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

Nonparametric Hypothesis Testing: Rank and Permutation Methods with Applications in R,
First Edition. Stefano Bonnini, Livio Corain, Marco Marozzi and Luigi Salmaso.

Companion website: http://www.wiley.com/go/hypothesis_testing

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 $\boldsymbol{X} = \{X_1, \dots, X_n\}$ from the unknown population under study. In the two-sample problems the numerical sample data fom the jth unknown population are $\boldsymbol{X}_j = \{X_{j1}, \dots, X_{jn_j}\}$, 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 ith statistical unit is denoted by $\boldsymbol{X}_i = \{X_{i1}, \dots, X_{iq}\}$ and, for the two-sample problem, the observation related to the ith statistical unit in the jth group is denoted by $\boldsymbol{X}_{ji} = \{X_{ji1}, \dots, X_{jiq}\}$.

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 ith 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 $\{X_1, \dots, X_n\}$ 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 $\{X_1, \dots, X_n\}$ 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 $(r - t + 1), (r - t + 2), \dots, r$. Rank transformation is non-bijective, in the sense that a given set of ranks $\{R_1, \dots, R_n\}$ may correspond to distinct sets of sample data.

Let us consider an example related to a pharmacological experiment. A pharmaceutical company needs to test whether a new experimental drug for lowering blood cholesterol levels is more effective than another drug already present in the pharmaceutical 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 hypothesis 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 $H_{01}, \dots, H_{0k}$ 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 $\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{X})$ be a k-dimensional vector of test statistics and each component $T_i(\boldsymbol{X})$ 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 $T_i(\boldsymbol{X})$. 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 $\boldsymbol{T}$: $\boldsymbol{T}_{obs} = [T_1(\boldsymbol{X}), \dots, T_k(\boldsymbol{X})]' = [T_{1(0)}, \dots, T_{k(0)}]'$.

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: $\boldsymbol{T}^*_{(1)} = [T_1(\boldsymbol{X}^*_{(1)}), \ldots, T_k(\boldsymbol{X}^*_{(1)})]'$.
3. Perform B independent repetitions of step (2) and obtain $\boldsymbol{T}^*_{(b)} = [T^*_{1(b)}, \ldots, T^*_{k(b)}]'$, $b = 1, \ldots, B$.
4. For each i compute an estimate of the significance level function $\Pr\{T^*_i \geq z\}$: $\widehat{L}_i(z) = \{\frac{1}{2} + \sum_r \mathbb{I}[T^*_{i(r)} \geq z]\}/(B+1)$, $i = 1, \ldots, k$.
5. For each b compute $\lambda^*_{i(b)} = \widehat{L}_i(T^*_{i(b)})$, $b = 1, \ldots, B$ and compute $\lambda_{i(0)} = \widehat{L}_i(T_{i(0)})$, $i = 1, \ldots, k$.
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 ψ.
7. Compute the estimate of the p-value of the test as $\lambda'' = \sum_b \mathbb{I}[T''^*_{(b)} \geq T''_{(0)}]/B$.

The final decision should be based on λ'' 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 ψ. 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 α level the critical value T''_α is assumed to be finite and strictly smaller than the supremum value. Some suitable combining functions are:

- the Fisher *omnibus* combining funtion: $T''_F = -2\sum_i \log(\lambda_i)$;
- the Liptak combining funtion: $T''_L = \sum_i \Phi^{-1}(1 - \lambda_i)$;
- 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: $F(x) = 1 - F(-x)$ $\forall x \in \mathcal{R}$. Let us consider the cited one-sample problems.

1.3.1 The Kolmogorov goodness-of-fit test

Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ 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 $\boldsymbol{X}$. 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

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

The EDF of $\boldsymbol{X}$ is $\hat{F}_n(x) = 1/n \sum_{i=1}^{n} \mathbb{I}(X_i \leq x)$ where $\mathbb{I}(X_i \leq x)$ takes value 1 if $X_i \leq x$ and 0 otherwise. The Kolmogorov test statistic is given by

$$D_n = \sup_{x \in \mathcal{R}} |\widehat{F}_n(x) - F_0(x)|.$$

In some problems, the alternative hypothesis is one-sided, that is the CDF $F(x)$ is supposed to be smaller than $F_0(x)$ or larger than $F_0(x)$. Formally the one-sided alternative might be

$$H_1 : F(x) \leq F_0(x)\ \forall x \in \mathcal{R} \text{ and } F(x) < F_0(x) \text{ for some } x,$$

or similarly

$$H_1 : F(x) \geq F_0(x)\ \forall x \in \mathcal{R} \text{ and } F(x) > F_0(x) \text{ for some } x,$$

thus the suitable test statistic is $D_n^+ = \max_{x \in \mathcal{R}}[F_0(x) - \widehat{F}_n(x)]$ and $D_n^- = \max_{x \in \mathcal{R}}[\widehat{F}_n(x) - F_0(x)]$, respectively. The tests reject the null hypothesis for large values of the test statistics.

A result shows that if $\boldsymbol{X}$ is a random sample from an absolutely continuous population with the CDF F_0, then the distribution of the statistic D_n does not depend on F_0 but only on the sample size n (Bagdonavicius *et al.* 2011). Therefore in the two-sided test, the hypothesis H_0 is rejected with a significance level α when $D_n > D_\alpha(n)$, where $D_\alpha(n)$ is the critical value of the statistic D_n, that is the $(1 - \alpha)$-quantile of the null distribution of D_n. Equivalently, the null hypothesis is rejected when the p-value of the test (probability that under H_0 the test statistic takes values greater than the observed value of D_n) is less than α. A similar procedure should be applied to the one-sided tests. Exact quantiles for D_n and approximate quantiles for D_n^+ and D_n^- have been tabulated. When $n > 40$ the asymptotic approximation may be used. Some computationally friendly representations of the distribution of the test statistics for sample sizes less than 100 and with no ties are proposed by Marsaglia *et al.* (2003) and Birnbaum and Tingey (1951). When $F_0(x)$ is discrete, a modification for the computation of the quantiles of the test statistics might be applied (Conover, 1972; Coberly and Lewis, 1973).

The basic package of *R* includes the function `ks.test`, which computes the Kolmogorov–Smirnov statistic for the one-sample or two-sample cases. The presence of ties in the case of noncontinuous variables generates a warning. If ties arise because of rounding, the test may be approximately valid, but even modest amounts of rounding can have an important effect on the computation of the test statistic.

Consider an industrial experiment in which we have a sample of $n = 10$ fabrics subjected to washing. The goal of the experiment is to analyze the performance of a new experimental detergent for clothes. Specifically, the response variable under study is the so called reflectance, that is the proportion of incident light which a given surface (of fabric) is able to reflect, which can be considered a measure related to the cleaning efficacy of the detergent. Suppose we wish to test, at the significance level

Table 1.1 Sample data of reflectance in the experiment on detergent performance.

Piece of fabric	1	2	3	4	5
Reflectance	0.608	0.533	0.912	0.498	0.885
Piece of fabric	6	7	8	9	10
Reflectance	0.291	0.805	0.436	0.868	0.721

$\alpha = 0.01$, whether the reflectance is uniformly distributed or not, namely whether it follows the distribution law $\mathcal{U}(0, 1)$. The observed sample is displayed in Table 1.1.

The null hypothesis of the testing problem is

$$H_0 : F(x) = F_0(x) = \begin{cases} 0 & \text{when} \quad x < 0 \\ x & \text{when} \quad 0 \leq x < 1 \\ 1 & \text{otherwise,} \end{cases}$$

where $F(x)$ is the CDF of the reflectance, here represented by a continuous random variable. The function $F_0(x)$ specified in the null hypothesis is the CDF of the uniform distribution in the interval $[0, 1]$. The alternative hypothesis is $H_1 : F(x) \neq F_0(x)$.

The *R* code for the analysis is the following:

```
> ref=c(0.608,0.533,0.912,0.498,0.885,0.291,0.805,0.436,0.868,0.721)
> plot(ecdf(ref),xlim=c(0,1),verticals=TRUE,xlab="Reflectance",
   main="")
> curve(punif,from=0,to=1,add=TRUE,lty="dashed",lwd=2)
> ks.test(ref,"punif",alternative="two.sided")
```

and the output of `ks.test` is

```
#######################################################################
#One-sample Kolmogorov--Smirnov test
#
#data: ref
#D = 0.336, p-value = 0.1651
#alternative hypothesis: two-sided
#######################################################################
```

The command `plot(ecdf(ref),xlim=c(0,1),verticals=TRUE,xlab="Reflectance",main="")` gives the EDF of the sample, and the argument `xlim` indicates the interval to be visualized on the x axis. With the command `curve(punif,from=0,to=1,add=TRUE,lty="dashed", lwd=2)` we can draw on the same graph the CDF of the uniform distribution. The first argument indicates the type of probability distribution and `punif` corresponds to the $\mathcal{U}(0, 1)$ distribution.

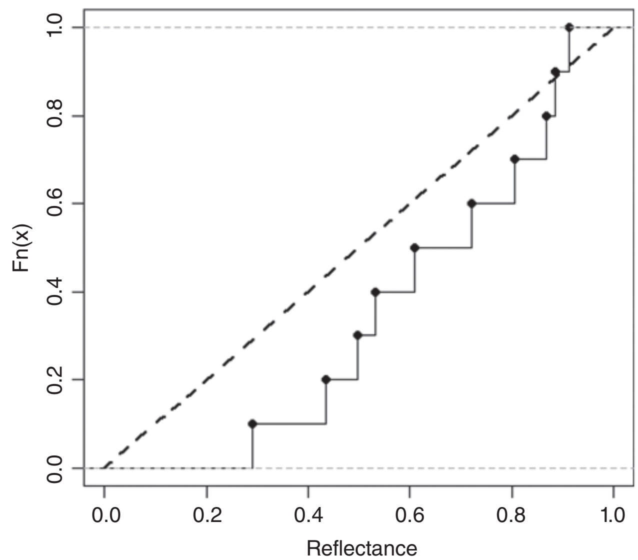

Figure 1.1 Representation of the EDF of reflectance (dots) and CDF of $U(0, 1)$ *(dashed line).*

The argument `lty` defines the line type. The graph is shown in Figure 1.1. It seems we are in the presence of an acceptable level of *goodness-of-fit*.

The first argument of the function `ks.test` indicates the observed data (`ref` in our example), the second one indicates the supposed distribution in the null hypothesis and the third indicates the type of alternative. The default is the two-sided alternative; the options `"less"` or `"greater"` correspond to the one-sided alternatives.

The observed value of the test statistic is `D = 0.336` and it corresponds to the p-value 0.1651, thus there is no empirical evidence to reject the null hypothesis of uniform distribution for the reflectance.

When $F_0(x)$ is not continuous the Kolmogorov test tends to be conservative (Sprent and Smeeton, 2007) but, as we previously noted, methods for computation of p-values have been proposed also for discrete distributions. For example, let us consider the case of a bank where the average waiting time of a customer at the counter is equal to 8 min. We wish to test whether the waiting time follows a $\mathcal{P}(8)$ distribution, that is a *Poisson* distribution with parameter $\tau = 8$, by observing a random sample of $n = 20$ waiting times. The significance level is $\alpha = 0.05$. The null hypothesis is

$$H_0 : F(x) = F_0(x) = \begin{cases} 0 & \text{when} \quad x < 0 \\ \displaystyle\sum_{k=0}^{x} \frac{e^{-8}8^k}{k!} & \text{when} \quad 0 \leq x < \infty \end{cases}$$

and the alternative is $H_1 : F(x) \neq F_0(x)$. The observed data are reported in Table 1.2.

Table 1.2 Random sample of $n = 20$ waiting times (in minutes) of customers at the counter in a bank.

Customer	1	2	3	4	5	6	7	8	9	10
Waiting time	9	8	5	4	4	7	12	6	9	11
Customer	11	12	13	14	15	16	17	18	19	20
Waiting time	10	6	11	7	7	8	11	13	9	9

The commands to perform the test are the following:

```
> wtime=c(9,8,5,4,4,7,12,6,9,11,10,6,11,7,7,8,11,13,9,9)
> ks.test(wtime,"ppois",8)
```

where in the `ks.test` command we have specified the CDF of a *Poisson* as tested distribution, and the value of the related parameter.

No indication about the alternative is specified so that the default option (two-sided) is considered. The output is

```
######################################################################
#One-sample Kolmogorov--Smirnov test
#
#data: x
#D = 0.2166, p-value = 0.305
#alternative hypothesis: two-sided
######################################################################
```

The observed value of the test statistic is 0.2166 and the p-value $= 0.305 > 0.05 = \alpha$ leads to the decision of no rejection of the null hypothesis that the distribution of the waiting time is $\mathcal{P}(8)$. Note that in this situation a warning is generated, due to the presence of ties.

For denoting the specific distribution $F_0(x)$ in the null hypothesis of the Kolmogorov–Smirnov test, the options displayed in Table 1.3 can be used.

1.3.2 A univariate permutation test for symmetry

Sometimes an asymmetric distribution of the observed values of a response variable might be a symptom of abnormalities of the phenomena under study. For example, in the statistical quality control of industrial processes an asymmetry of the distribution of the response may reveal the presence of some problems in the manufacturing process. Let us assume that we are given a random sample of $n = 24$ washers drawn by the whole production of a metallurgical factory. The data consist of differences between the measured external diameters of the washers and the target value equal to 10 μm (Table 1.4).

Table 1.3 Options for the main distributions that can be tested with the Kolmogorov procedure.

Discrete data		Continuous data	
Option	Distribution	Option	Distribution
pbinom	Binomial	pbeta	Beta
pgeom	Geometric	pcauchy	Cauchy
phyper	Hypergeometric	pchisq	Chi-square
ppois	Poisson	pexp	Exponential
		pf	F
		pgamma	Gamma
		pnorm	Normal
		pt	Student's t
		punif	Uniform
		pweibull	Weibull

Observing the density distribution of the observed data it is evident that we are in the presence of right or positive asymmetry (Figure 1.2) because some of the sampled washers present large external diameters (far from the target value) and this can cause a high percentage of waste. We are interested to test whether the whole production of the fabric is characterized by an asymmetric distribution of the external diameters at $\alpha = 0.10$.

By generalizing the problem, let F be the unknown distribution of a continuous variable, and let $\boldsymbol{X} = \{X_1, \dots, X_n\}$ be a random sample of observations from such a population. The test on the symmetry of F around the origin can be formalized by the hypotheses

$$H_0 : F(x) = 1 - F(-x)$$

against

$$H_1 : F(x) \neq 1 - F(-x).$$

To deal with this problem, we can consider the data as differences (Pesarin and Salmaso, 2010). In this case, the observed measurement of each unit is considered as if its sign was randomly assigned. In other words, under the null hypothesis the sign

Table 1.4 Differences (in micrometers) between the external diameters of washers and the target value (10 μm).

1.6	1	−0.8	−1.3	1.4	−0.1	1.1	−1	−0.1	−0.6	0.7	−0.6
2.1	−1	3.5	0.6	−0.2	0.5	0.5	4	1.9	−0.4	0.4	1.4

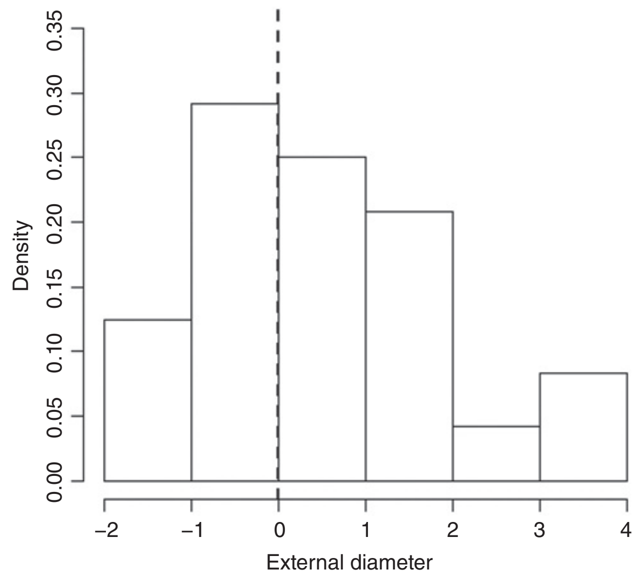

Figure 1.2 Frequency histogram for the differences between the external diameters of washers and the target value ($10\ \mu m$*).*

of each difference is considered as if it were randomly assigned with probability $\frac{1}{2}$. Thus, one way to solve the testing problem is to consider the test statistic given by the absolute value of the sum of the sample values:

$$T = \left| \sum_{i=1}^{n} X_i \right| .$$

Its distribution $F_T(t|\mathbf{X})$, conditional to a given set of observed values $X_1, \ldots, X_n$, is obtained under the assumption that H_0 is true by considering the random attribution in all possible ways of the plus or minus sign to each value with equal probability. Operationally, considering B permuted samples $\mathbf{X}^*$ obtained by randomly attributing the sign $+$ or $-$ to each X_i, and calculating for each permuted sample the corresponding permutation value T^* of T, we can estimate the null distribution of T^* according to the permutation distribution and compute the p-value of the test.

The *R* code for the problem of the external diameters of washers follows:

```
> source("t2p.r")
> source("permsign.r")
> exdiam=c(1.6,1,-0.8,-1.3,1.4,-0.1,1.1,-1,-0.1,-0.6,0.7,-0.6,
+ 2.1,-1,3.5,0.6,-0.2,0.5,0.5,4,1.9,-0.4,0.4,1.4)
> exdiam=array(exdiam,dim=c(length(exdiam),1))
> perm.sign(x=exdiam,B=10000)
```

The output is

```
########################################################################
# Permutation Test for Symmetry
# $p.value
# 0.04029597
########################################################################
```

The source files for the execution of the test are `"t2p.r"` and `"permsign.r"` and they can be loaded with the `source` command. The former includes the code for the computation of the significance level function and the p-values; the latter contains the code for the computation of the test statistic and its permutation distribution. This is done with the function `perm.sign` which computes the statistic T and its permutation distribution, through random permutations of the signs of the observed data. The function can be used also for the two-sample test for paired data. In the case of testing for symmetry it requires two arguments: the array of observed data and the number B of random permutations of signs. The dataset (`exdiam` in this case) has to be set up as array with the `array` command. We consider $B = 10\,000$ permutations. The default value for the number of permutations for B is 1000. The function `perm.sign` invokes the function `t2p` that computes the permutation p-value as the frequency of permutation values of the test statistic that are greater than or equal to the observed one. It is worth noting that the p-value results from the permutation distribution of the statistic of interest that is obtained through random permutations of the signs of the data. Considering that we do not use all the possible permutations but, for computational convenience, only a random sample of B permutations, then performing the test many times, especially for small values of B, one can obtain different permutation distributions, hence different p-values. To remember the sequence of the considered sampled permutations the function `set.seed(seed)` can be used. With this function we can specify a seed to be associated with the set of sampled permutations and use this seed to identify and retrieve the same set of permutations when necessary. For example, if we repeat the analysis without specifying the seed we may obtain a different (but similar) p-value because, by performing the test again, a different sample of 10 000 permutations is drawn:

```
> perm.sign(exdiam, B=10000)
########################################################################
# Permutation Test for Symmetry
# $p.value
# [1] 0.03779622
########################################################################
```

For storing the set of sampled permutations and use it again, it is possible to specify the seed value (eg 1234) before the execution of the permutation test:

```
> set.seed(1234)
> perm.sign(exdiam, B=10000)
```

```
########################################################################
# Permutation Test for Symmetry
# $p.value
# [1] 0.03969603
########################################################################
```

From this moment, every time the command `perm.sign(exdiam,B=10000)` will be preceded by `set.seed(1234)`, the procedure will always return the *p*-value 0.03969603.

In this example, the *p*-value leads to the rejection of H_0 because it does not exceed the significance level $\alpha = 0.10$. The distribution of the external diameters of the whole production is not symmetric thus there is a problem in the production process. The distribution of the external diameters of the whole production is not symmetric and data agree with the hypothesis that there is a problem in the production process. For this reason let us operate a stratification of the data distinguishing the washers coming from plant A (first 12 values) and the ones produced by plant B (last 12 values). Figure 1.3 shows that the asymmetry of the distribution seems to be caused by the production of plant B, because it is not in target and produces a high percentage of washers with too large external diameter. This hypothesis must be tested with a two-sample test on location (see Section 2.4).

The test for symmetry may also be used for testing location on one-sample problems. To be specific, let us suppose that the observed data $Y_1, \ldots, Y_n$ are symmetrically distributed around δ and that $H_0 : \delta = \delta_0$, so that the transformations $X_i = Y_i - \delta_0$ are symmetrically distributed around 0 if and only if H_0 is true.

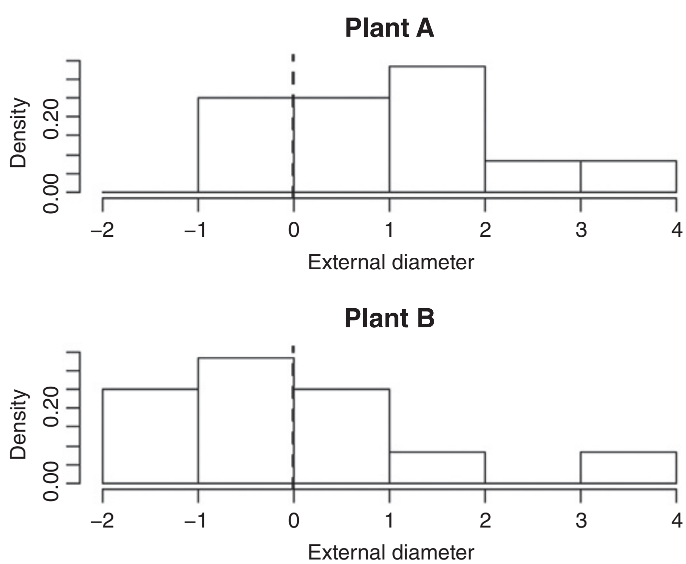

Figure 1.3 Frequency histograms for the differences between the external diameters of washers and the target value (10 μm) by production plant.

1.4 Multivariate one-sample tests

Consider the case of multivariate data. Let $\boldsymbol{X}$ be a multivariate dataset from a sample of size n and assume that the variable under study is q-dimensional. Formally the dataset is $\boldsymbol{X} = \{X_{ih}; i = 1, \ldots, n; h = 1, \ldots, q\}$ where X_{ih} denotes the ith observation of the hth variable. We assume that each of the nq-dimensional observations $\{X_{i1}, \ldots, X_{iq}\}$ comes from a population with CDF $F_i(\boldsymbol{x}; \boldsymbol{\theta})$, with $i = 1, \ldots, n$, $\boldsymbol{x} \in \mathcal{R}^q$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$ is a generic location (vector) parameter. In this section we consider one-sample tests concerning the location parameter $\boldsymbol{\theta}$ and a multivariate extension of the test on symmetry presented in the previous section.

1.4.1 Multivariate rank test for central tendency

The random variable $\boldsymbol{Z}$ taking values in $\mathcal{R}^q$ is said to be diagonally symmetric about $\mathbf{0}$ (q-dimensional vector of zeros) if both $\boldsymbol{Z}$ and $-\boldsymbol{Z}$ have the same CDF $F(z)$, $z \in \mathcal{R}^q$. For absolutely continuous CDFs with density function $f(z)$ the diagonal symmetry can be represented by $f(z) = f(-z)$, $\forall z \in \mathcal{R}^q$ (Puri and Sen, 1971).

Let us consider the multivariate location problem where the n q-variate populations with CDFs $F_1(x; \theta), \ldots, F_n(\boldsymbol{x}; \boldsymbol{\theta})$ are independent and diagonally symmetric about $\mathbf{0}$. We wish to test whether the location (vector) parameter is null, that is $H_0 : \boldsymbol{\theta} = \mathbf{0}$ against $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$. It is worth noting that the condition $F_1(\boldsymbol{x}; \boldsymbol{\theta}) = \ldots = F_n(\boldsymbol{x}; \boldsymbol{\theta})$ is not necessary (Puri and Sen, 1971).

For the general case of q-variate variables with not necessarily independent marginal components, let us consider the following transformation of $\boldsymbol{X}$: $g(\boldsymbol{X}) = \{s_i X_{ih}; i = 1, \ldots, n; h = 1, \ldots, q\}$ where $s_i = +1$ or $s_i = -1$. The number of possible vectors of signs $\boldsymbol{s} = (s_1, \ldots, s_n)'$ is 2^n hence, according to the basic permutation principle, the multivariate permutation distribution is spread over 2^n possible permutations. Under H_0 the permutation distribution is uniform because all the realizations are equally likely with probability equal to 2^{-n} (Puri and Sen, 1971). Hence we can obtain a distribution-free test for the present problem.

Let us now take into account a wide class of multivariate rank tests, useful to solve several different kinds of testing problems. Consider the $n \times q$ matrix $\boldsymbol{R} = [R_{ih}]$ whose generic element R_{ih} represents the rank of $|X_{ih}|$ in the set of values $\{|X_{1h}|, \ldots, |X_{nh}|\}$. No ties are admissible because of the continuity assumption. For each variable (that is for each column of $\boldsymbol{R}$) replace the ranks with the general scores $E^{(h)}(R_{ih})$. For each marginal variable consider the rank order statistics

$$T^{(h)} = \sum_{i=1}^{n} E^{(h)}\left(R_{ih}\right) c_{ih}, \quad \text{for } h = 1, \ldots, q.$$

The weights c_{ih} are the signs of the values X_{ih}, that is $c_{ih} = +1$ if $X_{ih} > 0$ and $c_{ih} = -1$ if $X_{ih} < 0$. Let us denote with $\mathbf{T}$ the q-dimensional vector of statistics $(T^{(1)}, \ldots, T^{(q)})'$. According to the permutation distribution $E(\boldsymbol{T}) = \mathbf{0}$ and $Var(\boldsymbol{T}) = E(\boldsymbol{T}\boldsymbol{T}') = n\boldsymbol{V}$. The matrix $\boldsymbol{V} = [v_{jk}]$ is assumed to be positive definite (if singular it can be replaced by the highest order nonsingular minor of $\boldsymbol{V}$) with elements

$$v_{jk} = (1/n) \sum_{i=1}^{n} E^{(j)}\left(R_{ij}\right) E^{(k)}\left(R_{ik}\right) c_{ij} c_{ik}.$$

The test statistic for the multivariate location problem under study is

$$S = \frac{1}{n}\mathbf{T}'\mathbf{V}^{-1}\mathbf{T},$$

where $\boldsymbol{V}^{-1}$ is the inverse of $\boldsymbol{V}$. Large values of S lead to the rejection of the null hypothesis in favor of the alternative hypothesis of non null central tendency.

According to the scores we can obtain different tests. Some examples are:

- $E^{(h)}(R) = 1$, $h = 1, \ldots, q$: multivariate sign test.
- $E^{(h)}(R) = R$, $h = 1, \ldots, q$: multivariate generalization of the one-sample Wilcoxon signed rank test.
- $E^{(h)}(R)$ is the expected value of the Rth smallest observation of a sample of size n from a chi-square distribution with 1 degree of freedom $h = 1, \ldots, q$: multivariate one-sample normal scores test.

We notice that all the considered tests cannot be applied to one-sided alternatives and they require the continuity assumption for the multivariate response variable. The solution proposed in this subsection for the present problem is the multivariate extension of the one-sample Wilcoxon signed rank test. The multivariate sign test does not require the symmetry, hence it is preferable to the signed rank test when this assumption is not realistic or not plausible. Furthermore, this is the only solution among these rank tests when only the signs of the differences are observed. Otherwise, under the symmetry assumption and when ranks of the sample differences can be determined, the Wilcoxon signed rank test is preferable because it uses more information than the sign test, and then it is more powerful under H_1, that is it rejects H_0 with higher probability when H_0 is not true. The normal score test is less flexible and it is preferable only in specific problems where it is reasonable to replace ordinary ranks with the related normal scores.

Consider a customer satisfaction survey about a recently opened shopping center. A sample of $n = 29$ customers was asked to evaluate 5 different aspects of the shopping center, such as the environmental temperature, the brightness, the presence of sales assistants, the range of products, and the background music volume. Note that these variables represent conditions that can make the shopping experience pleasant if present in the right amount, hence we can say that the best is 'neither too much nor too little'. Thus the evaluations are expressed on a scale from −100 ('too little') to +100 ('too much') and where 0 corresponds to 'just right'. We are interested to test if the mean values of the evaluations are equal to 0 or not at the significance level $\alpha = 0.05$. The sample data are reported in Table 1.5.

An R function, to perform the multivariate one-sample location test based on ranks, is in the package `ICSNP`. To perform the analysis, the following commands should be typed:

```
> library(ICSNP)
> data=read.csv("mall.csv",header=TRUE,sep=";";)
> rank.ctest(X=data,Y=NULL,mu=NULL,scores="rank")
```

Table 1.5 Customer satisfaction survey of a new shopping center.

Temperature	Brightness	Presence of sales assistants	Range of products	Background music volume
20	61	35	57	58
42	40	17	11	9
38	22	46	36	12
0	−1	10	16	−11
100	19	30	−31	−25
−20	−41	47	−14	−94
5	18	43	14	−26
−5	−21	34	48	−45
−40	−78	−68	−10	13
−61	−82	25	7	−88
−83	−76	−71	−89	−72
−77	−84	−30	−58	−86
99	59	56	44	−92
−79	−75	−36	−24	−73
3	−3	27	−22	−12
41	4	50	26	−9
−37	23	29	45	−46
21	−18	8	33	−34
60	62	64	65	66
−19	−42	−44	−13	−23
2	−6	15	32	−93
98	−29	−17	−63	−43
−2	6	−38	−33	−15
−80	−85	−87	−56	−70
1	−4	63	31	−27
−60	−74	−59	−16	−39
−81	−62	−90	28	−57
39	−7	−8	37	24
−35	−65	−91	−69	−28

The command `library(ICSNP)` is necessary to load the package `ICSNP`. Before installing `ICSNP`, the packages `mvtnorm` and `ICS` should be also loaded. The data of the present application can be loaded from the file `mall.csv` with the command `data=read.csv("mall.csv",header=TRUE,sep=";")`. The command `rank.ctest(X=data,Y=NULL,mu=NULL,scores="rank")` performs the test. The command requires a numeric data frame or matrix of data (`X`). The default value for the second argument `Y` is equal to `NULL`, thus a one-sample test is performed. The argument `mu` is a vector indicating the value of the location parameter under the null hypothesis. The default value is `NULL`, that represents the origin, thus for this problem we do not need to specify it. The argument `scores` indicates the score function we want to apply, and `scores="rank"` is the choice to perform a signed rank test. If `scores="sign"` a sign test is performed, whereas if `scores="normal"` a normal score test is performed.

The output is:

```
#########################################################################
# Marginal One Sample Signed Rank Test
#
# data: data
# T=15.7926, df=5, p-value=0.007462
# alternative hyp.: true location not equal to c(0,0,0,0,0)
#########################################################################
```

Thus the observed value of the test statistic is equal to 15.793 and the p-value of the test is equal to 0.007. Since the p-value is less than 0.05 we reject the null hypothesis in favor of the alternative that the vector of means is not equal to $(0, 0, 0, 0, 0)'$.

1.4.2 Multivariate permutation test for symmetry

Let us introduce now the multivariate extension of the permutation procedure for the test on symmetry. In this new problem the unknown distribution under investigation is multivariate. As stated in the univariate case, the problem of testing symmetry is a common problem in Statistical Quality Control where the goal of the analysis could be to test the symmetry of the distribution around the target of two or more characteristics of the product simultaneously considered. In other words, the interest is focused on the symmetry of the marginal distributions but without neglecting the multivariate nature of the problem and the possible dependence among the marginal variables. Formally let $f(\boldsymbol{x})$, with $\boldsymbol{x} \in \mathcal{R}^q$, denote the joint probability function (for discrete variables) or density function (for continuous variables), and $f_i(x)$ the analogous marginal function of the ith component variable. The null hypothesis of the problem is $H_0 : f(\boldsymbol{x}) = f(-\boldsymbol{x})$ and the alternative is $H_1 : f(\boldsymbol{x}) \neq f(-\boldsymbol{x})$. Practically this is equivalent to test the diagonal symmetry of the multivariate distribution.

To face this problem we use the nonparametric combination (NPC) methodology (Pesarin and Salmaso, 2010). We consider the null hypothesis as the intersection of q null sub-hypotheses of symmetry for each marginal distribution and we assume the global null hypothesis of symmetry to be true if each sub-hypothesis of marginal symmetry is true. Conversely the alternative hypothesis of the problem is true if at least one null sub-hypothesis is false. Hence the alternative hypothesis can be considered as the union of q alternative sub-hypotheses of asymmetry. Formally, according to Roy's union-intersection principle (Roy, 1953), we can write the null hypothesis as

$$H_0 = \bigcap_{i=1}^{q} H_{0i}$$

Table 1.6 External and internal diameters of washers (difference from the target value in micrometers).

External diameters											
1.6	1	−0.8	−1.3	1.4	−0.1	1.1	−1	−0.1	−0.6	0.7	−0.6
2.1	−1	3.5	0.6	−0.2	0.5	0.5	4	1.9	−0.4	0.4	1.4
Internal diameters											
−0.1	1	1.6	0.8	1	−0.2	0.1	0.5	−1.4	−0.3	2.4	−0.1
1.7	0.5	−0.1	0.5	1.5	1.2	0	1.4	0.4	−0.3	0.9	−1.1

against the alternative

$$H_1 = \bigcup_{i=1}^{q} H_{1i},$$

where $H_{0i} : f_i(x) = f_i(-x)$ and $H_{1i} : f_i(x) \neq f_i(-x)$.

Under the null hypothesis exchangeability of the signs holds, that is for each q-dimensional vector of observations $(X_{i1}, \ldots, X_{iq})'$, $i = 1, \ldots, n$, the signs can be permuted because the probability (for discrete variables) or density (for continuous variables) of observing $(X_{i1}, \ldots, X_{iq})'$ and $(-X_{i1}, \ldots, -X_{iq})'$ is the same. Let us consider again the industrial example of Section 1.3.2 and assume an interest in controlling both the external and internal diameters of washers. Let us assume to observe a random sample of $n = 24$ measures of differences from the target values of the external and internal diameters of washers drawn from the whole production. The data represent the difference of these measures (in micrometers) from the target values (Table 1.6 and Figure 1.4).

The *R* commands to perform the test are:

```
> source("t2p.r")
> source("comb.r")
> source("permsign.r")
> exdiam=c(1.6,1,-0.8,-1.3,1.4,-0.1,1.1,-1,-0.1,-0.6,0.7,-0.6,
+ 2.1,-1,3.5,0.6,-0.2,0.5,0.5,4,1.9,-0.4,0.4,1.4)
> indiam=c(-0.1,1,1.6,0.8,1,-0.2,0.1,0.5,-1.4,-0.3,2.4,-0.1,
+ 1.7,0.5,-0.1,0.5,1.5,1.2,0,1.4,0.4,-0.3,0.9,-1.1)
> x=array(c(exdiam,indiam),dim=c(length(exdiam),2))
> perm.sign(x,fun="F",B=10000)
```

With the `source()` command use of the functions included in the files `"t2p.r"`, `"comb.r"` and `"permsign.r"` is allowed. The code for performing the test is in `"permsign.r"`; the code for computing the significance level function is included in `"t2p.r"` and the code for the application of the NPC is in `"comb.r"`. The `perm.sign` function for the multivariate case requires three arguments. The first argument is the dataset `x` ($n \times q$ matrix of data) defined as an array. The argument (`fun`) is the combination

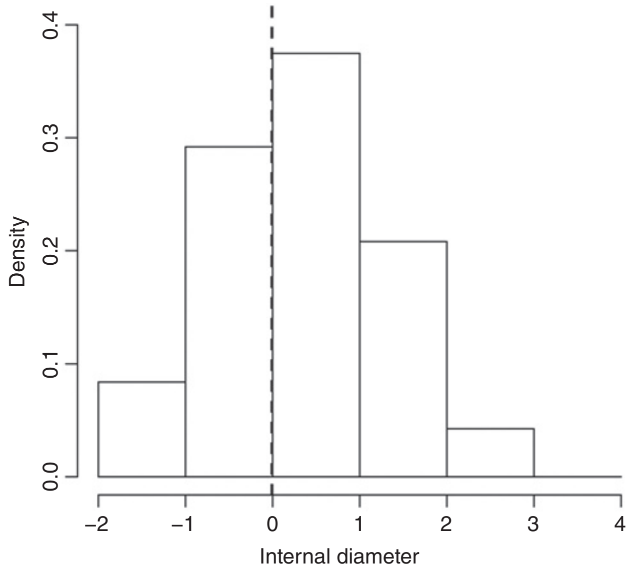

Figure 1.4 Frequency histogram for the differences between the internal diameters of washers and the target value (in micrometers).

function: `"F"`, `"L"` and `"T"` represent Fisher, Liptak and Tippett's combination, respectively. The third argument represents the number of permutations for estimating the null permutation distribution of the test statistic.

The output, using the Fisher combining function and 10 000 permutations, is:

```
########################################################################
# Multivariate Permutation Test for Symmetry
# Combination Function: Fisher
# $p.value
# 0.00549945
########################################################################
```

Thus the p-value of the global test on symmetry is 0.005 and leads to reject the null hypothesis of symmetry of the multivariate distribution at the significance level $\alpha = 0.01$.

1.5 Univariate two-sample tests

In this section we address the problem of comparing two independent samples in the presence of one numerical variable. The data consist of $n = n_1 + n_2$ observations, where n_j denotes the size of the jth sample ($j = 1, 2$). We are considering the most typical two-sample problem: the comparison of central tendencies.

Let us denote with $\boldsymbol{X}_1 = \{X_{1i}, i = 1, \dots, n_1\}$ the sample data from the first population and with $\boldsymbol{X}_2 = \{X_{2i}, i = 1, \dots, n_2\}$ the sample data from the second population. The most important nonparametric solutions for the two-sample location problem are the rank test proposed by Wilcoxon and the permutation two-sample test. These testing procedures are described in the following subsections.

1.5.1 The Wilcoxon (Mann–Whitney) test

The Wilcoxon rank sum test can be applied in the presence of independent random samples when the assumption of normal populations does not hold and the parametric *t*-test cannot be applied. The assumptions of this test are: (1) the data are realizations of continuous independent random variables and independence is assumed both between samples and within samples; (2) random variables generating data of the same sample are identically distributed. Moreover, the Wilcoxon test can also be applied in the presence of ordered categorical data, for example when the response variable represents categorical judges or it takes values in a Likert scale (Kvam and Vidakovic, 2007). The goal consists of comparing the central tendencies of the two samples, to test whether the locations of the respective populations are equal or not (two-sided test) or to test if one location is greater than the other (one-sided test). In some applications, especially (but not only) in medical or pharmacological studies, the problem consists of investigating the presence of a treatment effect represented by a shift of location. The null hypothesis is that of no treatment effect, that is, the samples can be thought as drawn from the same population.

An intuitive approach is to combine both samples into a single pooled ordered sample and then assign increasing ranks to the sample values, with no regard to which population each value comes from. Then the test statistic might be the sum of the ranks assigned to the first sample. Extreme values of the test statistic are empirical evidence in favor of the alternative hypothesis and the rejection region depends on the type of alternative, one-sided or two-sided (Conover, 1999).

By formalizing the problem, let us assume that the CDFs of the compared populations are $F_1(x)$ and $F_2(x)$ and let X_{ji} be generated by the random variable Z_{ji}. There are several ways of specifying the one-sided test. Two of them are the stochastic (or random) effect model and the fixed effect model. According to the former we have $Z_{1i} = \mu + \Delta_{1i} + \epsilon_{1i}$ and $Z_{2i} = \mu + \epsilon_{2i}$, $i = 1, \dots, n_j$, where μ is a constant, ϵ_{ji} $(j = 1, 2)$ are exchangeable random errors, with location equal to zero and scale parameter equal to σ, and Δ_{1i} are nonnegative random variables representing treatment effects. The latter is a special case where $\Delta_{1i} = \theta$ with probability one, with θ an unknown constant parameter. According to the fixed effects model the variances of the two compared populations are equal (homoscedasticity condition) and the two distributions may differ only in the location.

The one-sided test can be presented as a test on stochastic dominance or a test on location shift. In other words the hypothesis that, for example, the first population tends to assume greater values than the second one, can be represented as

$F_1(x) \leq F_2(x)\ \forall x \in \mathcal{R}$ and $F_1(x) < F_2(x)$ for some x or, given the location parameter θ, $F_1(x) = F_2\,(x - \theta)$ with $\theta > 0$. The null hypothesis of the problem is

$$H_0 : F_1(x) = F_2(x)\ \forall x \in \mathcal{R},$$

and the alternative hypothesis is the inequality between $F_1(x)$ and $F_2(x)$ just described or, according to the location-shift model we can write $H_0 : \theta = 0$ and $H_1 : \theta > 0$. The two representations are equivalent if we assume the fixed effect model. Henceforth in this subsection we will consider this model.

The test statistic of the Wilcoxon rank sum test (equivalent to the Mann–Whitney test) is based on the sum of the ranks of the elements of the first sample

$$W = \sum_{i=1}^{n_1} \mathbb{R}\left(X_{1i}\right),$$

where $\mathbb{R}(X_{1i})$ is the (increasing) rank of X_{1i} in the pooled sample $\{X_{11}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{2n_2}\}$, thus it is equal to 1 if X_{1i} is the smallest observed value, to 2 if X_{1i} is the second smallest observed value, and so on up to $n = n_1 + n_2$ for the largest value.

If the null hypothesis is true, the sum of the ranks of the first sample is expected to be similar to that of the second sample, hence when W assumes large values H_0 should be rejected in favor of H_1. Under H_0 the distribution of the statistic W does not depend on unknown parameters but depends on the sample sizes n_1 and n_2, because from the properties of ranks we obtain

$$P\left\{\left[\mathbb{R}\left(X_{11}\right), \ldots, \mathbb{R}\left(X_{1n_1}\right)\right] = (j_1, \ldots, j_{n_1})\right\} = \frac{n_2!}{n!}$$

for all $(j_1, \ldots, j_{n_1})$ obtained from n_1 different elements of the set $(1, 2, \ldots, n_1 + n_2)$. The minimum value of the statistic W is $w_{\min} = 1 + \cdots + n_1 = n_1(n_1 + 1)/2$ and the maximum value is $w_{\max} = (n_2 + 1) + (n_2 + 2) + \cdots + (n_2 + n_1) = n_1(2n_2 + n_1 + 1)/2$. Hence for all $k = w_{\min}, \ldots, w_{\max}$

$$P\{W = k\} = N_k \frac{n_2!}{n!}$$

where N_k is the number of vectors $(j_1, \ldots, j_{n_1})$ satisfying the condition $j_1 + \cdots + j_{n_1} = k$. The exact distribution of W can be computed and tabulated. Clearly when the alternative hypothesis is that the the first population takes smaller values than the second, the null hypothesis is rejected for small values of the test statistic. Finally the two-sided alternative hypothesis of inequality in distribution should be rejected if $W \leq c_1$ or $W \geq c_2$, where c_1 and c_2 are the maximum natural number and the minimum natural number, respectively, verifying the inequalities $\sum_{k=w_{\min}}^{c_1} \Pr\{W = k|H_0\} \leq \alpha/2$ and $\sum_{k=c_2}^{w_{\max}} P\{W = k|H_0\} \leq \alpha/2$. Upper-tail probabilities of the Wilcoxon rank sum test statistic are available from www.wiley.com/go/hypothesis_testing (Hollander and Wolfe, 1999). Under H_0 the means and variances of the sum of ranks W are

$$\mathbb{E}(W) = \frac{n_1(n+1)}{2}$$

and

$$\mathbb{V}(X) = Var(W) = \frac{n_1 n_2(n+1)}{12},$$

respectively. An important result (Bagdonavicius *et al.*, 2011) shows that if the probability distributions of the populations are absolutely continuous, then when $n \to \infty, n_1/n \to p \in (0,1)$ under the null hypothesis

$$Z_{(n_1,n_2)} = \frac{W - \mathbb{E}(W)}{\sqrt{\mathbb{V}(W)}} \overset{d}{\to} Z \sim N(0,1).$$

Hence for large sample sizes the normal approximation of the distribution of W can be used. The Mann–Whitney test is a testing procedure proposed for the same location problems described. For the one-sided test, where the first population is supposed to take greater values than the second in the alternative hypothesis, the test statistic is

$$U = \sum\nolimits_{i=1}^{n_1} \sum\nolimits_{s=1}^{n_2} D_{is},$$

where $D_{is} = I\left(X_{1i} > X_{2s}\right) = 1$ if $X_{1i} > X_{2s}$ and 0 otherwise. With a similar logic the Mann–Whitney test for the lower-tail one-sided test and the two-sided test can be derived. The Mann–Whitney test is equivalent to the Wilcoxon rank sum test.

The basic package of *R* contains the function `wilcox.test`, which computes the Wilcoxon rank sum test statistic and the *p*-value for the one-sample and the two-sample case. By default an exact *p*-value is computed if the sample sizes are less than 50 and there are no ties. Otherwise, a normal approximation is used.

Let us consider the following problem. Before being able to enrol in a first level degree course of Economics at some Italian Universities, students have to do an entrance test related to mathematical skills. The examination consists of a written test and, according to the final score, the students could be asked to participate to a preliminary remedial course. In Table 1.7 the test results for a university of two samples of candidates coming from scientific and classical studies backgrounds, respectively, are shown. Sample sizes are $n_1 = n_2 = 10$. We wish to test whether the mathematical skills of the two groups of students are equivalent against the alternative hypothesis that students coming from scientific studies are better prepared in Mathematics. Let $Score_{scient}$ and $Score_{class}$ denote the random variables representing the test result for a student from a scientific and from a classical high school, respectively. The hypotheses of the testing problem are $H_0 : Score_{scient} \overset{d}{=} Score_{class}$ and $H_1 : Score_{scient} \overset{d}{>} Score_{class}$.

Table 1.7 Results of the examination of mathematical skills for applicants enrolling in a university Economics course, coming from scientific and classical studies backgrounds.

Scientific studies									
82.261	81.191	74.902	87.119	84.410	81.551	90.806	82.818	71.843	82.504
Classical studies									
66.131	89.327	75.119	68.449	77.942	70.756	68.533	65.219	82.723	66.637

The symbol $\overset{d}{=}$ denotes equality in distribution, that is equality of the CDFs. The mathematical notation $\underset{d}{>}$ denotes stochastic dominance, that is the cumulative distribution of $Score_{scient}$ is less than or equal to the CDF of $Score_{class}$ (and the strict inequality is true for some subsets of $\mathcal{R}$). In other words, under the alternative hypothesis the scores of candidates coming from scientific studies tend to be distributed on greater values. The significance level is $\alpha = 0.01$.

Looking at the two-sample density histograms (Figure 1.5), it seems that the score distribution of students from scientific studies is shifted toward greater values than the score of students from classical studies. For testing whether this conclusion based on descriptive statistics can be extended to the corresponding populations we may apply the Wilcoxon–Mann–Whitney test.

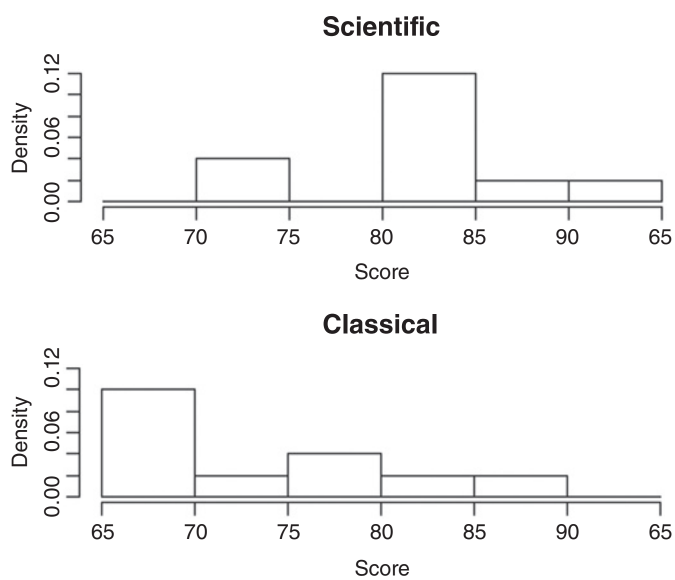

Figure 1.5 Histograms of the results of the examination of mathematical skills for applicants enrolling in a university Economics course, coming from scientific and classical studies backgrounds.

The *R* code for the analysis is the following:

```
> scient=c(82.261,81.191,74.902,87.119,84.410,81.551,90.806,
+ 82.818,71.843,82.504)
> class=c(66.131,89.327,75.119,68.449,77.942,70.756,68.533,
+ 65.219,82.723,66.637)
> wilcox.test(scient,class,alternative="greater")
```

Through the command `wilcox.test(x,y,alternative="greater")` we can perform the one-sided test of Wilcoxon–Mann–Whitney for comparing x and y (vectors of data of the first and second sample, respectively) for testing the hypothesis that the first population tends to take greater values than the second. For testing the opposite one-sided hypothesis the last argument should be `alternative="less"`. For the two-sided alternative of inequality in distribution the syntax is `alternative= "two.sided"`.

The final output after the application of the *R* code to the described problem is:

```
######################################################################
#Wilcoxon rank sum test
#
#data: scient and class
#W = 81, p-value = 0.009272
#alternative hypothesis: true location shift greater than 0
######################################################################
```

The observed value of the test statistic is $W = 81$ and the corresponding p-value is equal to 0.009 that leads to reject the null hypothesis in favor of the hypothesis that the scores of candidates coming from scientific studies tend to be greater at the significance level $\alpha = 0.01$.

A problem, similar from the statistical point of view but related to a completely different application, is the following. An experiment is designed to see if farmed fish exhibit a lower protein content than wild fish caught in the open sea. The experiment is performed on a species of saltwater fish. The goal consists of assessing whether there is a significant negative difference between the percentages of proteins in farmed fish and in wild fish. Let $Prot_{farmed}$ denote the percentage of proteins in farmed fish and $Prot_{sea}$ denote the percentage of proteins in wild fish. The null hypotheses of the problem is $H_0 : Prot_{farmed} \overset{d}{=} Prot_{sea}$ and the alternative is $H_1 : Prot_{farmed} \overset{d}{<} Prot_{sea}$. Two samples of healthy fish, similar in terms of age, gender, weight, etc., of sizes $n_1 = n_2 = 12$ were drawn from the respective populations (Table 1.8).

From a descriptive point of view sample data related to farmed fish tend to be greater than data related to the other sample, as shown in Figure 1.6.

Then the *R* code for this test is:

```
> farm=c(18.85,16.93,19.29,18.31,17.27,18.64,17.82,19.00,19.58,
+ 18.04,17.27,19.19)
```

JWST474-c01 JWST474-Bonnini Printer: Yet to Come June 9, 2014 14:6 Trim: 229mm × 152mm

Table 1.8 Percentage of proteins in two samples of farmed and wild fish.

Farmed fish											
18.85	16.93	19.29	18.31	17.27	18.64	17.82	19.00	19.58	18.04	17.27	19.19
Wild fish											
19.23	19.57	19.50	18.64	18.70	19.54	19.04	20.67	20.71	18.99	19.37	19.06

```
> sea=c(19.23,19.57,19.50,18.64,18.70,19.54,19.04,20.67,20.71,
+ 18.99,19.37,19.06)
> wilcox.test(farm,sea,alternative="less")
```

and the output is:

```
##########################################################################
#Wilcoxon rank sum test
#
#data: farm and sea
#W = 26.5, p-value = 0.004672
#alternative hypothesis: true location shift less than 0
##########################################################################
```

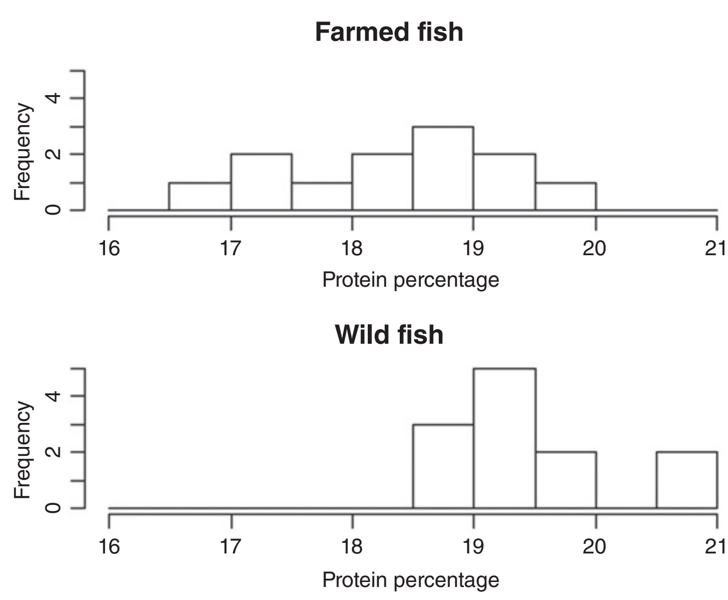

Figure 1.6 Histograms of protein percentage in two samples of farmed and wild fish.

The option for the alternative hypothesis now is `"less"`. The value of the test statistic is $W = 26.5$ and it corresponds to a p-value equal to 0.005, that leads to rejecting the null hypothesis in favor of the hypothesis of stochastic dominance, that is the amount of proteins in farmed fish is lower.

1.5.2 Permutation test on central tendency

The problem of comparing the central tendency of two independent samples in the presence of one numerical variable may be addressed also through a permutation solution. Let us assume homoscedasticity (i.e., equal variances) in the null hypothesis and denote with F_1 and F_2 the compared nondegenerate continuous distributions, both from the same family F. Consider the stochastic dominance problem where in the alternative hypothesis the first population is supposed to take greater values than the second. In other words, H_1 asserts the stochastic dominance of the first population on the second. Note also that H_0 implies exchangeability of observed data with respect to groups, and observed data may be viewed as if they were randomly assigned to two groups but they come from the same population.

The permutation solution does not need the assumption that means and variances of the response variables are finite. It only needs location parameters (mean, median, or others) to be finite and proper sampling indexes for them to be available. Unlike the Wilcoxon test, the permutation test does not require the continuity of the response variables and can be applied also in the presence of ties without any correction or approximation. Furthermore rank transformation is not one-to-one with respect to the dataset $\boldsymbol{X}$, hence the sufficiency property is not satisfied. A transformation of X is a sufficient statistic if it contains all the necessary information for solving the inferential problem on F. Hence the Wilcoxon rank sum test can have some power decay. Instead the permutation test is conditioned to the whole dataset $\boldsymbol{X}$ which is a sufficient statistic for F (Pesarin, 2001).

Let us consider $\boldsymbol{X} = \boldsymbol{X}_1 \biguplus \boldsymbol{X}_2$ with $\biguplus$ denoting vector concatenation, so that the two samples are pooled into one and the first n_1 elements of $\boldsymbol{X}$ correspond to $\boldsymbol{X}_1$ and the remaining n_2 elements to X_2.

A suitable permutation test statistic is $T^* = \overline{X}_1^* - \overline{X}_2^*$ where $\overline{X}_j^* = \sum_{i=1}^{n_j} X_{ji}^*/n_j$, $j = 1, 2$, are the sample means of the first n_1 elements and of the remaining n_2 elements of $\boldsymbol{X}^*$, respectively, and $\boldsymbol{X}^*$ is a permuted dataset, that is a vector obtained by changing the position of elements, or equivalently by randomly assigning n_1 of the observed values to the first sample and the remaining to the second. As a consequence of exchangeability, under the null hypothesis the distribution of T^* can be estimated by permuting the dataset B independent times and computing the value of the statistic corresponding to each permutation (Pesarin and Salmaso, 2010). The p-value of the test is $\lambda = \sum_{b=1}^{B} \mathbb{I}(T^*_{(b)} \geq T^0)/B$ (proportion of T^* permuted values greater than or equal to T^0), where $T^*_{(b)}$ is the value of the statistic related to the bth permutation and T^0 is the observed value of the statistic corresponding to the unpermuted dataset.

Consider the problem of the examination of mathematical skills of applicants enrolling in a university Economics course, coming from scientific and classical studies backgrounds (Table 1.7 and Figure 1.5). The *R* code for the application of the permutation test and the final result are:

```
> source("t2p.r")
> source("perm_2samples.r")
> data=c(scient,class)
> lab=rep(1:2,each=10)
> data_mat=cbind(lab,data)
> T=perm.2samples(data_mat,alt="greater",B=10000)
> T$p.value
[1] 0.00779922
```

The function `perm.2samples(data,alt,B)`, defined in the file `"perm_2samples.r"` (which requires the code in the file `"t2p.r"` for the computation of the p-value), computes the permutation distribution of the test statistic and it requires that data are arranged in a matrix where the first column contains the label of the groups and the second column contains the observed data. We can specify the type of alternative as usual with the options `alt="greater"`, `alt="less"` or `alt="two.sided"`. The default number of permutations is $B = 1000$ but a different number can be specified in the third argument (e.g., $B = 10\,000$). Here the matrix of data is `data_mat`:

```
        lab    data
 [1,]     1     82.261
 [2,]     1     81.191
 ...     ...     ...
[10,]     1     82.504
[11,]     2     66.131
[12,]     2     89.327
 ...     ...     ...
[20,]     2     66.637
```

The p-value is computed with the function `t2p` that computes the significance level function of the test statistic T according to the permutation distribution. The vector `t2p(T)` contains the significance levels corresponding to each of the permutation values of the test statistic and the first element is the significance level corresponding to the observed value of the test statistic, that is the p-value. The p-value can be obtained by typing `T$p.value`. The value 0.008 leads to the rejection of the null hypothesis in favor of the alternative that candidates from scientific schools are better prepared in mathematics. The result is then the same of the Wilcoxon test.

Also for the second problem of fish (Section 2.4.1), we can apply the permutation test:

```
> data=c(farm,sea)
> lab=rep(1:2,each=12)
> data_mat=cbind(lab,data)
```

```
> T=perm.2samples(data_mat,alt="less",B=10000)
> T$p.value
[1] 0.00089991
```

Even in this case the p-value (0.001) is less than α, hence the null hypothesis must be rejected in favor of the alternative hypothesis that the amount of proteins in farmed fish is less than in fish coming from the sea. The conclusion is then similar to that of the Wilcoxon test.

1.6 Multivariate two-sample tests

This section is dedicated to the multivariate extension of the two-sample location problem. The dataset here consists of $n = n_1 + n_2$ observations from two independent q-variate populations, where n_1 and n_2 are the sizes of the two samples. The ith multidimensional observation in the jth group is denoted by $\boldsymbol{X}_{ji} = \left\{X_{ji1}, \ldots, X_{jiq}\right\}$. A rank based and a permutation solution are considered.

1.6.1 Multivariate tests based on rank

Let us consider the problem of testing the identity of two multivariate distributions F_1 and F_2. We shall assume that F_1 and F_2 have a common unspecified form but possible different location vectors.

Let $F_j(\boldsymbol{x})$ be the CDF, belonging to the class of all continuous distribution functions, for the jth population with $j = 1, 2$. According to the fixed effect model and denoting with $\mathbf{Z}_{ji}$ the multivariate random variable from which $\boldsymbol{X}_{ji}$ is assumed to be generated, we have $\mathbf{Z}_{1i} = \boldsymbol{\mu} + \boldsymbol{\theta} + \boldsymbol{\epsilon}_{1i}$ and $\mathbf{Z}_{2i} = \boldsymbol{\mu} + \boldsymbol{\epsilon}_{2i}$, $i = 1, \ldots, n_j, j = 1, 2$, where $\mu \in \mathcal{R}^q$ is a constant vector, ϵ_{ji} are exchangeable q-variate random errors, with location equal to the null vector and variances/covariances matrix equal to $\boldsymbol{\Sigma}$, and $\boldsymbol{\theta} \in \mathcal{R}^q$ is the vector of parameters representing treatment fixed effects. The hypothesis to be tested is

$$H_0 : F_1(\boldsymbol{x}) = F_2(\boldsymbol{x}) = F(\boldsymbol{x}) \text{ for all } \boldsymbol{x} \in \mathcal{R}^q$$

against the general two-sided alternative

$$H_1 : F_1(\boldsymbol{x}) \neq F_2(\boldsymbol{x}).$$

For the translation-type alternatives, the null hypothesis can be written as

$$H_0 : \boldsymbol{\theta} = \mathbf{0}$$

against the alternative

$$H_1 : \boldsymbol{\theta} \neq \mathbf{0}.$$

The alternative hypothesis states that at least one element of the vector of effects θ is not equal to zero, that is at least for one component of the response variable we have a non-null effect.

Let R_{jih} be the rank of X_{jih} in the set $\{X_{11h}, \dots, X_{1n_1h}, X_{21h}, \dots, X_{2n_2h}\}$ for $h = 1, \dots, q$. Let R_h denote the observed n-dimensional vector of ranks related to the hth variable and consider the $n \times q$ matrix $\boldsymbol{R} = [\boldsymbol{R}_1, \dots, \boldsymbol{R}_q]$. Each column of this matrix can be considered as a permutation of the numbers $1, 2, \dots, n$. Thus $\boldsymbol{R}$ can be considered a realization of a $n \times q$ random matrix with $(n!)^q$ possible realizations. Since the q marginal components of the multivariate response are in general stochastically dependent, the joint distribution of the elements of the matrix of ranks will depend on the unknown distribution F even when $H_0 : F_1(\boldsymbol{x}) = F_2(\boldsymbol{x}) = F(\boldsymbol{x})$ holds. Under H_0, the distribution of the matrix of ranks conditional to the observed matrix $\boldsymbol{R}$ over the set $S(R^*)$ of all the possible permutations of the rows $\boldsymbol{R}^*$ under H_0 is uniform. The probability of observing one of the $n! = n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1$ possible realizations is $1/n!$.

A general class of rank scores can be defined as follows:

$$E^{(h)}(R) = g_h\left(\frac{R}{n+1}\right),$$

with $h = 1, \dots, q$ and $1 \le R \le n$. Hence the matrix $\boldsymbol{R}$ can be replaced by $\boldsymbol{E} = [\boldsymbol{E}_1, \dots, \boldsymbol{E}_q]$ where

$$\boldsymbol{E}_h = \left[E^{(h)}\left(R_{11h}\right), \dots, E^{(h)}\left(R_{1n_1h}\right), E^{(h)}\left(R_{21h}\right), \dots, E^{(h)}\left(R_{2n_2h}\right)\right]'$$

with $h = 1, \dots, q$. For each sample and for each of the q variables the average rank score can be computed as

$$T_{j\bullet h} = \frac{\sum_{i=1}^{n_j} E^{(h)}\left(R_{jih}\right)}{n_j}.$$

Under H_0 the average rank scores should be close to the total mean scores $\overline{T}_{\bullet\bullet h} = \left(n_1\overline{T}_{1\bullet h} + n_2\overline{T}_{2\bullet h}\right)/n$ and the contrasts $\left(\overline{T}_{j\bullet h} - \overline{T}_{\bullet\bullet h}\right)$ should stochastically be close to zero for $j = 1, 2$ and $h = 1, \dots, q$. In the presence of $C \ge 2$ groups a suitable test statistic for this problem might be

$$L = \sum_{j=1}^{C} n_j[\overline{\boldsymbol{T}}_j - \overline{\boldsymbol{T}}]'\boldsymbol{V}^{-1}[\overline{\boldsymbol{T}}_j - \overline{\boldsymbol{T}}],$$

where $\overline{\boldsymbol{T}}_j = [\overline{T}_{j\bullet 1}, \dots, \overline{T}_{j\bullet q}]'$, $\overline{\boldsymbol{T}} = [\overline{T}_{\bullet\bullet 1}, \dots, \overline{T}_{\bullet\bullet q}]'$ and $\boldsymbol{V}$ is the permutation covariance matrix of the contrasts $\overline{\boldsymbol{T}}_j - \overline{\boldsymbol{T}}$ under H_0 (see Section 3.5). L is Hotelling's type test statistic used in a parametric test which assumes that data are generated by a multivariate normal distribution. Hence it is a suitable test statistic under normality.

When the multivariate distribution is very far from the normal, this statistic may be not a valid choice. Hence for the two-sample problem a possible test statistic is

$$L = n_1[\overline{\boldsymbol{T}}_1 - \overline{\boldsymbol{T}}]'\boldsymbol{V}^{-1}[\overline{\boldsymbol{T}}_1 - \overline{\boldsymbol{T}}] + n_2[\overline{\boldsymbol{T}}_2 - \overline{\boldsymbol{T}}]'\boldsymbol{V}^{-1}[\overline{\boldsymbol{T}}_2 - \overline{\boldsymbol{T}}].$$

For large values of n and q asymptotic distributions for L are derived (Puri and Sen, 1971). The permutation distribution of L asymptotically, in probability, reduces to the chi-square distribution with $q(C-1)$ degrees of freedom. Thus for $C = 2$ the null hypothesis should be rejected when $L \geq \chi^2_{q;\alpha}$ where α denotes the significance level. According to the type of scores, different tests can be performed:

- For the *Multivariate Multisample Median Test* the score should be

$$E^{(h)}(R) = \begin{cases} 1 & \text{if } R \leq n/2 \\ 0 & \text{otherwise,} \end{cases}$$

  hence $T_{j\bullet h}$ is the proportion of values less than the median in the jth sample for the hth variable.

- For the *Multivariate Rank Sum Test* we define

$$E^{(h)}(R) = \frac{R}{n+1},$$

  hence the statistic $T_{j\bullet h}$ is equal to $(n+1)^{-1}$ times the average rank in the jth sample for the hth variable.

- For the *Normal Scores Test* we put $p = \frac{n-R+1}{n+1}$ and

$$E^{(h)}(R) = z_p,$$

  where z_p is the $(1-p)$th quantile of the standard normal distribution that is the value such that $\Phi(z_p) = 1 - p$, with Φ denoting the CDF of the standard normal distribution.

The median test is preferable when the interest is focused on median comparisons, that is when the median is the location parameter under study. The normal scores test may be used only in specific problems, when it is reasonable to replace ordinary ranks with the related normal scores. Otherwise, among the rank based procedures, the rank sum test is the better choice. All the procedures described in the present subsection are based on the assumption that responses are continuous variables and can be applied only for two-sided tests.

Let us consider again the example related to the entrance test to enrol in the Economics course (Subsection 2.4.1). In this application, the scores of a sample of 20 candidates to enroll in the Economics course are related to mathematical skills and to economic knowledge. Half of the 20 students come from scientific studies and the others come from classical studies. We want to test whether the bivariate distribution of the scores of the two groups are the same, against the alternative, that the distributions of the two groups differ. In other words the goal consists of testing

Table 1.9 Results of the examination of mathematical skills and economic knowledge for applicants enrolling in a university Economics course, coming from scientific and classical studies backgrounds.

	Scientific studies				
Mathematical skills	82.261	81.191	74.902	87.119	84.41
Economic knowledge	87.807	96.851	77.155	99.330	98.570
Mathematical skills	81.551	90.806	82.818	71.843	82.504
Economic knowledge	69.909	75.220	62.405	73.750	81.182
	Classical studies				
Mathematical skills	66.131	89.327	75.119	68.449	77.942
Economic knowledge	79.451	92.708	74.730	66.063	62.818
Mathematical skills	70.756	68.533	65.219	82.723	66.637
Economic knowledge	92.883	99.869	97.991	61.801	84.395

whether the mathematical skills and the economic knowledge of the groups are the same or not. The significance level is $\alpha = 0.10$. Formally by denoting with *Math* and *Econ* the variables representing the scores of the two tests, the null hypothesis is

$$H_0 : (Math_{scient}, Econ_{scient}) \stackrel{d}{=} (Math_{class}, Econ_{class}),$$

that is the null hypothesis is true if the bivariate distributions of the scores on Mathematics and Economics are equal between the groups. The alternative hypothesis is

$$H_1 : (Math_{scient}, Econ_{scient}) \stackrel{d}{\neq} (Math_{class}, Econ_{class}),$$

that is the scores of the two groups are different. Table 1.9 shows the data for the problem.

The *R* package `ICSNP` contains tools for nonparametric multivariate analysis. In particular in this package is the function `rank.ctest(X,Y,mu,scores)` that performs the C-sample location test (with $C \geq 2$) based on marginal ranks, for which the three described score functions are available. For the two-sample test the function requires the $n_j \times q$ matrices of sample observations (`X,Y`) and a vector indicating the difference in the means under the null hypothesis (`mu`). `NULL` indicates no difference between the group means. The argument `scores` requires the type of score test to be performed to be specified. It may be `"sign"` for a sign test, `"rank"` for a rank test or `"normal"` for a normal score test. The *R* code for the analysis is:

```
> library("ICSNP")
> data=read.csv("test_eco.csv",header=TRUE,sep=";")
> X=data[1:10,c(2,3)]
> Y=data[11:20,c(2,3)]
> rank.ctest(X,Y,mu=NULL,scores="rank")
```

The output is:

```
########################################################################
#    Marginal Two Sample Rank Sum Test
#
#data: X and Y
#T=5.8684, df=2, p-value=0.05317
#alternative hypothesis: true location difference is
#    not equal to c(0,0)
########################################################################
```

The function returns some information and in particular the observed value of the test statistic T and the p-value for the test (`p-value`). Note that only the two-sided alternative is available for this test. Thus for our multivariate example, we observe a p-value equal to $0.053 < 0.10$ that leads to the rejection of the null hypothesis of equal distributions in favor of the alternative that the two score distributions are different.

Another interesting problem is the multivariate extension of the application on fish also introduced in Section 2.4. Let us assume an interest in assessing whether there is a difference between farmed and wild fish in terms of percentage of proteins and lean body mass (lbm). The hypothesis we want to test is that there is no difference in the bivariate distribution between the two groups, against the general alternative that the two distributions differ. The observed data on the two samples of fish are reported in Table 1.10. The hypotheses of the problem are.

$$H_0 : \left(Prot_{farmed}, LBM_{farmed}\right) \overset{d}{=} \left(Prot_{sea}, LBM_{sea}\right),$$

against

$$H_1 : \left(Prot_{farmed}, LBM_{farmed}\right) \overset{d}{\neq} \left(Prot_{sea}, LBM_{sea}\right).$$

Table 1.10 Percentages of proteins and lean body mass in two samples of farmed and wild fish.

	Wild fish					
Proteins	20.67	19.34	18.67	19.33	19.42	19.80
Lean body mass	5.62	3.32	4.10	8.50	5.94	4.45
Proteins	19.01	18.91	18.51	19.09	18.99	19.63
Lean body mass	8.24	7.90	5.60	1.97	10.50	5.50
	Farmed fish					
Proteins	17.27	18.55	19.03	18.03	17.17	18.63
Lean body mass	0.98	2.74	7.82	1.33	1.56	5.47
Proteins	17.82	18.40	19.22	19.32	19.11	19.08
Lean body mass	7.98	5.20	1.78	1.23	4.78	5.91

Under the null hypothesis, the percentage of proteins and lbm levels are equal in the two groups of fish while under the alternative the two groups of fish differ.

The *R* code to perform the analysis is:

```
> library("ICSNP")
> bass_mv=read.csv("bass_mv.csv",header=TRUE,sep=";")
> X=bass_mv[1:12,c(2:3)] #first sample
> Y=bass_mv[13:24,c(2:3)] #second sample
> rank.ctest(X,Y,scores="rank")
```

The output is:

```
########################################################################
#    Marginal Two Sample Rank Sum Test
#
#data: X and Y
#T=8.8108, df=2, p-value=0.01221
#alternative hypothesis: true location difference is
#          not equal to c(0,0)
########################################################################
```

The resulting p-value is 0.012 hence at the significance level $\alpha = 0.05$ the null hypothesis should be rejected in favor of the alternative that the two populations of fish differ in the amount of proteins and/or lbm.

1.6.2 Multivariate permutation test on central tendency

A natural extension of the two-sample permutation test to multivariate problems is now presented. In this framework the data is q-dimensional ($q \geq 2$). Often in tests for complex hypotheses, in the presence of many response variables or when several aspects of the distribution are involved, the overall testing problem can be broken down into a finite set of $k > 1$ different partial tests. Note that the number q of responses does not always coincide with k, although for most multivariate location problems $k = q$. As in the problems considered before, the null hypothesis consists of the equality in distribution of two multivariate responses, for example the equality in distribution of each marginal variable. The NPC methodology can be applied. Even in this case H_0 may be properly and equivalently broken down into a finite set of sub-hypotheses H_{0i}, $i = 1, \dots, k$ each appropriate for a partial aspect of interest or for a marginal variable (Pesarin and Salmaso, 2010). Therefore H_0, also called the global null hypothesis, is true if all the H_{0i} are jointly true and thus it may be written as

$$H_0 : \bigcap_{i=1}^{k} H_{0i}.$$

The alternative hypothesis states that at least one of the null sub-hypotheses H_{0i} is not true. Hence the alternative may be represented by the union of k sub-alternatives as

$$H_1 : \bigcup_{i=1}^{k} H_{1i}$$

where each sub-hypothesis H_{1i} is the alternative of H_{0i}. Thus H_1 is true when at least one sub-alternative is true. In this framework, H_1 is called the global alternative hypothesis. For each univariate partial test on central tendency the difference of sample means may be a suitable test statistic and a univariate permutation test on central tendency may be applied (Section 1.5.2). Through the NPC methodology (Section 1.2.2) the partial p-values are combined to obtain a univariate test statistic suitable to solve the multivariate problem.

The main advantage of this procedure, besides the possibility of considering multivariate response variables neither assuming any specific distribution nor specifying the dependence structure among the component variables (but taking it into account implicitly), is in its great flexibility that allows the solution of very complex problems. No continuity assumption is needed, hence it may be applied to continuous, discrete or mixed multivariate variables. It can also be applied to one-sided alternatives and even to complex alternatives where some of the partial sub-hypotheses H_{1i} are two-sided and others one-sided with different possible directions.

Let us consider the example of the examination for students wishing to enroll in the Economics course. The scores in the Mathematics examination and in the Economics examination are considered for the sample of 20 students, 10 of which come from scientific studies and the others come from classical studies. To test whether the distributions of the two populations of students are the same, against the alternative that the distribution of the population coming from a scientific high school is stochastically greater, that is that students from scientific studies tend to get better results, using the multivariate permutation test with $B = 5000$ permutations, the following R code should be applied:

```
> source("dataperm.r")
> source("umultiaspect.r")
> source("t2p.r")
> source("comb.r")
> data=read.csv("test_eco.csv",header=TRUE,sep=";")
> lab=rep(c(1,2),c(10,10))
> data_rev=cbind(lab,data[,2:3])
> perm=dataperm(dataset=data_rev,B=5000)
> l=u_multi_aspect(perm,rep("DM",2),rep(1,2),maspt=0)$P
> T2=comb(l,"F")
> pv=t2p(T2)[1]
> pv
> [1] 0.01898102
```

The scripts are included in the files `"dataperm.r"`, `"umultiaspect.r"`, `"t2p.r"` and `"comb.r"` which can be loaded with the `source` command. The first is useful for permutating the dataset, the second for the calculation of the permutation multivariate distribution of the partial test statistics, the third for obtaining the significance level function and the fourth for the combination of partial tests. After the loading of data from the file `"test_eco.csv"` the dataset must be set in the form:

```
 var1  var2
1  x11  x12
1  x21  x22
...  ...
1  xn1  xn2
2  y11  y12
2  y21  y22
...  ...
2  ym1  ym2
```

that is with the vector of labels of the two groups (of sizes n and m) in the first column and with the variables of interest in the following columns.

The function `dataperm(dataset,B)` performs B random permutations of the multivariate dataset obtaining B permuted datasets. The function `u_multi_aspect` considers the observed and permuted datasets (output of `dataperm`), computes the multivariate distribution of partial test statistics and the corresponding significance levels according to the type of alternative. In this case the test statistics are the differences of means (`"DM"`) for both the partial tests (i.e., for both the variables) hence the second argument is `rep("DM",2)`. In the case of categorical variables, instead of difference of means, it is possible to use the Anderson–Darling statistic with the option `"AD"`. In the present problem the type of alternative, for each partial test, is *group1* > *group2* hence the next argument is `rep(1,2)` or equivalently `c(1,1)`. If we want consider the opposite alternative *group2* > *group1* for both the partial tests we have to specify `rep(-1,2)` or `c(-1,-1)` and for the two-sided alternative `rep(0,2)` or `c(0,0)`. Of course different alternatives can be specified and tested for the two partial tests using `c(-1,1), c(1,0)`, etc. To recover the multivariate significance level function the command is `l=u_multi_aspect(...)$P`. Thus we can obtain the combined test statistic combining the significance level functions and compute the corresponding p-value through the commands `T2=comb(l,"F")` and `l2=t2p(T2)[1]`, respectively. For the combination, the possible choices are `"F"` for the Fisher function, `"L"` for the Liptak rule and `"T"` for the Tippett formula.

The resulting p-value is $0.019 < \alpha = 0.10$ hence, as with the rank test, the null hypothesis should be rejected in favor of the alternative of better preparation of students coming from scientific high schools. The p-value of the permutation test is less than that of the rank test (0.053). This result is consistent with the greater power of the combination based procedure.

References

Bagdonavicius, V., Kruopis, J., Nikulin, M.S. (2011) Non-parametric Tests for Complete Data. John Wiley & Sons, Ltd.

Birnbaum, Z.W., Tingery, F.H. (1951) One-sided confidence contours for probability distribution functions. The Annals of Mathematical Statistics, 22, 4, 592–596.

Coberly, W.A., Lewis, T.O. (1973) A note on the one-sided Kolmogorov–Smirnov test of fit for discrete distribution functions. Annals of the Institute of Statistical Mathematics, 24, 183–187.

Conover, W.J. (1972) A Kolmogorov goodness-of-fit test for discontinuous distributions. Journal of the American Statistical Association, 67, 591–596.

Conover, W.J. (1999) Practical Nonparametric Statistics. John Wiley & Sons, Ltd.

Hollander, M., Wolfe, D.A. (1999) Nonparametric Statistical Methods. John Wiley & Sons, Ltd.

Kolmogorov, A.N. (1933) Sulla determinazione empirica di una legge di distribuzione. Giornale dell'Istituto Italiano degli Attuari, 4, 83–91.

Kvam, P.H., Vidakovic, B. (2007) Nonparametric Statistics with Applications to Science and Engineering. John Wiley & Sons, Ltd.

Marsaglia, G., Wai Wan, T., Jingbo, W. (2003), Evaluating Kolmogorov's distribution. Journal of Statistical Software, 8, 18.

Pesarin, F. (2001). Multivariate Permutation Tests with Applications in Biostatistics. John Wiley & Sons, Ltd.

Pesarin, F., Salmaso, L. (2010) Permutation Tests for Complex Data: Theory, Applications and Software. John Wiley & Sons, Ltd.

Puri, M.L., Sen, P.K. (1971) Nonparametric Methods in Multivariate Analysis. John Wiley & Sons, Ltd.

Roy, S.N. (1953) On a heuristic method of test construction and its use in multivariate analysis. Annals of Mathematical Statistics, 24, 220–238.

Smirnov, N.V. (1939) On the estimation of the discrepancy between empirical curves of distribution for two independent samples. Bulletin Moscow University, 2, 3–16.

Sprent, P., Smeeton, N.C. (2007) Applied NonParametric Statistical Methods. Chapman & Hall/CRC.