ and $\langle \bar{\mathbf{y}}_i, \bar{\mathbf{y}}_{i'}\rangle$, being unbiased estimators of $\|\boldsymbol{\mu}_i\|^2$ and $\langle \boldsymbol{\mu}_i, \boldsymbol{\mu}_{i'}\rangle$, can, respectively, be expressed as one- and two-sample U-statistics, denoted $\widehat{\|\boldsymbol{\mu}_i\|^2} = U_i$ and $\widehat{\langle \boldsymbol{\mu}_i, \boldsymbol{\mu}_{i'}\rangle} = U_{ii'}$, as

$$U_i = \frac{1}{r_i(r_i-1)} \sum_{\pi_{\neq}(k,r)}^{r_i} \sum^{r_i} h(\mathbf{X}_{ik}, \mathbf{X}_{ir}), \quad U_{ii'} = \frac{1}{r_i r_{i'}} \sum_{k=1}^{r_i} \sum_{l=1}^{r_{i'}} h(\mathbf{X}_{ik}, \mathbf{X}_{i'l}), \tag{7}$$

with symmetric kernels, $h(\mathbf{X}_{ik}, \mathbf{X}_{ir}) = \langle \mathbf{X}_{ik}, \mathbf{X}_{ir}\rangle$, $h(\mathbf{X}_{ik}, \mathbf{X}_{i'l}) = \langle \mathbf{X}_{ik}, \mathbf{X}_{i'l}\rangle$, $k \neq r$, $i, i' \in \{1,\dots,g\}$, $i \neq i'$. This gives $T_{ii'} = U_i + U_{i'} - 2U_{ii'}$ as an estimator of $\|\boldsymbol{\delta}_{ii'}\|^2$. Thus $\mathrm{E}(T_{ii'}) = \|\boldsymbol{\delta}_{ii'}\|^2$ and, using the moments of U-statistics (see e.g., Koroljuk and Borovskich, 1994; Ahmad, 2017, 2019),

$$\mathrm{Var}(T_{ii'}) = 2\|\boldsymbol{\Sigma}_{ii'}\|^2 + 4\langle \boldsymbol{\delta}_{ii'}, \boldsymbol{\Sigma}_{ii'}\boldsymbol{\delta}_{ii'}\rangle \tag{8}$$

$$\mathrm{Cov}(T_{ii'}, T_{ii''}) = \frac{2}{r_i^2}\|\boldsymbol{\Sigma}_i\|^2 + \frac{4}{r_i}\langle \boldsymbol{\delta}_{ii'}, \boldsymbol{\Sigma}_i\boldsymbol{\delta}_{ii''}\rangle \tag{9}$$

$$\mathrm{Cov}(T_{ii'}, T_{i'i''}) = \frac{2}{r_{i'}^2}\|\boldsymbol{\Sigma}_{i'}\|^2 + \frac{4}{r_{i'}}\langle \boldsymbol{\delta}_{ii'}, \boldsymbol{\Sigma}_{i'}\boldsymbol{\delta}_{i'i''}\rangle, \tag{10}$$

where $\boldsymbol{\Sigma}_{ii'} = \boldsymbol{\Sigma}_i/r_i + \boldsymbol{\Sigma}_{i'}/r_{i'}$, and $\mathrm{Cov}(T_{ii'}, T_{i''i'''}) = 0$ for $i \neq i''$ and $i' \neq i'''$. Further, $\|\mathbf{A}\| : \mathbb{R}^{p\times q} \to \mathbb{R}$ denotes the Frobenius norm for any positive semi-definite $\mathbf{A} \in \mathbb{R}^{p\times q}$, i.e., $\|\mathbf{A}\|^2 = \mathrm{tr}(\mathbf{A}^T\mathbf{A})$, or $\mathrm{tr}(\mathbf{A}^2)$ for symmetric $\mathbf{A} \in \mathbb{R}^p$. Under H_0, $\boldsymbol{\delta}_{ii'} = \mathbf{0}$ $\forall\ i, i'$, so that Eqs. (8)–(10) simplify to

$$\mathrm{Var}(T_{ii'}) = 2\|\boldsymbol{\Sigma}_{ii'}\|^2,\ \mathrm{Cov}(T_{ii'}, T_{ii''}) = \frac{2}{r_i^2}\|\boldsymbol{\Sigma}_i\|^2,\ \mathrm{Cov}(T_{ii'}, T_{i'i''}) = \frac{2}{r_{i'}^2}\|\boldsymbol{\Sigma}_{i'}\|^2. \tag{11}$$

Now, denoting $\binom{g}{2} T_{1M} = T_G$ in Eq. (6), we have

$$\mathrm{Var}(T_G) = \sum_{\pi_{\neq}(i,i')}^{g} \sum^{g} \mathrm{Var}(T_{ii'}) + \sum_{\pi_{\neq}\{(i,i'),(i'',i''')\}}^{g} \sum^{g} \sum^{g} \sum^{g} \mathrm{Cov}(T_{ii'}, T_{i''i'''}), \tag{12}$$

with $\mathrm{E}(T_G) = \delta = \sum_{i<i'} \|\boldsymbol{\delta}_{ii'}\|^2$ and, by substituting (8)–(10),

$$\begin{aligned}\mathrm{Var}(T_G) = {} & 2(g-1)^2 \sum_{i=1}^{g} \frac{\|\boldsymbol{\Sigma}_i\|^2}{r_i^2} + 4\sum_{\pi_<(i,i')}^{g} \sum^{g} \frac{\langle \boldsymbol{\Sigma}_i, \boldsymbol{\Sigma}_{i'}\rangle}{r_i r_{i'}} + 4\sum_{\pi_<(i,i')}^{g} \sum^{g} \langle \boldsymbol{\delta}_{ii}, \boldsymbol{\Sigma}_{ii'}\boldsymbol{\delta}_{ii'}\rangle \\ & + 4\sum_{\pi_<(i,i',i'')}^{g} \sum^{g} \sum^{g} \frac{1}{r_i}\langle \boldsymbol{\delta}_{ii'}, \boldsymbol{\Sigma}_i\boldsymbol{\delta}_{ii''}\rangle + 4\sum_{\pi_<(i,i',i'')}^{g} \sum^{g} \sum^{g} \frac{1}{r_{i'}}\langle \boldsymbol{\delta}_{ii'}, \boldsymbol{\Sigma}_{i'}\boldsymbol{\delta}_{ii''}\rangle.\end{aligned} \tag{13}$$

In particular, under H_0, $\mathrm{E}(T_G) = 0$, and

$$\mathrm{Var}(T_G) = 2(g-1)^2 \sum_{i=1}^{g} \frac{\|\boldsymbol{\Sigma}_i\|^2}{r_i^2} + 4\sum_{\pi_<(i,i')} \sum \frac{\langle \boldsymbol{\Sigma}_i, \boldsymbol{\Sigma}_{i'}\rangle}{r_i r_{i'}}. \tag{14}$$

We study the properties of T_{1M} under a general model, extending normality. With $\boldsymbol{\Lambda}_i = \boldsymbol{\Sigma}_i^{1/2}$, and using $\boldsymbol{\epsilon}_{ik} = \mathbf{y}_{ik} - \boldsymbol{\mu}_i$ from (3), $k \in \{1,\dots,r_i\}$, $i \in \{1,\dots,g\}$, $g \geq 2$, we define the model as

$$\mathbf{z}_{ik} = \boldsymbol{\Lambda}_i^{-1}\boldsymbol{\epsilon}_{ik}, \tag{15}$$

where $\mathbf{z}_{ik} \in \mathbb{R}^p$ has a non-degenerate distribution with $\mathrm{E}(\mathbf{z}_{ik}) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{z}_{ik}) = \mathbf{I}_p$. We further set the following assumptions, where γ, ρ_i are finite constants, and $n = \sum_{i=1}^{g} r_i$.

Assumption 2. $\mathrm{E}(\epsilon_{iks}^4) \leq \gamma < \infty$, $i \in \{1,\dots,g\}$, $\forall\ s \in \{1,\dots,p\}$, $\gamma \in \mathbb{R}^+$.

Assumption 3. As $r_i \to \infty$, $r_i/n \to \rho_i \in (0,\infty)$, $i \in \{1,\dots,g\}$.

Assumption 4. As $r_i, g \to \infty$, $r_i/g \to \vartheta_i \in (0,\infty)$, $i \in \{1,\dots,g\}$.

Assumption 5. $\|\boldsymbol{\Lambda}_i\|^2/p = O(1)$ and $\langle \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{i''}\boldsymbol{\mu}_{i'}\rangle \le O(p^2)$, $\forall\ i, i'i'' \in \{1, \ldots, g\}$.

We do not take $p \to \infty$, but Assumption 5 holds for a wide variety of parameters, whether $p < r_i$ or $p > r_i$. In fact, the exact bounds are mostly achieved only for spiked structures such as compound symmetry (see Section 3), otherwise they can be sharpened. For ex., for $\boldsymbol{\Sigma}_i$ as identity or sphericity, $\langle \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{i''}\boldsymbol{\mu}_{i'}\rangle$ is $O(p)$. The bounds thus help us simplify the steps for the limit of $\mathrm{T}_{1\mathrm{M}}$.

Now, consider $\mathrm{Var}(\mathrm{T}_{1\mathrm{M}}) = \binom{g}{2}^{-2}\mathrm{Var}(\mathrm{T}_G) = O(g^{-4})\,\mathrm{Var}(\mathrm{T}_G)$ with $\mathrm{Var}(\mathrm{T}_G)$ as in (13). Under the assumptions, all terms involving inner products in $\mathrm{Var}(\mathrm{T}_G)$ vanish, and the same terms vanish exactly under the null, simplifying $\mathrm{Var}(\mathrm{T}_G)$ to (14). Our main result, Theorem 1 below, gives the limit of $\widetilde{\mathrm{T}}_{1\mathrm{M}} = \binom{g}{2}^{-1}\widetilde{\mathrm{T}}_G$, with $\widetilde{\mathrm{T}}_G = \sum_{i<i'}(\widetilde{\mathrm{T}}_{ii'} - \mathrm{E}(\widetilde{\mathrm{T}}_{ii'}))/\sqrt{\mathrm{Var}(\widetilde{\mathrm{T}}_{ii'})}$, $\widetilde{\mathrm{T}}_{ii'} = n\,\mathrm{T}_{ii'}$, $n = \sum_{i=1}^{g} r_i$. Note that, $\widetilde{\mathrm{T}}_{ii'} - \mathrm{E}(\widetilde{\mathrm{T}}_{ii'}) = n(\mathrm{T}_{ii'} - \delta_{ii'}) = n\{U_i - \mathrm{E}(U_i)\} + n\{U_{i'} - \mathrm{E}(U_{i'})\} - 2n\{U_{ii'} - \mathrm{E}(U_{ii'})\}$.

Most properties of U-statistics follow from the projections of their kernels. These projections help compute the variances given in (8)–(10) and play a crucial role in determining the limit of a U-statistic. For details, see Koroljuk and Borovskich (1994) and Witting and Müller-Funk (1995).

The first-order projections of the kernel of U_i is $h_1(\mathbf{X}_{ik}) = \mathrm{E}[h(\cdot)|\mathbf{X}_{ik}] = \langle \mathbf{X}_{ik}, \boldsymbol{\mu}_i\rangle$; for $U_{ii'}$, the first-order projections with respect to each sample, are $\langle \mathbf{X}_{ik}, \boldsymbol{\mu}_{i'}\rangle$ and $\langle \boldsymbol{\mu}_i, \mathbf{X}_{i'r}\rangle$, respectively. Both being second-order U-statistics, the second-order projections are the kernels themselves. Denoting $\widetilde{h}(\cdot)$ and $\widetilde{h}_1(\cdot)$ as the kernel and first-order projection for $U_i - \mathrm{E}(U_i)$, and defining similar identities for $U_{ii'}$, it is obvious that U_i, $U_{i'}$ and $U_{ii'}$ are first-order degenerate as their first-order projections vanish under the null.

In Ahmad (2019), a proof of such degenerate U-statistics is given by first showing a convergence, for increasing sample size, to weighted Chi-square limits, and then, using a special corollary of Lindeberg–Feller central limit theorem, to a normal limit for increasing dimension. In our case, however, an alternative route, leading directly to a normal limit, can make the proof much easier. It uses martingale central limit theorem, the essential part of which is again a form of Lindeberg–Feller condition (see e.g., Koroljuk and Borovskich, 1994, ch. 4). A proof of this normal limit for a statistic which is precisely $\widetilde{\mathrm{T}}_{ii'}$ for any pair $i < i'$ is given in Ahmad (2025). The proof is given for $r_i, p \to \infty$ but it obviously holds for any given p as well. We proceed as following.

Theorem 1. *Let $\widetilde{\mathrm{T}}_{1M}$ be as defined above. Under Model (15) and Assumptions 1–5, $\widetilde{\mathrm{T}}_{1M} \xrightarrow{D} \mathrm{N}(0,1)$, as $r_i, g \to \infty$.*

Proof. Given $\widetilde{\mathrm{T}}_{ii'}$, let $\widetilde{\mathbf{T}} \in \mathbb{R}^G$ denote a vector with elements $(\widetilde{\mathrm{T}}_{ii'} - \mathrm{E}(\widetilde{\mathrm{T}}_{ii'}))/\sqrt{\mathrm{Var}(\widetilde{\mathrm{T}}_{ii'})}$, $i, i' = \{1, \ldots, g\}$, $i < i'$. Then $\widetilde{\mathrm{T}}_G = \mathbf{1}^T\widetilde{\mathbf{T}}_G$, with $\mathbf{1}$ a vector of G 1s. Now, take $i = 1$, $i' = 2$.

It is proved in Ahmad (2025) that $\widetilde{\mathrm{T}}_{12} \to \mathrm{N}(0,1)$ as $r_2, r_2 \to \infty$. Since the limit holds for any pair $i < i'$, and since $\mathrm{Cov}(\mathrm{T}_{ii'}, \mathrm{T}_{ii''})$ eventually vanish under the assumptions, it leaves $\widetilde{\mathrm{T}}_G$ as a vector of asymptotically independent $N(0,1)$ variables, with a diagonal covariance matrix, $\mathrm{Cov}(\widetilde{\mathbf{T}}_G) = \mathbf{D}$, with diagonal elements $2n^2\|\boldsymbol{\Sigma}_{ii'}\|^2$ which converges, under the assumptions, to $O(p^2)$.

It further implies that $\widetilde{\mathrm{T}}_G$ is a sum, hence $\widetilde{\mathrm{T}}_{1M}$ is an average, of $G = \binom{g}{2}$ asymptotically independent $N(0,1)$ variables. The final limit, for $g \to \infty$, now follows by Cramér–Wold device, which holds for $\mathbf{a}^T\widetilde{\mathbf{T}}_G$, for any non-zero $\mathbf{a} \in \mathbb{R}^G$, where in our case, $\mathbf{a} = \mathbf{1}_G$. □

We need to estimate the variance, essentially $\mathrm{Var}(\widetilde{\mathrm{T}}_{ii'})$ or, more simply, $\|\boldsymbol{\Sigma}_{ii'}\|^2$. We do this by plugging in efficient estimators of the components of $\|\boldsymbol{\Sigma}_{ii'}\|^2$, namely $\|\boldsymbol{\Sigma}_i\|^2$ and $\langle \boldsymbol{\Sigma}_i, \boldsymbol{\Sigma}_{i'}\rangle$, $i, i' \in \{1, \ldots, g\}$. The two estimators can be expressed as one- and two-sample U-statistics (Ahmad, 2017, 2019). We, however, prefer simpler, computationally efficient versions of these estimators. For the later, it follows directly, by the independence of samples, as $\mathrm{E}_{ii'} = \langle \widehat{\boldsymbol{\Sigma}}_i, \widehat{\boldsymbol{\Sigma}}_{i'}\rangle$.

For the former, such an estimator is given as $\mathrm{E}_i = \eta_i\{(r_i-1)(r_i-2)\|\widehat{\boldsymbol{\Sigma}}_i\|^2 + [\|\widehat{\boldsymbol{\Lambda}}_i\|^2]^2 - r_i Q_i\}$ with $\widehat{\boldsymbol{\Lambda}}_i = \widehat{\boldsymbol{\Sigma}}_i^{1/2}$, $Q_i = \sum_{k=1}^{r_i}\|\widetilde{\mathbf{X}}_{ik}\|^2/(r_i-1)$, $\widetilde{\mathbf{X}}_i = \mathbf{X}_{ik} - \overline{\mathbf{X}}_i$, $\eta_i = (r_i-1)/[r_i(r_i-2)(r_i-3)]$; see Himeno and Yamada (2014). It can be shown that $\mathrm{Var}(\mathrm{E}_i\,/\,\mathrm{E}(\mathrm{E}_i))$ and $\mathrm{Var}(\mathrm{E}_{ii'}\,/\,\mathrm{E}(\mathrm{E}_{ii'}))$ and bounded as $O(r_i^{-1})$ and $O(r_i^{-1} + r_{i'}^{-1})$, respectively. Plugging in these estimators gives a consistent estimator of $\|\boldsymbol{\Sigma}_{ii'}\|^2$, eventually of $\mathrm{Var}(\widetilde{\mathrm{T}}_{1M})$, say $\widehat{\mathrm{Var}}(\widetilde{\mathrm{T}}_{1M})$. We have

Corollary 1. *The limit in Theorem 1 remains valid by replacing $\mathrm{Var}(\widetilde{\mathrm{T}}_{1M})$ with $\widehat{\mathrm{Var}}(\widetilde{\mathrm{T}}_{1M})$ as defined above.*

To evaluate the power of test, it suffices to focus on $\widetilde{\mathrm{T}}_{ii'}$ since $\widetilde{\mathrm{T}}_{1M}$ averages them over all distinct pairs of independent samples. Let Z_α denote the quantile of $N(0,1)$-distribution, and denote $Z = (\widetilde{\mathrm{T}}_{ii'} - \mu)/\sigma$, with μ and σ^2 as mean and variance of $\widetilde{\mathrm{T}}_{ii'}$, to be denoted as μ_0, σ_0^2 under the null, and μ_1, σ_1^2 under the alternative. Then $\beta(\mu, \sigma) = \mathrm{P}(Z \ge Z_\alpha\,|\,\mathrm{H})$ is the power function, which gives the size, α, for H as H_0 and power, $1-\beta$, for H as H_1.

We have $\mu_0 = 0$, $\mu_1 = n\|\boldsymbol{\delta}_{ii'}\|^2$, $\sigma_0^2 = 2n^2\|\boldsymbol{\Sigma}_{ii'}\|^2$, and $\sigma_1^2 = 2n^2\|\boldsymbol{\Sigma}_{ii'}\|^2 + 4n^2\langle \boldsymbol{\delta}_{ii'}, \boldsymbol{\Sigma}_{ii'}\boldsymbol{\delta}_{ii'}\rangle$, where $n = \sum_{i=1}^{g} r_i$. We thus write $\mathrm{P}(\widetilde{\mathrm{T}}_{ii'} \ge \sigma_0 Z_\alpha\,|\,\mathrm{H}_0) = \alpha$, so that $1 - \beta = \mathrm{P}(Z_1 \ge [\sigma_0/\sigma_1]Z_\alpha - n\|\boldsymbol{\delta}_{ii'}\|^2\,|\,\mathrm{H}_1)$, with Z_1 as Z defined under H_1. It follows, under the assumptions, that σ_0/σ_1 is uniformly bounded since $\sigma_0^2 = c_{1i}O(p^2)$ and $\sigma_1^2 \le c_{2i}O(p^2)$, where, for convenience, c_{1i}, c_{2i} represent terms involving ρ_i's which are bounded under Assumption 3. Further, again under the assumptions, $\|\boldsymbol{\delta}_{ii'}\|^2 \le O(p^2)$, so that the power approximates $\mathrm{P}\{Z_1 \ge (Z_\alpha - n)O(p^2)\,|\,\mathrm{H}_1) = 1 - \mathrm{P}\{Z_1 \le (Z_\alpha - n)O(p^2)\,|\,\mathrm{H}_1\}$, and converges to 1 as $r_i \to \infty$.

3. Simulations

We evaluate $\mathrm{T}_{1\mathrm{M}}$ for size and power under different parameter settings. We consider Model (1) for $g = 2^c$ independent populations, with $c \in \{2, 3, \ldots, 8\}$, and evaluate T_{1M} for balanced and unbalanced designs. For balanced design, we generate $\mathbf{y}_{ik} \in \mathbb{R}^p$ vectors, from multivariate normal and t_5 distributions, with $r \in \{20, 40\}$ replications for each of g samples. We use $p \in \{5, 20, 50, 100\}$

Table 1
Estimated size of T_{1M} for normal distribution.

r_1	g	ID-CS				CS-AR			
		p: 5	20	50	100	5	20	50	100
20	4	0.065	0.062	0.069	0.062	0.069	0.063	0.061	0.061
	16	0.061	0.055	0.060	0.065	0.061	0.053	0.062	0.055
	64	0.057	0.049	0.063	0.058	0.064	0.060	0.051	0.057
	128	0.046	0.061	0.059	0.055	0.053	0.061	0.058	0.062
40	4	0.051	0.057	0.061	0.064	0.059	0.061	0.054	0.048
	16	0.058	0.054	0.059	0.060	0.057	0.057	0.060	0.056
	64	0.061	0.060	0.061	0.058	0.060	0.060	0.061	0.051
	128	0.055	0.059	0.057	0.52	0.049	0.053	0.050	0.059

Table 2
Estimated size of T_{1M} for t distribution.

r_1	g	ID-CS				CS-AR			
		p: 5	20	50	100	5	20	50	100
20	4	0.066	0.061	0.065	0.060	0.065	0.062	0.057	0.050
	16	0.064	0.063	0.060	0.058	0.053	0.057	0.062	0.061
	64	0.063	0.059	0.056	0.052	0.051	0.053	0.050	0.055
	128	0.051	0.066	0.059	0.049	0.060	0.0555	0.053	0.058
40	4	0.058	0.062	0.060	0.059	0.060	0.051	0.058	0.061
	16	0.059	0.066	0.064	0.052	0.055	0.057	0.055	0.053
	64	0.044	0.063	0.057	0.047	0.056	0.053	0.058	0.047
	128	0.062	0.057	0.066	0.055	0.063	0.060	0.063	0.057

Table 3
Estimated power of T_{1M} and Wilks' Λ for normal distribution.

r_1	g	ID-CS				CS-AR			
		p: 5	20	50	100	5	20	50	100
20	4	0.290	0.914	1.000	1.000	0.241	0.874	1.000	1.000
		0.236	0.841	0.997	NA	0.196	0.611	0.946	NA
	16	0.479	1.000	1.000	1.000	0.437	1.000	1.000	1.000
		0.716	1.000	1.000	1.000	0.679	1.000	1.000	1.000
	64	0.892	1.000	1.000	1.000	0.826	1.000	1.000	1.000
		1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000
	128	0.989	1.000	1.000	1.000	0.976	1.000	1.000	1.000
		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	4	0.468	0.999	1.000	1.000	0.405	0.984	1.000	1.000
		0.384	0.996	1.000	1.000	0.335	0.938	1.000	1.000
	16	0.827	1.000	1.000	1.000	0.748	1.000	1.000	1.000
		0.971	1.000	1.000	1.000	0.916	1.000	1.000	1.000
	64	0.999	1.000	1.000	1.000	0.998	1.000	1.000	1.000
		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	128	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

to represent small and large dimensions. For unbalanced design, we split each g into two subsets as $g_1 = g_2 = g/2$, and take $r_1 \in \{20, 40\}$ for g_1 subset of samples, and $r_2 = r_1 + 10$ for g_2 subset.

For each parameter setting, we associate three covariance structures, i.e., identity (ID), compound symmetry (CS), and AR(1), defined, respectively, as $\boldsymbol{\Sigma} = \mathbf{I}$, $\boldsymbol{\Sigma} = \sigma^2[(1-\rho)\mathbf{I} + \rho\mathbf{J}]$, and $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_{st})_{s,t}$ with $\boldsymbol{\Sigma}_{st} = \sigma^2\rho^{|s-t|}$, $s, t \in \{1, \ldots, p\}$, where $\mathbf{J}$ is a matrix of 1s, ρ denotes correlation coefficient, and $\sigma^2 > 0$ is a constant. Note that, for CS, $\rho \in (-1/(p-1), 1)$. We use $\rho = 0.5$ and $\sigma^2 = 1$ for both CS and AR(1).

We pair the covariance matrices for two subsets, as ID-CS and CS-AR, i.e., first split of g_1 populations is assigned ID, and second split of g_2 populations gets CS; likewise for CS-AR. As T_{1M} is location-invariant, for size, we take, without loss of generality, $\boldsymbol{\mu}_i = \mathbf{0}$ $\forall\ i$. For power, we perturb at least one of the samples by an amount of $\{1 : p\}/\sqrt{p}$ under the alternative hypothesis, leaving the rest same as under the null. For nominal $\alpha = 0.05$, we estimate size and power by averaging the appropriate probabilities over 1000 simulation runs.

For simplicity, we report results for $c \in \{2, 4, 6, 7\}$, which implies $g \in \{4, 16, 64, 128\}$, and only for unbalanced design. The results for other g values are very similar, and likewise for balanced case with relatively better accuracy particularly for small sample sizes. Tables 1–2 report estimated sizes for normal and t distributions, respectively, where Tables 3–4 report power in the same order.

Table 4
Estimated power of T_{1M} and Wilks' Λ for t distribution.

r_1	g	ID-CS				CS-AR			
		p: 5	20	50	100	5	20	50	100
20	4	0.278	0.922	1.000	1.000	0.264	0.882	0.999	1.000
		0.205	0.880	1.000	NA	0.188	0.657	0.978	NA
	16	0.485	1.000	1.000	1.000	0.441	0.998	1.000	1.000
		0.740	1.000	1.000	1.000	0.675	1.000	1.000	1.000
	64	0.882	1.000	1.000	1.000	0.810	1.000	1.000	1.000
		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	128	0.995	1.000	1.000	1.000	0.976	1.000	1.000	1.000
		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	4	0.461	0.996	1.000	1.000	0.415	0.986	1.000	1.000
		0.396	0.997	1.000	1.000	0.336	0.957	1.000	1.000
	16	0.804	1.000	1.000	1.000	0.743	1.000	1.000	1.000
		0.956	1.000	1.000	1.000	0.924	1.000	1.000	1.000
	64	0.999	1.000	1.000	1.000	0.997	1.000	1.000	1.000
		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	128	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

We observe an accurate size control for T_{1M} under both pairs of covariance structures and both distributions. We notice a slightly overestimated size for smaller sample size, but already by r_2 it improves reasonably well, and remains so for all other parameter settings, particularly for increasing g. We also notice a size accuracy for increasing dimension.

At even more appreciable level is the power performance of T_{1M}. As the alternative hypothesis mimics a contiguous case, and we observe a high power under this alternative, across the entire spectrum of parameter variants. In particular, the power increases, and rather rapidly, for increasing r_i, p, and specifically for increasing g. Moreover, this pattern remains same for both distributions and for both pairs of covariance structures. It may be added that the accurate size estimation and high power of T_{1M} under t distribution indicates its robustness to the normality assumption.

Although a detailed comparison of T_{1M}, particularly with its parametric counterparts, is relegated to an expanded work, we add a brief power comparison with the most popular classical statistic, i.e., Wilks' $\Lambda = |\mathbf{E}|/|\mathbf{H}+\mathbf{E}|$; see Section 2 for $\mathbf{H}$ and $\mathbf{E}$ matrices. Note that, with p fixed, we can use Λ but only for $p < n - g$, $n = \sum_{i=1}^{g} r_i$; thus, for the simulation parameters used for T_{1M}, Λ could not be computed for $g = 4$, $p = 100$. The second line under each (r_1, g) and (r_2, g) combination in Tables 3–4 show the power of Λ, where the first line remains for T_{1M}.

We observe a high power for Wilks' Λ for all covariance structures and distributions. We observe that the power of Λ competes well with that of T_{1M}, and, as expected of a likelihood ratio test, the power increases for increasing g and n. Note, however, that Λ cannot be used for every set of parameters as it rests on the non-singularity of $\mathbf{E}$, whereas T_{1M} does not need any such condition.

4. Discussion and outlook

A one-way MANOVA test is proposed for increasing number of independent samples, g. The dimension, p, is kept arbitrary, in that it may or may not be large, and may exceed g, or the sample size, n. The limiting normal distribution of the statistic is shown for increasing n and g. The formulation of the statistic is kept general, so that it can be easily adapted to other designs. Simulations are used to demonstrate the size accuracy and high power of the statistic.

Acknowledgments

The authors are thankful to the editors for their careful handling of the manuscript, and to two anonymous reviewers for their constructive comments which helped improve the original draft of the article.

Data availability

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

References

Ahmad, R, 2017. Location-invariant multi-sample U-tests for covariance matrices with large dimension. Scand. J. Stat. 44, 500–523.
Ahmad, MR, 2019. A unified approach to testing mean vectors with large dimensions. AStA: Adv Stat Anal 103, 593–618.
Ahmad, R, 2025. A test to determine the contributing subspace in high-dimensional classification. Math Meth Stat 34, 225–229.
Ahmad, R, 2026. Tests of parameter matrix in multivariate general linear model with large dimension. Statistics 60, 92–127.
Ahmed, E., Ahmad, R., 2023. Manova for large number of treatments. In: Pilz, J., Melas, V.B., Bathke, A. (Eds.), Statistical Modeling and Simulation for Experimental Design and Machine Learning Applications. In: SimStat 2019, Contributions to Statistics, Springer, Cham, pp. 17–30.
Anderson, T.W., 2003. Introduction To Multivariate Statistical Analysis, third ed. Wiley, NY.

Bathke, A., 2002. ANOVA for large number of treatments. Math Meth Stat 11, 118–132.
Bathke, A, Harrar, S, 2008. Nonparametric methods in multivariate factorial designs for large number of factor levels. J. Statist. Plann. Inference 138, 588–610.
Boos, D.D., Brownie, C., 1995. ANOVA and rank tests when the number of treatments is large. Statist. Probab. Lett. 23, 183–191.
Boos, D.D., Stefanski, L.A., 2013. Essnetials of Statistical Inference. Springer, NY.
Gupta, A.K., Harrar, S.W., Fujikoshi, Y, 2008. MANOVA for large hypothesis degrees of freedom under non-normality. Test 17, 120–137.
Harrar, S., Bathke, A., 2008. Nonparametric methods for unbalanced multivariate data and many factor levels.. J. Multiv. An 99, 1635–1664.
Himeno, T., Yamada, T., 2014. Estimations for some functions of covariance matrix in high dimension under non-normality and its applications. J. Multivariate Anal. 130, 27–44.
Jiménez-Gamero, M.D., Sillero-Denamiel, M.R., 2025. The k-sample problem using gini covariance for large k. J. Multiv. An 210, 105463.
Johnson, R.A., Wichern, D.W., 2007. Applied Multivariate Statistical Analysis, Sixth ed. Prentice Hall, NJ.
Koroljuk, V.S., Borovskich, Y.V., 1994. Theory of U-Statistics. Kluwer Press, Dordrecht.
Mardia, K.V., Kent, J.T., Taylor, C.C., 2024. Multivariate Analysis, second ed. Wiley, NY.
Park, J., Park, D., 2012. Testing the equality of a large number of normal population means. Comp Stat Data an 56, 1131–1149.
Wang, H., Akritas, M.G., 2004. Rank tests for ANOVA with large number of factor levels. Nonparam Stat. 16, 563–589.
Wang, L., Akritas, M.G., 2006. Two-way heteroscedastic ANOVA when the number of levels is large. Nonparam Stat. 16, 563–589.
Witting, H., Müller-Funk, U., 1995. Mathematische Statistik. Bd. II. Teubner, Stuttgart..