Nonparametric Hypothesis Testing
Rank and Permutation Methods
with Applications in R

Stefano Bonnini • Livio Corain
Marco Marozzi • Luigi Salmaso
Nonparametric Hypothesis Testing
WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: David J. Balding, Noel A.C. Cressie, Garrett M. Fitzmaurice, Geof H. Givens, Harvey Goldstein, Geert Molenberghs, David W. Scott, Adrian F.M. Smith, Ruey S. Tsay, Sanford Weisberg
Editors Emeriti: J. Stuart Hunter, Iain M. Johnstone, Joseph B. Kadane, Jozef L. Teugels

A complete list of the titles in this series appears at the end of this volume.
Nonparametric Hypothesis Testing

Rank and Permutation Methods
with Applications in R

Stefano Bonnini
University of Ferrara, Italy

Livio Corain
University of Padova, Italy

Marco Marozzi
University of Calabria, Italy

Luigi Salmaso
University of Padova, Italy

WILEY
The greatest value of a picture is when it forces us to notice what we never expected to see.

J. Tukey
Contents

Presentation of the book xi
Preface xiii
Notation and abbreviations xvii

1 One- and two-sample location problems, tests for symmetry and tests on a single distribution 1
  1.1 Introduction 1
  1.2 Nonparametric tests 2
    1.2.1 Rank tests 2
    1.2.2 Permutation tests and combination based tests 3
  1.3 Univariate one-sample tests 5
    1.3.1 The Kolmogorov goodness-of-fit test 6
    1.3.2 A univariate permutation test for symmetry 10
  1.4 Multivariate one-sample tests 15
    1.4.1 Multivariate rank test for central tendency 15
    1.4.2 Multivariate permutation test for symmetry 18
  1.5 Univariate two-sample tests 20
    1.5.1 The Wilcoxon (Mann–Whitney) test 21
    1.5.2 Permutation test on central tendency 27
  1.6 Multivariate two-sample tests 29
    1.6.1 Multivariate tests based on rank 29
    1.6.2 Multivariate permutation test on central tendency 34

References 37

2 Comparing variability and distributions 38
  2.1 Introduction 38
  2.2 Comparing variability 39
    2.2.1 The Ansari–Bradley test 40
    2.2.2 The permutation Pan test 43
    2.2.3 The permutation O’Brien test 46
  2.3 Jointly comparing central tendency and variability 49
    2.3.1 The Lepage test 50
    2.3.2 The Cucconi test 52
## CONTENTS

2.4 Comparing distributions 56  
2.4.1 The Kolmogorov–Smirnov test 56  
2.4.2 The Cramér–von Mises test 59  
References 61

3 Comparing more than two samples 65  
3.1 Introduction 65  
3.2 One-way ANOVA layout 66  
3.2.1 The Kruskal–Wallis test 67  
3.2.2 Permutation ANOVA in the presence of one factor 73  
3.2.3 The Mack–Wolfe test for umbrella alternatives 76  
3.2.4 Permutation test for umbrella alternatives 83  
3.3 Two-way ANOVA layout 87  
3.3.1 The Friedman rank test for unreplicated block design 87  
3.3.2 Permutation test for related samples 89  
3.3.3 The Page test for ordered alternatives 91  
3.3.4 Permutation analysis of variance in the presence of two factors 93  
3.4 Pairwise multiple comparisons 95  
3.4.1 Rank-based multiple comparisons for the Kruskal–Wallis test 96  
3.4.2 Permutation tests for multiple comparisons 98  
3.5 Multivariate multisample tests 99  
3.5.1 A multivariate multisample rank-based test 99  
3.5.2 A multivariate multisample permutation test 103  
References 105

4 Paired samples and repeated measures 107  
4.1 Introduction 107  
4.2 Two-sample problems with paired data 108  
4.2.1 The Wilcoxon signed rank test 108  
4.2.2 A permutation test for paired samples 114  
4.3 Repeated measures tests 116  
4.3.1 Friedman rank test for repeated measures 117  
4.3.2 A permutation test for repeated measures 120  
References 122

5 Tests for categorical data 124  
5.1 Introduction 124  
5.2 One-sample tests 125  
5.2.1 Binomial test on one proportion 125  
5.2.2 The McNemar test for paired data (or bivariate responses) with binary variables 128  
5.2.3 Multivariate extension of the McNemar test 131
CONTENTS ix

5.3 Two-sample tests on proportions or $2 \times 2$ contingency tables 134
  5.3.1 The Fisher exact test 135
  5.3.2 A permutation test for comparing two proportions 138
5.4 Tests for $R \times C$ contingency tables 139
  5.4.1 The Anderson–Darling permutation test for $R \times C$
    contingency tables 140
  5.4.2 Permutation test on moments 145
  5.4.3 The chi-square permutation test 148

References 151

6 Testing for correlation and concordance 153
  6.1 Introduction 153
  6.2 Measuring correlation 154
  6.3 Tests for independence
    6.3.1 The Spearman test 157
    6.3.2 The Kendall test 160
  6.4 Tests for concordance
    6.4.1 The Kendall–Babington Smith test 167
    6.4.2 A permutation test for concordance 172

References 174

7 Tests for heterogeneity 176
  7.1 Introduction 176
  7.2 Statistical heterogeneity 177
  7.3 Dominance in heterogeneity
    7.3.1 Geographical heterogeneity 180
    7.3.2 Market segmentation 184
  7.4 Two-sided and multisample test
    7.4.1 Customer satisfaction 189
    7.4.2 Heterogeneity as a measure of uncertainty 191
    7.4.3 Ethnic heterogeneity 194
    7.4.4 Reliability analysis 196

References 197

Appendix A Selected critical values for the null distribution of the
  peak-known Mack–Wolfe statistic 201

Appendix B Selected critical values for the null distribution of the
  peak-unknown Mack–Wolfe statistic 203

Appendix C Selected upper-tail probabilities for the null distribution
  of the Page $L$ statistic 206

Appendix D $R$ functions and codes 213

Index 219
Presentation of the book

The importance and usefulness of nonparametric methods for testing statistical hypotheses has been growing in recent years mainly due to their flexibility, their efficiency and their ease of application to several different types of problems, including most important and frequently encountered multivariate cases. By also taking account that with respect to parametric counterparts they are much less demanding in terms of required assumptions, these peculiarities of nonparametric methods are making them quite popular and widely used even by non-statisticians.

The growing availability of adequate hardware and software tools for their practical application, and in particular of free access to software environments for statistical computing like R, represents one more reason for the great success of these methods.

The recognized simplicity and good power behavior of rank and permutation tests often make them preferable to the classical parametric procedures based on the assumption of normality or other distribution laws. In particular, permutation tests are generally asymptotically as powerful as their parametric counterparts in the conditions for the latter. Moreover, when data exchangeability with respect to samples is satisfied in the null hypothesis, permutation tests are always exact in the sense that their null distributions are known for any given dataset of any sample size. On the other hand, those of parametric counterparts are often known only asymptotically. Thus for most sample sizes of practical interest, the related lack of efficiency of unidimensional permutation solutions may sometimes be compensated by the lack of approximation of parametric asymptotic competitors. For multivariate cases, especially when the number of processed variables is large in comparison with sample sizes, permutation solutions in most situations are more powerful than their parametric counterparts.

For these reasons in the specialized literature a book dedicated to rank and permutation tests, problem oriented with exhaustive but simple and easy to understand theoretical explanations, a practical guide for the application of the methods to most frequently encountered scientific problems, including related R codes and with many clearly discussed examples from several different disciplines, was lacking.
The present book fully satisfies these objectives and can be considered a prac-
tical and complete handbook for the application of the most important rank and
permutation tests. The presentation style is simple and comprehensible also for non-
statisticians with elementary education in statistical inference, but at the same time
precise and formally rigorous in the theoretical explanations of the methods.

Fortunato Pesarin

*Department of Statistics*
*University of Padova*
Preface

This book deals with nonparametric statistical solutions for hypotheses testing problems and codes for the software environment $R$ for the application of these solutions. In particular rank based and permutation procedures are presented and discussed, also considering real-world application problems related to engineering, economics, educational sciences, biology, medicine and several other scientific disciplines. Since the importance of nonparametric methods in modern statistics continues to grow, the goal of the book consists of providing effective, simple and user friendly instruments for applying these methods.

The statistical techniques are described mainly highlighting properties and applicability of the methods in relation to application problems, with the intention of providing methodological solutions to a wide range of problems. Hence this book presents a practical approach to nonparametric statistical analysis and includes comprehensive coverage of both established and recently developed methods. This ‘problem oriented’ approach makes the book useful also for non-statisticians. All the considered problems are real problems faced by the authors in their activities of academic counseling or found in the literature in their teaching and research activities. Sometimes data are exactly the same as in the original problem (and the data source is cited) but in most cases data are simulated and not real.

All $R$ codes are commented and made available through the book’s website www.wiley.com/go/hypothesis_testing, where data used throughout the book may also be downloaded. Part of the material, including $R$ codes, presented in the book is new and part is taken from existing publications from the literature and/or from websites of different authors providing suitable $R$ codes. We fully recognize the authorship of each $R$ code and a comprehensive list of useful websites is reported in Appendix D.

The book is mainly addressed to university students, in particular for undergraduate and postgraduate studies (i.e., PhD courses, Masters, etc.), statisticians and non-statisticians experts in empirical sciences, and it can also be used by practitioners with a basic knowledge in statistics interested in the same applications described in the book or in similar problems, or consultants/experts in statistics.

Chapter 1 deals with one-sample and two-sample location problems, tests for symmetry and tests on a single distribution. First of all an introduction to rank based testing procedures and to permutation testing procedures (including nonparametric combination methodology useful for multivariate or multiple tests) is presented.
Then in this chapter, according to the number of response variables and to the number of samples, we distinguish four kinds of methods: univariate one-sample tests, multivariate one-sample tests, univariate two-sample tests and multivariate two-sample tests. In the first category the Kolmogorov–Smirnov test and the permutation test for symmetry are considered; in the second group of procedures the multivariate rank test for central tendency and the multivariate extension of the permutation test on symmetry are presented; among the procedures included in the third family of solutions the Wilcoxon test and the permutation test on central tendency are described; finally the multivariate extensions of the two-sample test on central tendency both with the rank based and permutation approach are discussed.

Chapter 2 presents some tests for comparing variabilities and distributions. For problems of variability comparisons the Ansari–Bradley test, the permutation Pan test and the permutation O’Brien test are considered. For jointly comparing central tendency and variability the Lapage test and the Cucconi test are presented. For problems related to comparisons of distributions the Kolmogorov–Smirnov and the Cramer–von Mises proposals are taken into account.

Chapter 3 is dedicated to multisample tests. For the one-way analysis of variance (ANOVA) layout the following methods are presented: the Kruskal–Wallis test, the permutation one-way ANOVA, the Mack–Wolfe test and the permutation test for umbrella alternatives. As regards the two-way ANOVA layout, the considered procedures are the Friedman test, the permutation test for related samples, the Page test for ordered alternatives and the permutation two-way ANOVA. Multiple comparison procedures for the Kruskal–Wallis test and for a permutation test are also considered. For multivariate and multisample problems a rank based and a permutation approach are presented.

Chapter 4 concerns problems for paired samples and repeated measures. For the two-sample test with paired data the Wilcoxon signed rank proposal and the permutation test for two dependent samples are discussed. For repeated measures problems the Friedman rank based test and a permutation test are considered.

Chapter 5 deals with tests for categorical data. Among one-sample problems the binomial test on one proportion, the McNemar test for paired data with binary variables and its multivariate extension are illustrated. Then two-sample tests for proportion comparisons and in general tests for $2 \times 2$ contingency tables are discussed. In particular the Fisher exact test and the permutation test for comparison of proportions are examined. The considered solutions for general problems related to $R \times C$ contingency tables are: the Anderson–Darling type permutation test, the permutation test on moments, and the chi-square permutation test.

Chapter 6 studies correlation and concordance. First, the statistical relationship between two variables is considered. The Spearman test and the Kendall test for independence are presented. Secondly, the problem of whether a set of criteria or a group of judges is concordant in ranking some objects is addressed. The Kendall–Babington Smith test and a permutation test for concordance are presented.

Finally, Chapter 7 contains a wide range of application problems and methodological solutions concerning comparisons of heterogeneity for categorical variables. The
definition of statistical heterogeneity, the description of the testing problem of dominance in heterogeneity (two-sample one-sided test on heterogeneity), its two-sided and multisample extensions and the related permutation solutions are included.

We would like to express our thanks to Fortunato Pesarin, for stimulating discussions and helpful comments, to Rosa Arboretti, Eleonora Carrozzo and Iulia Cichi for helping with some R codes and in finding suitable reference literature. We also wish to thank Kathryn Sharples, Richard Davies and the John Wiley & Sons group in Chichester for their valuable publishing suggestions.

We welcome any suggestions for the improvement of the book and would be very pleased if the book provides users with new insights to the analysis of their data.

Stefano Bonnini
Department of Economics and Management
University of Ferrara

Marco Marozzi
Department of Economics, Statistics and Finance
University of Calabria

Livio Corain and Luigi Salmaso
Department of Management and Engineering
University of Padova
Notation and abbreviations

ANOVA: analysis of variance

$B$: the number of conditional Monte Carlo iterations

$\in$: belong(s) to

$\mathcal{B}(n, \theta)$: binomial distribution with $n$ trials and probability $\theta$ of success in one trial

CDF: cumulative distribution function

CMC: conditional Monte Carlo

$\text{Cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$: the covariance operator on $(X, Y)$

$\text{Cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$: the covariance operator on $(X, Y)$

CSP: constrained synchronized permutations

d.f.: degrees of freedom

EDF: empirical distribution function:

$\hat{F}_X(t) = \hat{F}(t|\mathcal{X}_X) = \sum_i \mathbb{I}(X_i \leq t)/n, t \in \mathcal{R}^1$

$\mathbb{E}(X) = \int_X x \cdot dF_X(x)$: the expectation operator (mean value) of $X$

$d$: equality in distribution: $X \overset{d}{=} Y \iff F_X(z) = F_Y(z), \forall z \in \mathcal{R}^1$

$d >$: stochastic dominance: $X \overset{d}{>} Y \iff F_X(z) \leq F_Y(z), \forall z$ and $\exists A : F_X(z) < F_Y(z), z \in A$, with $\Pr(A) > 0$

$\neq$: means $<$, or $\neq$, or $>$

$: distributed as, for example, $X \sim \mathcal{N}(0, 1)$ means $X$ is standard normal distributed

$\forall$: for every

$F_X(z) = F(z) = \Pr \{X \leq z\}$: the CDF of $X$

$F_{X,Y}(z, t) = F(z, t) = \Pr \{X \leq z, Y \leq t\}$: the CDF of $(X, Y)$

i.i.d.: independent and identically distributed

$\mathbb{I}(A)$: the indicator function, that is, $\mathbb{I}(A) = 1$ if $A$ is true, and $0$ otherwise

$\lambda = \Pr \{T \geq T^0|\mathcal{X}_X\}$: the attained $p$-value of test $T$ on dataset $\mathcal{X}$

$L_X(t) = L(t) = \Pr \{X \geq t\}$: the significance level function (same as the survival function)

$\mu = \mathbb{E}(X)$: the mean value of vector $\mathbf{X}$

MANOVA: multivariate analysis of variance
\textbf{Notation and Abbreviations}

- $M d(X) = \tilde{\mu}$: the median operator on variable $X$ such that $Pr\{X < \tilde{\mu}\} = Pr\{X > \tilde{\mu}\}$
- $n$: the (finite) sample size
- $\mathcal{N}(\mu, \sigma^2)$: Gaussian or normal distribution with mean $\mu$ and variance $\sigma^2$
- $P(\tau)$: Poisson distribution with parameter $\tau$
- $Pr\{A\}$: a probability statement relative to $A \in \mathcal{A}$
- $\mathbb{R}^n$: the set of $n$-dimensional real numbers
- $\mathbb{R}$: the rank operator
- $R_i = \mathbb{R}(X_i) = \sum_{1 \leq j \leq n} I(X_j \leq X_i)$ the rank of $X_i$ within $\{X_1, \ldots, X_n\}$
- $T^o = T_{obs} = T(X)$ the observed value of test statistic $T$ evaluated on $X$
- $\mathcal{U}(a, b)$: uniform distribution in the interval $(a, b)$
- $\mathbb{V}(X) = \mathbb{E}(X - \mu)^2 = \sigma^2$: the variance operator on variable $X$
- $X$ or $Z$: a univariate or multivariate random variable
- $X$: a sample of $n$ units, $X = \{X_i, i = 1, \ldots, n\}$
- $X^*$: a permutation of $X$
- $|X| = \{|X_i|, i = 1, \ldots, n\}$: a vector of absolute values
- $\mathbb{U}$: the operator for pooling (concatenation) of two datasets: $X = X_1 \mathbb{U} X_2$
1

One- and two-sample location problems, tests for symmetry and tests on a single distribution

1.1 Introduction

Many real phenomena can be represented by numerical random variables. Considering a given population and a random sample of it, for forecasting or improving the effectiveness of inferential techniques related to estimation and testing of hypothesis, it would be useful to know the functional form of the distribution of the data. Sometimes, the central interest of the statistical analysis is focused only on the symmetry or on the location of the distribution itself. Another very common statistical problem consists of comparing two independent populations in terms of central tendency. In the simpler cases the object of the analysis is a univariate population, but in some real applications we are in the presence of many variables and multivariate datasets.

The methods presented in this chapter consist of rank or permutation procedures for the tests of the hypotheses cited above. Section 1.2 is an introduction to rank and permutation tests. In Section 1.3, devoted to one-sample tests, the Kolmogorov procedure for testing whether the data are distributed according to an hypothesized
cumulative distribution function (CDF), and the permutation test on the symmetry of the distribution are taken into account. Section 1.4 deals with multivariate one-sample tests, and introduces the multivariate location problem and the multivariate test on symmetry. In Section 1.5 the univariate two-sample location problem is discussed. Section 1.6 considers the multivariate extension of the location problem for two independent populations and presents some solutions for it.

In the one-sample problems the data are a random sample of numerical data \( X = \{X_1, \ldots, X_n\} \) from the unknown population under study. In the two-sample problems the numerical sample data from the \( j \)th unknown population are \( X_j = \{X_{j1}, \ldots, X_{jn_j}\} \), with \( j = 1, 2 \) and \( n_1 + n_2 = n \). In the multivariate extensions, in the presence of \( q \) component variables, for the one-sample problem, the observation related to the \( i \)th statistical unit is denoted by \( X_i = \{X_{i1}, \ldots, X_{iq}\} \) and, for the two-sample problem, the observation related to the \( i \)th statistical unit in the \( j \)th group is denoted by \( X_{ji} = \{X_{ji1}, \ldots, X_{jiq}\} \).

### 1.2 Nonparametric tests

Traditional parametric testing methods are based on the assumption that data are generated by well-known distributions, characterized by one or more unknown population parameters (mean, median, variance, etc.) and the hypotheses of the problems are formulated as equalities/inequalities related to these unknown parameters. For example, the location problem can be formalized using the mean parameter, the scale problem can be expressed in terms of variance comparisons, etc.

In other words parametric methods are based on a modeling approach and on the introduction of stringent assumptions, often quite unrealistic, unclear and connected with the availability of inferential methods (Pesarin, 2001). Hence the critical values or alternatively the \( p \)-values can be computed according to the distribution of the test statistic under the null hypothesis, which can be derived from the assumptions related to the assumed underlying distribution of data. When the assumed distribution is not true, when we are not sure whether it is true or not or when it is not plausible, other methods, which ignore the true distribution of data, are needed. These methods are called nonparametric or distribution-free.

Since, when the parametric assumptions hold, the nonparametric procedures are only slightly less powerful than the parametric methods and they are the only valid solution when the parametric assumptions do not hold, nonparametric tests are in general more flexible and often more appropriate than parametric counterparts. Basically the nonparametric testing procedures can be classified into two kinds of methods: rank based tests and permutation tests.

#### 1.2.1 Rank tests

The main aspect which characterizes rank tests is that observations are transformed into their sample ranks. Hence in the rank tests we compare the observations based
on their ranks within the sample. Formally the rank of the $i$th observation with respect
to a set of $n$ data is given by

$$R_i = \mathbb{R}(X_i) = \sum_{1 \leq j \leq n} \mathbb{I}(X_j \leq X_i),$$

where $\mathbb{R}$ is the (increasing) rank operator, $X_i$ is the transformed observation, $\mathbb{I}(A)$ is
the indicator function of the event $A$, that is $\mathbb{I}(A) = 1$ when $A$ is true and $\mathbb{I}(A) = 0$
otherwise. Hence the rank of $X_i$ within $\{X_1, \ldots, X_n\}$ is equal to 1 if $X_i$ is the minimum
value, it is equal to 2 if $X_i$ is the second smaller value, up to $n$ if $X_i$ is the maximum.

Often in the case of ties, the midrank method is applied, that is the mean of the ranks
 corresponding to the positions in the sorted set of observations is assigned to the
tied observations. Formally if rank $r$ is assigned to $t$ observations equal to a certain
value $x$ ($t \leq r$), that is $r$ observations in the set $\{X_1, \ldots, X_n\}$ are less than or equal
to $x$, then the rank of these $t$ observations, according to the midrank rule, is adjusted
into the mean value of the $t$ ranks $(r - t + 1), (r - t + 2), \ldots, r$. Rank transformation
is non-bijective, in the sense that a given set of ranks $\{R_1, \ldots, R_n\}$ may correspond
to distinct sets of sample data.

Let us consider an example related to a pharmacological experiment. A pharma-
ceutical company needs to test whether a new experimental drug for lowering blood
cholesterol levels is more effective than another drug already present in the pharma-
ceutical market. A group of patients is treated with the new drug and another group
with the old drug. The null hypothesis consists of ‘no difference’ between the two
treatment effects; the alternative hypothesis states the superiority of the new drug, that
is the effect of the new drug is greater than the effect of the old one. Let us denote with
$n_1$ and $n_2$ the number of patients treated with the new and the old drug, respectively,
independent samples from populations with continuous probability function $F_1$ and $F_2$,
respectively. The null hypothesis of no difference between the effects of the two
treatments can be written as $H_0 : F_1 = F_2 = F$ with $F$ unknown. $H_0$ implies that the
two samples can be considered as just one sample from a unique distribution $F$. A
way to solve this problem is provided by the Wilcoxon rank sum test, a rank based
testing procedure which takes into account the ranking of the observations within
the pooled sample of $n_1 + n_2$ data and considers the sum of the ranks of the first
sample as test statistic. When $H_0$ is true, the test statistic tends to take neither too
large nor too small values. The distribution of the test statistic under the null hypo-
thesis can be computed considering all the possible rankings as equally likely and the
corresponding values of the statistic. Hence the computation of critical values and
$p$-values does not depend on the unknown distribution $F$. This is why it is considered
a distribution-free method.

1.2.2 Permutation tests and combination based tests

In many testing problems, the dataset can be seen as a partition into groups or samples
according to the treatment levels of a real or symbolic experiment. According to the
permutation testing principle, if two experiments characterized by the same sample
space (the set of all possible samples) give the same dataset, then the result of the testing procedure conditional on the dataset itself must be the same, provided that the exchangeability condition with respect to samples holds under the null hypothesis (Pesarin, 2001). This is the reason why inference based on permutation tests is also called conditional inference.

In real applications, random sampling, on which the parametric methods are based, is rarely achieved. Hence often the unconditional inferences associated with parametric tests are not applicable. In these situations permutation tests are suitable solutions. Furthermore some common assumptions of parametric methods, such as the existence of mean values and variances, or equal variances of responses (homoscedasticity) under the alternative hypothesis are not needed within the permutation testing procedures.

For example, for the two-sample test related to the pharmaceutical problem, under the null hypothesis observations are exchangeable among samples because they are supposed to come from the same population and their belonging to one group or to another is actually random. A suitable test statistic for the problem may be the difference of the two-sample means which is expected to take neither too large nor too small values when \( H_0 \) is true. The distribution of the test statistic under the null hypothesis, and then the \( p \)-value of the test, can be computed considering all the possible permutations (i.e., reallocations of the observations to the two groups) as equally likely and computing the corresponding values of the statistic for each permutation. Alternatively, for computational simplicity, a random sample of all the possible permutations can be considered and the null distribution of the test statistic can be well approximated by Conditional Monte Carlo (CMC) techniques.

1.2.2.1 Nonparametric combination methodology

A suitable method to perform multivariate permutation tests or multiple permutation test procedures is the so called nonparametric combination (NPC) of dependent permutation tests. Let us suppose that the null hypothesis \( H_0 \) of a testing problem can be broken down into \( k \) sub-hypotheses or partial hypotheses \( H_{01}, \ldots, H_{0k} \) such that \( H_0 \) is true if and only if all the sub-hypotheses are true, formally \( H_0 : \bigcap_{i=1}^{k} H_{0i} \). Similarly the alternative hypothesis \( H_1 \) is true if and only if at least one of the null sub-hypotheses is false, and consequently at least one of the alternative sub-hypotheses is true, briefly \( H_1 : \bigcup_{i=1}^{k} H_{1i} \). Let \( T = T(X) \) be a \( k \)-dimensional vector of test statistics and each component \( T_i(X) \) be a suitable test statistic for testing \( H_{0i} \) against \( H_{1i} \) and without loss of generality assume that \( H_{0i} \) is rejected for large values of \( T_i(X) \). Assuming as usual that each row of the dataset corresponds to a statistical unit, and considering for example a test for independent samples, the NPC method works as follows:

1. Compute the vector of the observed values of \( T \): \( T_{obs} = [T_1(X), \ldots, T_k(X)]' = [T_{1(0)}, \ldots, T_{k(0)}]'. \)
2. Consider a permutation of the rows of the dataset, that is a reallocation of the units to the groups, and compute the corresponding values of the test statistics: 
\[ T^*_1 = (X^*_1), \ldots, T^*_k = (X^*_k) \].

3. Perform \( B \) independent repetitions of step (2) and obtain \( T^*_b = (T^*_1, \ldots, T^*_k) \), \( b = 1, \ldots, B \).

4. For each \( i \) compute an estimate of the significance level function \( \Pr \{ T^*_i \geq z \} : \)
\[ \hat{L}_i(z) = \left( \frac{1}{2} + \sum_r I[T^*_i(r) \geq z] \right) / (B + 1), \]
\( i = 1, \ldots, k \).

5. For each \( b \) compute \( \lambda^*_i(b) = \hat{L}_i(T^*_i(b)), \)
\( b = 1, \ldots, B \) and compute \( \lambda_i(0) = \hat{L}_i(T^*_i(0)) \).

6. For each \( b \) compute the combined values \( T^*_{(b)} = \psi(\lambda^*_1(b), \ldots, \lambda^*_k(b)) \) and \( T^*_{(0)} = \psi(\lambda_1(0), \ldots, \lambda_k(0)) \) using a suitable combining function \( \psi \).

7. Compute the estimate of the \( p \)-value of the test as \( \lambda^* = \sum_b I[T^*_{(b)} \geq T^*_{(0)}] / B \).

The final decision should be based on \( \lambda^* \) in the sense that \( H_0 \) should be rejected in favor of \( H_1 \) if \( \lambda^* \leq \alpha \). The NPC method is very useful to solve complex problems, in particular multivariate problems or problems where a multivariate test statistic may be suitable. The main advantage with respect to other standard parametric methods is that the multivariate distribution of the test statistic does not need to be known or estimated, and in particular the dependence structure between the component variables does not need to be known or explicitly specified. The dependence is implicitly taken into account through the permutation strategy and the application of the combining function \( \psi \). The combining function must satisfy the following simple properties: (1) it must be nonincreasing in each argument; (2) it must attain its supremum even when only one argument tends to zero; and (3) for each \( \alpha \) level the critical value \( T^*_{\alpha} \) is assumed to be finite and strictly smaller than the supremum value. Some suitable combining functions are:

- the Fisher omnibus combining function: \( T^*_F = -2 \sum \log(\lambda_i) \);
- the Liptak combining function: \( T^*_L = \sum \Phi^{-1}(1 - \lambda_i) \);
- the Tippett combining function: \( T^*_T = \max_i (1 - \lambda_i) \).

Tippett combination provides powerful tests when one or a few but not all of the alternative sub-hypotheses are true; Liptak’s function has a more powerful behavior when all of the alternative sub-hypotheses are jointly true; Fisher’s solution is intermediate between the two.

1.3 Univariate one-sample tests

The basic assumption of an inferential problem is that the observed phenomena can be represented by random variables with unknown distributions. The goal of the
inferential study consists of investigating some aspects of the unknown distribution. Let us assume that the observed random sample has been drawn from a numerical population with unknown CDF $F(x)$. In order to test whether $F(x)$ is equal to a fully specified function (without any unknown nuisance parameter), a powerful and commonly used solution is provided by the procedure introduced by Kolmogorov (1933). Such a procedure is based on the comparison between the empirical distribution function (EDF) and the specified tested distribution (see Section 1.2.1). As it tests the distribution’s fit to a set of data, it is classified as a goodness-of-fit test. In this sense it can be considered an alternative for ordinal data to the goodness-of-fit chi-square test, valid for nominal categorical variables. An important difference between the two procedures is that, for continuous variables, the Kolmogorov test is exact even for small sample sizes (in the case of non continuity it is not distribution-free), while the chi-square test requires that $n$ is large enough so that the test statistic under the null hypothesis approximately follows a chi-square distribution (Conover, 1999).

In some applications the test involves only one or a few aspects of the functional form of $F(x)$, hence only a specific property of $F(x)$ is specified under $H_0$. This is the case of the test on symmetry, very useful in particular in the statistical quality control of industrial processes (see Section 1.2.2). For continuous variables, symmetry of the distribution around the origin is equivalent to the property: $F(x) = 1 - F(-x)$ $\forall x \in R$. Let us consider the cited one-sample problems.

1.3.1 The Kolmogorov goodness-of-fit test

Let $X = \{X_1, \ldots, X_n\}$ be a random sample from a population with unknown continuous CDF $F(x)$ and assume an interest in testing the hypothesis that $F(x)$ corresponds to a known and completely specified distribution $F_0(x)$ against the alternative that this is not true. The testing procedure proposed by Kolmogorov (1933) is based on the supremum of the vertical distance between $F_0(x)$ and the EDF based on the observed sample $X$. Smirnov (1939) proposed an extension of the Kolmogorov test for comparing the distributions of two independent populations. Statistics based on the vertical distance between $F_0(x)$ and the EDF are called Kolmogorov-type statistics, while similar statistics based on the vertical distance between two EDFs are called Smirnov-type statistics (Conover, 1999). The Kolmogorov goodness-of-fit test presented in this paragraph is also called the one-sample Kolmogorov–Smirnov test. Formally the problem consists of testing the null hypothesis

$$H_0 : F(x) = F_0(x)$$

against the alternative

$$H_1 : F(x) \neq F_0(x).$$