Review

Review about the Permutation Approach in Hypothesis Testing

Stefano Bonnini [1],*, **Getnet Melak Assegie** [2] **and Kamila Trzcinska** [3]

1   Department of Economics and Management, University of Ferrara, Via Voltapaletto 11, 44121 Ferrara, Italy
2   Department of Economics and Management, University of Parma, 43125 Parma, Italy; getnetm57@gmail.com
3   Department of Statistical Methods, University of Lodz, 41 Rewolucji 1905 r. St., 92-2014 Lodz, Poland; kamila.trzcinska@uni.lodz.pl
*   Correspondence: stefano.bonnini@unife.it

Abstract: Today, permutation tests represent a powerful and increasingly widespread tool of statistical inference for hypothesis-testing problems. To the best of our knowledge, a review of the application of permutation tests for complex data in practical data analysis for hypothesis testing is missing. In particular, it is essential to review the application of permutation tests in two-sample or multi-sample problems and in regression analysis. The aim of this paper is to consider the main scientific contributions on the subject of permutation methods for hypothesis testing in the mentioned fields. Notes on their use to address the problem of missing data and, in particular, right-censored data, will also be included. This review also tries to highlight the limits and advantages of the works cited with a critical eye and also to provide practical indications to researchers and practitioners who need to identify flexible and distribution-free solutions for the most disparate hypothesis-testing problems.

Keywords: conditional inference; exchangeability; hypothesis testing; nonparametric methods; permutation test

MSC: 62G09; 62G10; 62H15; 62J05; 62N03

1. Introduction to Permutation Tests

In many fields, the application of permutation tests in statistical inference has been increasing. For instance, many papers deal with the application of permutation tests in ecology, neurosciences, biostatistics, engineering, economics, and other disciplines [1–6]. However, in 1936, Fisher introduced the permutation test to a problem concerning paired agricultural data collected by Charles Darwin [7]. Since then, the permutation approach has become a valid solution for testing problems when the parametric methods are not powerful, feasible, or suitable due to the violations of the assumptions [8–12]. Permutation tests are flexible, powerful, and robust; hence, they are becoming very popular in several empirical disciplines, especially for complex data structures [13–29]. For instance, in multivariate permutation tests, the dependence structure of response variables is implicitly considered without modeling it or assuming a specific family for the joint probability distribution.

In 1959, Scheffé provided the following formal and concise definition of the permutation test [30]:

> "Permutation tests for a hypothesis exist whenever the joint distribution of the observations under the hypothesis has a certain kind of symmetry, namely when a set of permutations of the observations leave the distribution the same (the distributions are invariant under a group of permutations)."

In many cases, the permutation test is mainly based on the equality of distributions under the null hypothesis. Although the permutation test is carried out through a sort of re-sampling method similar to bootstrapping, unlike the bootstrap test, it is based on a sampling process that takes place by conditioning on the observed dataset [5,31,32].

For this reason, regarding permutation tests, we speak of conditioned inference. In other words, given the observed sample data, to estimate the null permutation distribution of the test statistic, the observations are randomly re-assigned to the samples to the units or, equivalently, resampled without replacement many times. Conversely, the nonparametric bootstrap is based on resampling with replacement. Extensive simulation results in previous studies revealed that, in general, the permutation tests are more powerful than the bootstrap counterparts [32–36].

Thanks to the improvement in computers' power and speed, computational limitations of permutation tests have been reduced [37]. Several *R* packages and source codes developed in the *R* programming environment have been dedicated to permutation tests. Traditional software, such as *SPSS*, *STATA*, and others, have also added an option for permutation testing. Some examples of *R* packages for permutation tests are *lmperm* [38], *coin* [39], and *flip* [40]. Some useful *R* packages and source codes for the application of permutation tests are described in [4,31].

Permutation tests are distribution-free. Let us assume that the *i*th observation x_i is a realization of the random variable X_i with $i = 1, 2, \ldots, n$. Typically, parametric approaches assume that the random variables $X_1, X_2, \ldots, X_n$ are independent and identically distributed (IID). It is well known that, in the permutation approach, the strong IID condition is not necessary as well, and it is replaced by the milder condition of exchangeability [41,42]. For example, let us consider the univariate two-sample test on the means of two populations, where the null hypothesis H_0 of equality of the two population distributions is tested against the alternative hypothesis H_1 that μ_1, the mean of the first population is not equal to (or it is greater than) μ_2, the mean of the second population. Let us *assume* that we have randomly selected a sample from each population (classical two-independent-sample problem). Without a loss of generality, let us also assume that the first n_1 observed sample data (arranged in the first n_1 rows of the dataset) refer to the sample selected from the first population, and the other n_2 observed data (arranged in the last n_2 rows of the dataset) refer to the sample selected from the second population, where n_1 and n_2 are the sample sizes, with $n = n_1 + n_2$. If the null hypothesis is true, the two populations have the same underlying distribution, and the sequence of observed data $x_1 \ldots, x_n$ has the same probability of occurrence of any other of the $\binom{n}{n_1}$ sequences obtained by permuting the observations to reassign the data to sample 1 and sample 2. Thus, the null distribution of the test statistic (for example, the difference between the sample means in the one-sided test and its absolute value in the two-sided test) can be obtained by considering all the permutations corresponding to data reassignments (or equivalently, a random sample of them) and the value of the test statistic computed for each permuted dataset. The *p*-value of the permutation test is the proportion of permutation values that are equal to or more extreme than the observed one, i.e., the one corresponding to the original non-permuted dataset. According to the classic decision rule, the null hypothesis is rejected in favor of the alternative hypothesis if the *p*-value is less than or equal to the significance level of the test. Good focused on the exchangeability condition and provided formal properties and examples [43].

Other conditions and properties of permutation tests, such as unbiasedness and consistency, were investigated by several authors (see, for instance, [44,45]). From a practical point of view, such properties are very important. In fact, the former ensures that the probability of rejecting the null hypothesis when it is false is greater than the probability of rejecting it when it is true (type I error). Formally,

$$P[T_{n_1,n_2} \in \mathcal{R}_\alpha | H_1, X] \geq \alpha$$

where T_{n_1,n_2} is the test statistic, α is the significance level, $\mathcal{R}_\alpha$ is the rejection region, H_1 is the alternative hypothesis, and X is the observed dataset. The latter implies that the probability

of rejecting the null hypothesis when it is false increases with sample sizes and tends to one when the sample sizes diverge. Equivalently,

$$P[T_{n_1,n_2} \in \mathcal{R}_\alpha | H_1, X] \overset{min(n_1,n_2) \to \infty}{\to} 1.$$

On the other hand, Konietschke et al. [46] argued that permutation tests can also be applied to non-exchangeable data. In particular, they considered tests for dependent samples, the Studentized test statistic, and different permutation strategies and proved that all these strategies lead to asymptotically valid tests, which are more powerful than the parametric *t*-test.

Exchangeability is a simple and weak condition. It is reasonable in many frameworks typical of both experimental and observational studies. For instance, in experimental designs, the random allocation of subjects to treatments is sufficient to justify this assumption. Furthermore, depending on the goal of analysis, and on the study design, different types of exchangeablilities are introduced by [43]: preserving transforms, asymptotic, partial, and weak exchangeability. Exchangeability and conditioning on sample data imply an important invariance property [43,45]. In fact, under exchangeability, the joint distribution of the observations (random variables) is invariant under reassignments of the subscripts. As a result, sufficient statistics can be computed with respect to the permutation distribution.

In the permutation test, the decision to reject or not the null hypothesis is based on the *p*-value obtained from the null permutation distribution. Anderson studied the univariate and multivariate permutation analysis of variance for one-way, multifactor, and complex designs by proposing asymptotic permutation *p*-values as an alternative to the permutation *p*-values [6,47]. Such asymptotic *p*-values are computed by approximating the permutation distribution of the test statistic with a linear combination of chi-square random variables. She suggests using the asymptotic *p*-values in particular when the number of possible permutations is low. The proposed test statistic is a function of the matrix of distances between couples of units. As a distance, several different types of dissimilarity measures may be used. Such a test is applicable to balanced designs, and its performance depends on the considered type of distance. Moreover, the author notes that exact permutation tests on the interaction effects in multifactorial problems are challenging. For such a problem, approximate permutation tests may be applied [12,17]. In this framework, the application of restricted and synchronized permutations provides appropriate and valid solutions [48,49].

Excluding asymptotic solutions, the *p*-values of the permutation tests are computed empirically and conditionally on the observed sample data. In the literature, especially when the number of possible permutations is large, for the computation of *p*-values, the conditional Monte Carlo method, according to the permutation principle, is very frequently used [48,50–55]. In other words, to determine the null permutation distribution and compute the *p*-values, instead of calculating the exact *p*-values by taking into account the set of all possible permutations (exact test), a random sample of permutations is drawn. Often, a number of conditional Monte Carlo permutations of 1000 should be sufficient to have a good approximation of the null permutation distribution and obtain *p*-values very close to the exact ones [56,57]. Some authors use 10,000 permutations [31,58]. This approach is particularly important because, as the sample sizes increase, the cardinality of the permutation space, and thus the total number of possible permutations, increases rapidly and can be very large, even for relatively small sample sizes. For example, going back to the two-independent-sample test mentioned above, when $n = 20$ and $n_1 = n_2 = 10$, the number of permutations of the exact test is $\binom{20}{10} = 184,756$, and computing the approximated *p*-value by generating 10,000 random permutations is more computationally convenient.

In [59], a simulation study to determine the appropriate number of permutations to obtain well-approximated *p*-values was proposed. It was found that the number of permutations required for estimating the power of the permutation test and the *p*-values for practical data analysis depends on the level of significance. For instance, 1000 and 5000 per-

mutations are sufficient for estimating power performance and *p*-values, respectively, for practical data analysis at a nominal level of 0.5. The permutation distributions of the test statistic obtained from all possible permutations, and with the conditional Monte Carlo method, asymptotically provide the same conclusion.

The permutation testing principle for two-sample or multi-sample tests, strictly connected with the concept of conditional inference, was defined by Pesarin in [5] (p. 8) as follows:

> "If two experiments, taking values on the same sample space and respectively with underlying distributions P_1 and P_2, both members of $\mathcal{P}$, give the same dataset *X*, then the two inferences conditional on *X* and obtained by using the same test statistic must be the same, provided that the exchangeability of data with respect to groups is satisfied in the null hypothesis. Consequently, if two experiments, with underlying distributions P_1 and P_2, give respectively X_1 and X_2, with $X_1 \neq X_2$, then the two conditional inferences may be different."

Another approach to the definition of permutation tests is proposed by [9] (p.633). According to this approach, permutations are seen as groups of transformations that preserve the null distribution of the test statistic [60] (pp. 3–4). Let *X* be the dataset and Ω the sample space. The finite set of transformations $g : \Omega \to \Omega$, denoted by *G*, is a group if the following properties hold: (1) *G* includes the identity transformation; (2) for every transformation in *G*, the inverse is also in *G*, i.e., $g \in G \Rightarrow g^{-1} \in G$; (3) the composition of every pair of transformations in *G* is also in *G*, i.e., g_1 and $g_2 \in G \Rightarrow g_1 \circ g_2 \in G$. A more detailed discussion of this approach is provided in Section 2.

This paper considers a review of permutation tests for comparative studies and linear regression models. The paper is structured as follows. Section 2 is dedicated to permutation tests for independent samples. The case of dependent samples is presented in Section 3. In Section 4, permutation solutions for regression models are presented. Section 5 covers the topic of permutation approaches for missing data. Finally, Section 6 comprises some final remarks.

2. Permutation Tests for Independent Samples

For tests of hypotheses concerning comparisons of independent samples in terms of location; scale; proportion; or other aspects of the distribution, numerical, categorical, or binary outcomes, the standard parametric approach might not be suitable in the presence of skewed distributions, large kurtosis, non-normal distributions, unequal variances, small sample sizes, missing data, and other situations. For such problems, a nonparametric solution is a valid and, in some cases, inevitable alternative. Within the category of nonparametric methods, permutation tests are often preferable for their flexibility, robustness, and power.

In general, the test statistic may be defined so as to reject the null hypothesis in favor of the alternative for large values. Let *B* denote the total number of permutations (either the exact number of reassignments or the number of random permutations), $t_{n_1,n_2}^{*(b)}$ is the value of the test statistic in the *b*th permutation, and $t_{n_1,n_2}^{(0)}$ is the observed value of the test statistic. The *p*-value of the permutation test is

$$p = \frac{\sum_{b=1}^{B} I_b \left(t_{n_1,n_2}^{(0)} \right)}{B}$$

where

$$I_b \left(t_{n_1,n_2}^{(0)} \right) = \begin{cases} 1 \text{ if } t_{n_1,n_2}^{*(b)} \geq t_{n_1,n_2}^{(0)} \\ 0 \text{ otherwise} \end{cases}.$$

Sometimes, the values 0.5 and 1 are added in the numerator and in the denominator, respectively, to make p strictly between 0 and 1. In the case of the two-sample test on means

mentioned above, a suitable test statistic is the difference between sample means; hence, the observed value of the test statistic is

$$t_{n_1,n_2}^{(0)} = \frac{\sum_{i=1}^{n_1} x_i}{n_1} - \frac{\sum_{i=n_1+1}^{n} x_i}{n_2} \text{ or } t_{n_1,n_2}^{(0)} = \left| \frac{\sum_{i=1}^{n_1} x_i}{n_1} - \frac{\sum_{i=n_1+1}^{n} x_i}{n_2} \right|,$$

depending on whether the test is one-sided or two-sided respectively. Similarly, the bth permutation value of the test statistic is

$$t_{n_1,n_2}^{*(b)} = \frac{\sum_{i=1}^{n_1} x_i^{*(b)}}{n_1} - \frac{\sum_{i=n_1+1}^{n} x_i^{*(b)}}{n_2} \text{ or } t_{n_1,n_2}^{*(b)} = \left| \frac{\sum_{i=1}^{n_1} x_i^{*(b)}}{n_1} - \frac{\sum_{i=n_1+1}^{n} x_i^{*(b)}}{n_2} \right|,$$

where $x_i^{*(b)}$ is the ith value of the variable under study in the bth permutation.

In the multivariate case, with a number $k \geq 2$ of variables under study, instead of the elements of the vector $(x_1, x_2, \ldots, x_n)'$, the rows of the dataset X must be permuted in order to reassign the profiles of values observed on the statistical units to the two samples. Taking into account the definition of permutations as groups of transformations [9,60], the bth dataset permutation can be defined as a transformation of X, which is a $n \times k$ matrix, obtained as a premultiplication of X by a suitable $n \times n$ matrix $\Gamma^{(b)}$. Formally,

$$X^{*(b)} = \Gamma^{(b)} X$$

with $b = 1, 2, \ldots, B$. To simplify the notation, let us remove the superscript (b). Each element of $\Gamma = [\gamma_{ji}]$ is either zero or one. In each column and each row of Γ, only one element is equal to one. Specifically, if in the permutation the ith row of X takes the place of the jth row of X^*, then

$$\gamma_{jr} = \begin{cases} 1 & \text{if } r = i \\ 0 & \text{otherwise.} \end{cases}$$

Such an approach also applies in the univariate case, when $k = 1$. The test statistics of the multivariate permutation tests depend on the specific problems and the hypotheses under test.

2.1. Two-Sample Location and/or Scale Problems

In two-sample location/scale problems, when the assumption of normality is violated, the permutation test is a suitable solution [61–65]. The standard Student's t-test for univariate responses and its multivariate extension, the Hotelling T^2 test, cannot provide reliable results when this assumption is not satisfied. Furthermore, when the alternative hypothesis is directional, there is no parametric solution to test the shift in location between two samples for multivariate response variables [4,66]. In other words, when there is a directional (positive or negative) treatment effect or a shift in the location parameter in the alternative hypothesis, the classic parametric Hotelling T^2 test, valid for the two-sided alternative hypothesis, has no counterpart for the one-tailed problem. The permutation approach provides an effective method for solving such a problem.

Permutation tests for two-sample problems, when the treatment effect is fixed, have been extensively investigated in the literature (see, for example, [4,67–69]). However, different solutions need to be applied when the treatment effect is not constant (heterogeneous treatment effects). For such a problem, a suitable solution, based on the permutation approach, is proposed by [70]. In this work, a permutation test for heterogeneous treatment effects when a nuisance parameter is present was introduced. This approach uses a martingale-transformed test statistic to determine asymptotic critical values for a pivotal statistic. Simulation results revealed that the method of [70] controls the type I error rate.

Similarly, many authors recommended the use of pivotal statistics in permutation tests [1,6,58]. However, since the permutation test is distribution-free, a pivotal quantity as

a test statistic is neither a necessary nor a sufficient condition for the application. However, some permutation tests also provide reliable results using pivotal test statistics.

The traditional parametric approach can not apply in the presence of mixed multivariate variables. On the other hand, the nonparametric condition of dependent permutation tests allows us to deal with different types of response variables (numeric and categorical) [31]. In fact, assuming that each response variable corresponds to a partial test, the general problem can be solved through a suitable combination of the *p*-values of the partial tests, regardless of the nature of the variables of the single partial tests [31].

The traditional parametric approach does not provide solutions for testing location and scale parameters jointly [4]. In other words, if the researcher is interested in jointly examining the treatment effects in the first and the second moment, that is, the change in both mean and variance, the traditional parametric test cannot face such a problem. In contrast, the permutation approach solves such a complex problem. For instance, Pesarin and Salmaso, in [31], used a multi-aspect permutation solution to jointly test a shift in location parameter and a change in variability between two populations in both univariate and multivariate problems. They tested the significance of location and scale parameters using the difference of means and difference of the sum of squares as a statistic for the test on location and scale, respectively, and then combined the *p*-values of the two partial tests. In addition, a multi-aspect permutation test for the first and second moment in a problem of shape analysis is presented in [71].

Moreover, Mielke and Berry, in 1994, proposed an interesting study for a completely randomized experimental design to jointly test a shift in location parameter and change in scale parameter for two samples [37]. Their method to test the significance of location and scale parameters is somehow different from that of [31]. They considered an omnibus statistic for the general problem, whereas Pesarin and Salmaso considered different test statistics for the different aspects and, as said, a final combination of *p*-values for the overall problem [31]. The three proposed omnibus permutation tests were the Euclidean commensuration, the Hotelling commensuration, and the permutation version of the Bartlett–Nanda–Pillai trace test [72–74]. A simulation study under various distributions was carried out to compare the proposed tests with the parametric Bartlett–Nanda–Pillai trace test. The simulation results revealed that the Euclidean commensuration permutation test was the most performant, irrespective of the considered distributions for a bivariate response. The power of the tests was not good for the other distance. In other words, the power performance of these methods depends on the choice of the distance metric. In addition, the Hotelling commensuration permutation test was not stable due to the loss of degrees of freedom.

A comparative study on location for numeric variables when we have unequal unknown variances of the populations, the so-called Behrens–Fisher problem, is proposed by [75]. Janssen, in 1997, proposed an asymptotic test [62] based on the permutation version of the Welch test [76], which is an extension of Pitman's permutation test for the Behrens–Fisher problem on the equality of two population means [11]. Although the permutation approach provides an exact test under the IID assumption in the null hypothesis, the Janssen method based on the Studentized statistic does not require the IID assumption to obtain an asymptotic test. In the paper, this method was applied to a composite null hypothesis against one-sided alternatives. The simulation results revealed that the permutation test is more powerful than the Welch test [77] for skewed distributions. The limitation of this method is that it is not stable for an unbalanced design.

According to [31], Studentized and non-Studentized test statistics are permutationally equivalent due to the conditioning on the sample data and invariance property, and they provide approximately similar results. Eliminating the denominator of the Studentized permutation test statistic may improve the computational aspect of this solution. Furthermore, other authors solved the Behrens–Fisher problem nonparametrically by using the permutation approach [75,78–80].

2.2. Multi-Sample Problems for Numeric Data

In addition to a vast literature on two-sample permutation tests, there are many scientific contributions (theoretical and applied) on the univariate and multivariate multi-sample extensions, such as the permutation analysis of variance (ANOVA) and the permutation multivariate analysis of variance (MANOVA) [34,35,68,69,81–84]. The standard parametric F test for the ANOVA requires equal variances, normality, continuous response, and IID observations. Similar assumptions are also required for the multivariate extension, i.e., the so-called MANOVA. In addition, even if the assumptions of (M)ANOVA are satisfied, there are some conditions in which the classical (M)ANOVA cannot be applied [72,85,86]. In such a problem, the permutation solution is possible, appropriate, and performant.

Similar to what happens in the two-sample multivariate location problem, the loss of degrees of freedom in the presence of a large number of response variables, especially with small sample sizes, is quite common in MANOVA. In this case, the parametric tests, as well as the nonparametric rank-based tests, cannot be applied when the number of variables exceeds the sample sizes [87]. In such a situation, the nonparametric solution based on the permutation approach is the unique possible choice. Permutation (M)ANOVA provides accurate and reliable (powerful) results [68]. In fact, for these problems on the difference between the means of three or more groups, the permutation test under the null hypothesis yields a type I error rate that respects the predetermined level of significance α and rejects the null hypothesis when the alternative is true with high probability. In this framework, permutation tests are exact, unbiased, and consistent [5].

In general, under the null hypothesis, to obtain the null distribution of the test statistic, the permutations of the statistical units for the one-way (M)ANOVA are carried out by exchanging the row vectors of the observations between treatment groups or factor levels irrespective of the outcome variable(s) [35,68,81]. However, the permutation principle for the two-way (M)ANOVA, or general factorial design, requires some restrictions to test the main effect of only one of the two factors. For this reason, researchers used synchronized permutations to permute row vectors of the observations between levels of one factor by keeping the other factors as blocks [4,49,88].

In addition, the synchronization process could be constrained or unconstrained [89,90]. In the case of constrained synchronized permutations, if we have an unbalanced design, the smallest sample size determines the number of units to be exchanged within treatment levels, and the position of the units must be fixed. Meanwhile, in unconstrained synchronized permutations, the position of the units doesn't matter when permuting the observations (see [91] for a detailed explanation).

In [91], a synchronized permutation test for unbalanced two-factor ANOVA with two levels was developed. Since the error terms are not exchangeable between the two factors under the null hypothesis, the authors used the synchronized permutation principle. However, the error terms are exchangeable within the levels of one factor by considering the other factor as a block to obtain an approximately exact p-value. They used the weighted sum of observations within each cell as a permutation test statistic to test the significance of factors' main effects and interaction effects. The conditional Monte Carlo simulations revealed that the power performance of the permutation test is somehow influenced by the permutation mechanism (constrained or unconstrained) and the weights of the test statistic. For instance, the constrained synchronized permutation mechanisms reduced the power of the permutation test. In addition, the test was conservative for restricted weighted and constrained synchronized permutation mechanisms. However, they proved that the power of the test using the constrained synchronized permutations is unstable for small sample sizes in the case of unbalanced designs.

The choice of the best test statistic, i.e., the one that allows obtaining the most powerful permutation test among these exact tests under H_0 for a given system of hypotheses, is one of the challenging problems [87]. In permutation MANOVA, different test statistics have been considered—for example, distance-based and specific (M)ANOVA statistics based on deviances. Thus, comparing the type I error rate and the power of the tests based on those

test statistics is crucial for selecting the appropriate statistics and improving the power performance of the permutation test. In line with this, there is an interesting comparative study about the powers of distance-based statistics and variable-based statistics in a multi-sample problem [92]. The statistics were classified as distance-based when the average distance of pairs of units or dissimilarity measures were used and variable-based statistics when summary statistics such as Wilks lambda, likelihood ratio type, and ANOVA F type statistic were considered. A distance-based statistic requires the transformation or standardization of the original data. On the other hand, a variable-based statistic is usually conceived in relation to the assumption of normality. The test statistics used in permutation tests do not require neither standardization nor the assumption of a specific underlying probability distribution. In general, with permutation tests, variable-based statistics are often preferable to distance-based ones. As far as the distance-based statistic is concerned, standardization and transformation of the original variables might impact the power of the test [92]. In addition to the type of transformation, another choice for distance-based statistics is that of the metric to be used. The most commonly used distances are the Euclidean, Bray–Curtis, and Manhattan [93].

According to the simulation study conducted by Warton and Hudson in 2004, the effect on the power of the preliminary transformation of the original data is particularly pronounced when the Bray–Curtis dissimilarity measure is used in the presence of unbalanced designs [92]. Other comparative studies on the permutation tests for ANOVA were proposed using different test statistics such as F or ANOVA-type statistics (ATS) [94], Wald type statistic (WTS) [35], Fisher–Pitman statistic [95], Studentized statistic [34,35,70], and distance-based statistics [68]. Although the Wald-type statistic is useful for comparing the means of multivariate outcomes among groups, it requires the computation of the sample covariance matrix. Furthermore, the convergence of its distribution to the asymptotic χ^2 law is slow [96]. Smaga, in 2020, introduced a modified form of the Wald-type statistic using a weight diagonal matrix, providing a test with good asymptotic and finite sample properties [34] (pp. 245–254). Regarding the use of distance metrics, the Euclidean distance-based test statistic seems more performant than other distance-based statistics [68,97]. A limitation of such statistics is that they cannot be applied to categorical responses. In 2001, Peres-Neto and Olden extended the application of distance-based permutation statistics to categorical outcomes in a permutation (M)ANOVA framework [42].

A permutation solution for the MANOVA problem based on eigenvalues by partitioning the total variability into the different contributions of the single factors was proposed by [68,97]. However, the eigenvalues obtained from the distance matrix sometimes are negative. The negative eigenvalues imply difficulty in interpreting the results, and some authors corrected it by adding a constant to the sum of squares [98]. McArdle and Anderson, in 2001, proposed a permutation version of the Euclidean-based statistic in hypothesis testing of MANOVA or MANCOVA [97]. Their solution was based on the partition of the total variability or dissimilarity because the pseudo F-statistic could be derived from the Euclidean distance matrix to test the significance of the factors. They carried out a simulation study to assess the accuracy of the proposed method, called distance-based redundancy analysis (dB-RDA), which does not require any correction to have positive Eigenvalues. Their method was compared with three competitors, including a test based on the Bray–Curtis distance; the approach of distance-based redundancy analysis; and axes analysis for testing fixed, random, and mixed effects. The simulation results showed that the test has a type I error rate approximately equal to the significance level α for the fixed-effect model, even if it is conservative for random and mixed linear models, especially for small sample sizes and a lognormal distribution. However, in general, the proposed distance-based redundancy analysis (dB-RDA) provided a well-approximated type I error rate. The simulation study may be extended to check the good power performance under an alternative hypothesis.

In one-way ANOVA design, when the assumptions of the F test are violated, a permutation version of the Fisher–Pitman test can be used for testing the equality of location

parameters among the groups [95]. The advantage of the Fisher–Pitman test lies in the way the permutation test statistic is computed. Since the total number of observations N, the number of groups S, and the total sample mean $\overline{Y}$ are permutationally constant due to the invariance property under exchangeability, the ANOVA F test could be simplified with the Fisher–Pitman permutation test statistic or the weighted sum of squares of the within-group sample means $\overline{Y}_g$:

$$T = \sum_{g=1}^{S} n_g \overline{Y}_g^2$$

Such a test statistic has computational advantages and provides efficient permutation testing procedures. In the balanced sample designs, the weights represented by the sample sizes can be eliminated.

Smaga, in his work of 2020, proposed a solution for the multivariate Behrens–Fisher problem of a two-way MANOVA [34]. He used the ANOVA-type statistic (ATS) to test the equality of population mean vectors under heteroscedastic error variances. The asymptotic null distribution of ATS is a mixture of central chi-square distributions. This method is robust against skewed distributions, unequal variances, and nested factorial designs. It was proved that the permutation version of ATS is consistent. However, although the ATS works well for finite sample sizes and small groups, it is affected by the number of groups, suggesting the standardized form. In addition, they derived the confidence region for the interval estimation of mean vectors. Moreover, it was proved to be more powerful than its competitor WTS and the bootstrap method, regardless of the distributions' variance structure, sample sizes, and skewness.

Similarly, Anderson, in 2001, developed a family of permutation tests on the equality of treatment effects with multivariate outcomes [6]. The proposed solution used the squared Euclidean distance as a test statistic. This method is useful for jointly testing the significance of both factors' main effects and interaction effects. Moreover, she developed permutation solutions for general factorial designs and complex models such as nested or hierarchical models based on a distance measure as a test statistic. One advantage of this test is that it considers the possible dependence among the components of multivariate outcomes in multifactorial designs, which is problematic to cope with using traditional parametric MANOVA. Nonetheless, attention must be paid when testing the interaction effects since error terms are not always exchangeable under the null hypothesis to obtain exact *p*-values. In addition, the distance-based statistic requires the transformation of the data. In this case, the interpretation of the results may change according to the data transformation. Moreover, if the number of groups is more than two, a further assessment may be carried out, to identify which pairs of means are significantly different, through permutation *t*-tests for pairwise comparison.

In a work published in 2015, Pauly et al. extended the application of permutation tests to general factorial designs with two or more factors, nested and hierarchical design in the presence of heteroscedasticity [35]. The Wald-type parametric test cannot provide a valid result for the heteroscedastic case and in the case of small sample sizes. Indeed, the distribution of the Wald-type test hardly approximates the chi-square distribution with small sample sizes. Furthermore, when the number of factors and levels in factorial designs are large the Wald-type test cannot efficiently cope with the test of hypotheses on treatment effects. In contrast, the proposed Wald-type permutation test statistic works well under small sample sizes, heteroscedastic errors, and large number of factors. In addition, the authors considered the standardized Wald-type statistic using the estimated variance obtained from permuted data. However, the computation of the error variance may slow down the execution of the permutation test, and it needs simplification. Finally, they proved that the permutation distribution of the Wald-type test statistic, which is conditional on the sample data, weakly converges in probability to the central chi-square distribution. Moreover, their simulation study revealed that the chi-square test, the parametric ANOVA, and the Wald-type parametric test provide a liberal type I error, especially under small

sample sizes and skewed error distributions. Conversely, the Wald-type permutation test, under H_0, respects the nominal α level regardless of the sample size, the covariance structure, and the distribution skewness, except for log-normal distribution with unequal variances, where the test shows slight inaccuracy.

2.3. Small Sample Multivariate Problems

Although the literature about solutions for multivariate tests with the number of variables greater than the number of observations is not rich, some authors have addressed the problem [31,63,64,99]. In two-sample or multi-sample tests, when sample sizes are smaller than the number of response variables, the classic parametric solutions, such as the Hotelling T^2 test, cannot be used due to the loss of degrees of freedom. In this case, permutation solutions are possible [31]. For these tests, the so-called finite-sample consistency is satisfied. This property implies that, under some conditions, the power of the test in the alternative hypothesis H_1 increases with the number of variables irrespective of the finite sample sizes.

Furthermore, the authors of [64] proposed an invariant inter-point distance method using the permutation principle to compare the location parameters for multivariate two-sample problems in the presence of high dimensional response variables with fixed small sample sizes. They carried out a simulation study to compare the power behavior of the permutation test with other nonparametric tests (such as the multivariate generalization of the run test [100], the nearest neighbor [101], and the Rosenbaum test [102] under normal and Laplace distributions. They found that the permutation test is more powerful than the competitors, whose power tends to zero when the dimension of the response variable diverges. The power of the permutation test based on the inter-point distance tends to one when the number of components of the multivariate response variable increases, keeping the sample sizes fixed. The simulations proved that the permutation test is consistent and powerful when the number of responses is diverges, with small fixed sample sizes, under the Behrens–Fisher problem as well. However, their method assumed the uniformly bounded forth moments and weak correlation among the components of the response variables.

A recent development to improve the power behavior of a multivariate version of the two-sample permutation test, the so-called combination-based permutation test, was proposed by [87]. The main property of the combination-based permutation tests is that they condense the statistical information of all response variables into one statistic, and they implicitly take into account the dependence structure between response variables without assuming a specific joint probability distribution and without modeling such a dependence.

Another contribution to the multivariate testing problems with a large number of response variables is [42]. Through Monte Carlo simulations, the authors also found that the power behavior of the Hotelling T^2 and the rank-based test decrease as the number of variables increases with fixed sample sizes, while the power of combination-based permutation tests increases monotonically as the number of responses under the alternative hypothesis increases with fixed sample sizes. The power of the combination-based permutation test was good for small shifts in the population means as well.

2.4. Tests for Categorical Data

Testing problems for categorical data, especially in the multivariate case, are not easy to solve. With multidimensional variables, a crucial and complex aspect concerns the dependence structure of the response components. For this reason, a distribution-free approach for such testing problems in the presence of multivariate variables with dependent components is useful and important.

Permutation solutions for multivariate stochastic ordering have been investigated by [103–106]. Statistical ordering is typical of various complex problems such as tests for restricted (directional) alternatives [10], multiple comparisons [107], and ranking of populations [103]. For instance, [108] applied multivariate permutation tests to compare

two populations in the presence of a multivariate categorical response. She focused on the so-called case–control study and used the odds ratio as the test statistic. If we consider the contingency table where rows correspond to treatments and columns to response categories (or vice versa), permutations make the joint absolute frequencies of the table change, keeping the marginal absolute frequencies constant. Investigating the treatment effect is equivalent to testing the association between responses and treatments. Unlike the classic chi-square test, in the permutation test, there are no restrictions on the minimum frequency value in the single cells. In [108] it was proved that the permutation test is robust, even for small sample sizes and, in some cases, small frequencies.

In practice, parametric tests such as the likelihood ratio test are not appropriate for testing the stochastic dominance for ordered categorical variables. On the other hand, an interesting application of permutation tests for univariate and multivariate ordered categorical variables under restricted alternatives was proposed by [105]. This method is based on transforming the categorical variables into numeric variables according to the rank transformation. The global null hypothesis consists of the treated and control group equality in distribution. First, H_0 is broken down into $k-1$ partial null hypotheses, where k is the number of categories of the ordinal variable. Then, partial permutation test statistics based on the difference of sample moments are used to test the sub-null hypotheses of equality of the first $k-1$ moments. Finally, the authors used simulated data from an ordinal multinomial distribution with four categories using the GenOrd *R* package to investigate the power behavior of the permutation test (Fisher and Liptak combining function) compared to the competitors' rank-based Wilcoxon test and the Brunnel–Munzel test. The simulations showed good power behavior under both null and alternative hypotheses. Under the alternative hypothesis, the permutation test based on the sample moments after the score transformation was the most powerful. The power of such a permutation test was proved to be affected by the score transformation. For example, symmetric or asymmetric scores can lead to different results. When the nature of the response variables is ordinal, for example, in Likert-scale questionnaires, tests for group comparisons can be carried out by replacing the ordered categories with numeric scores [67]. The Wilks lambda-type permutation statistic, based on rank transformation of units, was proven to be performant for both symmetric and skewed data. In the simulation study, the proposed solution was compared to the traditional MONAOVA test, based on Wilks lambda, the nonparametric Wilks, structural equation models for MANOVA tests, the test based on spatial signs with inner centering and outer centering, and others. Except for the structural equation models for MANOVA tests, all the tests respected the nominal α level under H_0.

One interesting problem with categorical data is to determine the ranking of several groups based on some characteristics. For instance, in a biomedical study, several groups of patients may be sorted based on ordinal categorical outcomes with respect to different dose levels. Hence, such a ranking problem can be broken down into several stochastic dominance pairwise comparison metric-free permutation tests for comparing and ranking multivariate populations. A solution based on the permutation approach was proposed by the authors of [103]. When the null hypothesis of equality of the multivariate populations is rejected, the need arises to determine which groups show significant differences to find the final ranking. A post hoc comparison is the solution for such a kind of problem. The Anderson–Darling-type statistic for the permutation pairwise comparisons was proposed. Simulations of multivariate ordinal categorical data were carried out to investigate the ranking of populations for the different number of groups, small sample sizes, and different skewed and normal distributions. The simulation results revealed that the estimate of correct ranking is close to the true one under the homogeneity assumptions, irrespective of the sample size and choice of error distributions. In addition, the bias was lower under the null hypothesis. Moreover, such a permutation solution for stochastic dominance and ranking can apply to mixed-response variables.

3. Permutation Tests for Dependent Samples

This section is dedicated to permutation approaches for comparative studies with dependent samples. Among them, we consider the permutation (M)ANOVA for longitudinal data and two-sample or multi-sample tests for paired data or repeated measures. We start with the simplest tests, and then we extend the review to more complex designs. It is well known that the typical parametric solution for the two-sample test on location for dependent samples, the so-called paired *t*-test, is not appropriate for small samples and non-normal data. When the underlying distribution is unspecified, the nonparametric approach based on the Wilcoxon signed-rank test is a valid alternative for such a problem. Nonetheless, the nonparametric rank test requires continuous distribution, so the probability that ties occur, i.e., to observe two identical data, is zero [4]. The re-sampling methods, in particular, permutation tests, do not require continuity; therefore, they are more flexible [71,109].

Let us consider the classic univariate two-sample test on the means of numeric variables for dependent samples and let $(x_{11}, x_{21}, \ldots, x_{n1})'$ and $(x_{12}, x_{22}, \ldots, x_{n2})'$ be the vectors of values of the response observed in the first and second sample respectively. For example, they could be the values of an outcome of interest observed before and after the treatment on a sample of n patients in a medical study. A suitable test statistic in such a problem is based on the sample mean of the differences, and the observed value of the test statistic is

$$t_n^{(0)} = \frac{\sum_{i=1}^n (x_{i1} - x_{i2})}{n} = \frac{\sum_{i=1}^n d_i}{n} \text{ or } t_n^{(0)} = \left| \frac{\sum_{i=1}^n (x_{i1} - x_{i2})}{n} \right| = \left| \frac{\sum_{i=1}^n d_i}{n} \right|$$

depending whether the alternative hypothesis is $H_1 : \mu_1 > \mu_2$ or $H_1 : \mu_1 \neq \mu_2$, where μ_1 and μ_2 are the means of the first and second population, respectively, i.e., the mean of the outcome before and after the treatment in the medical example.

Under the null hypothesis of equality of the means, the within-unit exchangeability holds. In the medical example, this means that, for the i th patient, conditional on the observed values x_{i1} and x_{i2}, the probability that the former is the values observed before the treatment and the latter is the one observed after the treatment is ½, the same as the probability that the chronological order of observation of the two values (assignment to sample 1 and sample 2) is the opposite, and this is true for all $i = 1, 2, \ldots, n$. Thus, for each i, x_{i1} and x_{i2} are exchangeable. The null permutation distribution of the test statistic can be obtained by permuting the pairs (x_{i1}, x_{i2}) in all possible ways on the n individuals and independently from one individual to another.

In the exact test, there are 2^n possible permutations. Hence, when $n \geq 10$ and consequently $2^n \geq 1024$, the approach of random permutations, with $B = 1000$, is computationally convenient. Random permutations of the pairs (x_{i1}, x_{i2}) are equivalent to random attributions of the signs of the differences d_i ($i = 1, 2, \ldots, n$). Thus, a random permutation on the ith individual or unit is equivalent to the realization of the random variable S_i, which takes either the value 1 or the value -1 with the same probability 0.5. It is worth noting that S_i can be defined as a function of a Bernoulli random variable and that there is independence across individuals. Formally,

$$\forall i \ Z_i \sim \mathcal{B}e(0.5) \text{ and } S_i = 2Z_i - 1 \text{ with } Z_1, Z_2, \ldots, Z_n \text{ independent}$$

The b th permutation value of the test statistic is

$$t_n^{*(b)} = \frac{\sum_{i=1}^n d_i^{*(b)}}{n} = \frac{\sum_{i=1}^n s_i^{*(b)} d_i}{n} \text{ or } t_n^{*(b)} = \left| \frac{\sum_{i=1}^n d_i^{*(b)}}{n} \right| = \left| \frac{\sum_{i=1}^n s_i^{*(b)} d_i}{n} \right|,$$

where $s_1^{*(b)}, s_2^{*(b)}, \ldots, s_n^{*(b)}$ are determinations of $S_1, S_2, \ldots, S_n$ respectively, i.e., the n signs randomly generated in the bth permutation. The multivariate extension of the null permutation distribution, assuming that there are k response variables, can be obtained by replacing

in the formulas the values $d_1, d_2, \ldots, d_n$ with the k-dimensional vectors $d_1, d_2, \ldots, d_n$. The computation of the *p*-values follows the same rule presented above for the independent-sample case.

Konietschke and Pauly [46] studied the permutation solution for paired data and compared it with the nonparametric Bootstrap test and the parametric paired *t*-test in the case of small sample sizes and skewed distributions. Under the null hypothesis, in the paired sample problem, the observations are exchangeable within the statistical units. Indeed, the permutation inference requires a mild condition of exchangeability under the null hypothesis [43,45]. However, in some problems, the permutation test can be applied under some conditions, even if the data are not exchangeable, because exchangeability holds approximately or asymptotically. See, for example, the permutation approach for paired samples studied by [46,110,111]. For example, Konietschke and Pauly [46] presented three simple conditions in which the permutation principle works to obtain an asymptotic *p*-value. Such conditions are the asymptotical independence of the resampling distribution from the distribution of the data, the existence of a limit of the resampling distribution, and the asymptotic equality between the distribution of the test statistic and the conditional resampling distribution [71,112,113]. A simulation study revealed that the standard paired *t*-test controls the type I error level only in the presence of symmetric distributions. The bootstrap method, for small sample sizes, provides a liberal type I error rate. Conversely, the permutation test is shown to be accurate regardless of the level of dependence among paired samples, the sample sizes, and the skewness of the distributions. Under the alternative hypothesis, the permutation test was proved to be more powerful than the paired *t*-test, especially with skewed distributions [46].

Similarly to the multi-aspect comparative study for independent samples, hypothesis testing on mean and variance jointly considered is also a challenging area of methodological research in the paired sample case. In fact, in addition to location shift, a change in variability connected to the treatment might arise in paired data and could be investigated. In this regard, Brombin and Salmaso [71] introduced the multi-aspect permutation procedure for testing mean and variance with paired data. Specifically, the treatment effect on the first and second moment of the response, in paired data problems, was investigated. The null hypothesis consists of the equality of the moments of the pre-treatment and the post-treatment group. The solution is based on two partial permutation tests whose statistics are the between group-difference of the sample moments and on the nonparametric combination of such partial tests [31,71]. A comparative simulation study was carried out by [71] to investigate the performance of the proposed solution compared to those of other competitors as a function of the dependence between the paired responses. The competitors taken into account are Pitman's test based on Pearson's product-moment correlation coefficient, the jackknife test based on the log-ratio of the sample variances, and Spearman's rank correlation coefficient. The simulation results proved the good power behavior of the permutation test, especially in the preference of high correlation among pre-treatment and post-treatment observations. Moreover, the simulation showed that the power of the test is also good under normality and for small sample sizes.

Parametric ANOVA for repeated-measure designs also requires certain assumptions (such as normality, constant variance, and no correlation between units) to test the mean difference among groups. For such a problem, the permutation approach can be applied in order to provide a robust and flexible distribution-free solution. Higgins and Noble [114] deeply explained the split-plot analysis for repeated-measure experiments when the assumption of constant correlation is satisfied. They noted that in case of a violation of parametric ANOVA assumptions, obtaining an exact *p*-value is not possible, and a nonparametric test is necessary.

However, the application of the permutation approach needs careful consideration of what to permute. Indeed, in repeated-measure designs, "...*The permutations should correspond to the randomization*" [114] (p. 241). In other words, for experimental designs, the permutation is possible under the null hypothesis if the observations are exchangeable due

to the random assignment of the treatment to units. For instance, in case–control studies of repeated-measure designs, the row vectors of the observations are permuted among case and control groups. However, the permutation of the measurements within statistical units is not possible since the randomization does not apply [114]. In some repeated-measure designs, the null hypothesis may be the equality of means for different follow-up time points. Hence, in this case, under the null hypothesis, the observations within statistical units are exchangeable, and permutations of data within statistical units are appropriate. Although the applicability of permutations based on randomization is restricted to the experimental framework, for observational studies or when there is no randomization, some conditions, such as independence of observations, ensure exchangeability and therefore the possibility to apply the permutation method.

In the literature, if the permutation test is exact, the probability of rejecting the null hypothesis of no treatment effect when it is true does not exceed the pre-assigned level of significance [12,32]. In repeated-(M)ANOVA design or general mixed models [48], the permutation test could provide an exact or approximate *p*-value depending on the exchangeability condition under the null hypothesis. They used different permutations in repeated-measure designs to obtain exact and approximate results, such as permuting reduced residuals and modified residuals. According to them, for a balanced design, the reduced residuals (i.e., the residuals of the reduced model) could be permuted under the null hypothesis if they result in zero expectation after removing the effect of confounding factors. In other words, the observations of the residuals are permuted in a row-wise fashion rather than permuting within individuals. They provide a necessary and sufficient condition to define the reduced residuals for a mixed linear model or repeated measures.

In addition, in the case of the non-spherical distribution of residuals, an exact *p*-value is obtained by modifying the reduced residuals and removing the correlation within subjects. In repeated-measure design or mixed linear model, the random effect does not affect the principle of permuting the observations. In other words, under the null hypothesis, the outcome variable is equal to the measurement error term and the random error term. As a result, since the random error term has zero expectation under the null hypothesis, it does not affect the permutations. In [48], a simulation study was carried out to compare three types of permutations (raw data, reduced residuals, and modified residuals) with the counterpart of the F test for a repeated-measures ANOVA under different distributions.

As a result, the simulations provided an approximate type I error rate close to the nominal *α* level for all permutation methods. Furthermore, the power behavior of the permutation test based on permuting the reduced residuals, the raw data, and modified residuals provided similar power performance regardless of non-normal distributions. Moreover, the power behavior of the permutation test was good over the standard F test for the repeated ANOVA, especially for non-normal distributions. However, the modified residuals permutation method needs to estimate the eigenvalues at each permutation to obtain an exact *p*-value, and it might increase the computation time.

Dealing with a comparative study with repeated measures and multivariate outcomes, Dragset [115] highlight the importance of taking into account not only the dependency between the components of the multivariate response but also that of the variables of interest over time, measured through the so-called intraclass correlations. Thus, a parametric approach for longitudinal analysis should consider and estimate such intraclass correlations. However, a steady correlation in follow-up studies is unlikely to hold since observations within statistical units are highly related over time. For this reason, assuming constant correlation within individuals seems implausible.

Friedrich et al. [116] proposed a permutation test, regardless of the time dependence, for testing treatment effects in multivariate repeated-measures or longitudinal MANOVA settings. The permutation procedure in their study consists of permuting the pooled data regardless of the dependency within subjects. They considered the permutation version of the Wald test on the difference in means between groups over time for the multivariate longitudinal dataset. They proved that the permutation distributions of the Wald-type

and of the Studentized Wald-type statistic are approximately the same distribution under the null hypothesis. Furthermore, through a simulation study, they showed that their permutation version of the Wald-type test and of the ANOVA maintained good asymptotic power behavior, but, above all, it controlled the type I error rate under the null hypothesis for small and moderate sample sizes. Such a property holds under various multivariate (even non-normal) distributions and covariance structures.

4. Permutation Tests for Regression Models

In the parametric approach, inference on the regression coefficients may be problematic, if not impossible, in the presence of several explanatory variables, small samples, or violation of some assumptions about the distribution of the model errors. For example, parametric tests on model parameters may not be performant when the assumption of uncorrelated normal errors does not hold. Interesting solutions to face such difficulties can be found within the methodology of maximum entropy estimation [117].

In the literature of nonparametric regression analysis, the vast majority of research has been dedicated to estimation problems for linear models and panel or longitudinal models when the underlying distribution of the error terms is unknown (see [118–121] for more details). One of the most popular methods of nonparametric estimation is the kernel smoother, which does not require the functional form of the regression model to be known [121]. However, nonparametric regression mainly relies on detecting the bandwidth parameter, and this is a computationally intensive problem. There is a variety of bandwidth calculation methods, such as the rule of thumb, plug-in methods, cross-validation, and bootstrap method [121]. Several R packages are dedicated to such a problem. They propose techniques to reduce the computationally intensive nature of bandwidth computation. Among the most diffused of these packages, we mention the np package [121], the kernel smoother package [122], and the npregfast package [123]. The extension to multivariate models might be difficult because it is demanding from the computational point of view.

The literature dedicated to permutation tests for regression models is less extensive but very important and with notable contributions. A non-exhaustive list of contributions on this topic includes [1,124–126]. However, some distinctions about the adopted approach must be made. For instance, some authors used permutations of raw data [81,126,127], while other researchers used the permutation of residuals of the full model [128,129]. Moreover, some of the researchers permuted the residuals of the reduced model and found approximate permutation p-values [130,131]. Some permutation approaches are based on permutations of the vector of observed values of the dependent variable, but permuting the design matrix is more frequently considered.

The inferential results based on permuting the raw data, the residuals of the full model, and the residuals of the reduced model provided similar conclusions, even if, for skewed data and in the presence of outliers, the performance of the various methods may differ [126,132]. Nonetheless, extensive simulation studies were performed to compare the power of different permutation approaches. According to such studies, the most accurate and reliable permutation test, which controls the type I error rate, was obtained using the Freedman–Lane permutation method [1,125,126].

4.1. Test on the Regression Coefficients

In some practical applications, testing the significance of regression coefficients is quite common, since the goal of the study is to detect a sub-set of effective explanatory variables. However, in multiple linear regression analysis, the permutation solution to such a problem is a complex task since under H_0 exchangeability does not apply, unless the effect of associated nuisance variables is eliminated.

If the test concerns all the explanatory variables, residuals are exchangeable under the null hypothesis, and the test on the whole regression coefficients is possible. In this case, we can obtain an exact permutation test. However, computing an exact p-value might be difficult in the case of the significance test on partial regression coefficients. An

approximate permutation solution is possible by estimating the nuisance parameters (non-tested regression coefficients) and permuting the residuals of the reduced model, similar to the Freedman–Lane approach [126].

In a multiple linear regression analysis, permutation partial tests on regression coefficients have been presented by several authors [81,128,131,133–138]. Before describing the most interesting proposals in this area, let us consider, for the moment, the easier case of the simple linear regression model, with only one explanatory variable. The model is as follows:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

with $i = 1, 2, \ldots, n$. The errors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are exchangeable random variables with a null mean and constant variance, i.e., $E[\varepsilon_i] = 0$ and $Var[\varepsilon_i] = E[\varepsilon_i^2] = \sigma^2 \ \forall i$. To test the null hypothesis $H_0 : \beta_1 = 0$ versus the alternative hypothesis $H_1 : \beta_1 \neq 0$, a suitable test statistic is the square of the Pearson linear correlation index:

$$T_n = \frac{\left[\sum_{i=1}^n (Y_i - \overline{Y})(x_i - \overline{x})\right]^2}{\sum_{i=1}^n (Y_i - \overline{Y})^2 \sum_{i=1}^n (x_i - \overline{x})^2} \, ,$$

where $\overline{Y} = \sum_i Y_i / n$ and $\overline{x} = \sum_i x_i / n$.

Conditional on the n observed pairs of values (x_i, y_i), $i = 1, 2, \ldots n$, all the possible $n!$ matchings between the values of the dependent variable $y_1, y_2, \ldots, y_n$, and the values of the predictor $x_1, x_2, \ldots, x_n$ observed on the n statistical units are equi-probable. Hence, the null distribution of the test statistic can be obtained by permuting the values $x_1, x_2, \ldots, x_n$ while keeping fixed the vector $(y_1, y_2, \ldots, y_n)'$ (or vice versa). The observed value of the test statistic is

$$t_n^{(0)} = \frac{\left[\sum_{i=1}^n (y_i - \overline{y})(x_i - \overline{x})\right]^2}{\sum_{i=1}^n (y_i - \overline{y})^2 \sum_{i=1}^n (x_i - \overline{x})^2} \, ,$$

and the value of the test statistic in the null distribution corresponding to the bth permutation is

$$t_n^{*(b)} = \frac{\left[\sum_{i=1}^n (y_i - \overline{y})\left(x_i^{*(b)} - \overline{x}\right)\right]^2}{\sum_{i=1}^n (y_i - \overline{y})^2 \sum_{i=1}^n \left(x_i^{*(b)} - \overline{x}\right)^2} \, ,$$

where $x_i^{*(b)}$ is the value of the dependent variable matched with y_i in the bth permutation of the vector $(x_1, x_2, \ldots, x_n)'$. As usual, the p-value is the proportion of permutation values of the test statistic greater than or equal to $t_n^{(0)}$ and, if the p-value is less than or equal to the significance level α, the null hypothesis must be rejected in favor of the alternative hypothesis; otherwise, H_0 must be not rejected.

Let us now consider the multiple linear regression model with two explanatory variables:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \xi_i$$

with $i = 1, 2, \ldots, n$, where $\xi_1, \xi_2, \ldots, \xi_n$ are exchangeable random variables such that $E[\xi_i] = 0$ and $Var[\xi_i] = E[\xi_i^2] = \sigma^2 \ \forall i$. The system of hypotheses of interest is

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases} .$$

We are interested in investigating the effect of the first explanatory variable X_1 on the response Y, whereas X_2 takes the role of concomitant variable. To this aim, let us consider the residuals of the regression of Y on X_2

$$e_i = y_i - c_0 - c_1 x_{i2}$$

and the residuals of the regression of X_1 on X_2

$$u_i = x_{i1} - \delta_0 - \delta_1 x_{i2}.$$

An appropriate test statistic, to control the concomitant effect of X_2, is the square of the linear correlation index between such residuals. The quantities

$$c_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_{i2} - \bar{x}_2)}{\sum_{i=1}^{n}(x_{i2} - \bar{x}_2)^2} \text{ and } c_0 = \bar{y} - c_1\bar{x}_2$$

are estimates of the coefficients γ_1 and γ_0, respectively, in the linear regression model $Y_i = \gamma_0 + \gamma_1 x_{i2} + \varepsilon_i$, and similarly

$$\delta_1 = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\sum_{i=1}^{n}(x_{i2} - \bar{x}_2)^2} \text{ and } \delta_0 = \bar{x}_1 - \delta_1\bar{x}_2 .$$

The observed value of the test statistic is

$$t_n^{(0)} = \frac{\left(\sum_{i=1}^{n} e_i u_i\right)^2}{\sum_{i=1}^{n} e_i^2 \sum_{i=1}^{n} u_i^2} .$$

If γ_0 and γ_1 were known, an exact test could be obtained by permuting $e_1, e_2, \ldots, e_n$, and iteratively fitting new values of the response, using such values to re-estimate as γ_0 and γ_1 and using such estimates to re-compute the residuals according to the following procedure:

1. Use $y_1, y_2, \ldots, y_n$ to compute c_0 and c_1 as estimates of γ_0 and γ_1, respectively.
2. Compute the residuals $e_i = y_i - c_0 - c_1 x_{i2}$ with $i = 1, 2, \ldots, n$.
3. Permute the residuals $e_1, e_2, \ldots, e_n$, obtaining $e_1^{*(1)}, e_2^{*(1)}, \ldots, e_n^{*(1)}$.
4. Compute new values of the dependent variable as a function of the permuted residuals $\widetilde{y}_i^{(1)} = \gamma_0 + \gamma_1 x_{i2} + e_i^{*(1)}$ with $i = 1, 2, \ldots, n$.
5. Use $\widetilde{y}_1^{(1)}, \widetilde{y}_2^{(1)}, \ldots, \widetilde{y}_n^{(1)}$, to compute $c_0^{(1)}$ and $c_1^{(1)}$ as estimates of γ_0 and γ_1, respectively.
6. Compute the residuals $\widetilde{e}_i^{(1)} = \widetilde{y}_i^{(1)} - c_0^{(1)} - c_1^{(1)} x_{i2}$, $i = 1, 2, \ldots, n$.
7. Repeat steps from 3 to 6 B times so that, after the bth permutation of $e_1, e_2, \ldots, e_n$

    a. You obtain the permuted residuals $e_1^{*(b)}, e_2^{*(b)}, \ldots, e_n^{*(b)}$;
    b. Use them, γ_0 and γ_1 to compute the values $\widetilde{y}_1^{(b)}, \widetilde{y}_2^{(b)}, \ldots, \widetilde{y}_n^{(b)}$;
    c. Use such values to re-estimate γ_0 and γ_1 obtaining $c_0^{(b)}$ and $c_1^{(b)}$;
    d. Finally, re-compute the residuals $\widetilde{e}_i^{(b)} = \widetilde{y}_i^{(b)} - c_0^{(b)} - c_1^{(b)} x_{i2}$, with $i = 1, 2, \ldots, n$ and $b = 1, 2, \ldots, B$.

The value of the test statistic in the null distribution corresponding to the bth permutation is

$$t_n^{(b)} = \frac{\left(\sum_{i=1}^{n} \widetilde{e}_i^{(b)} u_i\right)^2}{\sum_{i=1}^{n} \left(\widetilde{e}_i^{(b)}\right)^2 \sum_{i=1}^{n} u_i^2} .$$

Since γ_0 and γ_1 are unknown, the exact test cannot be carried out in practice, and alternative procedures must be applied. In a proposal by Freedman and Lane [131], at step 4, γ_0 and γ_1 are replaced with c_0 and c_1, their estimates in the observed dataset. Thus, according to this proposal, after the bth permutation of the values $e_1, e_2, \ldots, e_n$, the permuted residuals $e_1^{*(b)}, e_2^{*(b)}, \ldots, e_n^{*(b)}$ are used to compute the values

$$\widetilde{y}_i^{(b)} = c_0 + c_1 x_{i2} + e_i^{*(b)}$$

and these values are used to re-estimate γ_0 and γ_1, obtaining $c_0^{(b)}$ and $c_1^{(b)}$. Such estimates, in turn, are used to recalculate the residuals:

$$\widehat{e}_i^{(b)} = \widetilde{y}_i^{(b)} - c_0^{(b)} - c_1^{(b)} x_{i2}$$

The value of the test statistic in the null distribution corresponding to the bth permutation is

$$t_n^{(b)} = \frac{\left(\sum_{i=1}^n \widehat{e}_i^{(b)} u_i\right)^2}{\sum_{i=1}^n \left(\widehat{e}_i^{(b)}\right)^2 \sum_{i=1}^n u_i^2} \ .$$

The method proposed by Kennedy [138] is simpler because, instead of re-computing the values of the dependent variable and the residuals, it directly uses the permuted residuals $e_1^{*(b)}, e_2^{*(b)}, \ldots, e_n^{*(b)}$ in the calculation of the test statistic. Thus, at the bth permutation, we obtain

$$t_n^{(b)} = \frac{\left(\sum_{i=1}^n e_i^{*(b)} u_i\right)^2}{\sum_{i=1}^n \left(e_i^{*(b)}\right)^2 \sum_{i=1}^n u_i^2} \ .$$

Manly [81] proposed an alternative approach that differs from that of Freedman and Lane because it is based on the permutation of $y_1, y_2, \ldots, y_n$ instead of the permutation of residuals. Thus, after the bh permutation, the permuted values $y_1^{*(b)}, y_2^{*(b)}, \ldots, y_n^{*(b)}$ are used to compute the estimates $c_0^{(b)}$, $c_1^{(b)}$ and the residuals

$$\widehat{e}_i^{(b)} = y_i^{*(b)} - c_0^{(b)} - c_1^{(b)} x_{i2} \ ,$$

with $i = 1, 2, \ldots, n$. The value of the test statistic I, the null distribution corresponding to the bth permutation, is

$$t_n^{(b)} = \frac{\left(\sum_{i=1}^n \widehat{e}_i^{(b)} u_i\right)^2}{\sum_{i=1}^n \left(\widehat{e}_i^{(b)}\right)^2 \sum_{i=1}^n u_i^2} \ .$$

Another interesting method proposed by terBraak [128] is based on the permutation of the residuals of the full model $w_i = y_i - b_0 - b_1 x_{i1} - b_2 x_{i2}$. After the bth permutation, the permuted values $w_1^{*(b)}, w_2^{*(b)}, \ldots, w_n^{*(b)}$ are use to compute the new values:

$$\widehat{y}_i^{(b)} = b_0 + b_1 x_{i1} + b_2 x_{i2} + w_i^{*(b)}.$$

Such values are used to re-estimate the coefficients of the model, and the re-estimated coefficients $b_0^{(b)}, b_1^{(b)}$ and $b_2^{(b)}$ are used to re-compute the residuals

$$\widetilde{w}_i^{(b)} = \widehat{y}_i^{(b)} - b_0^{(b)} - b_1^{(b)} x_{i1} - b_2^{(b)} x_{i2} \ .$$

The value of the terBraak test statistic corresponding to the bth permutation is

$$t_n^{(b)} = \frac{\left(\sum_{i=1}^n \widetilde{w}_i^{(b)} u_i\right)^2}{\sum_{i=1}^n \left(\widetilde{w}_i^{(b)}\right)^2 \sum_{i=1}^n u_i^2} \ .$$

All these methods can be easily extended to the case of more than one variable under test and/or more than one concomitant variable. One advantage of this permutation test is that it can also be applied to one-sided alternative hypotheses (by considering the correlation coefficient instead of its square).

Anderson and Robinson carried out a comparative study to investigate the performance (power behavior) of different tests on a subset of regression coefficients in a multiple linear regression analysis [126]. They proved that the null permutation distribution of such a test, obtained from different permutation approaches, by permuting raw data, residuals of the reduced model, and residuals of the full model, approximates the normal distribution under the null hypothesis, regardless of the permutation approach. The simulation results revealed that the Kennedy permutation method inflated the type I error rate, especially for small sample sizes. Conversely, the Freedman–Lane permutation method provided a rejection rate similar to α under the null hypothesis, regardless of the number of predictor variables, the sample size, and the error distribution. However, the proposed methods only work for linear models.

One of the most interesting applications of permutation tests in generalized linear models concerns problems with confounding factors [125]. When the significance of regression coefficient estimates is tested, the confounding effect of covariates must be eliminated through a suitable permutation strategy. Hence, the test statistic is computed as a function of the observed data of the predictors. However, since the values of the regression coefficients of the confounding factors are unknown, the permutation solution provides approximate results. In this framework, the design matrix is partitioned into two parts: predictors and confounding factors. Accordingly, the vector of regression coefficients is also partitioned. The authors explained how the exchangeability condition is satisfied under the null hypothesis (null effects of the predictors under test) for different error term structures, such as independent symmetric errors, exchangeable errors, and block-dependent error terms. In the case of dependency of the error terms due to blocks, the exchangeability could be applied by permuting units within blocks or exchanging the whole blocks (see the formal reasoning in [125]).

The permutation test statistic could also be based on the estimated coefficients [125]. However, the authors recommended a pivotal test statistic similar to Anderson's, such as the F statistic, t-test statistic, Pearson's correlation coefficient, and coefficient of determination. Furthermore, they considered an extensive simulation study to compare the power behavior of different permutation methods with the parametric method based on the F test under different scenarios, such as skewed distributions, balanced and unbalanced sample sizes, equal and unequal variances, and even the presence of outliers. The permutation approach to account for the outliers is interesting. Under the null hypothesis that the model coefficients are zero, the Still–White and Kennedy solutions are conservative, especially for small sample sizes. On the other hand, Freedman–Lane and Smith's proposals seem to perform well under the null hypothesis, respecting the nominal level α. Moreover, the power behavior of the permutation test based on the Freedman–Lane approach under the alternative hypothesis outperforms those of other permutation methods and the parametric F test regardless of sample sizes, error distributions, and other conditions.

Apart from the violation of assumptions, in some cases, the parametric tests are not feasible or not suitable in the presence of outliers, with small sample sizes, mixed error distributions, and nonrandom sampling. To overcome these problems, the authors of [139] proposed an R package called lmPerm, which is based on the permutation approach, for computing the p-values of the tests on the significance of regression coefficients. This package is also suitable for analyzing polynomial models and multivariate responses. They noted that permuting rows of observations is possible even in the presence of interaction effects. They carried out a simulation study to investigate the power behavior of the permutation test based on the lmp function and the standard F test. The simulations revealed that the F test is conservative and less powerful for small sample sizes, especially under non-normal distributions. Regarding the outliers, the power of the parametric F test decreases as the percentage of outliers increases and the sample size decreases. The permutation test is performant even with a high percentage of outliers and small sample sizes. In the case of heterogeneous variances, the two methods have the same power performance. However, a limitation of the permutation method is that it is computationally

intensive when an exact permutation test is considered (all permutations of the data are taken into account). To overcome this problem, the authors of [139] proposed other methods based on stopping the sampling upon meeting the given criteria, for example, when the estimated standard deviation of *p*-values is less than 0.1. However, the estimation of standard deviation may not be convenient for some reasons, and it is better to approximate the permutation distributions via the conditional Monte Carlo procedure.

4.2. Tests on Models for Longitudinal Data

Due to heteroscedastic and autocorrelated errors, estimating and testing the significance of coefficients of the panel, longitudinal, and mixed-effects models using standard parametric methods may be problematic unless strong assumptions are introduced in the testing procedure. In contrast, the permutation approach provides approximate but flexible solutions for such problems [140]. However, sometimes, exact results based on the permutation approach for autocorrelated error terms might be challenging since the exchangeability condition may not work in this circumstance.

On the other hand, in the case of non-autocorrelated error terms, for instance, for the general fixed-effect model, an exact permutation solution can be obtained, even in the presence of heteroscedastic errors. Basso and Finos, in 2012, proposed an exact multivariate permutation solution for mixed variables within the fixed-effect framework, even with heteroscedasticity [141]. Their permutation solution could account for the model misspecification.

One of the advantages of the permutation tests is their robustness to the departure from the error normality and their capacity to consider the multivariate responses' dependence structure. Their power behavior under different scenarios, by considering different multicollinearity levels, different sample sizes, different types, and intensities of dependence between responses, with both equal or unequal variances and balanced or unbalanced designs, has been proved through several extended simulation studies.

In the linear mixed model, one concern is the inclusion or exclusion of the random component, which accounts for the individual heterogeneity in the longitudinal study. The test on the random effects is crucial. In what follows, the standard parametric tests such as the score test, Wald test, and likelihood test might not be applicable under some conditions, for instance, with small sample sizes [142]. In addition, as Lee and Braun explained, the parametric tests do not have some approximate distributions under the null hypothesis, such as a chi-square distribution, since the variance component of the random effect is zero under the null hypothesis. Moreover, the significance test of random effect coefficients, using a parametric test, is not robust to the distribution of the random component.

However, the permutation approach can be applied to test the significance of random effect parameters. Lee and Braun [142] proposed a permutation test statistic based on the best linear unbiased predictions (BLUPs) and the restricted likelihood ratio test for testing the random effect coefficients in the linear mixed model (LMM). When they are concerned with single random-effect parameters, the problem is equivalent to testing whether the variance component is zero or not. Under the hypothesis of null variance, they proposed the permutation of the residuals. Investigating the significance of some random-effect parameters, the effect of nuisance components was eliminated by taking the appropriate weight matrix. The simulation results revealed that the permutation test based on BLUP and the restricted likelihood ratio test respected α under the null hypothesis. In contrast, the parametric test based on the asymptotic likelihood ratio test in some cases, especially with small sample sizes, is anti-conservative because of the inflated type I error rate. The power behavior under the hypothesis of non-zero variance component or non-zero random effect revealed that the permutation test is better than the asymptotic likelihood ratio test, regardless of error distributions and sample sizes. Nonetheless, the BLUP permutation test statistic is not a reasonable solution when we have two or more random coefficients. Moreover, a simultaneous testing method was proposed based on the restricted likelihood ratio test approach that requires estimating the variance component at each permutation.

In addition, hypothesis testing on generalized mixed-model coefficients in the case of a multivariate response is another challenging problem within the standard parametric approach based on the likelihood ratio test [140]. The challenge might arise due to the dependency among the response variables. Finos and Basso [140] proposed a permutation strategy for testing the significance of vector generalized linear mixed-model (VGLM) parameters in the presence of factors regardless of the known multivariate distributions. In particular, they proposed a permutation method to test the significance of the interaction effect for the family of VGLM. The proposed method can be applied for testing coefficients of linear models, nonlinear models, generalized linear models, and generalized mixed models, whereas the permutation solution of [126] could not be applied to VGLM parameters.

The permutation solution is flexible and applicable to any model (linear or nonlinear model, with a continuous or binary response). For example, a combined permutation test could test the significance of the estimates of the coefficients of both fixed effects and random effects of a general (linear) mixed model. When testing for the interaction effects with the permutation approach, some of the regression coefficients are nuisance parameters, and their confounding effect can be eliminated through suitable residualization. The exact permutation test is possible when the exchangeability condition is fulfilled where we have two or more factors so that an approximate significance test is possible. Such a condition is not satisfied when not all the regression coefficients are being tested. The mentioned residualization allows obtaining approximation permutation tests. Finos and Basso, in 2014, carried out simulations for both univariate and multivariate two-sample tests, under different distributions, to compare the power behavior of the exact permutation test, the asymptotic permutation test, and the likelihood ratio test [140]. They found that both the exact and asymptotic permutation tests provide rejection rates closer to α under H_0 for balanced samples. In comparison, the likelihood test is too conservative. Under H_1, the power of the permutation test is greater than that of the likelihood ratio test, regardless of the error variance and underlying probability law.

5. Permutation Approach and Missing Data

One of the most challenging problems in statistical data analysis, and in particular in hypothesis testing, is the missingness of observations. In particular, this is very common in follow-up studies and, therefore, in longitudinal data, panel data, survival analysis, and time-series data. However, the missingness issue is not rare, even in cross-sectional studies. Missing observations can be the consequence of censored as well as truncated data [143].

In the literature, parametric solutions for missing data problems are mainly based on the stringent assumption of missingness completely at random (MCAR) [31]. Under such a condition, the probability of missingness is unrelated to treatments, the censoring process, and observed values of the outcome variables [144–146]. Moreover, when the missingness is not at random (MNAR), the parametric methods are not flexible and effective in dealing with missing data. Meanwhile, distribution-free testing methods such as permutation tests overcome the problem of missing data conveniently and efficiently [27,147,148].

In paired samples with missing data, the commonly used nonparametric log-rank test might not be a general solution in some situations, such as small sample sizes. For this situation, some researchers proposed the use of the permutation approach [147]. They used a permutation test for a comparative study of pre-treatment and post-treatment effects for incomplete paired data. First, they considered two test statistics computed from paired and incomplete data using the mean differences. Then, they took the linear combination of the two statistics to compute the overall test statistic of the permutation test. Finally, they simulated data under normal and skewed distributions to investigate the power behavior of the proposed permutation test by comparing it with the standard paired *t*-test and the Bhoj *t*-test for paired and unpaired data [149]. The simulation results revealed that the parametric paired *t*-test does not control the type I error rate. On the contrary, the permutation test and the Bhoj *t*-test respect the nominal level and are more powerful regardless of the underlying distribution and sample size. However, the proposed permutation method requires the

estimation of the combination weights of the two test statistics. An alternative solution could be based on the nonparametric combination approach [31].

The right censoring is quite common in survival analysis. For such a problem, the permutation method is appropriate [27,150,151]. An interesting permutation test for comparing the survival curves of two independent samples of patients in the presence of right-censored data was proposed by [148]. The distribution of right censoring under the alternative hypothesis might be equal or unequal in the two or more independent samples. If the distributions of the right censoring are equal, the censoring occurs with the same probability in all the samples, and it can be ignorable [148]. However, unequal censoring distributions under the alternative hypothesis may affect the reliability of the inferential results if not taken into account. The solution proposed by [148] considers the null composite hypothesis of equality of survival curves for the time to the event and censored values. If n_o is the number of observed data about the time to event and n_c is the number of missing data (right censoring) in the pooled sample, with $n = n_o + n_c$, then the comparison between the survival curves is broken down into n_o sub-problems, making the test multivariate. Each sub-problem corresponds to an observed time to event t_i, with $i = 1, 2, \ldots, n_o$ and consists of the comparison of the joint probability of no censoring and of surviving at time t_i between the two or more samples. Given that exchangeability under H_0 is satisfied, the permutation approach can be applied, and the null distribution of the multivariate test statistic can be obtained by randomly permuting the rows of the dataset, i.e., randomly reassigning units to samples. To solve the overall problem, the multivariate statistic is transformed into a univariate one through the application of a nonparametric combination. Such a solution is valid and effective for any kind of missing data mechanism and, consequently, for both equal-censoring and unequal-censoring processes [31]. To prove the good performance of the proposed permutation method, the authors of [148] carried out an interesting simulation study by comparing the proposed solution with the log-rank test and the weighted Kaplan–Meier test for both equal-censoring and unequal-censoring cases. The simulation results proved that the permutation test is exact and robust, especially for small sample sizes. The permutation test is performant when facing the censoring issue: it does not assume a specific underlying distribution and models the dependency structure between the survival time and censoring.

6. Conclusions

The use of the permutation approach for testing hypotheses in many fields has been increasing. It is a valid statistical tool for analyzing complex data when parametric tests are not applicable or not powerful due to the violation of some assumptions. Permutation tests also tackle the Behrens–Fisher problem. The use of pivotal statistics is neither necessary nor sufficient for a permutation test. Moreover, in the permutation approach, the mild assumption of exchangeability is a sufficient condition, making such an approach flexible and robust. The permutation method can be applied in the paired sample case by permuting the signs of the differences. Meanwhile, in the presence of independent samples, the permutations should correspond to the reallocation of the dataset rows (units) to the samples [114]. Furthermore, permutation tests can also be applied to tests of hypotheses on the coefficients of linear models, nonlinear models, general linear models, mixed models, and general mixed models [82]. Furthermore, the permutation tests are suitable for directional hypotheses, umbrella alternatives, location-scale problems, multivariate tests, and other complex problems. Finally, suitable and valid solutions to testing problems with missing data, including when the missingness process is MNAR (Missing Not At Random), are available in the literature [11].

The evidence of the growing interest in this methodology being applied in various empirical disciplines is confirmed by recent literature. The following references regarding scientifically relevant works represent an important but not exhaustive list. Regarding psychometrics and biostatistics, a recent interesting work concerning a regression analysis to investigate mental health in the COVID-19 lockdown period is proposed in [127]. Func-

tional connectivity in adolescents with depression is studied in [152]. Work-related stress and organizational well-being are the subjects of a work dedicated to multivariate analysis of variance in the presence of big data [99]. A permutation test on the Spearman correlation coefficient is proposed in [153] and applied to two biomedical case studies related to breast cancer data and prostate-specific antigen values.

Permutation tests are also very popular in the field of neuroimaging. In this framework, a permutation-based method for the analysis of partial least squares is studied in [154]. Another interesting work about neuroimaging concerns spatial-extent inference for testing variance components is [155].

Permutation-based inference is increasingly being applied in the study of reliability in general and in the field of structural reliability and the reliability of engineering systems in particular [156–159]. A permutation test for a regression analysis is applied in [160] to study the reliability of solid-state drivers of computers.

In applied economics, recent complex empirical works solved via permutation solutions include [161], dedicated to a multi-sample test for non-monotonic hypotheses, and [162], concerning one-sided hypotheses in the presence of multivariate binary data. The case studies in both these works are focused on the circular economy. Multi-sample comparative evaluations on customer satisfaction are addressed in [163] and in [164]. The impact of public policy incentives on the propensity to adopt Industry 4.0 technological innovations is faced in [165] through the use of an innovative permutation test on the goodness-of-fit of a generalized linear model (GLM) for count data.

The number of contributions in statistics applied to sports data, a field already flourishing in the United States and other countries, is also growing in Europe. Recent works dedicated to sports analytics, in which permutation tests have been applied to basketball data, include [166,167].

All these applications confirm the utility and relevance of the permutation approach in the modern sciences and, consequently, its scientific impact in the real world. Despite the undoubted usefulness and advantages that derive from the application of permutation tests, in particular, those presented in the paper, there remain some limitations of conditional inference and some areas of research where further developments are possible and can be the subject of future works. For example, the computational problem has been successfully scaled down with the practice of random permutations, thanks to the power and efficiency of modern computers. However, some permutation procedures, in particular for big data problems, are still computationally demanding.

Another problem with permutation tests is the choice of the significance level in the presence of very small samples. There is a minimum α value compatible with the cardinality of the permutation space that depends on the sample sizes. In the two-independent-sample exact test, the minimum possible p-value, i.e., minimum proportion of values of the test statistic greater than or equal to the observed one, is $\binom{n}{n_1}^{-1}$. For example, if $n = 8$ and $n_1 = n_2 = 4$, then such a minimum value is

$$\binom{8}{4}^{-1} = 70^{-1} = 0.0143$$

Therefore, setting a significance level α of 0.01 makes no sense because, in this case, the p-value cannot be less than or equal to α.

In general, even if permutation tests are more flexible and powerful than parametric methods, especially when the conditions for the application of the latter are not met, and performant when the conditions are satisfied, when it happens, the parametric tests are preferable. Another limitation of permutation tests concerns multivariate problems. The power is a decreasing function of the correlation between the components of the multivariate response variable. When such a correlation is high, the probability of rejecting the null hypothesis, when it is false and must be rejected, tends to be low. As said, in

multiple tests, typical, for example, of the combined permutation test approach, attributing the eventual significance of the overall test to any of the partial tests requires adjusting the partial *p*-values. However, the most commonly used adjustment methods are too conservative, and most of the partial null hypotheses are not rejected, even if the original unadjusted *p*-values are small. The development of new adjustment methods that solve this problem is one of the interesting challenges in multiple permutation testing research. Finding new combining functions more performant in terms of power or, in general, new powerful combined permutation tests for complex problems is also an interesting, useful, and complex line of research.

Finally, a reflection on the comparison between permutation tests and the other main non-parametric methods for hypothesis testing, i.e., the bootstrap and rank tests, is necessary. As said, the permutation and the bootstrap approach are both considered resampling methods. In the two-independent-sample problem, a permutation, i.e., a reassignment of units (rows of the dataset) to the samples, is equivalent to the sample without repetition n_1 units from the n total units and assigning them to sample 1. In this sense, we speak of "resampling" in the permutation approach. In the case of bootstrapping, instead of permuting, resampling with replacement is carried out. In each resampling, n_1 units selected with repetition from the total n units are reassigned to sample 1, and n_2 units selected with repetition from the total n are reassigned to sample 2. Hence, the null distribution of the test statistic in permutation tests is obtained through resampling without repetition, and in bootstrap tests is obtained through resampling with repetition. As said, permutation tests require the condition of exchangeability to be satisfied in the null hypothesis and are usually more powerful than their bootstrap counterparts. The permutation approach can be used to compute confidence intervals but, to this aim, it requires more assumptions that are not always plausible. Conversely, the boostrap method is suitable for confidence intervals; thus, it is preferable to the permutation approach for interval estimation. Rank tests are, in general, less flexible than permutation tests because they require more assumptions. Furthermore, the distributions of the test statistics are tabulated for small samples. In practical applications, approximate distributions are often used, which are valid under asymptotic properties. Apart from the fact that for large samples, the performances of the three approaches tend to converge if the approximation of the distribution of a rank test is not good, the validity of the inferential results is compromised.

References

1. Hemerik, J.; Thoresen, M.; Finos, L. Permutation testing in high-dimensional linear models: An empirical investigation. *J. Stat. Comput. Simul.* **2021**, *91*, 897–914. [CrossRef]
2. Frossard, J.; Renaud, O. Permutation tests for regression, ANOVA, and comparison of signals: The permuco package. *J. Stat. Softw.* **2021**, *99*, 1–32. [CrossRef]
3. Ding, P.; Feller, A.; Miratrix, L. Randomization inference for treatment effect variation. *J. R. Stat. Soc. Ser. B* **2016**, *78*, 655–671. [CrossRef]
4. Bonnini, S.; Corain, L.; Marozzi, M.; Salmaso, L. *Nonparametric Hypothesis Testing: Rank and Permutation Methods with Applications in R*; Wiley: Chichester, UK, 2014.
5. Pesarin, F. *Multivariate Permutation Tests: With Applications in Biostatistics*; Wiley: Chichester, UK, 2001.

6. Anderson, M.J. Permutation tests for univariate or multivariate analysis of variance and regression. *Can. J. Fish. Aquat. Sci.* **2001**, *58*, 626–639. [CrossRef]
7. Fisher, R.A. Design of experiments. *Br. Med. J.* **1936**, *1*, 554. [CrossRef]
8. Flandin, G.; Friston, K.J. Analysis of family-wise error rates in statistical parametric mapping using random field theory. *Hum. Brain Mapp.* **2019**, *40*, 2052–2054. [CrossRef]
9. Lehmann, E.L.; Romano, J.P.; Casella, G. *Testing Statistical Hypotheses*; Springer: New York, NY, USA, 2005.
10. Westfall, P.H.; Young, S.S. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*; Wiley: Hoboken, NJ, USA, 1993.
11. Pitman, E.J.G. Significance tests which may be applied to samples from any populations. *Suppl. J. R. Stat. Soc.* **1937**, *4*, 119–130. [CrossRef]
12. Anderson, M.; Ter Braak, C. Permutation tests for multi-factorial analysis of variance. *J. Stat. Comput. Simul.* **2003**, *73*, 85–113. [CrossRef]
13. Christensen, W.F.; Zabriskie, B.N. When your permutation test is doomed to fail. *Am. Stat.* **2021**, *76*, 53–63. [CrossRef]
14. de Andrade, B.M.; de Gois, J.S.; Xavier, V.L.; Luna, A.S. Comparison of the performance of multiclass classifiers in chemical data: Addressing the problem of overfitting with the permutation test. *Chemom. Intell. Lab. Syst.* **2020**, *201*, 104013. [CrossRef]
15. Berrett, T.B.; Wang, Y.; Foygel Barber, R.; Samworth, R.J. The conditional permutation test for independence while controlling for confounders. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2020**, *82*, 175–197. [CrossRef]
16. Mo, Q.; Yao, H.; Cao, F.; Chang, Z.; Qin, C. Reversible data hiding in encrypted image based on block classification permutation. *Comput. Mater. Contin.* **2019**, *59*, 119–133. [CrossRef]
17. Rao, K.; Drikvandi, R.; Saville, B. Permutation and Bayesian tests for testing random effects in linear mixed-effects models. *Stat. Med.* **2019**, *38*, 5034–5047. [CrossRef]
18. Samuh, M.H.; Pesarin, F. Applications of conditional power function of two-sample permutation test. *Comput. Stat.* **2018**, *33*, 1847–1862. [CrossRef]
19. DiCiccio, C.J.; Romano, J.P. Robust permutation tests for correlation and regression coefficients. *J. Am. Stat. Assoc.* **2017**, *112*, 1211–1220. [CrossRef]
20. Gagnon-Bartsch, J.; Shem-Tov, Y. The classification permutation test: A nonparametric test for equality of multivariate distributions. *arXiv* **2016**, arXiv:1611.06408.
21. Filgueiras, P.R.; Alves, J.C.L.; Sad, C.M.S.; Castro, E.V.R.; Dias, J.C.M.; Poppi, R.J. Evaluation of trends in residuals of multivariate calibration models by permutation test. *Chemom. Intell. Lab. Syst.* **2014**, *133*, 33–41. [CrossRef]
22. Hahn, U. A studentized permutation test for the comparison of spatial point patterns. *J. Am. Stat. Assoc.* **2012**, *107*, 754–764. [CrossRef]
23. Basso, D.; Salmaso, L. A permutation test for umbrella alternatives. *Stat. Comput.* **2011**, *21*, 45–54. [CrossRef]
24. Phipson, B.; Smyth, G.K. Permutation P-values should never be zero: Calculating exact P-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.* **2010**, *9*, e1585. [CrossRef] [PubMed]
25. McIntee, E.; Viglino, E.; Kumor, S.; Rinke, C.; Ni, L.; Sigman, M.E. Non-parametric permutation test for the discrimination of float glass samples based on LIBS spectra. *J. Chemom.* **2010**, *24*, 312–319. [CrossRef]
26. Bonnini, S.; Corain, L.; Munaò, F.; Salmaso, L. Neurocognitive effects in welders exposed to aluminium: An application of the NPC test and NPC ranking methods. *Stat. Methods Appl.* **2006**, *15*, 191–208. [CrossRef]
27. Bonnini, S.; Salmaso, L.; Solari, A. Multivariate permutation tests for evaluating effectiveness of universities through the analysis of student dropouts. *Stat. Appl.* **2005**, *3*, 37–44.
28. Finos, L.; Salmaso, L. Nonparametric multi-focus analysis for categorical variables. *Commun. Stat.-Theory Methods* **2004**, *33*, 1931–1941. [CrossRef]
29. Corain, L.; Salmaso, L. An empirical study on new product development process by Nonparametric Combination (NPC) testing methodology and post-stratification. *Statistica* **2003**, *63*, 335–357.
30. Scheffé, H. *The Analysis of Variance*; Wiley: New York, NY, USA, 1959; pp. 25–50.
31. Pesarin, F.; Salmaso, L. *Permutation Tests for Complex Data: Theory, Applications and Software*; Wiley: Chichester, UK, 2010.
32. Janssen, A.; Thorsten, P. A Monte Carlo comparison of studentized bootstrap and permutation tests for heteroscedastic two-sample problems. *Comput. Stat.* **2005**, *20*, 369–383. [CrossRef]
33. Wang, R.; Wangli, X. An approximate randomization test for high-dimensional two-sample Behrens-Fisher problem under arbitrary covariances. *Biometrika* **2021**, *109*, 1117–1132. [CrossRef]
34. Smaga, L. Inference for General MANOVA Based on ANOVA-Type Statistic. In *Advanced Studies in Classification and Data Science*; Springer: Singapore, 2020; pp. 241–254.
35. Pauly, M.; Brunner, E.; Konietschke, F. Asymptotic permutation tests in general factorial designs. *J. R. Stat. Soc. Ser. B* **2015**, *77*, 461–473. [CrossRef]
36. RIC Chris, F.; Manly, B.F.J. Bootstrap calibration to improve the reliability of tests to compare sample means and variances. *Environmetr. Off. J. Int. Environmetr. Soc.* **2001**, *12*, 713–729.
37. Mielke, J.P.W.; Kenneth, J.B. Permutation tests for common locations among samples with unequal variances. *J. Educ. Stat.* **1994**, *19*, 217–236.

38. Wheeler, B.; Torchiano, M.; Torchiano, M.M. *Package 'lmPerm'. R Package*, Version 2.1.0; R Foundation for Statistical Computing: Vienna, Austria, 2016.
39. Achim, Z.; Wiel, M.A.; Hornik, K.; Hothorn, T. Implementing a class of permutation tests: The coin package. *J. Stat. Softw.* **2008**, *28*, 1–23.
40. Finos, L.; Finos, M.L.; Rcpp, I. *Package 'flip'*, Version 2.5.0; R Foundation for Statistical Computing: Vienna, Austria, 2011.
41. Myles, H.; Douglas, A. *Wolfe, and Eric Chicken. Nonparametric Statistical Methods*; Wiley: Hoboken, NJ, USA, 2013.
42. Peres-Neto, P.R.; Olden, J.D. Assessing the robustness of randomization tests: Examples from behavioural studies. *Anim. Behav.* **2001**, *61*, 79–86. [CrossRef]
43. Good, P.I. Extensions of the concept of exchangeability and their applications. *J. Mod. Appl. Stat. Methods* **2002**, *1*, 34. [CrossRef]
44. Spencer, L.M.; Schooley, M.W.; Anderson, L.A.; Kochtitzky, C.S.; DeGroff, A.S.; Devlin, H.M.; Mercer, S.L. Peer reviewed: Seeking best practices: A conceptual Framework for planning and improving evidence-based practices. *Prev. Chronic Dis.* **2013**, *10*, E207. [CrossRef] [PubMed]
45. Arboretti, R.; Brombin, C. Overview of NonParametric combination-based permutation tests for multivariate multi-sample problems. *Statistica* **2014**, *74*, 233–246.
46. Konietschke, F.; Pauly, M. Bootstrapping and permuting paired t-test type statistics. *Stat. Comput.* **2014**, *24*, 283–296. [CrossRef]
47. Anderson, M.J. Permutational multivariate analysis of variance. *Dep. Stat. Univ. Auckl. Auckl.* **2005**, *26*, 32–46.
48. Kherad-Pajouh, S.; Renaud, O. A general permutation approach for analyzing repeated measures ANOVA and mixed-model designs. *Stat. Pap.* **2015**, *56*, 947–967. [CrossRef]
49. Basso, D.; Pesarin, F.; Salmaso, L.; Solari, A. Synchronized permutation tests in two-way ANOVA. In *Permutation Tests for Stochastic Ordering and ANOVA*; Springer: New York, NY, USA, 2009; pp. 133–172.
50. Bertanha, M.; Chung, E.Y. Permutation tests at nonparametric rates. *J. Am. Stat. Assoc.* **2023**, *118*, 2833–2846. [CrossRef]
51. Giacalone, M.; Zirilli, A.; Cozzucoli, P.C.; Alibrandi, A. Bonferroni-Holm and permutation tests to compare health data: Methodological and applicative issues. *BMC Med. Res. Methodol.* **2018**, *18*, 81. [CrossRef]
52. Arboretti Giancristofaro, R.; Bonnini, S.; Corain, L.; Salmaso, L. Dependency and truncated forms of combinations in multivariate combination-based permutation tests and ordered categorical variables. *J. Stat. Comput. Simul.* **2016**, *86*, 3608–3619. [CrossRef]
53. Arboretti, G.R.; Bonnini, S. Moment-based multivariate permutation tests for ordinal categorical data. *J. Nonparametr. Stat.* **2008**, *20*, 383–393.
54. Hayasaka, S.; Nichols, T.E. Combining voxel intensity and cluster extent with permutation test framework. *Neuroimage* **2004**, *23*, 54–63. [CrossRef]
55. Akkartal, E.; Mendes, M.; Mendes, E. Determination of suitable permutation numbers in comparing independent group means: A monte carlo simulation study. *J. Sci. Ind. Res.* **2010**, *69*, 422–425.
56. Marozzi, M. Some notes on nonparametric inferences and permutation tests. *Metron Int. J. Stat.* **2002**, *60*, 139–151.
57. Wan, Y.; Cohen, J.; Guerra, R. A permutation test for the robust sib-pair linkage method. *Ann. Hum. Genet.* **1997**, *61*, 77–85.
58. Kennedy, P.E.; Cade, B.S. Randomization tests for multiple regression. *Commun. Stat.-Simul. Comput.* **1996**, *25*, 923–936. [CrossRef]
59. Marozzi, M. Some remarks about the number of permutations one should consider to perform a permutation test. *Statistica* **2004**, *64*, 193–201.
60. Hemerik, J.; Goeman, J. Exact testing with random permutations. *Test* **2018**, *27*, 811–825. [CrossRef]
61. Mardia, K.V. A non-parametric test for the bivariate two-sample location problem. *J. R. Stat. Soc. Ser. B Methodol.* **1967**, *29*, 320–342. [CrossRef]
62. Janssen, A. Studentized permutation tests for non-iid hypotheses and the generalized Behrens-Fisher problem. *Stat. Probab. Lett.* **1997**, *36*, 9–21. [CrossRef]
63. Neuhaus, G.; Zhu, L. Permutation tests for multivariate location problems. *J. Multivar. Anal.* **1999**, *69*, 167–192. [CrossRef]
64. Biswas, M.; Ghosh, A.K. A nonparametric two-sample test applicable to high dimensional data. *J. Multivar. Anal.* **2014**, *123*, 160–171. [CrossRef]
65. Pesarin, F.; Salmaso, L.; Carrozzo, E.; Arboretti, R. Union–intersection permutation solution for two-sample equivalence testing. *Stat. Comput.* **2016**, *26*, 693–701. [CrossRef]
66. Bertoluzzo, F.; Pesarin, F.; Salmaso, L. On multi-sided permutation tests. *Commun. Stat.-Simul. Comput.* **2013**, *42*, 1380–1390. [CrossRef]
67. Finch, W.H. Comparison of multivariate means across groups with ordinal dependent variables: A Monte Carlo simulation study. *Front. Appl. Math. Stat.* **2016**, *2*, 2. [CrossRef]
68. Pillar, V. How accurate and powerful are randomization tests in multivariate analysis of variance? *Community Ecol.* **2013**, *14*, 153–163. [CrossRef]
69. Székely, G.J.; Rizzo, M.L. Testing for equal distributions in high dimension. *InterStat* **2004**, *5*, 1249–1272.
70. Chung, E.Y.; Olivares, M. Permutation test for heterogeneous treatment effects with a nuisance parameter. *J. Econom.* **2021**, *225*, 148–174. [CrossRef]
71. Brombin, C.; Salmaso, L. Multi-aspect permutation tests in shape analysis with small sample size. *Comput. Stat. Data Anal.* **2009**, *53*, 3921–3931. [CrossRef]
72. Pillai, K.C.S. Some new test criteria in multivariate analysis. *Ann. Math. Stat.* **1955**, *26*, 117–121. [CrossRef]
73. Bartlett, M.S. The standard errors of discriminant function coefficients. *Suppl. J. R. Stat. Soc.* **1939**, *6*, 169–173. [CrossRef]

74. Nanda, D.N. Distribution of the sum of roots of a determinantal equation under a certain condition. *Ann. Math. Stat.* **1950**, *21*, 432–439. [CrossRef]
75. Neubert, K.; Brunner, E. A studentized permutation test for the non-parametric Behrens–Fisher problem. *Comput. Stat. Data Anal.* **2007**, *51*, 5192–5204. [CrossRef]
76. Welch, B.L. The significance of the difference between two means when the population variances are unequal. *Biometrika* **1938**, *29*, 350–362. [CrossRef]
77. Welch, B.L. The generalization of 'STUDENT'S' problem when several different population varlances are involved. *Biometrika* **1947**, *34*, 28–35. [CrossRef]
78. Welz, A.; Ruxton, G.D.; Neuhäuser, M. A non-parametric maximum test for the Behrens–Fisher problem. *J. Stat. Comput. Simul.* **2018**, *88*, 1336–1347. [CrossRef]
79. Proschan, M.A. A permutation approach to the Behrens-Fisher problem. *J. Korean Stat. Soc.* **2004**, *33*, 79–97.
80. Anderson, M.J.; Walsh, D.C.; Clarke, K.R.; Gorley, K.R.; Guerra-Castro, E. Some solutions to the multivariate Behrens–Fisher problem for dissimilarity-based analyses. *Aust. N. Z. J. Stat.* **2017**, *59*, 57–79. [CrossRef]
81. Manly, B.F.J. *Randomization, Bootstrap and Monte Carlo Methods in Biology: Texts in Statistical Science*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2018.
82. Mantel, N.; Valand, R.S. A technique of nonparametric multivariate analysis. *Biometrics* **1970**, *26*, 547–558. [CrossRef]
83. Westfall, P.H.; Randall, D.T.; Wolfinger, R.D. *Multiple Comparisons and Multiple Tests Using SAS*; SAS Institute: Arlington, VA, USA, 2011.
84. Kherad-Pajouh, S.; Renaud, O. An exact permutation method for testing any effect in balanced and unbalanced fixed effect ANOVA. *Comput. Stat. Data Anal.* **2010**, *54*, 1881–1893. [CrossRef]
85. Wilks, S.S. Certain generalizations in the analysis of variance. *Biometrika* **1932**, *24*, 471–494. [CrossRef]
86. Hotelling, H. Relations between two sets of variates. In *Breakthroughs in Statistics*; Springer: New York, NY, USA, 1992; pp. 162–190.
87. Corain, L.; Salmaso, L. Improving power of multivariate combination-based permutation tests. *Stat. Comput.* **2015**, *25*, 203–214. [CrossRef]
88. Basso, D.; Chiarandini, M.; Salmaso, L. Synchronized permutation tests in replicated I× J designs. *J. Stat. Plan. Inference* **2007**, *137*, 2564–2578. [CrossRef]
89. Arboretti, R.; Ceccato, R.; Corain, L.; Ronchi, F.; Salmaso, L. Multivariate small sample tests for two-way designs with applications to industrial statistics. *Stat. Pap.* **2018**, *59*, 1483–1503. [CrossRef]
90. Corain, L.; Salmaso, L. A critical review and a comparative study on conditional permutation tests for two-way ANOVA. *Commun. Stat.—Simul. Comput.* **2007**, *36*, 791–805. [CrossRef]
91. Hahn, S.; Salmaso, L. A comparison of different synchronized permutation approaches to testing effects in two-level two-factor unbalanced ANOVA designs. *Stat. Pap.* **2017**, *58*, 123–146. [CrossRef]
92. Warton, D.I.; Hudson, H.M. A MANOVA statistic is just as powerful as distance-based statistics, for multivariate abundances. *Ecology* **2004**, *85*, 858–874. [CrossRef]
93. Ricotta, C.; Podani, J. On some properties of the Bray-Curtis dissimilarity and their ecological meaning. *Ecol. Complex.* **2017**, *31*, 201–205. [CrossRef]
94. Brunner, E.; Puri, M.L. Nonparametric methods in factorial designs. *Stat. Pap.* **2001**, *42*, 1–52. [CrossRef]
95. Berry, K.J.; Mielke, P.W., Jr.; Mielke, H.W. The Fisher-Pitman permutation test: An attractive alternative to the F test. *Psychol. Rep.* **2002**, *90*, 495–502. [CrossRef]
96. Hahn, S.; Konietschke, F.; Salmaso, L. A comparison of efficient permutation tests for unbalanced ANOVA in two by two designs--and their behavior under heteroscedasticity. In *Topics in Statistical Simulation Research Papers from the 7th International Workshop on Statistical Simulation*; Springer: New York, NY, USA, 2013; pp. 257–269.
97. McArdle, B.H.; Anderson, M.J. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology* **2001**, *82*, 290–297. [CrossRef]
98. Gower, J.C.; Legendre, P. Metric and Euclidean properties of dissimilarity coefficients. *J. Classif.* **1986**, *3*, 5–48. [CrossRef]
99. Bonnini, S.; Assegie, G.M. Advances on Permutation Multivariate Analysis of Variance for big data. *Stat. Transit. New Ser.* **2022**, *23*, 163–183. [CrossRef]
100. Friedman, J.H.; Rafsky, L.C. Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *Ann. Stat.* **1979**, *7*, 697–717. [CrossRef]
101. Henze, N. A multivariate two-sample test based on the number of nearest neighbor type coincidences. *Ann. Stat.* **1988**, *16*, 772–783. [CrossRef]
102. Rosenbaum, P.R. An exact distribution-free test comparing two multivariate distributions based on adjacency. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 515–530. [CrossRef]
103. Arboretti, R.; Bonnini, S.; Corain, L.; Salmaso, L. A permutation approach for ranking of multivariate populations. *J. Multivar. Anal.* **2014**, *132*, 39–57. [CrossRef]
104. Bonnini, S. Testing for heterogeneity with categorical data: Permutation solution vs. bootstrap method. *Commun. Stat.-Theory Methods* **2014**, *43*, 906–917. [CrossRef]

105. Arboretti Giancristofaro, R.; Bonnini, S. Some new results on univariate and multivariate permutation tests for ordinal categorical variables under restricted alternatives. *Stat. Methods Appl.* **2009**, *18*, 221–236. [CrossRef]

106. Arboretti, G.R.; Bonnini, S. Permutation tests for heterogeneity comparisons in presence of categorical variables with application to university evaluation. *Metod. Zv.* **2007**, *4*, 21.

107. Gupta, S.S.; Panchapakesan, S. *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2002.

108. Arboretti, G.R. Multivariate permutation tests in genetics. *Statistica* **2002**, *62*, 681–694.

109. Kaiser, J. An exact and a Monte Carlo proposal to the Fisher–Pitman permutation tests for paired replicates and for independent samples. *Stata J.* **2007**, *7*, 402–412. [CrossRef]

110. Konietschke, F.; Bathke, A.C.; Solomon, W.H.; Pauly, M. Parametric and nonparametric bootstrap methods for general MANOVA. *J. Multivar. Anal.* **2015**, *140*, 291–301. [CrossRef]

111. Chung, E.Y.; Romano, J.P. Exact and asymptotically robust permutation tests. *Ann. Stat.* **2013**, *41*, 484–507. [CrossRef]

112. Markus, P. Discussion about the quality of F-ratio resampling tests for comparing variances. *Test* **2011**, *20*, 163–179.

113. Omelka, M.; Pauly, M. Testing equality of correlation coefficients in two populations via permutation methods. *J. Stat. Plan. Inference* **2012**, *142*, 1396–1406. [CrossRef]

114. Higgins, J.J.; Noble, W. A permutation test for a repeated measures design. *Conf. Appl. Stat. Agric.* **1993**. [CrossRef]

115. Dragset, I.G. Analysis of Longitudinal Data with Missing Values.: Methods and Applications in Medical Statistics. Master's Thesis, Institutt for Matematiske Fag, Trondheim, Norway, 2009.

116. Friedrich, S.; Brunner, E.; Pauly, M. Permuting longitudinal data despite all the dependencies. *arXiv* **2015**, arXiv:1509.05570.

117. Golan, M.; Judge, G.G.; Miller, D. *Maximum Entropy*; Wiley: Somerset, UK, 1996.

118. Mami, A.M.; Mansour, M.M.; Jaber, A.M. On using Multivariate Nonparamet Regression Methond. *Sch. J. Phys. Math. Stat.* **2020**, *7*, 11–29. [CrossRef]

119. Nathaniel, H.E. *Package 'Multiway'*, Version 1.0-6; R Foundation for Statistical Computing: Vienna, Austria, 2019.

120. Rodriguez-Poo, J.M.; Soberón, A. Nonparametric estimation of fixed effects panel data varying coefficient models. *J. Multivar. Anal.* **2015**, *133*, 95–122. [CrossRef]

121. Hayfield, T.; Racine, J.S. Nonparametric econometrics: The np package. *J. Stat. Softw.* **2008**, *27*, 1–32. [CrossRef]

122. R Core Team. *R: A Language and Environment for Statistical Computing*, Version 201; R Foundation for Statistical Computing: Vienna, Austria, 2013.

123. Sestelo, M.; Villanueva, N.M.; Meira-Machado, L.; Roca-Pardiñas, J. npregfast: An R package for nonparametric estimation and inference in life sciences. *J. Stat. Softw.* **2017**, *82*, 1–27. [CrossRef]

124. Nyblom, J. Permutation tests in linear regression. In *Modern Nonparametric, Robust and Multivariate Methods*; Springer: Cham, Switzerland, 2015; pp. 69–90.

125. Winkler, A.M.; Ridgway, G.R.; Webster, M.A.; Smith, S.M.; Nichols, T.E. Permutation inference for the general linear model. *Neuroimage* **2014**, *92*, 381–397. [CrossRef]

126. Anderson, M.J.; Robinson, J. Permutation tests for linear models. *Aust. N. Z. J. Stat.* **2001**, *43*, 75–88. [CrossRef]

127. Bonnini, S.; Borghesi, M. Relationship between Mental Health and Socio-Economic, Demographic and Environmental Factors in the COVID-19 Lockdown Period—A Multivariate Regression Analysis. *Mathematics* **2022**, *10*, 3237. [CrossRef]

128. Ter Braak, C.J.F. Permutation versus bootstrap significance tests in multiple regression and ANOVA. In *Bootstrapping and Related Techniques*; Springer: Berlin/Heidelberg, Germany, 1992; pp. 79–85.

129. Anderson, M.J.; Legendre, P. An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *J. Stat. Comput. Simul.* **1999**, *62*, 271–303. [CrossRef]

130. Edgington, E.S. Randomized single-subject experimental designs. *Behav. Res. Ther.* **1996**, *34*, 567–574. [CrossRef]

131. Freedman, D.; Lane, D. A nonstochastic interpretation of reported significance levels. *J. Bus. Econ. Stat.* **1983**, *1*, 292–298. [CrossRef]

132. Legendre, P. Comparison of permutation methods for the partial correlation and partial Mantel tests. *J. Stat. Comput. Simul.* **2000**, *67*, 37–73. [CrossRef]

133. Brown, B.M.; Maritz, J.S. Distribution-free methods in regression. *Aust. J. Stat.* **1982**, *24*, 318–331. [CrossRef]

134. Collins, M.F. A permutation test for planar regression. *Aust. J. Stat.* **1987**, *29*, 303–308. [CrossRef]

135. Oja, H. On permutation tests in multiple regression and analysis of covariance problems. *Aust. J. Stat.* **1987**, *29*, 91–100. [CrossRef]

136. Gail, M.H.; Tan, W.Y.; Piantadosi, S. Tests for no treatment effect in randomized clinical trials. *Biometrika* **1988**, *75*, 57–64. [CrossRef]

137. Welch, W.J. Construction of permutation tests. *J. Am. Stat. Assoc.* **1990**, *85*, 693–698. [CrossRef]

138. Kennedy, P.E. Randomization tests in econometrics. *J. Bus. Econ. Stat.* **1995**, *13*, 85–94. [CrossRef]

139. Wheeler, R.E.; Torchiano, M. Permutation tests for linear models in R. *Compr. R Arch. Netw.* **2010**, *1*, 1–36.

140. Finos, L.; Basso, D. Permutation tests for between-unit fixed effects in multivariate generalized linear mixed models. *Stat. Comput.* **2014**, *24*, 941–952. [CrossRef]

141. Basso, D.; Finos, L. Exact multivariate permutation tests for fixed effects in mixed-models. *Commun. Stat.-Theory Methods* **2012**, *41*, 2991–3001. [CrossRef]

142. Lee, O.E.; Braun, T.M. Permutation tests for random effects in linear mixed models. *Biometrics* **2012**, *68*, 486–493. [CrossRef]

143. Prinja, S.; Gupta, N.; Verma, R. Censoring in clinical trials: Review of survival analysis techniques. *Indian J. Community Med. Off. Publ. Indian Assoc. Prev. Soc. Med.* **2010**, *35*, 217. [CrossRef]
144. Little, R.J.; Rubin, D.B.; Zangeneh, S.Z. Conditions for ignoring the missing-data mechanism in likelihood inferences for parameter subsets. *J. Am. Stat. Assoc.* **2017**, *112*, 314–320. [CrossRef]
145. Little, R.J.A.; Rubin, D.B. *Statistical Analysis with Missing Data*; Wiley: Hoboken, NJ, USA, 2019.
146. Rubin, D.B. Inference and missing data. *Biometrika* **1976**, *63*, 581–592. [CrossRef]
147. Yu, D.; Lim, J.; Liang, F.; Kim, K.; Kim, B.S.; Jang, W. Permutation test for incomplete paired data with application to cDNA microarray data. *Comput. Stat. Data Anal.* **2012**, *56*, 510–521. [CrossRef]
148. Arboretti Giancristofaro, R.; Bolzan, M.; Campigotto, F.; Corain, L.; Salmaso, L. Combination-based permutation testing in survival analysis. *Quad. Di Stat.* **2010**, *12*, 21–44.
149. Bhoj, D.S. Testing equality of means of correlated variates with missing observations on both responses. *Biometrika* **1978**, *65*, 225–228. [CrossRef]
150. Brendel, M.; Janssen, A.; Mayer, C.D.; Pauly, M. Weighted logrank permutation tests for randomly right censored life science data. *Scand. J. Stat.* **2014**, *41*, 742–761. [CrossRef]
151. Ditzhaus, M.; Friedrich, S. More powerful logrank permutation tests for two-sample survival data. *J. Stat. Comput. Simul.* **2020**, *90*, 2209–2227. [CrossRef]
152. Camp, C.C.; Noble, S.; Scheinost, D.; Stringaris, A. Test-Retest Reliability of Functional Connectivity in Adolescents with Depression. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **2024**, *9*, 21–29. [CrossRef]
153. Yu, H.; Hutson, A.D. A robust Spearman correlation coefficient permutation test. *Commun. Stat.—Theory Methods* **2024**, *53*, 2141–2153. [CrossRef]
154. Danyluik, M.; Zeighami, Y.; Mukora, A.; Lepage, M.; Sha, J.; Joober, R. Evaluating permutation-based inference for Partial Least Squares analysis of neuroimaging data. *bioRxiv* **2024**. [CrossRef]
155. Pan, R.; Dickie, E.; Hawco, W.; Reid, N.; Voineskos, A.N.; Park, J.Y. Spatial extent inference for testing variance components in reliability and heritability studies. *Imaging Neurosci.* **2024**, *2*, 1–15. [CrossRef]
156. Hasofer, A.M.; Lind, N.C. Exact and invariant Second-Moment Code Format. *J. Eng. Mech. Div.* **1974**, *100*, 111–121. [CrossRef]
157. Zhang, L.W.; Dang, C.; Zhao, Y.G. An efficient method for accessing structural reliability indeces via power transforation family. *Reliab. Eng. Syst. Saf.* **2023**, *233*, 109097. [CrossRef]
158. Hosseinzadeh, S.; Bahari, M.R.; Abyani, M. Reliability assessment for pipelines corroded by longitudinally aligned defects. *Ocean Eng.* **2024**, *310*, 118625. [CrossRef]
159. Xu, L.H.; Mutze, T.; Konietschke, F.; Friede, T. A studentized permutation test in group sequential designs. *arXiv* **2024**, arXiv:2404.12483.
160. Borghesi, M.; Zambelli, C.; Micheloni, R.; Bonnini, S. Modeling 3D NAND Flash with Nonparametric Inference on regression Coefficients for Reliability Solid-State Storage. *Future Internet* **2023**, *13*, 319. [CrossRef]
161. Bonnini, S.; Borghesi, M.; Giacalone, M. Advances on multisample permutation tests for "V-shaped" and "U-shaped" alternatives with applications to circular economy assessments. *Ann. Oper. Res.* **2023**, 1–16. [CrossRef]
162. Bonnini, S.; Borghesi, M.; Giacalone, M. Simultaneous marginal homogeneity versus directional alternatives for multivariate binary data with application to circular economy assessments. *Appl. Stoch. Models Bus. Ind.* **2024**, *40*, 389–407. [CrossRef]
163. Corain, L.; Arboretti, R.; Bonnini, S.; Ceccato, R.; Salmaso, L. Different views of the multivariate ranking problem. *Electron. J. Appl. Stat. Anal.* **2023**, *16*, 323–338.
164. Bonnini, S. Multivariate approach for comparative evaluations of customer satisfaction with application to transport services. *Commun. Stat. Simul. Comput.* **2016**, *45*, 1554–1568. [CrossRef]
165. Bonnini, S.; Borghesi, M.; Giacalone, M. Semi-parametric approach for modeling overdispersed count data with application to Industry 4.0. *Socio-Econ. Plan. Sci.* **2024**, *95*, 101976. [CrossRef]
166. Bonnini, S.; Corain, L.; Pesarin, F.; Salmaso, L. Multivariate permutation McNemar's test with application to performance evaluation of basket players. *Stat. Appl.—Ital. J. Appl. Stat.* **2023**, *35*, 11–29. [CrossRef]
167. Corain, L.; Arboretti, R.; Salmaso, L.; Ceccato, R.; Ronchi, F. Testing and ranking on round-robin design for data sport analytics with application to basketball. *Stat. Model.* **2019**, *19*, 5–27. [CrossRef]