Statistical Papers **42**, 1–52 (2001)

Statistical Papers

Nonparametric methods in factorial designs

Edgar Brunner and Madan L. Puri

Received: October 13, 1999; revised version: June 26, 2000

SUMMARY

In this paper, we summarize some recent developments in the analysis of nonparametric models where the classical models of ANOVA are generalized in such a way that not only the assumption of normality is relaxed but also the structure of the designs is introduced in a broader framework and also the concept of treatment effects is redefined. The continuity of the distribution functions is not assumed so that not only data from continuous distributions but also data with ties are included in this general setup. In designs with independent observations as well as in repeated measures designs, the hypotheses are formulated by means of the distribution functions. The main results are given in a unified form. Some applications to special designs are considered, where in simple designs, some well known statistics (such as the Kruskal-Wallis statistic and the χ^2-statistic for dichotomous data) come out as special cases. The general framework presented here enables the nonparametric analysis of data with continuous distribution functions as well as arbitrary discrete data such as count data, ordered categorical and dichotomous data.

Key words: Rank Tests, Factorial Designs, Repeated Measures, Unbalanced Designs, Ordered Categorical Data, Count Data

Contents

1 Independent Observations

1.1 Introduction

The analysis of factorial desings is one of the most important and frequently encountered problems in statistics. In the past, numerous models and procedures were developed under more or less restrictive assumptions on the underlying distribution functions of the observations. These assumptions were relaxed and the models were generalized under different requirements of the applications. For a historical overview, we refer to Brunner and Puri (1996, 2000).

With the exception of a few special cases, the analysis of factorial designs in a nonparametric setup was mainly restricted to designs with one fixed factor, where by *nonparametric* we mean that no specific parametric class of distribution functions is assumed. Thus, there are no parameters by which treatment effects can be defined and hypotheses can be formulated. Therefore, one of the main problems in a nonparametric setup is the formulation of the hypotheses in factorial designs beyond the one-way layout. In some papers, interactions are excluded from the model and only main effects in a two-way layout were considered (Mack and Skilling, 1980; Rinaman, 1983; Groggle and Skillings, 1986; Thopmson and Ammann, 1989). In other papers, main effects and interactions are amalgamated in the hypotheses (Hora and Iman, 1988; Thopmson and Ammann, 1990; Akritas, 1990; Thompson, 1991). Other approaches are restricted to special designs (Patel and Hoel, 1973; Brunner and Neumann, 1986a; Boos and Brownie, 1992; Brunner Puri and Sun, 1995; Marden and Muyot, 1995). All these approaches are of more or less limited meaning for the analysis of real data sets in general factorial designs. In many cases, the terminology 'main effects' and 'interaction' is used in the sense of a linear model.

The problem to define interactions and main effects in nonparametric factorial designs and to formulate hypotheses remained open until Akritas and Arnold (1994) provided the simple idea to formulate the hypotheses in a two-way repeated measures model by contrasts of the distribution functions. This idea seems to be a breakthrough towards a purely nonparametric formulation of hypotheses in higher-way layouts. Several important points should be noted:

(1) This formulation of the hypotheses is a straightforward generalization of the nonparametric hypothesis in the one-way layout.

(2) The hypotheses in the linear model are implied by these nonparametric hypotheses.

(3) The nonparametric hypotheses are not restricted to continuous distribution functions and models with discrete observations are included in this setup.

(4) Under these hypotheses, the asymptotic covariance matrix of a vector of linear contrasts of the rank means has a simple form and can be estimated by the ranks of the observations.

(5) This formulation is not restriced to independent observations and is also valid for repeated measures.

To handle the case of continuous and discontinuos distribution functions in a unified form, Ruymgaart (1980) suggested the use of the so-called normalized version of the true and the empirical distribution function. The combination of this technique with the formulation of the hypotheses by contrasts of the distribution functions provides the basis for the derivation of asymptotic results in general nonparametric factorial designs. The results derived by this approach give a new insight into the so-called 'rank transform method' (Conover and Iman, 1976, 1981 and Lemmer, 1980) and it can easily be seen when the heuristic technique of the 'rank transform' fails.

In this section, we combine the results of some recent papers in this area in a unified form. Not only new procedures for designs with independent observations are derived from this general approach but also some well-known procedures in simple designs come out as special cases. Designs with repeated measures are considerd separately in Section 2.

1.2 Models and Notations

1.2.1 Notations

For a convenient formulation of hypotheses and the relevant statistics in factorial designs, the following matrix notations are used throughout the paper.

Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)'$ be a d-dimensional vector of constants. Hypotheses concerning the components of $\boldsymbol{\mu}$ are formulated by contrast matrices where a matrix $\boldsymbol{C}_{r\times d}$ is called a *contrast matrix* if $\boldsymbol{C}_{r\times d}\mathbf{1}_d = \mathbf{0}_{r\times 1}$ where $\mathbf{1}_d = (1, \ldots, 1)'$ denotes the d-dimensional vector of 1's. In particular, we use the contrast matrix (sometimes called *centering matrix*)

$$\boldsymbol{P}_d = \boldsymbol{I}_d - \tfrac{1}{d}\boldsymbol{J}_d \tag{1.1}$$

where $\boldsymbol{I}_d$ is the d-dimensional unit matrix and $\boldsymbol{J}_d = \mathbf{1}_d\mathbf{1}_d'$ is the $d \times d$ matrix of 1's. Note that $\boldsymbol{P}_d$ is a d-dimensional projection matrix of rank $d-1$, i.e. $\boldsymbol{P}_d^2 = \boldsymbol{P}_d$ and $\boldsymbol{P}_d' = \boldsymbol{P}_d$.

For a technically simple formulation of hypotheses and test statistics in two- and higher-way layouts, we use the Kronecker-product (direct product) and the Kronecker-sum (direct sum) of matrices. The Kronecker-product of two matrices

$$\boldsymbol{A}_{p\times q} = \begin{pmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & & \vdots \\ a_{p1} & \cdots & a_{pq} \end{pmatrix} \quad \text{and} \quad \boldsymbol{B}_{r\times s} = \begin{pmatrix} b_{11} & \cdots & b_{1s} \\ \vdots & & \vdots \\ b_{r1} & \cdots & b_{rs} \end{pmatrix}$$

is defined as

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{pmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1q}\boldsymbol{B} \\ \vdots & & \vdots \\ a_{p1}\boldsymbol{B} & \cdots & a_{pq}\boldsymbol{B} \end{pmatrix}_{pr\times qs}$$

where $\boldsymbol{A} = \boldsymbol{A}_{p\times q}$, $\boldsymbol{B} = \boldsymbol{B}_{r\times s}$ and the Kronecker-product of the matrices $\boldsymbol{A}_i$, $i = 1, \ldots, a$ is written as $\bigotimes_{i=1}^{a} \boldsymbol{A}_i$.

The Kronecker-sum of the two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined as

$$\boldsymbol{A} \oplus \boldsymbol{B} = \left(\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{B} \end{array} \right)_{(p+r)\times(q+s)}$$

and the Kronecker-sum of the matrices $\boldsymbol{A}_i$, $i = 1, \ldots, a$ is written as $\bigoplus_{i=1}^{a} \boldsymbol{A}_i$.

Note that

$$\left(\bigotimes_{i=1}^{a} \boldsymbol{A}_i \right) \left(\bigotimes_{i=1}^{a} \boldsymbol{B}_i \right) = \bigotimes_{i=1}^{a} \boldsymbol{A}_i \boldsymbol{B}_i \quad \text{and} \quad \left(\bigoplus_{i=1}^{a} \boldsymbol{A}_i \right) \left(\bigoplus_{i=1}^{a} \boldsymbol{B}_i \right) = \bigoplus_{i=1}^{a} \boldsymbol{A}_i \boldsymbol{B}_i$$

if the matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are conformable with respect to multiplication.

Factors (in the sense of experimental design) are denoted with capital letters $A, B, C, \ldots$ and the levels of A are numbered by $i = 1, \ldots, a$, the levels of B are numbered by $j = 1, \ldots, b$, etc. If factor B is nested under factor A, this is denoted by $B(A)$.

1.2.2 Nonparametric Model

We consider independent random variables

$$X_{ij} \sim F_i(x), \quad i = 1, \ldots, d, \; j = 1, \ldots, n_i, \tag{1.2}$$

where $F_i(x) = \frac{1}{2}\left[F_i^+(x) + F_i^-(x)\right]$ denotes the normalized-version of the distribution function while $F_i^+(x) = P(X_i \leq x)$ is the right continuous version and $F_i^-(x) = P(X_i < x)$ is the left continuous version of the distribution function. Here, and in the sequel, $X_{ij} \sim F_i(x)$ means that X_{ij} is distributed according to the distribution function $F_i(x)$. This definition of the distribution function includes the case of ties and, moreover ordered categorical data are included in this setup. The normalized version of the distribution function dates back to Lévy (1925) and Kruskal (1952) and was later on used by Ruymgaart (1980), Brunner, Puri and Sun (1995) and Munzel (1999a) among others to derive asymptotic results for rank statistics including the case of ties in a unified form. We use the normalized-version of the distribution function, the empirical distribution function and the counting function. In the sequel, we will drop the expression 'normalized-version' for brevity and when using the above quoted functions, the 'normalized-version' is understood unless stated otherwise. The vector of the distribution functions is denoted by $\boldsymbol{F} = (F_1, \ldots, F_d)'$.

A two-way or a higher-way layout, is described by putting a structure on the index i, i.e. $i = 1, \ldots, d$ is split into $i_1 = 1, \ldots, i_{d_1}$ and $i_2 = 1, \ldots, i_{d_2}$, etc. and the distribution functions $F_1, \ldots, F_d$ are a lexicographic ordering of the higher-way layout distribution functions, e.g., $F_{11}, \ldots, F_{d_1 d_2}$ such that the second index i_2 is changed first.

1.3 Relative Treatment Effects and Hypotheses

1.3.1 Relative Treatment Effects

Since no parameters are involved in the general model (1.2), we use the distribution functions $F_i(x)$ to describe a treatment effect. To this end, we consider the so-called *relative treatment effects*

$$p_i \;=\; \int H(x)dF_i(x), \quad i = 1, \ldots, d, \tag{1.3}$$

where $H(x) = N^{-1}\sum_{i=1}^d n_i F_i(x)$ is the weighted average of all distribution functions in the experiment. The p_i's can be regarded as 'relative effects' with respect to the weighted average $H(x)$. They describe a *tendency* (Kruskal, 1952) of $F_i(x)$ with respect to $H(x)$. If $H(x) \equiv x$, then $p_i = \mu_i = \int x dF_i(x)$ is the expectation (if it exists). In this sense, $p_i = \int H(x)dF_i(x)$ or shortly $p_i = \int H dF_i$ is a generalized expectation.

Since the random variables $X_{rk} \sim F_r(x), k = 1, \ldots, n_r,\ r = 1, \ldots, d$ are independent and identically distributed the relative effect p_i can also be written as a weighted average of the probabilities $P(X_{j1} < X_{i1})$, where the average is taken over $j = 1, \ldots, d$ and i is fixed. Note that $F_j(x) = P(X_{j1} < x) + \frac{1}{2}P(X_{j1} = x)$ and $H(x) = \frac{1}{N}\sum_{j=1}^d n_j \left[P(X_{j1} < x) + \frac{1}{2}P(X_{j1} = x)\right]$. Thus, by (1.3),

$$p_i \;=\; \frac{1}{N}\sum_{j=1}^d n_j \left[P(X_{j1} < X_{i1}) + \frac{1}{2}P(X_{j1} = X_{i1})\right].$$

If the distribution functions $F_1(x), \ldots, F_d(x)$ are continuous then $P(X_{j1} = X_{i1}) = 0$ and p_i reduces to $p_i = \frac{1}{N}\sum_{j=1}^d n_j P(X_{j1} \le X_{i1})$.

We denote by $\boldsymbol{p} = (p_1, \ldots, p_d)' = \int H d\boldsymbol{F}$, the vector of the relative treatment effects. Note that in general, $\boldsymbol{p}$ depends on the sample sizes n_i through $H(x)$. To avoid the dependence on sample sizes, the function $H(x)$ is replaced by the unweighted mean $H^*(x) = \frac{1}{d}\sum_{i=1}^d F_i(x)$ of all distribution functions in the experiment. Thus, the relative effects

$$\pi_i \;=\; \int H^* dF_i, \quad i = 1, \ldots, d, \tag{1.4}$$

do not depend on the sample sizes n_i. In some sense, they correspond to parameters of distribution functions and may be used to formulate nonparametric hypotheses. The vector of these (unweighted) relative effects is denoted by $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_d)'$. Clealy, $\boldsymbol{p} = \boldsymbol{\pi}$ if all sample sizes are equal. For brevity, we shall only discuss the weighted relative treatment effects p_i in this paper.

1.3.2 Hypotheses

In the nonparametric setup introduced above, hypotheses may be either formulated by the distribution functions F_i or by the (unweighted) relative treatment effects π_i. Since we are only discussing the (weighted) relative treatment effects p_i we only shall consider the hypotheses which are formulated by the distribution functions. Let $\boldsymbol{C}$ denote a contrast matrix as given in (1.1). Then these nonparametric hypotheses in their most general form are written as $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$.

For example, the simplest hypothesis is the hypothesis that there is no treatment effect at all. This hypothesis is formulated as $H_0^F : F_1 = \cdots = F_d$ which can formally be written as $H_0^F : \boldsymbol{P}_d\boldsymbol{F} = \boldsymbol{0}$, where $\boldsymbol{P}_d$ is given in (1.1) and $\boldsymbol{0}$ denotes a $d \times 1$ vector of functions which are identically $\boldsymbol{0}$.

More complex hypotheses or hypotheses in higher-way layouts may be formulated by a suitable contrast matrix $\boldsymbol{C}$ as $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$. This formulation of the hypotheses in a nonparametric setup is analogous to the formulation of the hypotheses in the theory of linear models where the hypotheses are formulated in terms of the expectations $\mu_i = \int x dF_i$, i.e. $H_0^\mu : \boldsymbol{C\mu} = \boldsymbol{0}$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)'$. Note that in general,

$$H_0^F : \boldsymbol{CF} = \boldsymbol{0} \;\Rightarrow\; H_0^\mu(\boldsymbol{C}) : \boldsymbol{C\mu} = \boldsymbol{0}$$

since $\boldsymbol{C\mu} = \boldsymbol{C} \int x d\boldsymbol{F} = \int x d(\boldsymbol{CF})$.

The nonparametric hypotheses H_0^F which are based on the distribution functions have been introduced by Akritas and Arnold (1994) and have been further developed and discussed by Akritas, Arnold and Brunner (1997), Akritas and Brunner (1997), Brunner and Puri (1996, 2000) and Brunner, Munzel and Puri (1999). For details, we refer to these papers. Some examples for nonparametric hypotheses are given in Section 1.6.

1.3.3 Estimators

The relative treatment effects p_i are estimated by replacing the distribution functions $F_i(x)$ by their empirical counterparts

$$\widehat{F}_i(x) \;=\; \frac{1}{2}\left[\widehat{F}_i^+(x) + \widehat{F}_i^-(x)\right] = \frac{1}{n_i}\sum_{j=1}^{n_i} c(x - X_{ij}) \tag{1.5}$$

where $c(u) = \frac{1}{2}\left[c^+(u) + c^-(u)\right]$ denotes the counting function and $c^+(u) = 0$ or 1 according as $u <$ or ≥ 0 and $c^-(u) = 0$ or 1 according as $u \leq$ or > 0. The vector of the empirical distribution functions is denoted by $\widehat{\boldsymbol{F}}(x) = (\widehat{F}_1(x), \ldots, \widehat{F}_d(x))'$ or shortly by $\widehat{\boldsymbol{F}} = (\widehat{F}_1, \ldots, \widehat{F}_d)'$. The combined empirical distribution function of the $N = \sum_{i=1}^d n_i$ random variables $X_{11}, \ldots, X_{dn_d}$ is denoted by

$$\widehat{H}(x) \;=\; \frac{1}{N}\sum_{i=1}^{d} n_i \widehat{F}_i(x) = \frac{1}{N}\sum_{i=1}^{d}\sum_{j=1}^{n_i} c(x - X_{ij}) \tag{1.6}$$

and an unbiased estimator for p_i is given by

$$\widehat{p}_i = \int \widehat{H} d\widehat{F}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \frac{1}{N}\left(R_{ij} - \frac{1}{2}\right) = \frac{1}{N}\left(\overline{R}_{i\cdot} - \frac{1}{2}\right) \tag{1.7}$$

where

$$R_{ij} = \frac{1}{2} + N\widehat{H}(X_{ij}) = \frac{1}{2} + \sum_{r=1}^{d}\sum_{s=1}^{n_r} c(X_{ij} - X_{rs}) \tag{1.8}$$

is the (mid)-rank of the random variable X_{ij} among all the N observations. Note that $1/2$ has to be added to $N\widehat{H}(X_{ij})$ in order to get the 'position numbers' of the ordered observations in case of no ties, since $c(0) = 1/2$. Note also that R_{ij} is the midrank in case of ties. Thus, the vector of the relative treatment effects $\boldsymbol{p} = (p_1, \ldots, p_d)'$ is estimated unbiasedly by

$$\widehat{\boldsymbol{p}} = \int \widehat{H} d\widehat{\boldsymbol{F}} = \frac{1}{N}\left(\overline{\boldsymbol{R}}_{\cdot} - \tfrac{1}{2}\boldsymbol{1}_d\right) = \frac{1}{N}\begin{pmatrix} \overline{R}_{1\cdot} - \frac{1}{2} \\ \vdots \\ \overline{R}_{d\cdot} - \frac{1}{2} \end{pmatrix} = \frac{1}{N}\begin{pmatrix} \widehat{p}_1 \\ \vdots \\ \widehat{p}_d \end{pmatrix}, \tag{1.9}$$

where $\overline{\boldsymbol{R}}_{\cdot} = (\overline{R}_{1\cdot}, \ldots, \overline{R}_{d\cdot})'$ denotes the vector of the rank means $\overline{R}_{i\cdot} = n_i^{-1}\sum_{k=1}^{n_i} R_{ik}$. The notation given in (1.9) enables a simple and short presentation of the asymptotic theory of rank statistics in nonparametric factorial designs.

1.4 Asymptotic Theory

1.4.1 Basic Results and Assumptions

In this Section, some asymptotic properties of the statistic $\widehat{\boldsymbol{p}}$ are given and the asymptotic normality of $\sqrt{N}\boldsymbol{C}\widehat{\boldsymbol{p}}$ is derived under the hypothesis $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$ where $\boldsymbol{C}$ is a suitable contrast matrix to formulate the hypothesis. To derive the asymptotic results, the following weak regularity conditions are needed.

ASSUMPTIONS 1.1

(a) $N = \sum_{i=1}^{d} n_i \to \infty$,

(b) $N/n_i \leq N_0 < \infty, \quad i = 1, \ldots, d$,

(c) $\sigma_i^2 = Var[H(X_{i1})] > 0$, $i = 1, \ldots, d$, *where* $H(x) = N^{-1}\sum_{i=1}^{d} n_i F_i(x)$.

Below, it will be stated separately for each theorem or proposition which of these assumptions are needed to prove the results. First, conditions for the consistency of the estimators $\widehat{p}_i$ are given.

Proposition 1.2 (Consistency) *Let $X_{ij} \sim F_i(x)$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$ be independent random variables and let p_i and $\widehat{p}_i$ be as given in (1.3) and (1.7) respectively.. Then, under the assumptions 1.1 (a) and (b), $E(\widehat{p}_i - p_i)^2 \to 0$, $i = 1, \ldots, d$, as $n \to \infty$.*

Proof: see Brunner and Puri (2000). □

The next theorem is one of the basic results in the theory of rank tests. It provides a sequence of independent random variables which has, asymptotically, the same distribution as a certain sequence of non-independent random variables. Here and in the sequel, the asymptotic equivalence of two sequences of random variables U_N and T_N is denoted by $U_N \doteq T_N$.

Theorem 1.3 (Asymptotic Equivalence) *Let $X_{ij} \sim F_i(x)$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$ be independent random variables. Then, under the assumptions 1.1 (a) and (b),*

$$\sqrt{N} \int \widehat{H} d\left(\widehat{\boldsymbol{F}} - \boldsymbol{F}\right) \;\; \doteq \;\; \sqrt{N} \int H d\left(\widehat{\boldsymbol{F}} - \boldsymbol{F}\right) \; = \; \sqrt{N}(\overline{\boldsymbol{Y}}_{\cdot} - \boldsymbol{p}),$$

where $\overline{\boldsymbol{Y}}_{\cdot} = (\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{d\cdot})'$ is a vector of independent (unobservable) random variables $\overline{Y}_{i\cdot} = n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}$, $i = 1, \ldots, d$, and where $Y_{ij} = H(X_{ij})$.

Proof: See Akritas, Arnold and Brunner (1997) or Brunner and Puri (2000). □

The quantity $Y_{ij} = H(X_{ij})$ is called *asymptotic rank transform (ART)* because Y_{ij} is asymptotically equivalent to $\widehat{Y}_{ij} = \widehat{H}(X_{ij})$, (Akritas, 1990). We note that $\sqrt{N} \int H d\widehat{\boldsymbol{F}} = \sqrt{N}\, \overline{\boldsymbol{Y}}_{\cdot}$ is a vector of independent (unobservable) random variables and thus, the covariance matrix of $\sqrt{N}\, \overline{\boldsymbol{Y}}_{\cdot}$ is a diagonal matrix, viz.

$$\boldsymbol{V}_N \;=\; Cov\left(\sqrt{N}\, \overline{\boldsymbol{Y}}_{\cdot}\right) = N \bigoplus_{i=1}^{d} \frac{1}{n_i} \sigma_i^2, \tag{1.10}$$

where σ_i^2 is given in the assumption 1.1 (c).

1.4.2 Asymptotic Normality under $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$

It should be pointed out that the statement of Theorem 1.3 is that $\sqrt{N}(\overline{\boldsymbol{Y}}_{\cdot} - \boldsymbol{p})$ is asymptotically equivalent to the random vector $\sqrt{N}\left(\widehat{\boldsymbol{p}} - \int \widehat{H} d\boldsymbol{F}\right)$ where $\int \widehat{H} d\boldsymbol{F}$ is unobservable. However, under the hypothesis $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$ it follows that

$$\sqrt{N}\boldsymbol{C}\left(\widehat{\boldsymbol{p}} - \int \widehat{H} d\boldsymbol{F}\right) \;=\; \sqrt{N}\boldsymbol{C}\widehat{\boldsymbol{p}} - \sqrt{N} \int \widehat{H} d(\boldsymbol{CF}) \;=\; \sqrt{N}\boldsymbol{C}\widehat{\boldsymbol{p}}.$$

Note that the random vector $\int \widehat{H} d\boldsymbol{F}$ vanishes under $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$ and thus, by Theorem 1.3,

$$\sqrt{N}\boldsymbol{C}\widehat{\boldsymbol{p}} \;\; \doteq \;\; \sqrt{N}\boldsymbol{C}\overline{\boldsymbol{Y}}_{\cdot}$$

under $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$ where $\sqrt{N}\boldsymbol{C}\widehat{\boldsymbol{p}}$ is a vector of (observable) linear rank statistics. These considerations are the key point for the derivation of rank tests in factorial designs. They were introduced by Akritas and Arnold (1994). Finally, the asymptotic normality of $\sqrt{N}\,\overline{\boldsymbol{Y}}_.$ follows immediately from the Central Limit Theorem since the components of $\overline{\boldsymbol{Y}}_.$ are means of independent and identically distributed random variables which are also uniformly bounded. We summarize the above considerations in the following

THEOREM 1.4 (ASYMPTOTIC NORMALITY) *Let $X_{ij} \sim F_i(x)$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$, be independent random variables and assume that $\sigma_i^2 \geq \sigma_0^2 > 0$ where σ_i^2 is given in the assumption 1.1 (c). Let $\boldsymbol{V}_N$ be as given in (1.10). Then, under the assumptions 1.1 (a), (b) and (c) and under the hypothesis $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$, the statistic $\sqrt{N}\boldsymbol{C}\widehat{\boldsymbol{p}}$ has, asymptotically, a multivariate normal distribution with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{CV}_N\boldsymbol{C}'$.*

1.4.3 Estimation of the Asymptotic Variances

The asymptotic variances σ_i^2, $i = 1, \ldots, d$, are unknown. They can easily be estimated from the ranks R_{ij}. An L_2-consistent estimator of σ_i^2 is given in the following theorem.

THEOREM 1.5 (VARIANCE ESTIMATORS) *Let $X_{ij} \sim F_i(x)$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$, be independent random variables and assume that $\sigma_i^2 \geq \sigma_0^2 > 0$. Then, under the assumptions 1.1 (a), (b) and (c), $E(\widehat{\sigma}_i^2/\sigma_i^2 - 1)^2 \to 0$ as $N \to \infty$, where*

$$\widehat{\sigma}_i^2 \;=\; \frac{1}{N^2(n_i - 1)} \sum_{j=1}^{n_i} \left(R_{ij} - \overline{R}_{i\cdot}\right)^2, \quad \overline{R}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} R_{ij}\,, \quad i = 1, \ldots, d, \qquad (1.11)$$

where R_{ij} is the rank of X_{ij} among all the N observations. Moreover, $\widehat{\boldsymbol{V}}_N \boldsymbol{V}_N^{-1} \xrightarrow{p} \boldsymbol{I}_d$ where $\widehat{\boldsymbol{V}}_N = N \bigoplus_{i=1}^{d} \frac{1}{n_i} \widehat{\sigma}_i^2$.

PROOF: see Brunner and Puri (2000). □

It should be noted that in some special cases, all or some of the variances σ_i^2 may be equal under H_0^F. Thus, the corresponding estimators may be pooled to have a better estimator for the common variance. For example, in the one-way layout under the hypothesis $H_0^F : \boldsymbol{P}_d\boldsymbol{F} = \boldsymbol{0}$, it follows that $\sigma_1^2 = \cdots = \sigma_d^2 = \sigma^2$ which is estimated consistently by

$$\widehat{\sigma}_N^2 \;=\; \frac{1}{N^2(N-1)} \sum_{i=1}^{d} \sum_{j=1}^{n_i} \left(R_{ij} - \frac{N+1}{2}\right)^2 \qquad (1.12)$$

since $\overline{R}_{\cdot\cdot} = (N + 1)/2$. In case of no ties, $\widehat{\sigma}_N^2$ reduces to $\widehat{\sigma}_N^2 = (N + 1)/(12N)$. This means that it is not necessary to give a 'correction for ties' since $\widehat{\sigma}_i^2$ given in (1.11) and $\widehat{\sigma}_N^2$ given in (1.12) automatically accomodate for ties.

1.4.4 Asymptotic Normality under Fixed Alternatives

The results of this subsection are used to derive confidence intervals for relative treatment effects. Therefore, we restict the considerations to the component i of the vector $\widehat{\boldsymbol{p}}$. The asymptotic multivariate distribution $\sqrt{N}\ \boldsymbol{C}\widehat{\boldsymbol{p}}$ has a rather involved covariance matrix which we shall not consider in this paper. We refer to Puri (1964).

THEOREM 1.6 (ASYMPTOTIC NORMALITY UNDER FIXED ALTERNATIVES) *Let $X_{ij} \sim F_i(x)$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$, be independent random variables. Furthermore, let*

$$\begin{aligned} Z_{ik} &= \frac{1}{N}[NH(X_{ik}) - n_i F_i(X_{ik})], \\ Z_{rs}^{(-i)} &= \frac{1}{N}[NH(X_{rs}) - (N - n_i)H^{(-i)}(X_{rs})], \quad r \neq i, \end{aligned}$$

where

$$H^{(-i)}(x) = \frac{1}{N - n_i}\sum_{r \neq i}^{d} n_r F_r(x) \tag{1.13}$$

denotes the weighted average of all distribution functions without the distribution function $F_i(x)$. Finally, let $\sigma_i^2 = Var(Z_{i1})$ and $\tau_{r:i}^2 = Var(Z_{r1}^{(-i)})$ and assume that $\sigma_i^2, \tau_{r:i}^2 \geq \sigma_0^2 > 0$, $i \neq r = 1, \ldots, d$. Then, under the assumptions 1.1 (a) and (b), the statistic $\sqrt{N}(\widehat{p}_i - p_i)$ has, asymptotically, a normal distribution with expectation 0 and variance

$$s_i^2 = \frac{N}{n_i}\sigma_i^2 + \frac{N}{n_i^2}\sum_{r \neq i}^{d} n_r \tau_{r:i}^2, \quad i = 1, \ldots, d. \tag{1.14}$$

PROOF: The proof follows easily from Theorem 1.3 by noting that the random variables Z_{ik} and $Z_{rs}^{(-i)}$ are independent, $k = 1, \ldots, n_i$, $s = 1, \ldots, n_r$, $i \neq r = 1, \ldots, d$. □

The unknown variances σ_i^2 and $\tau_{r:i}^2$ in (1.14) can be estimated easily by using three different types of rankings. The estimators are given in the following theorem.

THEOREM 1.7 (VARIANCE ESTIMATOR FOR s_i^2) *Let $R_{ik}^{(i)}$ denote the rank of X_{ik} among all the n_i observations within treatment level i (within-ranks), $i = 1, \ldots, d$, and let $R_{rs}^{(-i)}$ denote the rank of X_{rs} among all the $(N - n_i)$observations without the observations $X_{i1}, \ldots, X_{in_i}$ within treatment level i (partial ranks). Further let*

$$\widehat{\sigma}_i^2 = \frac{1}{N^2(n_i - 1)}\sum_{k=1}^{n_i}\left(R_{ik} - R_{ik}^{(i)} - \overline{R}_{i\cdot} + \frac{n_i + 1}{2}\right)^2, \tag{1.15}$$

$$\widehat{\tau}_{r:i}^2 = \frac{1}{N^2(n_r - 1)}\sum_{s=1}^{n_r}\left(R_{rs} - R_{rs}^{(-i)} - \overline{R}_{r\cdot} + \overline{R}_{r\cdot}^{(-i)}\right)^2, \quad r \neq i, \tag{1.16}$$

where $\overline{R}_{r\cdot}^{(-i)} = n_r^{-1}\sum_{s=1}^{n_r} R_{rs}^{(-i)}$ denotes the mean of the partial ranks $R_{rs}^{(-i)}$ within treatment level i. If $\sigma_i^2, \tau_{r:i}^2 \geq \sigma_0^2 > 0$, $i \neq r = 1, \ldots, d$, then, under the assumptions 1.1 (a) and (b),

$$\widehat{s}_i^2 = \frac{N}{n_i}\widehat{\sigma}_i^2 + \frac{N}{n_i^2}\sum_{r\neq i}^{d} n_r \widehat{\tau}_{r:i}^2 \tag{1.17}$$

is a consistent estimator of s_i^2 given in (1.14) in the sense that $E(\widehat{s}_i^2/s_i^2 - 1)^2 \to 0$ as $N \to \infty$.

PROOF: The proof follows by using the same techniques as in the proof of Theorem 3.3 of Brunner and Puri (2000) and is therefore omitted. □

1.5 Statistics

To test the nonparametric hypothesis $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$, three different types of statistics are considered.

1. Two quadratic forms (explained below in (a) and (b)) based on $\sqrt{N}\boldsymbol{C}\widehat{\boldsymbol{p}}$ are used to detect *general alternatives* of the form $\boldsymbol{Cp} \neq \boldsymbol{0}$.
   (a) The so-called Wald-type statistic (WTS) uses a generalized inverse of the covariance matrix $\boldsymbol{CV}_N\boldsymbol{C}'$ to generate the quadratic form, where the unknown covariance matrix $\boldsymbol{V}_N$ is replaced by the consistent estimator $\widehat{\boldsymbol{V}}_N$ given in Theorem 1.5.
   (b) The so-called ANOVA-type statistic (ATS) is also used to detect general alternatives of the form $\boldsymbol{Cp} \neq \boldsymbol{0}$. Compared with WTS, its small sample properties are more desirable.
2. Linear rank statistics of the form $\sqrt{N}\boldsymbol{w}'\boldsymbol{C}\widehat{\boldsymbol{p}}$ are used to detect the special *patterned alternatives* of the form $\boldsymbol{w}'\boldsymbol{Cp} \neq 0$, where $\boldsymbol{w} = (w_1, \ldots, w_d)'$ is a vector of known constants corresponding to the conjectured pattern. In particular, this includes the cases of *ordered alternatives* for $\boldsymbol{w} = (1, 2, 3, \ldots, d)'$ or $\boldsymbol{w} = (d, \ldots, 3, 2, 1)'$. Moreover, linear rank statistics are used to derive confidence intervals for relative treatment effects. In view of the discussion in Sections 1.3.1 and 1.3.2, we restrict ourselves either to the case of $d = 2$ treatments or to equal sample sizes $n_1 = \cdots = n_d = n$.

1.5.1 Quadratic Forms

A nonparametric hypothesis of the form $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$ can be tested by a quadratic form

$$\begin{aligned} Q_N^*(\boldsymbol{C}) &= \sqrt{N}(\boldsymbol{C}\widehat{\boldsymbol{p}})'\,\boldsymbol{A}\,\sqrt{N}(\boldsymbol{C}\widehat{\boldsymbol{p}}) \\ &= N\cdot\widehat{\boldsymbol{p}}'\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}\widehat{\boldsymbol{p}}, \end{aligned}$$

where $\boldsymbol{C}$ is the contrast matrix by which the hypothesis is formulated and $\boldsymbol{A}$ is a suitable symmetric matrix. Both matrices depend on the hypothesis and on the structure of the design.

Wald-Type Statistics In this paragraph, we consider a quadratic form to test hypotheses of the form $H_0^F : \boldsymbol{CF} = \mathbf{0}$ in an arbitrary experimental design where $\boldsymbol{C}$ has to be chosen appropriately. First, consider

$$Q_N^*(\boldsymbol{C}) \;=\; N \cdot \widehat{\boldsymbol{p}}'\boldsymbol{C}'[\boldsymbol{C}\boldsymbol{V}_N\boldsymbol{C}']^-\boldsymbol{C}\widehat{\boldsymbol{p}},$$

where $[\boldsymbol{C}\boldsymbol{V}_N\boldsymbol{C}']^-$ denotes a g-inverse of $\boldsymbol{C}\boldsymbol{V}_N\boldsymbol{C}'$ where the covariance matrix $\boldsymbol{V}_N$ is given in (1.10) and is assumed to be of full rank according to the assumption 1.1 (c). Thus, under $H_0^F : \boldsymbol{CF} = \mathbf{0}$, the quadratic form Q_N^* has, asymptotically, a χ_f^2-distribution with $f = r(\boldsymbol{C})$ degrees of freedom. Since $\boldsymbol{V}_N$ is unknown in general, it is replaced by the consistent estimator $\widehat{\boldsymbol{V}}_N$ given in Theorem 1.5. The quadratic form

$$Q_N(\boldsymbol{C}) \;=\; N \cdot \widehat{\boldsymbol{p}}'\boldsymbol{C}'[\boldsymbol{C}\widehat{\boldsymbol{V}}_N\boldsymbol{C}']^-\boldsymbol{C}\widehat{\boldsymbol{p}} \tag{1.18}$$

is called *Wald-type statistic (WTS)* (or rank version of WTS) and has, asymptotically, also a χ_f^2-distribution with $f = r(\boldsymbol{C})$ degrees of freedom. However, very large sample sizes are needed to achieve a good approximation by this distribution. Therefore, another quatratic form should be used for medium or small sample sizes which is considered below.

ANOVA-Type Statistics The hypothesis $H_0^F : \boldsymbol{CF} = \mathbf{0}$ can be formulated equivalently as $H_0^F : \boldsymbol{MF} = \mathbf{0}$ where $\boldsymbol{M} = \boldsymbol{C}'(\boldsymbol{CC}')^-\boldsymbol{C}$ is a projection matrix. Note that all elements of $\boldsymbol{M}$ are known constants and $\boldsymbol{M}$ does not depend on the special choice of the g-inverse $(\boldsymbol{CC}')^-$. In many cases (in all complete crossed-classified designs, for example), the contrast matrix $\boldsymbol{C}$ can be chosen such that all diagonal elements of $\boldsymbol{M}$ are identical to m, say. This leads to the simplified form of the Approximation Procedure given below.

To test the hypothesis $H_0^F : \boldsymbol{CF} = \mathbf{0}$, consider the quadratic form $Q_N(\boldsymbol{M}) = N \cdot \widehat{\boldsymbol{p}}'\,\boldsymbol{M}\,\widehat{\boldsymbol{p}}$. The asymptotic distribution of Q_N under the hypothesis is given in Theorem 1.8 and a small sample approximation is given in Approximation Procedure 1.9.

THEOREM 1.8 *Let $\boldsymbol{M} = \boldsymbol{C}'(\boldsymbol{CC}')^-\boldsymbol{C}$ and let $\boldsymbol{V}_N$ be as given in (1.10). Then, under the assumptions 1.1 (a) and (b) and under the hypothesis $H_0^F : \boldsymbol{CF} = \mathbf{0}$, the quadratic form $Q_N(\boldsymbol{M}) = N \cdot \widehat{\boldsymbol{p}}'\,\boldsymbol{M}\,\widehat{\boldsymbol{p}}$ has, asymptotically, a weighted χ^2-distribution, i.e. the same distribution as of $\sum_{i=1}^d \lambda_i U_i$ where the U_i are independent random variables each having a χ_1^2-distribution and the λ_i are the eigenvalues of $\boldsymbol{M}\boldsymbol{V}_N\boldsymbol{M}$.*

PROOF: First note that $\boldsymbol{MF} = \mathbf{0} \iff \boldsymbol{CF} = \mathbf{0}$ since $\boldsymbol{C}'(\boldsymbol{CC}')^-$ is a generalized inverse of $\boldsymbol{C}$. Thus, under H_0^F, by Theorem 1.4, $\sqrt{N}\,\boldsymbol{M}\widehat{\boldsymbol{p}}$ has, asymptotically, a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{M}\boldsymbol{V}_N\boldsymbol{M}$. From this, it follows that $Q_N(\boldsymbol{M}) = N \cdot (\boldsymbol{M}\widehat{\boldsymbol{p}})'\,\boldsymbol{M}\,\boldsymbol{M}\,\widehat{\boldsymbol{p}} = N \cdot \widehat{\boldsymbol{p}}'\,\boldsymbol{M}\,\widehat{\boldsymbol{p}}$ has, asymptotically, the same distribution as of $\sum_{i=1}^d \lambda_i U_i$, which has a weighted χ^2-distribution (see e.g. Graybill, 1976, p.136). □

APPROXIMATION PROCEDURE 1.9 *Let $M = C'(CC')^{-}C$ and assume that the diagonal elements m_{ii} of M are identical to m, say, i.e. $m_{ii} \equiv m$. Further let $\Lambda_d = diag\{n_1, \ldots, n_d\}$. Then, under the assumptions of Theorem 1.8, the distribution of the statistic*

$$F_N(M) = \frac{N}{m \cdot tr(\widehat{V}_N)} \cdot \widehat{p}' M \widehat{p} = \frac{Q_N(M)}{m \cdot tr(\widehat{V}_N)} \tag{1.19}$$

can be approximated by the central $F(\widehat{f}_1, \widehat{f}_0)$-distribution with estimated degrees of freedom

$$\widehat{f}_1 = m^2 \cdot \frac{\left[tr(\widehat{V}_N)\right]^2}{tr(M\widehat{V}_N M\widehat{V}_N)} = (Nm)^2 \cdot \frac{\left(\sum_{i=1}^d \widehat{\sigma}_i^2/n_i\right)^2}{tr(M\widehat{V}_N M\widehat{V}_N)} \quad and \tag{1.20}$$

$$\widehat{f}_0 = \frac{\left[tr(\widehat{V}_N)\right]^2}{tr\left(\widehat{V}_N^2(\Lambda_d - I_d)^{-1}\right)} = \frac{\left(\sum_{i=1}^d \widehat{\sigma}_i^2/n_i\right)^2}{\sum_{i=1}^d \widehat{\sigma}_i^4/[n_i^2(n_i-1)]}, \tag{1.21}$$

where $\widehat{\sigma}_i^2$ is given in (1.11) and $tr(\cdot)$ denotes the trace of a square matrix.

For the derivation of this approximation procedure, see Brunner, Dette and Munk (1997) where also the the more general case is considered where M does not have identical diagonal elements and the accuracy of the approximation is verified by some simulation studies.

1.5.2 Patterned Alternatives

The method of Page (1963) and Hettmansperger and Norton (1987) is used to derive test statistics which are especially sensitive against a conjectured patterned alternative. The estimated treatment effects are weighted by a set of constants $w_1, \ldots, w_d$ reproducing the conjectured pattern of the alternative which has to be specified in advance. Let $w = (w_1, \ldots, w_d)'$ denote the vector of the weights w_i. Then under $H_0^F : CF = 0$, the linear rank statistic

$$L_N(w) = \sqrt{N}\, w' C \widehat{p} \tag{1.22}$$

has, asymptotically, a normal distribution with mean 0 and variance

$$\sigma_N^2 = w' C V_N C' w \,,$$

which can be estimated consistently by

$$\widehat{\sigma}_N^2 = w' C \widehat{V}_N C' w \,,$$

where $\widehat{V}_N$ is given in Theorem 1.5. Thus the statistic $T_N(w) = L_N(w)/\widehat{\sigma}_N$ has, asymptotically, a standard normal distribution under $H_0^F : CF = 0$. For small sample sizes, the distribution of the statistic $L_N(w)/\widehat{s}_N$ may be approximated by a central $t_{\widehat{f}}$-distribution with

$$\widehat{f} = \frac{\left(\sum_{i=1}^d q_i^2 \widehat{\sigma}_i^2/n_i\right)^2}{\sum_{i=1}^d \left(q_i^2 \widehat{\sigma}_i^2/n_i\right)^2/(n_i-1)}, \tag{1.23}$$

degrees of freedom. Here, the quantities q_i are the components of $\boldsymbol{q}' = \boldsymbol{w}'\boldsymbol{C}$. In case of no ties, this approximation is rather accurate for $n_i \geq 7$, while in case of ties, the quality of the approximation apparently depends on the number and the size of the ties.

1.5.3 The Rank Transform Property

In this subsection, the rank transform (RT)-technique suggested by Conover and Iman (1976, 1981) is discussed. This technique has been criticized by Blair, Sawilowski and Higgens (1987), Akritas (1990, 1991), Thompson and Ammann (1990) and Thompson (1991) and was studied in detail by Brunner and Neumann (1986*a*). The following considerations shall clarify the question when the method of the RT-technique works and when it fails. Moreover, the procedures described in this paper shall be distinguished from the usual RT-statistics and the expression *RT-technique* should be replaced by the terminology *rank transform property of a rank statistic* which is explained in detail below.

First note that with the statistic $\widehat{\boldsymbol{p}}$, the original observations are replaced by the (mid)-ranks among all observations. Recall that $\overline{\boldsymbol{Y}}_{\cdot} = (\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{d\cdot})'$ is the mean vector of the asymptotic rank transform $Y_{ij} = H(X_{ij})$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$. It follows from Theorem 1.3 that $\sqrt{N}\boldsymbol{C}\widehat{\boldsymbol{p}}$ is asymptotically equivalent to $\sqrt{N}\boldsymbol{C}\overline{\boldsymbol{Y}}_{\cdot}$ if $\boldsymbol{CF} = \boldsymbol{0}$, i.e. if the hypotheses are formulated in terms of the distribution functions. In some cases, the hypotheses in the linear model are equivalent to the corresponding nonparametric hypotheses (see e.g. Brunner and Puri, 2000, Proposition 5.1). Note that most counter examples for the RT-technique use linear hypotheses $\boldsymbol{C\mu} = \boldsymbol{0}$ such that $\boldsymbol{CF} \neq \boldsymbol{0}$ where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)'$ is the vector of the expectations.

Next, consider the covariance matrix

$$\boldsymbol{V}_N \;=\; Cov\left(\sqrt{N}\,\overline{\boldsymbol{Y}}_{\cdot}\right) \;=\; N \cdot diag\{n_1^{-1}\sigma_1^2, \ldots, n_d^{-1}\sigma_d^2\}$$

and note that in general the diagonal elements $\sigma_i^2 = Var(Y_{i1})$ of $\boldsymbol{V}_N$ are not necessarily all equal, even if homoscedasticity is assumed for the X_{ij}'s, since $H(\cdot)$ is a non-linear transformation (Akritas, 1990). However, some of the diagonal elements in $\boldsymbol{V}_N$ may be equal under the hypothesis $H_0 : \boldsymbol{CF} = \boldsymbol{0}$ and the corresponding estimators given in (1.11) can be pooled to estimate $\boldsymbol{V}_N$ consistently.

Now let $U_{ij} \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, d$, be independent normally distributed random variables where $\mu_i = E(Y_{i1})$ and $\sigma_i^2 = Var(Y_{i1})$. Let $\overline{\boldsymbol{U}}_{\cdot} = (\overline{U}_{1\cdot}, \ldots, \overline{U}_{d\cdot})'$ denote the mean vector of the U_{ij}'s. Then, by definition, the statistics $\sqrt{N}\,\overline{\boldsymbol{U}}_{\cdot}$ and $\sqrt{N}\,\overline{\boldsymbol{Y}}_{\cdot}$ have, asymptotically, the same multivariate normal distribution. Furthermore, define $\widetilde{\sigma}_i^2 = (n_i - 1)^{-1}\sum_{j=1}^{n_i}(U_{ij} - \overline{U}_{i\cdot})^2$ and let $\widetilde{\boldsymbol{V}}_N$ denote the matrix $\boldsymbol{V}_N$ with σ_i^2 replaced by $\widetilde{\sigma}_i^2$. Then $\widetilde{\boldsymbol{V}}_N$ is consistent for $\boldsymbol{V}_N$ and from Theorem 1.11, $\widehat{\boldsymbol{V}}_N$ is consistent for $\boldsymbol{V}_N$ in the sense that $\widetilde{\boldsymbol{V}}_N\boldsymbol{V}_N^{-1} \xrightarrow{p} \boldsymbol{I}_d$ and $\widehat{\boldsymbol{V}}_N\boldsymbol{V}_N^{-1} \xrightarrow{p} \boldsymbol{I}_d$, respectively. Note that $\widetilde{\sigma}_i^2$ is derived from $\widehat{\sigma}_i^2$ by replacing the ranks R_{ij} by the corresponding normally distributed random variables U_{ij}. Thus, the statistic $\widehat{\boldsymbol{p}}$ is a 'rank transform' of the statistic $\overline{\boldsymbol{U}}_{\cdot}$ and it follows from the above considerations that under $H_0 : \boldsymbol{CF} = \boldsymbol{0}$, the statistics $\sqrt{N}\boldsymbol{C}\widehat{\boldsymbol{p}}$ and $\sqrt{N}\boldsymbol{C}\overline{\boldsymbol{U}}_{\cdot}$ have, asymptotically, a multivariate normal distribution

$N(\mathbf{0}, \boldsymbol{C}\boldsymbol{V}_N\boldsymbol{C}')$. This property of the rank statistic $\sqrt{N}\boldsymbol{C}\widehat{\boldsymbol{p}}$ shall be called *rank transform property (RTP) with respect to the normal theory statistic $\sqrt{N}\boldsymbol{C}\overline{\boldsymbol{U}}_{\cdot}$*. Note that the expectations μ_i and the variances σ_i^2 of the corresponding normally distributed random variables U_{ij} are determined from the ART under $H_0 : \boldsymbol{CF} = \mathbf{0}$.

For testing $H_0 : \boldsymbol{CF} = \mathbf{0}$, there are two useful ways to define a statistic from $\sqrt{N}\boldsymbol{C}\widehat{\boldsymbol{p}}$. One possibility is to define the quadratic form $Q_N(\boldsymbol{C}) = N\widehat{\boldsymbol{p}}'(\boldsymbol{C}\widehat{\boldsymbol{V}}_N\boldsymbol{C}')^{-1}\widehat{\boldsymbol{p}}$ which is the rank version of WTS. It follows from Theorem 1.4 that under $H_0 : \boldsymbol{CF} = \mathbf{0}$, the quadratic form $Q_N(\boldsymbol{C})$ has, asymptotically, a central χ_f^2-distribution with $f = rank(\boldsymbol{C})$ degrees of freedom. The other possibility is to define the quadratic form $F_N(\boldsymbol{M}) = N\widehat{\boldsymbol{p}}'\boldsymbol{M}\widehat{\boldsymbol{p}}'$ which has the RTP with respect to the normal theory statistic $N\overline{\boldsymbol{U}}_{\cdot}'\boldsymbol{M}\overline{\boldsymbol{U}}_{\cdot}$ where $\boldsymbol{M} = \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}$ is a projection matrix which is taken from the ANOVA models with equal sample sizes. Under $H_0 : \boldsymbol{CF} = \mathbf{0}$, the asymptotic distribution of $N\overline{\boldsymbol{U}}_{\cdot}'\boldsymbol{M}\overline{\boldsymbol{U}}_{\cdot}$ is a weighted χ^2-distribution because the variances σ_i^2 are not necessarily equal, in general. For small sample sizes, $N\overline{\boldsymbol{U}}_{\cdot}'\boldsymbol{M}\overline{\boldsymbol{U}}_{\cdot}$ is approximated by a scaled F-distribution with estimated degrees of freedom.

1.6 Applications to Special Designs

In this section, the general theory described in the previous subsections is applied to some special factorial designs. Some explicit statistics shall be derived from the general approach where it turns out that several known rank statistics which have been proposed for some simple designs, come out as special cases. In particular, we consider the one-factor design where the Kruskal-Wallis statistic (1952, 1953) and the rank-transform statistic (Conover and Iman, 1981) come out as special cases. Moreover, for $d = 2$ treatments, the Wilcoxon-Mann-Whitney (WMW) statistic also comes out as a special case and if the observations have a Bernoulli-distribution, the χ^2-statistic for comparing proportions comes out as the square of the WMW statistic. Cross-classified models are considered as examples for two-way layouts. The extension to higher-way layouts is straightforward.

1.6.1 One-Way-Layout

Kruskal-Wallis-Test In the one-way layout, we observe independent random variables $X_{ij} \sim F_i = \frac{1}{2}[F_i^+ + F_i^-]$, $i = 1, \ldots, a$, $j = 1, \ldots, n_i$. Let $\boldsymbol{P}_a = \boldsymbol{I}_a - \frac{1}{a}\boldsymbol{J}_a$ be the contrast matrix defined in (1.1). Then the hypothesis for the one-way layout is written as $H_0^F : F_1 = \cdots = F_a$ or equivalently as $\boldsymbol{P}_a\boldsymbol{F} = \mathbf{0}$. Let $\boldsymbol{p} = (p_1, \ldots, p_a)' = \int Hd\boldsymbol{F}$ be the vector of the relative treatment effects $p_i = \int HdF_i$ where $H(x) = \frac{1}{N}\sum_{i=1}^a n_i F_i(x)$. The vector $\boldsymbol{p}$ is estimated consistently by $\widehat{\boldsymbol{p}} = \int \widehat{H}d\widehat{\boldsymbol{F}} = (\widehat{p}_1, \ldots, \widehat{p}_a)'$ where $\widehat{p}_i = \frac{1}{N}\left(\overline{R}_{i\cdot} - \frac{1}{2}\right)$ and $\overline{R}_{i\cdot} = n_i^{-1}\sum_{j=1}^{n_i} R_{ij}$ is the mean of the (mid-)ranks R_{ij} of X_{ij} among all the $N = \sum_{i=1}^a n_i$ observations.

Under H_0^F, the statistics $\sqrt{N}\boldsymbol{P}_a\widehat{\boldsymbol{p}}$ and $\sqrt{N}\boldsymbol{P}_a\overline{\boldsymbol{Y}}_{\cdot}$ are asymptotically equivalent (see Theorem 1.3) where $\overline{\boldsymbol{Y}}_{\cdot} = (\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{a\cdot})'$, $\overline{Y}_{i\cdot} = n_i^{-1}\sum_{j=1}^{n_i} H(X_{ij})$. Let $\boldsymbol{\Lambda} = diag\{n_1, \ldots, n_a\}$ denote the diagonal matrix of the sample sizes. Then under $H_0^F : \boldsymbol{P}_a\boldsymbol{F} = \mathbf{0}$, it follows that

$\sigma_1^2 = \cdots = \sigma_a^2 = \sigma^2$ and $\boldsymbol{V}_N = Cov\left(\sqrt{N}\,\overline{\boldsymbol{Y}}_{\cdot}\right) = N\sigma^2\boldsymbol{\Lambda}^{-1}$. A consistent estimator of σ^2 follows immediately from Theorem 1.5 by pooling the estimators $\widehat{\sigma}_i^2$, viz.

$$\widehat{\sigma}_N^2 = \frac{1}{N^2(N-a)}\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(R_{ij} - \overline{R}_{i\cdot}\right)^2. \tag{1.24}$$

To test the hypothesis H_0^F, consider the quadratic form Q_N given in (1.18) and let $\widehat{\boldsymbol{W}}_N = \widehat{\boldsymbol{V}}_N^{-1}(\boldsymbol{I}_a - \boldsymbol{J}_a\widehat{\boldsymbol{V}}_N^{-1}/tr(\widehat{\boldsymbol{V}}_N^{-1}))$, where $\widehat{\boldsymbol{V}}_N^{-1} = \boldsymbol{\Lambda}/(N\widehat{\sigma}_N^2)$. Note that $\widehat{\boldsymbol{W}}_N$ is a g-inverse of $\boldsymbol{P}_a\widehat{\boldsymbol{V}}_N\boldsymbol{P}_a$ and that $\boldsymbol{P}_a\widehat{\boldsymbol{W}}_N\boldsymbol{P}_a = \widehat{\boldsymbol{W}}_N$. Then, the quadratic form

$$\begin{aligned} Q_N &= N\widehat{\boldsymbol{p}}'\boldsymbol{P}_a[\boldsymbol{P}_a\widehat{\boldsymbol{V}}_N\boldsymbol{P}_a]^-\boldsymbol{P}_a\widehat{\boldsymbol{p}} = N\widehat{\boldsymbol{p}}'\widehat{\boldsymbol{W}}_N\widehat{\boldsymbol{p}} \\ &= \frac{1}{\widehat{\sigma}_N^2}\widehat{\boldsymbol{p}}'\left(\boldsymbol{\Lambda} - \frac{1}{N}\boldsymbol{\Lambda}\boldsymbol{J}_a\boldsymbol{\Lambda}\right)\widehat{\boldsymbol{p}} \\ &= \frac{N-a}{\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(R_{ij} - \overline{R}_{i\cdot}\right)^2}\sum_{i=1}^{a} n_i\left(\overline{R}_{i\cdot} - \frac{N+1}{2}\right)^2 \end{aligned} \tag{1.25}$$

is a WTS which has, asymptotically, a central χ^2_{a-1}-distribution under H_0^F. For small samples, the distribution of the statistic $Q_N/(a-1)$ may be approximated by the central $F(f_1, f_2)$-distribution where $f_1 = a-1$ and $f_2 = N-a-1$. Note that under H_0^F, the variances are equal, i.e. $\sigma_1^2 = \cdots = \sigma_a^2$ and thus, it is not necessary to apply the small sample approximation considered in the previous section. It is well known that for continuous distributions, the approximation is quite accurate if $a \geq 3$ and $n_i \geq 6$.

REMARK 1.1 *Q_N given in (1.25) has the so called rank transform (RT) property, i.e. if the ranks R_{ij} are replaced by independent normally distributed random variables, then the corresponding normal theory statistic has, asymptotically, the same distribution as Q_N. Note that the statistic Q_N was called 'rank transform statistic' by Conover and Iman (1981).*

If $\widehat{\sigma}_N^2$ given in (1.24) is replaced by

$$\widehat{\sigma}_0^2 = \frac{1}{N^2(N-1)}\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(R_{ij} - \frac{N+1}{2}\right)^2, \tag{1.26}$$

then the quadratic form Q_N given in (1.25) becomes the Kruskal-Wallis statistic which accomodates automatically for ties if the mid-ranks are used. If the distribution functions are continuous, then,

$$\widehat{\sigma}_0^2 = \frac{1}{N^2(N-1)}\sum_{i=1}^{N}\left(i - \frac{N+1}{2}\right)^2 = \frac{N+1}{12N},$$

under H_0^F. Then, Q_N given in (1.25) becomes the Kruskal-Wallis H-statistic (Kruskal and Wallis, 1952, 1953). Note that both the variance estimators $\widehat{\sigma}_N^2$ and $\widehat{\sigma}_0^2$ are consistent estimators of σ^2 under H_0^F.

Wilcoxon-Mann-Whitney- and χ^2-Test Straightforward computations show that, for $a = 2$, Q_N given in (1.25) reduces to the square of the Wilcoxon-Mann-Whitney (WMW) statistic

$$W_N^2 = \frac{N-1}{N} \cdot n_1 n_2 \cdot \frac{(\overline{R}_{1\cdot} - \overline{R}_{2\cdot})^2}{\sum_{i=1}^{2} \sum_{j=1}^{n_i} \left(R_{ij} - \frac{N+1}{2} \right)}, \tag{1.27}$$

which in case of no ties becomes

$$W_N^2 = \frac{12 n_1 n_2}{N(N+1)} \left(\overline{R}_{1\cdot} - \overline{R}_{2\cdot} \right)^2 .$$

Now assume that the observations are dichotomous having Bernoulli-distributions, i.e. $X_{ij} \sim B(q_i)$, $i = 1, 2,\ j = 1, \ldots, n_i$. Generally, the results of such a trial are arranged in a contingency table where n_{i0} denotes the number of 0's and n_{i1} denotes the number of 1's, $i = 1, 2$, within treatment i.

Outcome	Treatment 1	Treatment 2	Total
$X_{ij} = 0$	n_{10}	n_{20}	$n_{\cdot 0}$
$X_{ij} = 1$	n_{11}	n_{21}	$n_{\cdot 1}$
	n_1	n_2	N

The hypothesis $H_0^F : F_1 = F_2$ is equivalent to $H_0 : q_1 = q_2$ since $B(q_i)$ is completely determined by q_i, $i = 1, 2$. Note that with dichotomous data, only two different mid-ranks occur, namely

$$R_{ij} = \begin{cases} (1 + n_{\cdot 0})/2, & \text{if} \quad X_{ij} = 0, \\ (n_{\cdot 0} + N + 1)/2, & \text{if} \quad X_{ij} = 1 . \end{cases}$$

Thus, the rank means and their differencs are

$$\begin{aligned} \overline{R}_{1\cdot} &= \frac{1 + n_{\cdot 0}}{2} + \frac{N}{2} \cdot \frac{n_{11}}{n_1}, \\ \overline{R}_{2\cdot} &= \frac{1 + n_{\cdot 0}}{2} + \frac{N}{2} \cdot \frac{n_{21}}{n_2}, \\ \overline{R}_{1\cdot} - \overline{R}_{2\cdot} &= \frac{N}{2} \cdot \left(\frac{n_{11}}{n_1} - \frac{n_{21}}{n_2} \right) . \end{aligned}$$

Moreover, the quantity $N\widehat{\sigma}_N^2$ reduces to

$$N\widehat{\sigma}_n^2 = \frac{1}{N-1} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left(R_{ij} - \frac{N+1}{2} \right)^2 = \frac{N}{N-1} \cdot \frac{n_{\cdot 0} n_{\cdot 1}}{4}$$

and W_N^2 given in (1.27) finally becomes

$$W_N^2 = (N-1) \cdot \frac{(n_{10} n_{21} - n_{11} n_{20})^2}{n_1 n_2 n_{\cdot 0} n_{\cdot 1}}, \tag{1.28}$$

which is the well known χ^2-contingency table statistic, up to a factor $N/(N-1)$. Thus, for dichotomuous data, the WMW test is asymptotically equivalent to the χ^2-contingency table tst.

1.6.2 Two-Way-Layout

Next, we consider the two-way cross classification where factor A has $i = 1, \ldots, a$ levels and factor B has $j = 1, \ldots, b$ levels with $k = 1, \ldots, n_{ij}$ replications per cell (i, j) and the independent random variables X_{ijk} have distribution functions $F_{ij}(x) = \frac{1}{2}[F_{ij}^+ + F_{ij}^-]$. Let

$$\boldsymbol{F} \;=\; (F_{11}, \ldots, F_{1b}, \ldots, F_{a1}, \ldots, F_{ab})'$$

denote the vector of the distribution functions. Let $\boldsymbol{C}_A = \boldsymbol{P}_a \otimes \frac{1}{b}\boldsymbol{1}_b'$, $\boldsymbol{C}_B = \frac{1}{a}\boldsymbol{1}_a' \otimes \boldsymbol{P}_b$ and $\boldsymbol{C}_{AB} = \boldsymbol{P}_a \otimes \boldsymbol{P}_b$ where $\boldsymbol{P}_a$ and $\boldsymbol{P}_b$ are given in (1.1). Then the nonparametric hypotheses of 'no main effect A', 'no main effect B' or 'no interaction AB' are formulated as

$$H_0^F(A) : \boldsymbol{C}_A\boldsymbol{F} = \boldsymbol{0}, \quad H_0^F(B) : \boldsymbol{C}_B\boldsymbol{F} = \boldsymbol{0}, \quad H_0^F(AB) : \boldsymbol{C}_{AB}\boldsymbol{F} = \boldsymbol{0}.$$

REMARK 1.2 *In a linear model without interaction (i.e. where the main effects are well defined), the hypotheses of no nonparametric main effect A or B, respectively are equivalent to the parametric hypotheses of no main effect A or B, respectively (in the usual linear model). For a further discussion of nonparametric hypotheses, see Akritas and Arnold (1994), Akritas, Arnold and Brunner (1997), Brunner, Puri and Sun (1995) and Brunner and Puri (1996).*

Let $\widehat{\boldsymbol{F}}(x) = (\widehat{F}_{11}(x), \ldots, \widehat{F}_{ab}(x))'$ denote the vector of the empirical distribution functions $\widehat{F}_{ij}(x) = n_{ij}^{-1}\sum_{k=1}^{n_{ij}} c(x - X_{ijk})$ and let $\widetilde{R}_{i\cdot\cdot} = b^{-1}\sum_{j=1}^{b} \overline{R}_{ij\cdot}$, $i = 1, \ldots, a$, denote the unweighted means of the cell means $\overline{R}_{ij\cdot} = n_{ij}^{-1}\sum_{k=1}^{n_{ij}} R_{ijk}$ where R_{ijk} is the rank of X_{ijk} among all the $N = \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}$ observations. To test the hypotheses $H_0^F(\cdot)$ formulated above, consider the statistic $\widehat{\boldsymbol{p}} = \int \widehat{H} d\widehat{\boldsymbol{F}} = \frac{1}{N}(\overline{R}_{11\cdot} - \frac{1}{2}, \ldots, \overline{R}_{ab\cdot} - \frac{1}{2})'$ under the hypothesis $H_0^F : \boldsymbol{C}\boldsymbol{F} = \boldsymbol{0}$ using the contrast matrices $\boldsymbol{C}_A$, $\boldsymbol{C}_B$ and $\boldsymbol{C}_{AB}$. Let

$$\widehat{\sigma}_{ij}^2 = \frac{1}{N^2(n_{ij}-1)}\sum_{k=1}^{n_{ij}}(R_{ijk} - \overline{R}_{ij\cdot})^2, \qquad \widehat{\boldsymbol{V}}_N = N\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{\widehat{\sigma}_{ij}^2}{n_{ij}}, \tag{1.29}$$

$$\widehat{\tau}_i^2 = \frac{1}{b^2}\sum_{j=1}^{b}\frac{\widehat{\sigma}_{ij}^2}{n_{ij}}, \qquad \widehat{\boldsymbol{\Sigma}}_a = \bigoplus_{i=1}^{a}\widehat{\tau}_i^2.$$

First we consider the WTS for this design. Let $\widehat{\boldsymbol{W}}_a = N^{-1}\widehat{\boldsymbol{\Sigma}}_a^{-1}(\boldsymbol{I}_a - \boldsymbol{J}_a\widehat{\boldsymbol{\Sigma}}_a^{-1}/\boldsymbol{1}_a'\widehat{\boldsymbol{\Sigma}}_a^{-1}\boldsymbol{1}_a)$ and note that $\widehat{\boldsymbol{W}}_a$ is a g-inverse of $\boldsymbol{C}_A\widehat{\boldsymbol{V}}_N\boldsymbol{C}_A' = N\boldsymbol{P}_a\widehat{\boldsymbol{\Sigma}}_a\boldsymbol{P}_a$ and that $\boldsymbol{P}_a\widehat{\boldsymbol{W}}_a\boldsymbol{P}_a = \widehat{\boldsymbol{W}}_a$. Then, under $H_0^F(A)$, it follows from Theorem 1.4 that the quadratic form

$$\begin{aligned} Q_N(\boldsymbol{C}_A) &= N\widehat{\boldsymbol{p}}'\boldsymbol{C}_A'(\boldsymbol{C}_A\widehat{\boldsymbol{V}}_N\boldsymbol{C}_A')^-\boldsymbol{C}_A\widehat{\boldsymbol{p}} = N\widehat{\boldsymbol{p}}'\left(\widehat{\boldsymbol{W}}_a \otimes \tfrac{1}{b}\boldsymbol{J}_b\right)\widehat{\boldsymbol{p}} \\ &= \sum_{i=1}^{a}\frac{1}{\widehat{\tau}_i^2}\left(\widetilde{R}_{i\cdot\cdot} - \frac{1}{\sum_{r=1}^{a}(1/\widehat{\tau}_r^2)}\sum_{r=1}^{a}\frac{\widetilde{R}_{r\cdot\cdot}}{\widehat{\tau}_r^2}\right)^2, \end{aligned} \tag{1.30}$$

has asymptotically a central χ_f^2-distribution with $f = a - 1$ degrees of freedom. Because of the symmetry, rows and columns are interchangeable in this design and the quadratic form $Q_N(\boldsymbol{C}_B)$ for testing $H_0^F(B)$ is obtained from $Q_N(\boldsymbol{C}_A)$ by interchanging the indices i and j.

The statistic for testing the hypothesis $H_0^F(AB)$ of no nonparametric interaction, namely

$$Q_N(\boldsymbol{C}_{AB}) = N\widehat{\boldsymbol{p}}'\boldsymbol{C}'_{AB}(\boldsymbol{C}_{AB}\widehat{\boldsymbol{V}}_N\boldsymbol{C}'_{AB})^{-}\boldsymbol{C}_{AB}\widehat{\boldsymbol{p}}$$

is also derived from Theorem 1.4 and $Q_N(\boldsymbol{C}_{AB})$ has, asymptotically, a central χ^2_f-distribution with $f = (a-1) \times (b-1)$ degrees of freedom under $H_0^F(AB)$.

Simulation studies showed (see Brunner, Dette and Munk, 1997) that the approximation by the limiting χ^2distributions of the quadratic forms given above are rather poor and large sample sizes are needed to have acceptable approximations. Therefore, we derive the ANOVA-type statistics described in subsection 1.5.1.

In the two-way layout, the nonparametric hypothesis of no main effect A is equivalently restated as $H_0^F(A) : \boldsymbol{M}_A\boldsymbol{F} = \boldsymbol{0}$ where $\boldsymbol{M}_A = \boldsymbol{P}_a \otimes \frac{1}{b}\boldsymbol{J}_b$ is a projection matrix with constant diagonal elements $m_a = (a-1)/(ab)$. Let

$$\begin{aligned}
\widetilde{p}_{i\cdot} &= \frac{1}{b}\sum_{j=1}^{b}\widehat{p}_{ij} = \frac{1}{b}\sum_{j=1}^{b}\frac{1}{N}\left(\overline{R}_{ij\cdot} - \frac{1}{2}\right) = \frac{1}{N}\left(\widetilde{R}_{i\cdot\cdot} - \frac{1}{2}\right), \\
\widetilde{p}_{\cdot\cdot} &= \frac{1}{N}\left(\widetilde{R}_{\cdot\cdot\cdot} - \frac{1}{2}\right),
\end{aligned}$$

where $\widetilde{R}_{i\cdot\cdot} = b^{-1}\sum_{j=1}^{b}\overline{R}_{ij\cdot}$ and $\widetilde{R}_{\cdot\cdot\cdot} = a^{-1}\sum_{i=1}^{a}\widetilde{R}_{i\cdot\cdot}$. Then, under $H_0^F(A)$, the statistic

$$\begin{aligned}
F_N(\boldsymbol{M}_A) &= \frac{N}{tr(\boldsymbol{M}_A\widehat{\boldsymbol{V}}_N)}\,\widehat{\boldsymbol{p}}'\boldsymbol{M}_A\,\widehat{\boldsymbol{p}} = \frac{Nab}{(a-1)tr(\widehat{\boldsymbol{V}}_N)}\sum_{i=1}^{a}\sum_{j=1}^{b}(\widetilde{p}_{i\cdot} - \widetilde{p}_{\cdot\cdot})^2 \\
&= \frac{ab^2}{N^2(a-1)\sum_{i=1}^{a}\sum_{j=1}^{b}\widehat{\sigma}_{ij}^2/n_{ij}}\sum_{i=1}^{a}\left(\widetilde{R}_{i\cdot\cdot} - \widetilde{R}_{\cdot\cdot\cdot}\right)^2
\end{aligned}$$

has, asymptotically, a central $F(\widehat{f}_A, \widehat{f}_0)$-distribution where the degrees of freedom $\widehat{f}_A$ and $\widehat{f}_0$ are derived from (1.20) and (1.21) respectively by replacing $\boldsymbol{M}$ with $\boldsymbol{M}_A = \boldsymbol{P}_a \otimes \frac{1}{b}\boldsymbol{J}_b$, m with $(a-1)/(ab)$ and $\widehat{\sigma}_{ij}^2$ and $\widehat{\boldsymbol{V}}_N$ are given in (1.29). Because of the symmetry, rows and columns are interchangeable in this design, the quadratic form $F_N(\boldsymbol{M}_B)$ for testing $H_0^F(B)$ is obtained from $F_N(\boldsymbol{M}_A)$ by interchanging the indices i and j.

Finally, the nonparametric hypothesis of no interaction is restated equivalently as $H_0^F(AB)$: $\boldsymbol{M}_{AB}\boldsymbol{F} = \boldsymbol{0}$ where $\boldsymbol{M}_{AB} = \boldsymbol{P}_a \otimes \boldsymbol{P}_b$ is a projection matrix with constant diagonal elements $m_{ab} = (a-1)(b-1)/(ab)$. Let $\widetilde{p}_{\cdot j} = a^{-1}\sum_{i=1}^{a}\widehat{p}_{ij} = \frac{1}{N}(\widetilde{R}_{\cdot j\cdot} - \frac{1}{2})$, where $\widetilde{R}_{\cdot j\cdot} = a^{-1}\sum_{i=1}^{a}\overline{R}_{ij\cdot}$. Then, under $H_0^F(AB)$, the statistic

$$\begin{aligned}
F_N(\boldsymbol{M}_{AB}) &= \frac{N}{tr(\boldsymbol{M}_{AB}\widehat{\boldsymbol{V}}_N)}\,\widehat{\boldsymbol{p}}'\boldsymbol{M}_{AB}\,\widehat{\boldsymbol{p}} \\
&= \frac{Nab}{(a-1)(b-1)tr(\widehat{\boldsymbol{V}}_N)}\sum_{i=1}^{a}\sum_{j=1}^{b}(\widehat{p}_{ij} - \widetilde{p}_{i\cdot} - \widetilde{p}_{\cdot j} + \widetilde{p}_{\cdot\cdot})^2 \\
&= \frac{ab}{N^2(a-1)(b-1)\sum_{i=1}^{a}\sum_{j=1}^{b}\widehat{\sigma}_{ij}^2/n_{ij}}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{R}_{ij\cdot} - \widetilde{R}_{i\cdot\cdot} - \widetilde{R}_{\cdot j\cdot} + \widetilde{R}_{\cdot\cdot\cdot}\right)^2
\end{aligned}$$

has, asymptotically, a central $F(\widehat{f}_{AB}, \widehat{f}_0)$-distribution where the degrees of freedom $\widehat{f}_{AB}$ and $\widehat{f}_0$ are derived from (1.20) and (1.21) respectively by replacing $\boldsymbol{M}$ with $\boldsymbol{M}_{AB} = \boldsymbol{P}_a \otimes \boldsymbol{P}_b$, m with $(a-1)(b-1)/(ab)$. The quantities $\widehat{\sigma}_{ij}^2$ and $\widehat{\boldsymbol{V}}_N$ are given in (1.29).

1.6.3 Higher-Way-Layouts

In this Section, it is explained how to extend the methods presented in the previous Sections to higher-way layouts. This will be done by means of the three-way layout (cross-classification).

Example: Three-Way Layout

The observations $X_{ijk\ell} \sim F_{ijk}(x)$ are assumed to be independent with distribution functions $F_{ijk}(x)$, $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, c$ and the index $\ell = 1, \ldots, n_{ijk}$ denotes the independent and identically distributed replications. The hypotheses for the nonparametric main effects are expressed as

$$\begin{aligned} H_0^F(A) : \overline{F}_{1\cdot\cdot} = & \cdots & = \overline{F}_{a\cdot\cdot}, \\ H_0^F(B) : \overline{F}_{\cdot 1\cdot} = & \cdots & = \overline{F}_{\cdot b\cdot}, \\ H_0^F(C) : \overline{F}_{\cdot\cdot 1} = & \cdots & = \overline{F}_{\cdot\cdot c}, \end{aligned}$$

where $\overline{F}_{i\cdot\cdot}$ denotes the mean over all bc distribution functions within level i of factor A, $\overline{F}_{\cdot j\cdot}$ denotes the mean over all ac distribution functions within level j of factor B and $\overline{F}_{\cdot\cdot k}$ denotes the mean over all ab distribution functions within level k of factor C. Then, the hypotheses can be written as

$$\begin{aligned} H_0^F(A)&: (\boldsymbol{P}_a \otimes \tfrac{1}{b}\mathbf{1}_b' \otimes \tfrac{1}{c}\mathbf{1}_c')\boldsymbol{F} &= \boldsymbol{C}_A\boldsymbol{F} &= \mathbf{0}, \\ H_0^F(B)&: (\tfrac{1}{a}\mathbf{1}_a' \otimes \boldsymbol{P}_b \otimes \tfrac{1}{c}\mathbf{1}_c')\boldsymbol{F} &= \boldsymbol{C}_B\boldsymbol{F} &= \mathbf{0}, \\ H_0^F(C)&: (\tfrac{1}{a}\mathbf{1}_a' \otimes \tfrac{1}{b}\mathbf{1}_b' \otimes \boldsymbol{P}_c)\boldsymbol{F} &= \boldsymbol{C}_C\boldsymbol{F} &= \mathbf{0}, \end{aligned}$$

where $\boldsymbol{F} = (F_{111}, \ldots, F_{abc})'$ denotes the vector of the distribution functions.

The hypothesis of no nonparametric AB-interaction is usually expressed as

$$H_0^F(AB) : \overline{F}_{ij\cdot} + \overline{F}_{\cdots} \;=\; \overline{F}_{i\cdot\cdot} + \overline{F}_{\cdot j\cdot},\; i = 1, \ldots, a,\; j = 1, \ldots, b,$$

where $\overline{F}_{ij\cdot}$ denotes the mean over all c distribution functions within the levels i of factor A and level j of factor B and $\overline{F}_{\cdots}$ denotes the mean over all abc distribution functions in the experiment. In matrix notation, this hypothesis is written as

$$H_0^F(AB) : (\boldsymbol{P}_a \otimes \boldsymbol{P}_b \otimes \tfrac{1}{c}\mathbf{1}_c')\boldsymbol{F} \;=\; \boldsymbol{C}_{AB}\boldsymbol{F} = \mathbf{0}.$$

In the same way, the hypotheses for the other nonparametric interactions are formulated as

$$\begin{aligned} H_0(AC)&: (\boldsymbol{P}_a \otimes \tfrac{1}{b}\mathbf{1}_b' \otimes \boldsymbol{P}_c)\boldsymbol{F} &= \boldsymbol{C}_{AC}\boldsymbol{F} &= \mathbf{0}, \\ H_0(BC)&: (\tfrac{1}{a}\mathbf{1}_a' \otimes \boldsymbol{P}_b \otimes \boldsymbol{P}_c)\boldsymbol{F} &= \boldsymbol{C}_{BC}\boldsymbol{F} &= \mathbf{0}, \\ H_0(ABC)&: (\boldsymbol{P}_a \otimes \boldsymbol{P}_b \otimes \boldsymbol{P}_c)\boldsymbol{F} &= \boldsymbol{C}_{ABC}\boldsymbol{F} &= \mathbf{0}. \end{aligned} \tag{1.31}$$

The contrast matrices $\boldsymbol{C}_A$, $\boldsymbol{C}_B$, ..., $\boldsymbol{C}_{ABC}$ by which the hypotheses are formulated are used to write the statistics for testing these hypotheses. Let R_{ijkl} denote the rank of X_{ijkl} among all $N = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} n_{ijk}$ observations and let $\overline{\boldsymbol{R}}_{\cdot} = (\overline{R}_{111\cdot}, \ldots, \overline{R}_{abc\cdot})'$ denote the vector of the rank means $\overline{R}_{ijk\cdot} = n_{ijk}^{-1}\sum_{\ell=1}^{n_{ijk}} R_{ijk\ell}$ within all abc treatment combinations. Finally, denote by

$$\widehat{\boldsymbol{S}}_N = diag\left\{\frac{s_{111}^2}{n_{111}}, \ldots, \frac{s_{abc}^2}{n_{abc}}\right\}$$

the diagonal matrix of the variance estimators

$$s_{ijk}^2 = \frac{1}{n_{ijk}-1}\sum_{\ell=1}^{n_{ijk}}\left(R_{ijk\ell} - \overline{R}_{ijk\cdot}\right)^2$$

divided by the sample sizes n_{ijk}. Then, for large sample sizes, the statistic

$$Q_N(\boldsymbol{C}) = \overline{\boldsymbol{R}}_{\cdot}'\boldsymbol{C}'[\boldsymbol{C}\widehat{\boldsymbol{S}}_N\boldsymbol{C}']^{-1}\boldsymbol{C}\overline{\boldsymbol{R}}_{\cdot} \tag{1.32}$$

has, approximately, a central χ_f^2-distribution with $f = rank(\boldsymbol{C})$ degrees of freedom.

To test the hypothesis $H_0^F(A)$ of no nonparametric main effect A, for example, the matrix $\boldsymbol{C}$ in (1.32) is replaced by $\boldsymbol{C}_A = \boldsymbol{P}_a \otimes \frac{1}{b}\boldsymbol{1}_b' \otimes \frac{1}{c}\boldsymbol{1}_c'$ where $rank(\boldsymbol{C}_A) = a-1$. In the same way, the statistics for testing the other hypotheses can be derived from (1.32) by using the contrast matrices given above to formulate the hypotheses.

For small sample sizes, let

$$\begin{aligned}
\boldsymbol{T}_A &= \boldsymbol{C}_A'[\boldsymbol{C}_A\boldsymbol{C}_A']^{-1}\boldsymbol{C}_A &&= \boldsymbol{P}_a \otimes \tfrac{1}{b}\boldsymbol{1}_b' \otimes \tfrac{1}{c}\boldsymbol{1}_c', && \text{–for } H_0^F(A),\\
\boldsymbol{T}_B &= \boldsymbol{C}_B'[\boldsymbol{C}_B\boldsymbol{C}_B']^{-1}\boldsymbol{C}_B &&= \tfrac{1}{a}\boldsymbol{1}_a' \otimes \boldsymbol{P}_b \otimes \tfrac{1}{c}\boldsymbol{1}_c', && \text{–for } H_0^F(B),\\
\boldsymbol{T}_C &= \boldsymbol{C}_C'[\boldsymbol{C}_C\boldsymbol{C}_C']^{-1}\boldsymbol{C}_C &&= \tfrac{1}{a}\boldsymbol{1}_a' \otimes \tfrac{1}{b}\boldsymbol{1}_b' \otimes \boldsymbol{P}_c, && \text{–for } H_0^F(C),\\
\boldsymbol{T}_{AB} &= \boldsymbol{C}_{AB}'[\boldsymbol{C}_{AB}\boldsymbol{C}_{AB}']^{-1}\boldsymbol{C}_{AB} &&= \boldsymbol{P}_a \otimes \boldsymbol{P}_b \otimes \tfrac{1}{c}\boldsymbol{1}_c', && \text{–for } H_0^F(AB),\\
\boldsymbol{T}_{AC} &= \boldsymbol{C}_{AC}'[\boldsymbol{C}_{AC}\boldsymbol{C}_{AC}']^{-1}\boldsymbol{C}_{AC} &&= \boldsymbol{P}_a \otimes \tfrac{1}{b}\boldsymbol{1}_b' \otimes \boldsymbol{P}_c, && \text{–for } H_0^F(AC),\\
\boldsymbol{T}_{BC} &= \boldsymbol{C}_{BC}'[\boldsymbol{C}_{BC}\boldsymbol{C}_{BC}']^{-1}\boldsymbol{C}_{BC} &&= \tfrac{1}{a}\boldsymbol{1}_a' \otimes \boldsymbol{P}_b \otimes \boldsymbol{P}_c, && \text{–for } H_0^F(BC),\\
\boldsymbol{T}_{ABC} &= \boldsymbol{C}_{ABC}'[\boldsymbol{C}_{ABC}\boldsymbol{C}_{ABC}']^{-1}\boldsymbol{C}_{ABC} &&= \boldsymbol{P}_a \otimes \boldsymbol{P}_b \otimes \boldsymbol{P}_c, && \text{–for } H_0^F(ABC).
\end{aligned}$$

Here, the matrices $\boldsymbol{P}_a = \boldsymbol{I}_a - \frac{1}{a}\boldsymbol{J}_a$, $\boldsymbol{P}_b = \boldsymbol{I}_b - \frac{1}{b}\boldsymbol{J}_b$ and $\boldsymbol{P}_c = \boldsymbol{I}_c - \frac{1}{c}\boldsymbol{J}_c$ are the centering matrices of dimensions a, b and c respectively.

For small samples sizes, the hypotheses $H_0^F : \boldsymbol{C}\boldsymbol{F} = \boldsymbol{0}$ are tested by the statistic

$$F_N(\boldsymbol{T}) = \frac{1}{tr(\boldsymbol{T}\widehat{\boldsymbol{S}}_N)}\overline{\boldsymbol{R}}_{\cdot}'\boldsymbol{T}\,\overline{\boldsymbol{R}}_{\cdot} \tag{1.33}$$

by replacing in (1.33) the matrix $\boldsymbol{T}$ by one of the matrices $\boldsymbol{T}_A$, $\boldsymbol{T}_B$, ..., corresponding to the hypothesis to be tested. In (1.33), $tr(\cdot)$ denotes the trace of a square matrix. Under the hypothesis H_0^F, the statistic $F_N(\boldsymbol{T})$ has approximately a central $F(\widehat{f}_T, \widehat{f}_0)$-distribution where

$$\widehat{f}_T = \frac{\left[tr(\boldsymbol{T}\widehat{\boldsymbol{S}}_N)\right]^2}{tr(\boldsymbol{T}\widehat{\boldsymbol{S}}_N\boldsymbol{T}\widehat{\boldsymbol{S}}_N)} \quad \text{and} \quad \widehat{f}_0 = \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} s_{ijk}^2/n_{ijk}\right)^2}{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(s_{ijk}^2/n_{ijk})^2/(n_{ijk}-1)}.$$

1.7 Example and Software

1.7.1 Two-way Layout with Count Data

Fertility Trial In this subsection, we apply some of the procedures discussed in the previous subsections to an example with count data in a two-way layout. The statistics and the approximations are given in subsection 1.6.2. The authors are grateful to Dr. Beuscher (Schaper & Brümmer, Inc., Salzgitter, Germany) for making available the data.

In a fertility trial, three groups of female Wistar rats were treated with three different dosages (placebo, dosage 1 and 2) of a drug (factor A). Among other fertility parameters, the number of corpora lutea from rat ovaries was counted after a section of the animals. The same trial was repeated one year later with three new groups of rats. The results of the trial for the two years (factor B) and the three groups with $n_{11} = 9, n_{12} = 13, n_{21} = 9, n_{22} = 8, n_{31} = 8, n_{32} = 12$ animals are given in Table 1.1.

TABLE 1.1 *Number of corpora lutea from Wistar rats in a fertility trial.*

Group	Year 1	Year 2
Placebo	13, 12, 11, 11, 14, 14, 13, 13, 13	12, 16, 9, 14, 15, 12, 12, 11, 13, 14, 12, 13, 12
Dosage 1	15, 12, 11, 11, 14, 13, 14, 14, 12	9, 12, 11, 15, 11, 10, 13, 11
Dosage 2	15, 12, 13, 14, 11, 14, 17, 15	15, 13, 17, 14, 14, 13, 13, 13, 9, 12, 15, 14

The rank means $\overline{R}_{ij\cdot}$, $i = 1,2,3$; $j = 1,2$, within the three treatment groups and the two years as well as the unweighted means $\widetilde{R}_{i\cdot\cdot}$ within the treatment groups and $\widetilde{R}_{\cdot j\cdot}$ within the two years are displayed in Table 1.2.

TABLE 1.2 *Rank means $\overline{R}_{ij\cdot}$, $\widetilde{R}_{i\cdot\cdot}$ and $\widetilde{R}_{\cdot j\cdot}$ and relative treatment effects for the number of the corpora lutea of the Wistar rats in the fertility trial.*

	Ranks			Relative Treatment Effects		
Group	Year 1	Year 2	$\widetilde{R}_{i\cdot\cdot}$	Year 1	Year 2	$\widetilde{p}_{i\cdot}$
Placebo	27.6	28.1	27.8	0.46	0.47	0.46
Dosage 1	30.1	17.0	23.5	0.50	0.28	0.39
Dosage 2	38.8	36.7	37.7	0.65	0.61	0.63
$\widetilde{R}_{\cdot j\cdot}$	32.1	27.7.		0.54	0.46	

The results of the analysis by the WTS $Q_N(\boldsymbol{C})$ and the ATS $F_N(\boldsymbol{M})$ given in subsection 1.6.2 along with the resulting p-values are displayed in Table 1.3

TABLE 1.3 *Test statistics and p-values for the nonparametric main effects and interaction in the fertility trial. The results of the test statistics obtained by the WTS with the resulting p-values are given in the left part and the results obtained by the ATS with the resulting p-values are given in the right part of the table.*

	Wald-Type Statistic		ANOVA-Type Statistic	
Hypothesis	$Q_N(\boldsymbol{C})$	p-Value	$F_N(\boldsymbol{M})$	p-Value
$H_0^F(A)$	6.91	0.032	3.80	0.031
$H_0^F(B)$	1.28	0.257	1.28	0.264
$H_0^F(AB)$	1.78	0.412	0.92	0.403

The large p-value ($p = 0.403$) for $H_0^F(AB)$ indicates that the results are quite homogeneous within the two years (no interaction). However, a significant treatment effect for the drug is proved at the 5% level ($p = 0.031$) and there is no evidence for an effect of the year ($p = 0.264$).

1.7.2 Software

Regarding software for the computation of the statistics described in this section, we note that the statistics $Q_N(\boldsymbol{C})$, $F_N(\boldsymbol{M})$ and $L_N(\boldsymbol{w})$ have the rank transform property under H_0^F. Therefore, it is only necessary to rank all the data and to identify the special heteroscedastic parametric model from the ART under H_0^F (see Subsection 1.5.3). Thus, any statistical software package which provides

1. the mid-ranks of the observations,
2. the analysis of heteroscedastic factorial designs

can be used to compute the statistics $Q_N(\boldsymbol{C})$, $F_N(\boldsymbol{M})$ and $L_N(\boldsymbol{w})$. Below, we provide the necessary statements for the Statistical Analysis System (SAS), where the DATA-step and the procedures RANK and MIXED are used.

Data Input The input of the data is handled in the same way as for the data of a parametric model, i.e. factors are treated as 'classifying variables'.

Ranking The procedure PROC RANK is used to assign the mid-ranks among all observations to the data. Note that the assignment of mid-ranks is the default with this procedure in SAS.

Heteroscedastic Model The procedure PROC MIXED provides the possibility to define the structure of the covariance matrix of the 'cell means' by the option 'TYPE=···' within the 'REPEATED' statement. Moreover, the 'GRP=···' option within the 'REPEATED' statement defines the factor levels (or combinations of them) where different variances are allowed. Note that many types of covariance matrices can be defined by these options (including diagonal matrices) so that the notation 'MIXED' of this SAS-procedure may be somewhat misleading. The WTS $Q_N(\boldsymbol{C})$ and the resulting p-values are printed out by adding the option 'CHISQ' after the slash '/' in the MODEL statement.

For independent observations, the covariance matrix has a diagonal structure which is defined by 'TYPE=UN(1)'. In general, for the nonparametric main effects and all interactions, the variances in this diagonal matrix may be different for all factor level combinations. Thus, the highest interaction term must be assigned in the 'GRP' option. For example, in a three-way layout with factors A, B and C, this option is 'GRP=A*B*C'.

Starting with version 8.0 (which should be available after the end of the year 1999), the option 'ANOVAF' can be added somewhere in the line of the PROC MIXED statement in order to print out the ATS $F_N(\boldsymbol{M})$ and the resulting p-values. The use of the ATS is recommended for small and medium numbers of replications.

Example: Fertility Trial (Results, see Table 1.3)

```
DATA fert;
INPUT   treat$ year number;
CARDS;
PL   1   13
PL   1   12
:
PL   2   12
D1   1   15
:
D2   2   14
;
RUN;
```

```
PROC RANK  DATA=fert  OUT=fert;
VAR    number;
RANKS    r;
RUN;

PROC MIXED  DATA=fert  ANOVAF;
CLASS    treat  year;
MODEL    r = treat | year / CHISQ;
REPEATED  /  TYPE=UN(1)    GRP= treat*year;
RUN;
```

The computation of the variances for the confidence intervals (see Subsection 1.6) needs some more involved rankings of the data which may be performed by a special macro where DATA steps and different types of rankings are used. Unfortunately, these computations are not yet available with a SAS standard procedure or with any statistical software package, to the best of our knowledge.

2 Repeated Measures

In a repeated measures model, randomly chosen subjects are observed repeatedly under the same or under different treatments. Such designs occur in many biological experiments and medical or psychological studies. They include growth curves, longitudinal data or repeated measures designs where a special structure for the dependencies of the multivariate observations, e.g. the compound symmetry, may or may not be assumed.

Nonparametric hypotheses and tests for the mixed model have already been considered by Sen (1967), Koch and Sen (1968) and by Koch (1969, 1970). In the latter article, a complex split-plot design is considered and different types of ranks are given to aligned and original observations and the asymptotic distributions of univariate and multivariate rank statistics are given. Mainly joint hypotheses in the linear model are considered, i.e., main effects and certain interactions are tested together. However, no unified theory for the derivation of rank tests in repeated measures models is presented in these papers. Moreover, some of the statistics are aligned rank statistics (not pure rank statistics) and therefore, they are restricted to linear models.

First ideas to use the so-called *marginal model* to define treatment effects in a nonparametric mixed model date back to Hollander, Pledger and Lin (1974) and Govindarajulu (1975) and were extended later on and studied in more detail by Brunner and Neumann (1982), Thompson (1990, 1991) and Brunner and Denker (1994). In this marginal model, a treatment effect is defined through the marginal distributions F_s, $s = 1, \ldots, d$ of $\boldsymbol{X}_k = (X_{k1}, \ldots, X_{kd})'$ where $\boldsymbol{X}_k$ is the vector of observations for subject k. The observations X_{ks} and $X_{k's'}$ coming from different subjects k and k' are assumed to be independent while the observations X_{ks} and $X_{ks'}$ from the same subject may be dependent.

A general formulation of hypotheses in the nonparametric marginal model was suggested by Akritas and Arnold (1994) who introduced the idea to formulate the hypotheses in terms of the distribution functions. They derived the relevant asymptotic distribution theory under the assumption of the continuity of the distribution functions which means that ties were not allowed. This is rather an unrealistic assumption for applications. Based on the idea of the normalized version of the marginal distribution function $F(x) = \frac{1}{2}[F^+(x) + F^-(x)]$ and of the empirical marginal distribution function (see Ruymgaart, 1980), Akritas and Brunner (1997) provided a unified approach to nonparametric repeated measures models using the concept of Akritas and Arnold (1994) to formulate nonparametric hypotheses. Brunner, Munzel and Puri (1999) generalized these results to the case of (randomly) missing values, singular covariance matrices, score functions with a bounded second derivative. Note that singular covariance matrices may appear quite often in models with ordered categorical data. Moreover, they derived an approximation of the distribution under the hypothesis for the ANOVA-type statistic in this setup. This approximation is particularly useful for small samples.

In what follows, we provide a summary of the main results of the above papers in a unified form such that procedures for particular problems or special designs may be derived from the general framework presented here.

2.1 Nonparametric Marginal Model

In the general repeated measures (or mixed) model, r treatment groups (the so-called whole-plot factor) are considered where every treatment group i contains $k = 1, \ldots, n_i$ independent (randomly chosen) subjects. These $n = \sum_{i=1}^{r} n_i$ subjects are observed repeatedly under $s = 1, \ldots, d$ different (fixed) situations (levels of the 'treatment factor', the so-called sub-plot factor) with $\ell = 1, \ldots, m_{iks}$ replications for subject k under the treatment combination(i, s). Thus, there are $M_{ik} = \sum_{s=1}^{d} m_{iks}$ repeated measures for each subject where the subjects are repeatedly observed under the same treatment as well as under different treatments. This general mixed model can be written by independent random vectors

$$
\begin{aligned}
\boldsymbol{X}_{ik} &= (\boldsymbol{X}'_{ik1}, \ldots, \boldsymbol{X}'_{ikd})', \ i = 1, \ldots, r, \ k = 1, \ldots, n_i \quad \text{where} \\
& \boldsymbol{X}_{iks} = (X_{iks1}, \ldots, X_{iksm_{iks}})', \ s = 1, \ldots, d
\end{aligned} \tag{2.1}
$$

and where $X_{iks\ell} \sim F_{is}(x) = \frac{1}{2}\left[F_{is}^{+}(x) + F_{is}^{-}(x)\right], i = 1, \ldots, r, s = 1, \ldots, d, k = 1, \ldots, n_i$, $\ell = 1, \ldots, m_{iks}$ (the sign $\sim$ means 'is distributed as'). To derive the general results, no particular structure is assumed for the dependencies between the components of the vectors $\boldsymbol{X}_{ik}$. It is only assumed that the vectors $\boldsymbol{X}_{ik}$ are independent, $i = 1, \ldots, r$, $k = 1, \ldots, n_i$ and that the bivariate marginal distribution functions of $(X_{iks\ell}, X_{iks'\ell'})$ do not depend on k, ℓ and ℓ', i.e. $(X_{iks\ell}, X_{iks'\ell'}) \sim F_{iss'}(x, y)$. This assumption is reasonable since the observations with $k \neq k'$ are independent replications and the observations with $\ell \neq \ell'$ for the same s and k are dependent replications of the same experiment. The dependencies between the observations on the same subject are considered as 'nuisance parameters' and their impact on the asymptotic distribution of the statistics to be derived in this section has to be estimated separately.

The rather general notation introduced above, covers a lot of designs which are commonly used in practice.

1. *Paired samples design:* This design is derived from (2.1) by letting $r = 1, n_i = n, s = 2, m_{iks} \equiv 1$. Here, n independent pairs of random variables $\boldsymbol{X}_k = (X_{k1}, X_{k2})'$, $k = 1, \ldots, n$ are observed, where $F_{is}(x) \equiv F_s(x)$, $s = 1, 2$.

2. *Simple repeated measures design:* Here, $r = 1$ group of $k = 1, \ldots, n$ subjects is observed under $s = 1, \ldots, d$ treatments and $F_{is} \equiv F_s$, $s = 1, \ldots, d$.

3. *Split-plot design:* In this design, $i = 1, \ldots, r \geq 2$ groups of $k = 1, \ldots, n_i$ independent subjects are observed.

4. *Two-fold nested design:* In this design, where $i = 1, \ldots, r$ treatments are applied, $k = 1, \ldots, n_i$ independent subjects are observed within each treatment group i. Every subject receives only one treatment but it is observed repeatedly $\ell = 1, \ldots, m_{ik}$ times under the same treatment in order to get a more accurate measurement for the variable of interest. In total, there are $N = \sum_{i=1}^{r} \sum_{k=1}^{n_i} m_{ik}$ observations of $n = \sum_{i=1}^{r} n_i$ independent subjects.

5. *Higher-Way layouts:* Higher-way layouts with repeated measures or longitudinal data are covered by the general model defined in (2.1) by splitting the indices i or s into sub-indices $i', i'', \ldots$ or $s', s'', \ldots$, respectively.

Note that in most cases with longitudinal data, $m_{iks} = 1$ or $m_{iks} = 0$ (if the observation is missing). The case of $m_{iks} \geq 1$ typically occurs when some material, tissue or a set of individuals is split into several homogeneous parts and the compound symmetry model can be used as an appropriate model for this design.

To introduce the ideas, to define treatment effects, to formulate hypotheses and to derive test procedures in the nonparametric marginal model, we consider only the case where $m_{iks} = 1$ in order to keep the notation simple. This means that we do not consider missing values and dependent replications. Regarding these cases, we refer to the literature (Brunner and Puri, 1996; Brunner, Munzel and Puri, 1999).

2.2 Relative Effects, Hypotheses and Estimators

Since no parameters are involved in the nonparametric model (2.1), the marginal distribution functions $F_{is}(x)$ are used to describe an effect (e.g. time effect or treatment effect). To this end, the so-called relative marginal effects $p_{is} = \int H(x)dF_{is}(x)$ are considered where $H(x) = N^{-1}\sum_{i=1}^{r}\sum_{s=1}^{d} n_i F_{is}(x)$ is the average of all $N = d \cdot \sum_{i=1}^{r} n_i$ distribution functions in the experiment. Let $\boldsymbol{F} = (F_{11}, \ldots, F_{1d}, \ldots, F_{r1}, \ldots, F_{rd})'$ the vector of the marginal distribution functions and let $\boldsymbol{p} = \int Hd\boldsymbol{F} = (p_{11}, \ldots, p_{1d}, \ldots, p_{r1}, \ldots, p_{rd})'$, the vector of the relative (marginal) effects.

In the nonparametric setup introduced above, hypotheses are formulated by the distribution functions $F_{is}(x)$ in the same way as for independent observations. Let $\boldsymbol{C}$ denote a covariance matrix (see section 1.2.1). Then a nonparametric hypothesis for a mixed model in its most general form is written as $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$. Some examples for nonparametric hypotheses are given in Section 2.5.

The vector of the relative marginal effects is estimated by replacing $F_{is}(x)$ and $H(x)$ by the empirical functions

$$\widehat{F}_{is}(x) = \frac{1}{n_i}\sum_{k=1}^{n_i} c(x - X_{iks}), \qquad \widehat{H}(x) = \frac{1}{N}\sum_{i=1}^{r}\sum_{s=1}^{d}\sum_{k=1}^{n_i} c(x - X_{iks}). \tag{2.2}$$

Here, $c(u) = \frac{1}{2}\left[c^+(u) + c^-(u)\right]$ denotes the normalized version of the counting function. Then, the relative marginal effects p_{is} are estimated by

$$\widehat{p}_{is} = \int \widehat{H}d\widehat{F}_{is} = \frac{1}{n_i}\sum_{k=1}^{n_i}\widehat{H}(X_{iks}) = \frac{1}{n_i}\sum_{k=1}^{n_i}\frac{1}{N}\left(R_{iks} - \frac{1}{2}\right), \tag{2.3}$$

where R_{iks} is the mid-rank of X_{iks} among all N observations. Let $\overline{\boldsymbol{R}}_{i\cdot} = n_i^{-1}\sum_{k=1}^{n_i}\boldsymbol{R}_{ik}$ denote the mean of the vectors $\boldsymbol{R}_{ik} = (R_{ik1}, \ldots, R_{ikd})'$, $i = 1, \ldots, r$, within level i of the whole-plot factor and let $\overline{\boldsymbol{R}}_{\cdot} = (\overline{\boldsymbol{R}}_{1\cdot}', \ldots, \overline{\boldsymbol{R}}_{d\cdot}')'$. Then, the vector $\boldsymbol{p}$ of the relative marginal effects is estimated by

$$\widehat{\boldsymbol{p}} = \int \widehat{H} d\widehat{\boldsymbol{F}} = \frac{1}{N}\left(\overline{\boldsymbol{R}}_{\cdot} - \tfrac{1}{2}\mathbf{1}_r \otimes \mathbf{1}_d\right) = \frac{1}{N}\begin{pmatrix} \overline{\boldsymbol{R}}_{1\cdot} - \frac{1}{2}\mathbf{1}_d \\ \vdots \\ \overline{\boldsymbol{R}}_{r\cdot} - \frac{1}{2}\mathbf{1}_d \end{pmatrix} \tag{2.4}$$

$$= \frac{1}{N}\begin{pmatrix} \overline{R}_{1\cdot 1} - \frac{1}{2} \\ \vdots \\ \overline{R}_{1\cdot d} - \frac{1}{2} \\ \vdots \\ \overline{R}_{r\cdot 1} - \frac{1}{2} \\ \vdots \\ \overline{R}_{r\cdot d} - \frac{1}{2} \end{pmatrix} = \frac{1}{N}\begin{pmatrix} \widehat{p}_{11} \\ \vdots \\ \widehat{p}_{1d} \\ \vdots \\ \widehat{p}_{r1} \\ \vdots \\ \widehat{p}_{rd} \end{pmatrix}.$$

The notation given in (2.4) enables a simple and short presentation of the asymptotic theory of rank statistics in nonparametric factorial designs with repeated measures.

We note that $\widehat{p}_{is}$ is only an asymptotically unbiased estimator of p_{is}. To derive the exact expectation of $\widehat{p}_{is}$, let $H^{(i)}(x) = \frac{1}{d}\sum_{\ell=1}^{d} F_{i\ell}(x)$ denote the mean distribution function within the level i of the whole-plot factor and let $p_s^{(i)} = \int H^{(i)} dF_{is}$. Further let $\Delta_s^{(i)} = \frac{1}{d}\sum_{\ell=1}^{d} \Delta_{\ell s}^{(i)}$ denote the mean of the within-subjects probabilities

$$\Delta_{\ell s}^{(i)} = P(X_{i1\ell} < X_{i1s}) + \frac{1}{2}P(X_{i1\ell} = X_{i1s})$$

and let $n = \sum_{i=1}^{r} n_i$ denote the total number of subjects. Then,

$$E(\widehat{p}_{is}) = p_{is} + \frac{1}{n}\left(\Delta_s^{(i)} - p_s^{(i)}\right), \quad i = 1, \ldots, r; \ s = 1, \ldots, d.$$

Obviously, $\widehat{p}_{is}$ is unbiased if $\Delta_s^{(i)} = p_s^{(i)}$. Let $\overline{R}_{i\cdot s}^{(i)} = n_i^{-1}\sum_{k=1}^{n_i} R_{iks}^{(i)}$, where $R_{iks}^{(i)}$ denotes the (mid-)rank of X_{iks} among all $n_i d$ observations within level i of the whole-plot factor and let $\overline{R}_{i\cdot s}^{*} = n_i^{-1}\sum_{k=1}^{n_i} R_{iks}^{*}$, where R_{iks}^{*} is the (mid-)rank of X_{iks} among all d observations within subject k in level i of the whole-plot factor. Note that $1 \le R_{iks}^{(i)} \le n_i d$ and $1 \le R_{iks}^{*} \le d$. Then, in practice, the bias can be checked by comparing

$$\widehat{\Delta}_s^{(i)} = \frac{1}{d}\left(\overline{R}_{i\cdot s}^{*} - \frac{1}{2}\right) \quad \text{and} \quad \widehat{p}_s^{(i)} = \frac{1}{n_i d}\left(\overline{R}_{i\cdot s}^{(i)} - \frac{1}{2}\right).$$

2.3 Asymptotic Theory

In this Section, the asymptotic distribution of $\sqrt{n}\boldsymbol{C}\widehat{\boldsymbol{p}} = \sqrt{n}\boldsymbol{C}(\widehat{p}_{11}, \ldots, \widehat{p}_{rd})'$ is derived under the hypothesis $H_0^F : \boldsymbol{CF} = \mathbf{0}$. Moreover, consistent estimators of the covariance matrix are provided for the compound symmetry model as well as for the multivariate model.

2.3.1 Basic Results and Assumptions

The asymptotic results are derived under the following assumptions:

ASSUMPTIONS 2.1

(a) $\min n_i \to \infty, i = 1, \dots, r,$

(b) $n/n_i \leq N_0 < \infty, i = 1, \dots, r,$ *where* $n = \sum_{i=1}^{r} n_i$ *is the total number of the subjects.*

The first result is that $\widehat{p}_{is}$ given in (2.3), is consistent for $p_{is} = \int H dF_{is}$ in the sense given in the following Proposition.

PROPOSITION 2.2 *Let* $\boldsymbol{X}_{ik} = (X_{ik1}, \dots, X_{ikd})'$ *be independent and identically distributed random vectors. Then, under the assumption 2.1 (a),* $E(\widehat{p}_{is} - p_{is})^2 \to 0$ *as* $n_i \to \infty$, $i = 1, \dots, r$, $s = 1, \dots, d$.

PROOF: see Brunner, Munzel and Puri (1999), for example. □

Next, the basic asymptotic equivalence for the mixed model is stated.

THEOREM 2.3 *Let* $\boldsymbol{X}_{ik}$ *be as given in Proposition 2.2 and let* $\boldsymbol{F} = (F_{11}, \dots, F_{rd})'$ *denote the vector of the marginal distributions and* $\widehat{\boldsymbol{F}} = (\widehat{F}_{11}, \dots, \widehat{F}_{rd})'$, *the vector of the empirical marginal distributions as given in (2.2). Then, under the assumptions 2.1 (a) and (b),*

$$\sqrt{n} \int \widehat{H} d\left(\widehat{\boldsymbol{F}} - \boldsymbol{F}\right) \doteq \sqrt{n} \int H d\left(\widehat{\boldsymbol{F}} - \boldsymbol{F}\right) = \sqrt{n}\left(\overline{\boldsymbol{Y}}_{\cdot} - \int H d\boldsymbol{F}\right)$$

where

$$\begin{aligned} \overline{\boldsymbol{Y}}_{\cdot} &= \left(\overline{\boldsymbol{Y}}'_{1\cdot}, \dots, \overline{\boldsymbol{Y}}'_{r\cdot}\right)', \quad \overline{\boldsymbol{Y}}_{i\cdot} = (\overline{Y}_{i\cdot 1}, \dots, \overline{Y}_{i\cdot d})' = \frac{1}{n_i}\sum_{k=1}^{n_i} \boldsymbol{Y}_{ik}, \\ \boldsymbol{Y}_{ik} &= (Y_{ik1}, \dots, Y_{ikd})', \quad Y_{iks} = H(X_{iks}), \; s = 1, \dots, d. \end{aligned} \tag{2.5}$$

PROOF: see Brunner, Munzel and Puri (1999), for example. □

2.3.2 Asymptotic Normality

To establish the asymptotic normality of $\boldsymbol{C}\widehat{\boldsymbol{p}}$ (or $\sqrt{n}\boldsymbol{C}\widehat{\boldsymbol{p}}$ to be precise), a further regularity assumption is needed. First note that the vectors $\overline{\boldsymbol{Y}}_{i\cdot}$ are independent. Thus,

$$\boldsymbol{V}_n = Cov(\sqrt{n}\,\overline{\boldsymbol{Y}}_{\cdot}) = \bigoplus_{i=1}^{r} \frac{n}{n_i} \boldsymbol{V}_i\,, \tag{2.6}$$

where $\boldsymbol{V}_i = Cov(\boldsymbol{Y}_{i1})$ and $\boldsymbol{Y}_{ik}$ is given in (2.5). Let $\rho_m(i)$ denote the smallest characteristic root of $\boldsymbol{V}_i$.

ASSUMPTION 2.1

(c) $\rho_m(i) \geq \rho_0 > 0, i = 1, \dots, r.$

THEOREM 2.4 *Let X_{ik} be as in Proposition 2.2 and let V_n be as given in (2.6). Then, under the assumptions 2.1 (a), (b) and (c) and under H_0^F : $CF = 0$, the statistic $\sqrt{n}C\widehat{p} = \sqrt{n}C \int \widehat{H} d\widehat{F}$ has, asymptotically, a multivariate normal distribution with mean 0 and covariance matrix CV_nC'.*

PROOF: The proof follows easily from the assumptions, from Theorem 2.3 and the Central Limit Theorem by noting that the random vectors Y_{ik} are independent and identically distributed. □

Regarding the asymptotic equivalence of the random vectors $\sqrt{n}C\widehat{p}$ and $\sqrt{n}C\overline{Y}_{\cdot}$ under H_0^F : $CF = 0$, the same considerations apply as given in subsection 1.4.2 for the case of independent observations.

2.3.3 Estimation of the Asymptotic Covariance Matrix

In most practical examples, the covariance matrices V_i defined in (2.6) are unknown and must be estimated from the data. To derive a consistent estimator of V_i, $i = 1, \ldots, r$, two models are distinguished. The *multivariate model* does not assume any special pattern for the bivariate marginal distribution functions while the *compound symmetry model* assumes the equality of all covariances under the hypothesis. This is stated in details in the last part of this subsection. In what follows, the estimators for V_i are provided for both models and the assumptions are given under which the consistency of these estimators follows.

Multivariate Model In the multivariate model, let $R_{ik} = (R_{ik1}, \ldots, R_{ikd})'$ denote the vector of the ranks R_{iks} of X_{iks} among all the $N = nd$ observations and let $\overline{R}_{i\cdot} = n_i^{-1} \sum_{k=1}^{n_i} R_{ik}$ denote the mean of these rank vectors within the treatment level i of the whole-plot factor, $i = 1, \ldots, r$. Finally, let

$$\widehat{V}_i = \frac{1}{N^2(n_i - 1)} \sum_{k=1}^{n_i} \left(R_{ik} - \overline{R}_{i\cdot}\right)\left(R_{ik} - \overline{R}_{i\cdot}\right)' \tag{2.7}$$

denote the sample covariance matrix of $\frac{1}{N}R_{ik}$, $k = 1, \ldots, n_i$, $i = 1, \ldots, r$, and let

$$\widehat{V}_n = \bigoplus_{i=1}^{r} \frac{n}{n_i} \widehat{V}_i \tag{2.8}$$

denote an estimator of V_n.

THEOREM 2.5 *Let V_i and V_n be as defined in (2.6) and let $\widehat{V}_i$ and $\widehat{V}_n$ be as given in (2.7) and (2.8) respectively. Then, under the assumptions 2.1 (a), (b) and (c),*

1. $\|\widehat{V}_i - V_i\| \xrightarrow{p} 0, \quad i = 1, \ldots, r,$ *and*

2. $\|\widehat{V}_n - V_n\| \xrightarrow{p} 0$,

where $\|\cdot\|$ *denotes the Euclidean norm of a matrix.*

PROOF: see Akritas and Brunner (1997). □

REMARK 2.1 *A stronger result, namely the* L_2*-consistency was shown by Brunner, Munzel and Puri (1999) in a more general setup. For details, we refer to this paper.*

It should be emphasized that $\boldsymbol{V}_n$ is the covariance matrix of $\sqrt{n}\,\overline{\boldsymbol{Y}}.$ and not of $\sqrt{n}\,\widehat{\boldsymbol{p}}$. The matrix $\boldsymbol{V}_n$ is only needed to compute the asymptotic covariance matrix $\boldsymbol{C}\boldsymbol{V}_n\boldsymbol{C}'$ of $\sqrt{n}\,\boldsymbol{C}\widehat{\boldsymbol{p}}$ under the hypothesis $H_0^F : \boldsymbol{C}\boldsymbol{F} = \boldsymbol{0}$.

Compound Symmetry Model In the compound symmetry model, it is assumed that under $H_0^F : F_{i1} = \cdots = F_{id}$, the bivariate marginal distribution functions of $(X_{iks}, X_{iks'})$ do not depend on k, s and s', i.e. $(X_{iks}, X_{iks'}) \sim F_{iss'}(x, y) = F_i^*(x, y)$, $s \neq s' = 1, \ldots, d$. Thus, under H_0^F, the variances and the covariances are given by

$$\sigma_i^2 \equiv \sigma_{is}^2 = Var(Y_{i1s}),\ s = 1, \ldots, d,\ i = 1, \ldots, r \tag{2.9}$$

$$c_i^* = Cov(Y_{i1s}, Y_{i1s'}),\ s \neq s' = 1, \ldots, d,\ i = 1, \ldots, r \tag{2.10}$$

and it follows that $\boldsymbol{V}_i = (\sigma_i^2 - c_i^*)\boldsymbol{I}_d + c_i^*\boldsymbol{J}_d$.

Compound symmetry is only assumed for hypotheses regarding the sub-plot factor, i.e. for hypotheses which can be written as $H_0^F : (\boldsymbol{I}_r \otimes \boldsymbol{C}_d)\boldsymbol{F} = \boldsymbol{0}$ where $\boldsymbol{C}_d$ is a suitable contrast matrix for the sub-plot factor. Thus, it is only necessary to estimate $\tau_i = \sigma_i^2 - c_i^*$, $i = 1, \ldots, r$, since $(\boldsymbol{I}_r \otimes \boldsymbol{C}_d)\bigoplus_{i=1}^{r} c_i^*\boldsymbol{J}_d = \boldsymbol{0}$.

THEOREM 2.6 *Let* $\tau_i = \sigma_i^2 - c_i^*$, $i = 1, \ldots, r$, *where* σ_i^2 *and* c_i^* *are defined in (2.9) and (2.10) respectively, and let*

$$\widehat{\tau}_i = \frac{1}{N^2 n_i(d-1)} \sum_{s=1}^{d} \sum_{k=1}^{n_i} (R_{iks} - \overline{R}_{ik\cdot})^2, \tag{2.11}$$

where $\overline{R}_{ik\cdot} = \frac{1}{d}\sum_{s=1}^{d} R_{iks}$ *is the mean of the ranks* R_{iks} *within subject* k *and* R_{iks} *is the rank of* X_{iks} *among all the* $N = nd$ *observations. Then, in the compound symmetry model, under the assumptions 2.1 (a), (b) and (c) and under the hypothesis* $H_0^F : F_{i1} = \cdots = F_{id}$, *the estimator* $\widehat{\tau}_i$ *is consistent for* τ_i *in the sense that* $E(\widehat{\tau}_i/\tau_i - 1)^2 \to 0$ *as* $n_i \to \infty$. *Moreover,*

$$\boldsymbol{C}\widehat{\boldsymbol{V}}_n\boldsymbol{C}' = (\boldsymbol{I}_r \otimes \boldsymbol{C}_d)\,\widehat{\boldsymbol{V}}_n\,(\boldsymbol{I}_r \otimes \boldsymbol{C}_d') = \bigoplus_{i=1}^{r} \widehat{\tau}_i \boldsymbol{C}_d \boldsymbol{C}_d'\,. \tag{2.12}$$

PROOF: see Brunner, Munzel and Puri (1999) where the more general case of $m_{iks} \geq 1$ is considered. □

2.4 Statistics

To test the nonparametric hypothesis $H_0^F : \boldsymbol{CF} = \mathbf{0}$, the rank versions of the WTS and of the ATS are considered to detect general alternatives while a linear ranks statistic is used to detect special patterned alternatives. Other statistics which are commonly used in multivariate analysis are not discussed here since they require the equality of the covariance matrices. In a nonparametric setup, however, this assumption is only justified in a few special cases. Note that in general any assumed homoscedasticity of the parent distribution functions is not transferred to the asymptotic rank transform $Y_{iks} = H(X_{iks})$ because $H(\cdot)$ is a non-linear transformation.

2.4.1 Quadratic Forms

Wald-Type Statistics (WTS) Let $\widehat{\boldsymbol{V}}_n$ denote the consistent estimator of $\boldsymbol{V}_n$ which is given in (2.8) and let $[\boldsymbol{C}\widehat{\boldsymbol{V}}_n\boldsymbol{C}']^-$ denote a g-inverse of $\boldsymbol{C}\widehat{\boldsymbol{V}}_n\boldsymbol{C}'$. If $\boldsymbol{V}_n \to \boldsymbol{V} \neq \mathbf{0}$ such that $rank(\boldsymbol{CV}_n) = rank(\boldsymbol{CV})$, then under $H_0^F : \boldsymbol{CF} = \mathbf{0}$, it follows from Theorem 2.4 and Theorem 2.5 that the rank version of the WTS

$$Q_n^W(C) \quad = \quad n\widehat{\boldsymbol{p}}'\boldsymbol{C}' \left[\boldsymbol{C}\widehat{\boldsymbol{V}}_n\boldsymbol{C}'\right]^- \boldsymbol{C}\widehat{\boldsymbol{p}} \qquad (2.13)$$

has, asymptotically, a central χ^2-distribution with $f = rank(\boldsymbol{CV})$ degrees of freedom. However, extremely large sample sizes are needed to achieve an acceptable approximation by this distribution. Therefore, the ANOVA-type statistic is considered in the following paragraph (as in the case of independent observations in subsection 1.5.1).

ANOVA-Type Statistics (ATS) Let $\boldsymbol{M} = \boldsymbol{C}'[\boldsymbol{CC}']^-\boldsymbol{C}$ where $[\boldsymbol{CC}']^-$ denotes some g-inverse of $\boldsymbol{CC}'$. Then, the rank version of the ATS is defined by

$$Q_n^A(C) \quad = \quad n\widehat{\boldsymbol{p}}'\boldsymbol{M}\widehat{\boldsymbol{p}}. \qquad (2.14)$$

Note that $\boldsymbol{M}$ is a projection matrix and that $\boldsymbol{MF} = \mathbf{0} \iff \boldsymbol{CF} = \mathbf{0}$ because $\boldsymbol{C}'[\boldsymbol{CC}']^-$ is a generalized inverse of $\boldsymbol{C}$. Thus, it is also reasonable to use $Q_n^A(C)$ as a test statistic for testing the hypothesis $H_0^F : \boldsymbol{CF} = \mathbf{0}$. The asymptotic distribution of $Q_n^A(C)$ is given in the next Theorem.

THEOREM 2.7 *Let $\boldsymbol{M} = \boldsymbol{C}'[\boldsymbol{CC}']^-\boldsymbol{C}$ and let $\boldsymbol{V}_n$ and $\widehat{\boldsymbol{V}}_n$ be as in (2.6) and (2.8) respectively. Then, under the hypothesis $H_0^F : \boldsymbol{CF} = \mathbf{0}$ and under the assumptions 2.1 (a), (b) and (c), the statistic $Q_n^A(C)$ given in (2.14) has, asymptotically, the same distribution as the random variable $\sum_{i=1}^r \sum_{s=1}^d \lambda_{is} Z_{is}$, where the λ_{is} are the characteristic roots of $\boldsymbol{MV}_n\boldsymbol{M}$ and the Z_{is} are independent random variables each having a central χ_1^2-distribution.*

PROOF: The proof follows from Theorem 2.4 and well known theorems on the distribution of quadratic forms (see e.g. Mathai and Provost, 1992, Chapter 4). □

The distribution of $\sum_{i=1}^{r}\sum_{s=1}^{d}\lambda_{is}Z_{is}$ can be approximated by a scaled χ^2-distribution in the same way as discussed in the case of independent observations (see subsection 1.5.1, approximation procedure 1.9).

APPROXIMATION PROCEDURE 2.8

1. *Assume that* $tr(\boldsymbol{M}\boldsymbol{V}_n) \geq t_0 > 0$. *Then, under* $H_0^F : \boldsymbol{C}\boldsymbol{F} = \boldsymbol{0}$, *the first two moments of the asymptotic distribution of* $Q_n^A(\boldsymbol{C})/tr(\boldsymbol{M}\boldsymbol{V}_n)$ *and of the* $F(f,\infty)$*-distribution coincide for* $f = [tr(\boldsymbol{M}\boldsymbol{V}_n)]^2/tr(\boldsymbol{M}\boldsymbol{V}_n\boldsymbol{M}\boldsymbol{V}_n)$.

2. *The unknown traces* $tr(\boldsymbol{M}\boldsymbol{V}_n)$ *and* $tr(\boldsymbol{M}\boldsymbol{V}_n\boldsymbol{M}\boldsymbol{V}_n)$ *can be estimated consistently by replacing* $\boldsymbol{V}_n$ *with* $\widehat{\boldsymbol{V}}_n$ *given in (2.8) and (2.12) respectively. This finally leads to the statistic*

$$F_n(\boldsymbol{C}) = \frac{1}{tr(\boldsymbol{M}\widehat{\boldsymbol{V}}_n)} Q_n^A(\boldsymbol{C}) = \frac{n}{tr(\boldsymbol{M}\widehat{\boldsymbol{V}}_n)} \widehat{\boldsymbol{p}}' \boldsymbol{M} \widehat{\boldsymbol{p}} \;\dot{\sim}\; F(\widehat{f},\infty), \tag{2.15}$$

where

$$\widehat{f} = \frac{[tr(\boldsymbol{M}\widehat{\boldsymbol{V}}_n)]^2}{tr(\boldsymbol{M}\widehat{\boldsymbol{V}}_n\boldsymbol{M}\widehat{\boldsymbol{V}}_n)}. \tag{2.16}$$

(Here, the sign $\dot{\sim}$ *means 'approximately distributed as'.)*

DERIVATION: see Brunner, Munzel and Puri (1999). □

REMARK 2.2 *The approximation procedure goes back to Box (1954) and turns out to be quite accurate for independent observations (see Brunner, Dette and Munk, 1997). For repeated measures,* $\widehat{f}$ *in (2.16) may be biased for small sample sizes. In all cases where the factor 'time' is not involved (i.e. tests for the whole-plot factors and their interactions), the approximation in (2.15) can be improved by estimating the second degree of freedom in a similar way as for independent observations in (1.21). The details are omitted for brevity.*

Comparison of the WTS and the ATS The main advantage of the WTS $Q_n^W(\boldsymbol{C})$ is that its asymptotic distribution under H_0^F is a known distribution function, namely a χ^2-distribution. The general drawback of $Q_n^W(\boldsymbol{C})$ is that it converges extremely slowly to its asymptotic distribution resulting in rather liberal decisions for small or moderate sample sizes. Moreover, the restrictive assumption that $\boldsymbol{V}_n \to \boldsymbol{V}$ such that $rank(\boldsymbol{C}\boldsymbol{V}_n) = rank(\boldsymbol{C}\boldsymbol{V})$ cannot be checked.

The ATS $F_n(\boldsymbol{C})$ has the main disadvantage that its asymptotic distribution under H_0^F contains unknown quantities, namely the characteristic roots of $\boldsymbol{M}\boldsymbol{V}_n\boldsymbol{M}$ which are unknown in general and must be estimated where the Box-approximation is used to approximate the distribution of the ANOVA-type statistic. However, note that even in the asymptotic case, the χ_f^2/f-distribution is an approximation of the true distribution of the statistic under H_0^F. One advantage is that it is neither necessary to assume the convergence of the covariance matrix $\boldsymbol{V}_n$ to a constant matrix $\boldsymbol{V}$ nor that the rank of $\boldsymbol{C}\boldsymbol{V}_n$ is preserved in the limit $\boldsymbol{C}\boldsymbol{V}$. The only additional assumption to the assumptions 2.1 (a), (b) and (c) which is needed for the ATS, is that $tr(\boldsymbol{M}\boldsymbol{V}_n) \neq 0$ which means that - regarding the hypothesis of interest - there is at least some variation among the observations of the experiment. This is close to a trivial assumption.

The main advantage of the ATS $F_n(\boldsymbol{C})$ is that the approximation by the χ_f^2/f-distribution works also fairly well for rather small sample sizes (for details, see e.g. Brunner and Langer, 1999) and can be recommended for small and moderate sample sizes.

2.4.2 Patterned Alternatives

As in the case of independent observations, the method of Page (1963) and Hettmansperger and Norton (1987) is also used for repeated measures to derive test statistics which are especially sensitive against a conjectured patterned alternative. The estimated treatment effects are weighted by a set of constants $w_{11}, \ldots, w_{rd}$ reproducing the conjectured pattern of the alternative which has to be specified in advance. Let $\boldsymbol{w} = (w_{11}, \ldots, w_{rd})'$ denote the vector of the weights w_{is}. Then under $H_0^F : \boldsymbol{C}\boldsymbol{F} = \boldsymbol{0}$, the linear rank statistic

$$L_n(\boldsymbol{w}) \;=\; \sqrt{n}\,\boldsymbol{w}'\boldsymbol{C}\,\widehat{\boldsymbol{p}} \tag{2.17}$$

has, asymptotically, a normal distribution with mean 0 and variance

$$\sigma_w^2 \;=\; \boldsymbol{w}'\boldsymbol{C}\boldsymbol{V}_n\boldsymbol{C}'\boldsymbol{w}\,,$$

which can be estimated consistently by

$$\widehat{\sigma}_w^2 \;=\; \boldsymbol{w}'\boldsymbol{C}\widehat{\boldsymbol{V}}_n\boldsymbol{C}'\boldsymbol{w}\,,$$

where $\widehat{\boldsymbol{V}}_n$ is given in the Theorems 2.5 and 2.6 respectively. Then, under $H_0^F : \boldsymbol{C}\boldsymbol{F} = \boldsymbol{0}$, the statistic $T_n(\boldsymbol{w}) = L_n(\boldsymbol{w})/\widehat{\sigma}_w$ has, asymptotically, a standard normal distribution.

Approximations for small samples have to be derived separately for the special designs. The considerations are similar to those in subsection 1.5.2 and are therefore omitted. Some examples can be found in Akritas and Brunner (1996).

2.4.3 The 'Rank Transform' (RT) Property for Repeated Measures

The WTS-statistics given in (2.13) can formally be derived from the parametric MANOVA statistics by replacing the original observations X_{iks} by their ranks R_{iks}. However, one must be careful with the assumptions of the model. First of all, the underlying testing problem must be identified from the asymptotic equivalence of the rank statistic $\sqrt{n}\boldsymbol{C}\int \widehat{H}d\widehat{\boldsymbol{F}}$ and the random vector $\sqrt{n}\boldsymbol{C}\int H d\widehat{\boldsymbol{F}} = \sqrt{n}\boldsymbol{C}\widehat{\boldsymbol{Y}}$. which is a contrast vector of the asymptotic rank transform. It follows from Theorems 2.3 and 2.4 that these two random vectors are asymptotically equivalent if the hypothesis is formulated in terms of the distribution functions, i.e. $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$. Furthermore, it has been pointed out by Akritas (1990) that any assumed homoscedasticity of the random variables X_{iks} is not transferred to the ART $Y_{iks} = H(X_{iks})$ in general. Thus, in the nonparametric marginal model, the homoscedasticity of the covariance matrices $\boldsymbol{V}_i$ can only be assumed in very special cases under the hypothesis and therefore, MANOVA-statistics for multivariate heteroscedastic models are required. A rank statistic corresponding to such a parametric MANOVA-statistic, is said to have the *rank transform property (RTP).*

If a rank statistic has the RTP then this is of importance for computational purposes. The parametric counterpart of a RT-statistic which may be available in a statistical software package can be applied to the ranked data. Only the quality of approximation to the asymptotic distribution or some finite approximation has to be taken into account. In any case, it is necessary to identify the properties (like independence and heteroscedasticity) of the ART under the hypothesis. The RT should not be regarded as a technique to derive statistics rather than a property of a statistic which can be useful for computational purposes.

2.5 Applications to Special Designs

In this section, the general theory derived in the previous subsections is applied to some special factorial designs with repeated measures. Some explicit statistics are given and it is shown that several known rank statistics which are given in the literature, follow as special cases from the general approach. In particular, the paired samples design, the simple repeated measures design and the so-called split-plot design are considered.

2.5.1 Paired Samples Design

In this special case, there is only one group of observations ($r = 1$) and the vector $\boldsymbol{X}_k$ has two components ($t = 2$). Thus, we observe independent random vectors $\boldsymbol{X}_k = (X_{k1}, X_{k2})'$, $k = 1, \ldots, n$, where $X_{ks} \sim F_s(x)$, $s = 1, 2$. Let $p = \int F_1 dF_2$, then the relative treatment effects $p_s = \int H dF_s$, $s = 1, 2$, are linearly dependent since $p = \frac{3}{2} - 2p_1 = 2p_2 - \frac{1}{2}$. Let R_{ks} denote the rank of X_{ks} among all the $N = 2n$ observations, let $\overline{R}_{\cdot s} = \frac{1}{n}\sum_{k=1}^n R_{ks}$, $s = 1, 2$, and let

$$S_{n,0}^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(R_{k2} - R_{k1} - \overline{R}_{\cdot 2} + \overline{R}_{\cdot 1}\right)^2 .$$

Then the statistic

$$T_n^F = \sqrt{n}\,\frac{\overline{R}_{\cdot 2} - \overline{R}_{\cdot 1}}{S_{n,0}}$$

has, asymptotically, a standard normal distribution under the hypothesis $H_0^F : F_1 = F_2$ which is equivalently written as $H_0^F : \boldsymbol{C}\boldsymbol{F} = \boldsymbol{0}$, where $\boldsymbol{C} = (-1, 1)'$. This follows easily from the Theorems 2.4 and 2.5. The statistic T_n^F is the statistic of the 'paired-ranks test' which has been considered in the literature under different assumptions by Mehra and Puri (1967), Govindarajulu (1975), Raviv (1978), Brunner and Neumann (1986*b*), Brunner and Puri (1996) and by Munzel (1999*b*), among possibly others.

2.5.2 Simple Repeated Measures Design

In the simple repeated measures design, n subjects are repeatedly observed under t treatments. Thus, we have independent random vectors

$$\boldsymbol{X}_k = (X_{k1}, \ldots, X_{kt})', \quad k = 1, \ldots, n,$$

where $X_{ks} \sim F_s(x)$, $s = 1, \ldots, t$. To test the hypothesis $H_0^F : F_1 = \cdots = F_t$, let $\boldsymbol{C} = \boldsymbol{P}_t = \boldsymbol{I}_t - \frac{1}{t}\boldsymbol{J}_t$ denote the t-dimensional centering matrix. Let R_{ks} denote the rank of X_{ks} among all the tn observations and let $\overline{R}_{\cdot s}$ denote the mean of the ranks within treatment level s. Two models are distinguished: the multivariate model, where no special structure of the covariance matrix $\boldsymbol{V}_n$ is assumed, and the compound symmetry model where it is assumed under H_0^F that $\sigma_1^2 = \cdots = \sigma_s^2 = \sigma^2$ and that all the covariances are equal to $c^* = Cov(H(X_{11}), H(X_{12}))$, where $H(X_{1\ell}) = Y_{11\ell}$, $\ell = 1, 2$, as given in (2.9) and (2.10) respectively. Compound symmetry can be assumed if all permutations of the observations $X_{k1}, \ldots, X_{kt}$ within subject k are equally likely under the hypothesis H_0^F. This assumption is not justified if the 'treatments' are the time points of time curves. Generally, observations which are more closer are higher correlated than more distant observations. In the latter case, the so-called multivariate model is used. Both models differ only in the structure, and thus in the estimators of the covariance matrix $\boldsymbol{V}_n$. Note that compound symmetry is assumed to derive the well-known Friedman-statistic and its distribution under the hypothesis. Thus, the Friedman-statistic cannot be used for the analysis of time curves.

Compound Symmetry In the compound symmetry model of the simple repeated measures design, it suffices to estimate the quantity $\tau = \sigma^2 - c^*$. From Theorem 2.6, it follows that

$$\widehat{\tau}_n = \frac{1}{(nt)^2 n(t-1)} \sum_{k=1}^{n} \sum_{s=1}^{t} (R_{ks} - \overline{R}_{k\cdot})^2$$

is a consistent estimator of τ and the statistic

$$Q_n^{cs} = \frac{n^2(t-1)}{\sum_{k=1}^{n} \sum_{s=1}^{t} (R_{ks} - \overline{R}_{k\cdot})^2} \sum_{s=1}^{t} \left(\overline{R}_{\cdot s} - \frac{nt+1}{2} \right)^2$$

has, asymptotically, a central χ_{t-1}^2-distribution under the hypothesis H_0^F. This statistic has been given in literature by Brunner and Neumann (1982) and by Kepner and Robinson (1988).

Multivariate Model In the general case of the multivariate model, let $\boldsymbol{R}_k = (R_{k1}, \ldots, R_{kt})'$, $k = 1, \ldots, n$, denote the vector of the ranks within subject k and let $\overline{\boldsymbol{R}}_{\cdot}$ denote the mean of the vectors $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$. Then,

$$\widehat{\boldsymbol{V}}_n = \frac{1}{(nt)^2(n-1)} \sum_{k=1}^{n} \left(\boldsymbol{R}_k - \overline{\boldsymbol{R}}_{\cdot}\right) \left(\boldsymbol{R}_k - \overline{\boldsymbol{R}}_{\cdot}\right)' \tag{2.18}$$

is a consistent estimator of $\boldsymbol{V}_n$ (c.f. Theorem 2.5) and the WTS

$$Q_n(\boldsymbol{P}_t) = n\widehat{\boldsymbol{p}}'\boldsymbol{P}_t[\boldsymbol{P}_t\widehat{\boldsymbol{V}}_n\boldsymbol{P}_t]^-\boldsymbol{P}_t\widehat{\boldsymbol{p}} = n\widehat{\boldsymbol{p}}'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{p}} \tag{2.19}$$

has, asymptotically, a central χ^2_{t-1}-distribution under the hypothesis H_0^F. Here,

$$\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{V}}_n^{-1}\left(\boldsymbol{I}_t - \boldsymbol{J}_t\widehat{\boldsymbol{V}}_n^{-1}/\boldsymbol{1}_t'\widehat{\boldsymbol{V}}_n^{-1}\boldsymbol{1}_t\right)$$

is a g-inverse of $\boldsymbol{P}_t\widehat{\boldsymbol{V}}_n\boldsymbol{P}_t$. Note also that $\boldsymbol{P}_t\widehat{\boldsymbol{W}}\boldsymbol{P}_t = \widehat{\boldsymbol{W}}$. This statistic was considered in the literature by Thompson (1991) and by Akritas and Arnold (1994).

For small sample sizes, the statistic $(n-t+1) \cdot Q_n(\boldsymbol{P}_t)/[(t-1)(n-1)]$ is compared with the central F-distribution with $f_1 = t-1$ and $f_2 = n-t+1$ degrees of freedom. This small sample approximation is motivated by the distribution of Hotelling's T^2-statistic under the assumption of multivariate normality, where the hypothesis $\mu_1 = \cdots = \mu_t$ is tested. A comprehensive simulation study shows (see Brunner and Langer, 1999) that this approximation is rather accurate, also for small numbers of subjects. The power of this statistic, however, compared with the power of the ATS (to be considered below) is rather poor and thus, the ATS should be preferred in case of small samples.

Next, the ATS given in (2.15) is considered in the multivariate model.

To derive the ANOVA-type statistic for the simple repeated measures design, note that $\boldsymbol{M} = \boldsymbol{P}_t'(\boldsymbol{P}_t\boldsymbol{P}_t')^-\boldsymbol{P}_t = \boldsymbol{P}_t$ since $\boldsymbol{P}_t$ is a projection matrix. Then, the ATS is derived from (2.15) and is given by

$$\begin{aligned} F_n(\boldsymbol{P}_t) &= \frac{n}{tr(\boldsymbol{P}_t\widehat{\boldsymbol{V}}_n)}\widehat{\boldsymbol{p}}'\boldsymbol{P}_t\widehat{\boldsymbol{p}} \\ &= \frac{n}{N^2 tr(\boldsymbol{P}_t\widehat{\boldsymbol{V}}_n)} \sum_{s=1}^{t}\left(\overline{R}_{\cdot s} - \frac{N+1}{2}\right)^2, \end{aligned} \tag{2.20}$$

where $\widehat{\boldsymbol{V}}_n$ is given in (2.18). Under $H_0^F : \boldsymbol{P}_t\boldsymbol{F} = \boldsymbol{0}$, the ATS $F_n(\boldsymbol{P}_t)$ is approximated by the central $F(\widehat{f}, \infty)$-distribution with

$$\widehat{f} = \left[tr\left(\boldsymbol{P}_t\widehat{\boldsymbol{V}}_n\right)\right]^2 / tr\left(\boldsymbol{P}_t\widehat{\boldsymbol{V}}_n\boldsymbol{P}_t\widehat{\boldsymbol{V}}_n\right)$$

degrees of freedom.

A linear rank statistic which is sensitive to the special patterned alternative descibed by the pattern $\boldsymbol{w} = (w_1, \ldots, w_t)'$, is derived from (2.17) and, in the simple repeated measures design, it reduces to

$$\begin{aligned} T_n(\boldsymbol{w}) &= L_n(\boldsymbol{w})/\widehat{\sigma}_w = \frac{\sqrt{n}}{\widehat{\sigma}_w}\boldsymbol{w}'\boldsymbol{P}_t\widehat{\boldsymbol{p}} \\ &= \frac{\sqrt{n}}{N\widehat{\sigma}_w}\sum_{k=1}^{n}\sum_{s=1}^{t} w_s\left(R_{ks} - \frac{N+1}{2}\right) \end{aligned}$$

and the variance estimator $\widehat{\sigma}_w$ is given by

$$\widehat{\sigma}_w^2 = \boldsymbol{w}'\boldsymbol{P}_t\widehat{\boldsymbol{V}}_n\boldsymbol{P}_t\boldsymbol{w} = \frac{1}{N^2(n-1)}\sum_{k=1}^{n}(U_k - \overline{U}.)^2,$$

where $U_k = \sum_{s=1}^{t}(w_s - \overline{w}.)R_{ks}$ and $\overline{w}. = t^{-1}\sum_{s=1}^{t} w_s$ denotes the mean of the weights $w_1, \ldots, w_s$.

2.5.3 Split-Plot Design

The so-called split-plot design is one of the most frequently used designs with repeated measures. If a different groups of subjects are repeatedly observed under t different treatments (generally t time points), then this design is appropriate to model the data of the experiment. Let

$$\boldsymbol{X}_{ik} = (X_{ik1}, \ldots, X_{ikt})', i = 1, \ldots, a, \ k = 1, \ldots, n_i,$$

denote the observation vector of subject k within group i. These vectors are assumed to be independent and the components X_{iks}, $s = 1, \ldots, t$, are assumed to have the distribution functions $F_{is}(x)$, $k = 1, \ldots, n_i$.

The hypothesis of no group effect (factor A) means that the averages $\overline{F}_{i\cdot} = t^{-1}\sum_{s=1}^{t} F_{is}$ of the distributions F_{is} over the t treatments are the same for all groups $i = 1, \ldots, a$. In matrix notation, this hypothesis is written as $H_0^F(A) : (\boldsymbol{P}_a \otimes \frac{1}{t}\boldsymbol{1}_t')\boldsymbol{F} = \boldsymbol{0}$.

Let $g_i = t^{-1}\sum_{s=1}^{t} p_{is}$ denote the mean of the relative marginal effects p_{is} which is estimated by

$$\widehat{g}_i = \frac{1}{t}\sum_{s=1}^{t}\widehat{p}_{is} = \frac{1}{t}\sum_{s=1}^{t}\frac{1}{N}\left(\overline{R}_{i\cdot s} - \frac{1}{2}\right) = \frac{1}{N}\left(\overline{R}_{i\cdot\cdot} - \frac{1}{2}\right).$$

According to the hypothesis $H_0^F(A) : (\boldsymbol{P}_a \otimes \frac{1}{t}\boldsymbol{1}_t')\boldsymbol{F} = \boldsymbol{0}$, the contrast vector is given by

$$\sqrt{n}\left(\boldsymbol{P}_a \otimes \frac{1}{t}\boldsymbol{1}_t'\right)\widehat{\boldsymbol{p}} = \sqrt{n}\,\boldsymbol{P}_a\left(\boldsymbol{I}_a \otimes \frac{1}{t}\boldsymbol{1}_t'\right)\widehat{\boldsymbol{p}} = \sqrt{n}\,\boldsymbol{P}_a\widehat{\boldsymbol{g}}$$

and its asymptotic distribution is derived from Theorem 2.4. It follows that $\sqrt{n}\ \boldsymbol{P}_a\,\widehat{\boldsymbol{g}}$ has, asymptotically, a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{P}_a \boldsymbol{\Sigma}_n \boldsymbol{P}_a$, where $\boldsymbol{\Sigma}_n$ is estimated consistently by

$$\widehat{\boldsymbol{\Sigma}}_n = \bigoplus_{i=1}^{a} \frac{n}{n_i}\widehat{\sigma}_i^2, \quad \widehat{\sigma}_i^2 = \frac{1}{N^2(n_i-1)}\sum_{k=1}^{n_i}(\overline{R}_{ik\cdot} - \overline{R}_{i\cdot\cdot})^2. \tag{2.21}$$

Here, $\overline{R}_{ik\cdot} = t^{-1}\sum_{s=1}^{t} R_{iks}$ denotes the mean of the ranks R_{iks} over the t treatments for subject k within group i and $\overline{R}_{i\cdot\cdot} = n_i^{-1}\sum_{k=1}^{n_i}\overline{R}_{ik\cdot}$ denotes the mean of all tn_i ranks within group i. Then the WTS follows from (2.13) and is given by

$$Q_n(A) = \sum_{i=1}^{a}\frac{n_i}{\widehat{\sigma}_i^2}\left(\overline{R}_{i\cdot\cdot} - \frac{1}{\sum_{\ell=1}^{a}(n_\ell/\widehat{\sigma}_\ell^2)}\sum_{\ell=1}^{a}\frac{n_\ell}{\widehat{\sigma}_\ell^2}\overline{R}_{\ell\cdot\cdot}\right)^2, \tag{2.22}$$

which has, asymptotically, a central χ^2_{a-1}-distribution under $H_0^F(A)$.

The ATS and the approximation by the central $F(\widehat{f},\infty)$-distribution is easily derived from (2.15) and reduces to

$$F_n(A) = \frac{n}{tr(\boldsymbol{P}_a\widehat{\boldsymbol{\Sigma}}_n)}\,\widehat{\boldsymbol{g}}'\,\boldsymbol{P}_a\,\widehat{\boldsymbol{g}} = \frac{a}{(a-1)\sum_{i=1}^{a}\widehat{\sigma}_i^2/n_i}\sum_{i=1}^{a}(\overline{R}_{i\cdot\cdot} - \widetilde{R}_{\cdot\cdot\cdot})^2, \tag{2.23}$$

and under $H_0^F(A)$, the distribution of $F_n(A)$ is approximated by the central $F(\widehat{f}_A, \widehat{f}_0)$-distribution with

$$\widehat{f}_A = \frac{(a-1)^2}{1 + a(a-2)\left[\sum_{i=1}^{a}(\widehat{\sigma}_i^2/n_i)^2\Big/\left(\sum_{i=1}^{a}\widehat{\sigma}_i^2/n_i\right)^2\right]} \tag{2.24}$$

and

$$\widehat{f}_0 = \frac{\left(\sum_{i=1}^{a}\widehat{\sigma}_i^2/n_i\right)^2}{\sum_{i=1}^{a}(\widehat{\sigma}_i^2/n_i)^2/(n_i-1)} \tag{2.25}$$

degrees of freedom. Here, $\widetilde{R}_{\cdot\cdot\cdot} = a^{-1}\sum_{i=1}^{a}\overline{R}_{i\cdot\cdot}$ denotes the unweighted mean of the rank means $\overline{R}_{i\cdot\cdot}$, $i = 1,\dots,a$.

The hypothesis of no treatment effect (time effect) (factor T) means that the averages $\overline{F}_{\cdot s}$ of the distributions F_{is} over all $i = 1,\dots,a$ groups at time point s are the same, i.e. $H_0^F(T) : \overline{F}_{\cdot 1} = \cdots = \overline{F}_{\cdot t}$. In matrix notation, this hypothesis is equivalently written as $H_0^F(T) : \left(\frac{1}{a}\boldsymbol{1}_a' \otimes \boldsymbol{P}_t\right)\boldsymbol{F} = \boldsymbol{0}$. Let $\boldsymbol{R}_{ik} = (R_{ik1},\dots,R_{ikt})'$, $i = 1,\dots,a$, $k = 1,\dots,n_i$, denote the vector of the ranks R_{iks} for the subject k within group i and let

$$\overline{\boldsymbol{R}}_{i\cdot} = \frac{1}{n_i}\sum_{k=1}^{n_i}\boldsymbol{R}_{ik} = \frac{1}{n_i}\sum_{k=1}^{n_i}(R_{ik1},\dots,R_{ikt})'$$

denote the mean of $\boldsymbol{R}_{i1}, \ldots, \boldsymbol{R}_{in_i}$ within group i and let finally

$$\widetilde{\boldsymbol{R}}_{\cdot\cdot} = \frac{1}{a}\sum_{i=1}^{a} \overline{\boldsymbol{R}}_{i\cdot} = \frac{1}{a}\sum_{i=1}^{a} (\overline{R}_{i\cdot 1}, \ldots, \overline{R}_{i\cdot t})'$$

denote the unweighted mean of $\overline{\boldsymbol{R}}_{1\cdot}, \ldots, \overline{\boldsymbol{R}}_{a\cdot}$ over all a groups. Then, consistent estimators for the covariance matrices $\boldsymbol{V}_i$ and $\boldsymbol{V}_n$ follow from (2.7) and (2.8) and are given by

$$\widehat{\boldsymbol{V}}_i = \frac{n}{N^2 n_i(n_i-1)} \sum_{k=1}^{n_i} (\boldsymbol{R}_{ik} - \overline{\boldsymbol{R}}_{i\cdot})(\boldsymbol{R}_{ik} - \overline{\boldsymbol{R}}_{i\cdot})', \quad \widehat{\boldsymbol{V}}_n = \bigoplus_{i=1}^{a} \widehat{\boldsymbol{V}}_i \tag{2.26}$$

where $n = \sum_{i=1}^{a} n_i$ denotes the total number of subjects and $N = n \cdot t$ the total number of observations. Furthermore, let $\widehat{\boldsymbol{V}}_t = \left(\frac{1}{a}\boldsymbol{1}_a' \otimes \boldsymbol{I}_t\right) \widehat{\boldsymbol{V}}_n \left(\frac{1}{a}\boldsymbol{1}_a \otimes \boldsymbol{I}_t\right) = a^{-2}\sum_{i=1}^{a} \widehat{\boldsymbol{V}}_i$. Then, the WTS for testing the hypothesis $H_0^F(T)$ follows from (2.13) and is given by

$$Q_n(T) = \frac{n}{N^2} \widetilde{\boldsymbol{R}}_{\cdot\cdot}' \boldsymbol{P}_t [\boldsymbol{P}_t \widehat{\boldsymbol{V}}_t \boldsymbol{P}_t]^- \boldsymbol{P}_t \widetilde{\boldsymbol{R}}_{\cdot\cdot}\,.$$

Under $H_0^F(T)$, the statistic $Q_n(T)$ has, asymptotically, a central χ^2_{t-1}-distribution. However, large sample sizes may be necessary to achieve a satisfactory approximation by the limiting χ^2-distribution. For small sample sizes, the ATS should be used to test the hypothesis $H_0^F(T)$. Let $\widetilde{R}_{\cdot\cdot s} = a^{-1}\sum_{i=1}^{a} \overline{R}_{i\cdot s}$ denote the unweighted mean of the rank means $\overline{R}_{i\cdot s}$ and let $\widetilde{R}_{\cdot\cdot\cdot} = t^{-1}\sum_{s=1}^{t} \widetilde{R}_{\cdot\cdot s}$. Then the ATS follows from (2.15) and is given by

$$F_n(T) = \frac{n}{N^2 tr(\boldsymbol{P}_t \widehat{\boldsymbol{V}}_t)} \widetilde{\boldsymbol{R}}_{\cdot\cdot}' \boldsymbol{P}_t \widetilde{\boldsymbol{R}}_{\cdot\cdot} = \frac{n}{N^2 tr(\boldsymbol{P}_t \widehat{\boldsymbol{V}}_t)} \sum_{s=1}^{t} (\widetilde{R}_{\cdot\cdot s} - \widetilde{R}_{\cdot\cdot\cdot})^2.$$

Under $H_0^F(T)$, the distribution of $F_n(T)$ is approximated by the central $F(\widehat{f}_T, \infty)$-distribution with $\widehat{f}_T = [tr(\boldsymbol{P}_t \widehat{\boldsymbol{V}}_t)]^2 / tr(\boldsymbol{P}_t \widehat{\boldsymbol{V}}_t \boldsymbol{P}_t \widehat{\boldsymbol{V}}_t)$ degrees of freedom.

In most cases when a split-plot design with repeated measures is conducted, it is mainly of interest to investigate an interaction between groups (factor A) and time (factor T). For example, if a placebo is applied in group 1 and the active treatment is given to group 2, then the distribution functions at the start of the trail (time point 1) F_{11} and F_{21} are identical if the subjects are randomly assigned to the two treatment groups of factor A. Then, an effect of the active treatment will produce non parallel time curves of the measurements. This means that there is an interaction between factor A and factor T. In a nonparametric setup, the hypothesis of no interaction is formulated as $H_0^F(AT) : F_{is} = \overline{F}_{i\cdot} + \overline{F}_{\cdot s} - \overline{F}_{\cdot\cdot}, i = 1, \ldots, a,\ s = 1, \ldots, t$. In matrix notation, this hypothesis is equivalently written as

$$H_0^F(AT) : \boldsymbol{C}_{AT}\boldsymbol{F} = \begin{pmatrix} F_{11} - \overline{F}_{1\cdot} - \overline{F}_{\cdot 1} + \overline{F}_{\cdot\cdot} \\ \vdots \\ F_{at} - \overline{F}_{a\cdot} - \overline{F}_{\cdot t} + \overline{F}_{\cdot\cdot} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \boldsymbol{0},$$

where $\boldsymbol{C}_{AT} = \boldsymbol{P}_a \otimes \boldsymbol{P}_t$ and $\boldsymbol{P}_a$ and $\boldsymbol{P}_t$ denote the a- and t-dimensional centering matrices respectively (c.f. (1.1)). Let $\overline{\boldsymbol{R}}. = (\overline{R}_{1\cdot 1}, \ldots, \overline{R}_{a\cdot t})'$ denote the vector of all at rank means. Then, the WTS

$$Q_n(AT) = \frac{n}{N^2} \overline{\boldsymbol{R}}.' \boldsymbol{C}'_{AT} [\boldsymbol{C}_{AT} \widehat{\boldsymbol{V}}_n \boldsymbol{C}'_{AT}]^- \boldsymbol{C}_{AT} \overline{\boldsymbol{R}}.$$

follows from (2.13) where $\widehat{\boldsymbol{V}}_n$ is given in (2.26). Under $H_0^F(AT)$, the statistic $Q_n(AT)$ has, asymptotically, a central $\chi^2_{(a-1)(t-1)}$-distribution for which very large sample sizes are needed for a satisfactory approximation.

The ATS is derived from (2.15) by letting $\boldsymbol{M}_{AT} = \boldsymbol{C}'_{AT}(\boldsymbol{C}_{AT}\boldsymbol{C}'_{AT})^- \boldsymbol{C}_{AT} = \boldsymbol{C}_{AT} = \boldsymbol{P}_a \otimes \boldsymbol{P}_t$ since $\boldsymbol{P}_a$ and $\boldsymbol{P}_t$ are both projection matrices. Let $\widetilde{R}_{\cdot\cdot s} = a^{-1}\sum_{i=1}^a \overline{R}_{i\cdot s}$ and $\widetilde{R}_{\cdots} = t^{-1}\sum_{s=1}^t \widetilde{R}_{\cdot\cdot s}$ denote the unweighted means over the $a \cdot t$ rank means $\overline{R}_{i\cdot s}$, $i = 1, \ldots, a$, $s = 1, \ldots, t$. Then, under $H_0^F(AT)$, the distribution of the ATS

$$F_n(AT) = \frac{n}{N^2 tr(\boldsymbol{M}_{AT}\widehat{\boldsymbol{V}}_n)} \sum_{i=1}^a \sum_{s=1}^t (\overline{R}_{i\cdot s} - \overline{R}_{i\cdot\cdot} - \widetilde{R}_{\cdot\cdot s} + \widetilde{R}_{\cdots})^2$$

can be approximated by the central $F(\widehat{f}_{AT}, \infty)$-distribution where $\widehat{f}_{AT}$ is given by

$$\widehat{f}_{AT} = \frac{[\mathrm{tr}(\boldsymbol{M}_{AT}\widehat{\boldsymbol{V}}_n)]^2}{\mathrm{tr}(\boldsymbol{M}_{AT}\widehat{\boldsymbol{V}}_n\boldsymbol{M}_{AT}\widehat{\boldsymbol{V}}_n)}.$$

2.6 Example and Software

2.6.1 Split-Plot Design with Ordered Categorical Data

The Roof-Experiment A rain cleaning experiment was performed in a low-mountain range (Solling) in Lower-Saxony (Germany). In two different areas of the forest, 300 m^2 each, the ground was covered by a roof at a height of $3m$ in order to protect the soil of the forest from the contaminated rain water. The precipitation was collected in tanks, and after preparation, it was re-splashed under the roofs. For the *control-roof* (D_2) which contained 23 trees, the water was not chemically purified and the same water which had been collected was re-splashed under the roof. Only the dust and larger sedimentary particles were filtered out so that the nozzles could not be blocked. For the *clean-rain-roof* (D_1) with 27 trees, the precipitation was re-splashed after a chemical purification. A third area of 300 m^2 without any roof (D_0) served as a control to examine the effect of the filtration for the control-roof. This area contained 22 trees.

The vitality of the trees was measured on a grading scale ranging from 1 (excellent) to 10 (dead). It was judged at the region of the treetops by means of a crane which was fixed in the center of the three areas. The experiment started in 1993 and longitudinal observations on the trees were taken in the years 1994, 1995, and 1996. The data are displayed in Table 2.1.

Several detailed questions had to be answered: (1) Is there any trend (increasing or decreasing) in the time curves of the vitality scores for the different areas D_0, D_1 and D_2? (2) Is there any effect of the clean-rain-roof? (3) Do the time curves of the three areas have the same shapes?

TABLE 2.1 *Vitality scores for the trees in the roof-experiment for the areas D_0, D_1 and D_2.*

Area D_0					Area D_2					Area D_1				
Tree	Year				Tree	Year				Tree	Year			
no.	93	94	95	96	no.	93	94	95	96	no.	93	94	95	96
569	2	2	2	2	547	8	4	4	5	646	2	3	2	1
570	1	1	1	1	549	1	1	1	1	647	6	4	4	5
589	3	1	2	2	551	4	4	4	3	648	3	2	2	2
590	2	1	1	3	561	4	3	3	3	649	1	1	1	1
592	5	4	3	4	562	2	1	1	2	650	4	5	4	2
593	1	1	1	2	564	5	3	3	3	651	6	5	5	3
601	4	3	3	4	566	3	4	4	3	652	8	7	6	5
602	4	4	4	4	567	4	3	3	2	682	3	2	2	2
611	1	1	2	3	596	5	4	4	4	683	3	2	2	2
613	3	2	2	2	597	2	1	2	2	684	5	4	4	5
618	4	2	3	3	599	5	2	2	3	685	2	2	2	3
619	6	5	4	4	614	7	5	5	5	686	3	3	1	2
620	2	1	2	2	615	6	4	5	6	687	5	4	3	2
636	3	3	4	2	616	6	6	3	3	693	6	4	4	4
638	3	2	1	3	617	4	3	5	3	694	8	7	8	7
639	1	1	2	1	626	5	4	3	3	695	5	3	2	3
653	6	7	6	5	627	1	2	2	2	696	4	1	1	2
655	1	1	1	1	628	2	1	1	1	697	3	2	3	2
656	6	3	3	3	629	6	4	4	5	698	4	4	4	4
657	1	1	1	2	630	3	2	2	1	723	4	4	4	3
659	8	5	6	4	631	4	3	3	2	724	6	4	4	4
681	1	2	1	1	632	2	1	1	1	725	5	4	3	2
					633	3	4	3	3	726	3	3	1	1
										733	4	4	5	4
										735	4	4	4	2
										736	3	3	2	1
										737	6	5	5	4

First we note that the observations are ordered categorical data. Thus, all results must be invariant under the choice of the grading scores $1, 2, \ldots, 10$, i.e. the results must be invariant under strictly monotone transformations of the data. This is a well known property of rank statistics and thus, the procedures considered in this section are especially appropriate for the analysis of ordered categorical data with repeated measures. The rank means $\overline{R}_{i \cdot s}$ and the estimated relative marginal effects $\widehat{p}_{is}$, $i = 1, 2, 3$; $j = 1, \ldots, 4$, for the three areas within the four years are displayed in Table 2.2.

TABLE 2.2 *Rank means and estimated relative marginal effects of the vitality scores for the trees within the three areas during the years 1993, 94, 95 and 96 in the clean rain experiment.*

	Rank Means $\overline{R}_{i \cdot s}$				Relative Marginal Effects			
	Years				Years			
Area	93	94	95	96	93	94	95	96
D0	135.2	103.7	110.1	122.0	0.47	0.36	0.38	0.42
D1	198.0	168.2	150.4	132.2	0.69	0.58	0.52	0.46
D2	183.2	141.7	140.0	132.3	0.63	0.49	0.48	0.46

The results of the analysis by the WTS $Q_n(C)$ and the ATS $F_n(M)$ given in subsection 2.5.3 along with the resulting p-values are displayed in Table 2.3.

TABLE 2.3 *Statistics and p-values for the main effects and the interaction in the clean-rain experiment.*

Factor	Wald-Type Statistics $Q_n(C)$	d.f.	p-Value	ANOVA-Type Statistics $F_n(M)$	$\widehat{f}_1$	f_0	p-Value
Area	4.51	2	0.1049	2.35	1.97	∞	0.0960
Year	58.06	3	$< 10^{-5}$	21.39	2.73	∞	$< 10^{-5}$
Interaction	14.82	6	0.0217	3.11	5.35	∞	0.0068

The difficulty with this example is that the trees could not be randomized to the three treatments and, at the beginning of the trial, the relative marginal effects for both the experimental areas D1 and D2 seem to be somewhat larger than for the area D0 without a roof (D0: $\widehat{p}_{11} = 0.47$, D1: $\widehat{p}_{21} = 0.69$, D2: $\widehat{p}_{31} = 0.63$). The p-value obtained by the Kruskal-Wallis test for the vitality scores on the first time point is $p = 0.066$ which indicates that the vitality scores may not have the same distribution for all three areas. Thus, the result of the test for the main effect of the area is difficult to interpret and the question of a potential treatment effect can only be answered by the analysis of the interaction between the areas and the years. The p-value of $p = 0.0068$ for the interaction is significant on the 1%-level and the interpretation is that the time curves of the relative marginal effects within the three areas are not parallel (see Figure 2.1). For a more detailed analysis including pairwise comparisons of the three areas and tests for decreasing trends, we refer to Brunner and Langer (1999), Section 8.3.7.

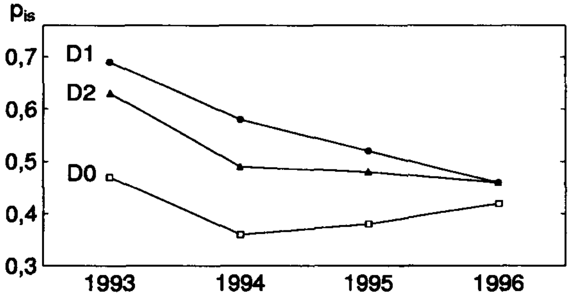

FIGURE 2.1 *Time curves of the relative marginal effects for the vitality scores in the three areas D0 (area without roof), D1 (clean-rain-roof) and D2 (control-roof) during the years 1993 - 1996.*

2.6.2 Software

Similar to the discussion of the software in Subsection 1.7.2 for independent observations, the statistics $Q_n(\boldsymbol{C})$, $F_n(\boldsymbol{M})$ and $T_n(\boldsymbol{w})$ have the rank transform property under H_0^F. Therefore, only mid-ranks have to be assigned to the data and the analysis can be performed on these ranks if the special heteroscedastic model of the ART under H_0^F is used (see Subsection 2.4.3). Then, the computation of the statistics $Q_n(\boldsymbol{C})$, $F_n(\boldsymbol{M})$ and $T_n(\boldsymbol{w})$ can be performed by any statistical software package which provides

1. the mid-ranks of the observations,
2. the analysis of heteroscedastic factorial designs with repeated measures and unspecified covariance matrices.

The Statistical Analysis System (SAS) provides these computations by the procedure 'MIXED'. Below, we provide the necessary statements for the DATA-step and the procedures SORT, RANK and MIXED.

Data Input The input of the data is handled in the same way as for the data of a parametric model, i.e. factors are treated as 'classifying variables'. Note, however, that PROC MIXED needs the data first sorted by the subjects and then by the repeated measures.

Ranking The procedure PROC RANK is used to assign the mid-ranks among all observations to the data. Note that the assignment of mid-ranks is the default with this procedure in SAS.

Heteroscedastic Model The procedure PROC MIXED provides the possibility to define the structure of the covariance matrix of the 'cell means' within the levels of the repeated measures factor by the option TYPE=$\cdots$ within the REPEATED statement. Moreover, the GRP=$\cdots$ option within the REPEATED statement defines the factor levels (or combinations of them) of the whole-plot factor(s) where different covariance matrices are allowed. The WTS $Q_N(\boldsymbol{C})$ and the resulting p-values are printed out by adding the option CHISQ after the slash (/) in the MODEL statement.

In general, for the nonparametric main effects and all interactions of the whole-plot factor(s), the covariance matrices may be different for all factor level combinations. Thus, the highest interaction term of the whole-plot factors must be assigned in the GRP option. For example, in a repeated measures design with two whole-plot factors A and B and with one sub-plot factor C (repeated measures), this option is GRP=A*B.

Starting with version 8.0 the option ANOVAF can be added somewhere in the line of the PROC MIXED statement in order to compute the ATS $F_N(\boldsymbol{M})$ and the resulting p-values. The use of the ATS is recommended for small and medium numbers of replications. To avoid computational difficulties by using the method REML (default) to estimate the unstructured covariance matrix (see SAS online documentation for the procedure MIXED) it is recommended to use the minimum variance quadratic unbiased estimation method by adding the option METHOD=MIVQUE0 after the option ANOVAF in the line of the PROC MIXED statement.

Example: Clean-Rain-Experiment

TABLE 2.4 *ANOVA-type statistics and p-values for the main effects and the interaction in the clean-rain experiment computed by the SAS-procedures 'RANK' and 'MIXED' where the statements are provided below.*

Factor	ANOVA-Type Statistics				
	$F_n(M)$	$\widehat{f}_1$	$\widehat{f}_0$	p-Value (PROC MIXED)	p-Value (corrected)
Area	2.35	1.97	64.4	0.1041	–
Year	21.39	2.73	179	$< 10^{-5}$	$< 10^{-5}$
Interaction	3.11	5.35	179	0.0086	0.0068

```
DATA roof;
INPUT area$ tree s1-s4;
  ARRAY ss{4} s1-s4;
  DO year=1 to 4;
    score=ss{year};
    OUTPUT;
  END;
  DROP s1-s4;
DATALINES;
D0 569 2 2 2 2
:
D2 737 6 5 5 4
;
RUN;
```

```
PROC RANK DATA=roof OUT=roof;
VAR score;
RANKS r;
RUN;

PROC SORT DATA=roof OUT=roof;
BY tree year;
RUN;

PROC MIXED DATA=roof ANOVAF METHOD=MIVQUE0;
CLASS tree area year;
MODEL r = area | year / CHISQ;
REPEATED year / TYPE=UN GRP=year SUB=tree;
RUN;
```

The results for the Wald-type statistics produced by the SAS-procedure 'MIXED' are identical to those displayed in Table 2.3 with the exception that the second degree of freedom $\widehat{f}_0$ is taken (automatically) from the option DDFM=KR. To obtain a better approximation for the p-values of the tests for 'Year' and 'Interaction', a separate DATA step must be added to compute the p-values $p = 1 - F(21.39|2.73, \infty) < 10^{-5}$ and $p = 1 - F(3.11|5.35, \infty) = 0.0068$. The results are displayed in Table 2.4.

3 Further Developments

3.1 Adjustment for Covariates

In applications, quite often the variable of interest $X_{ij}^{(0)} \sim F_i^{(0)}(x)$ depends on one or more covariates $X_{ij}^{(1)} \sim F_i^{(1)}, \ldots, X_{ij}^{(m)} \sim F_i^{(m)}$, $i = 1, \ldots, d$. Then, the motivation of the adjustment for covariates is twofold:

1. If the distributions $F_1^{(\ell)}, \ldots, F_d^{(\ell)}$, $\ell = 1, \ldots, m$, are assumed to be all equal to $F_0^{(\ell)}$, say, then the adjustment is intended to reduce the variances of the estimators $\widehat{p}_i^{(0)}$, $i = 1, \ldots, d$. The assumption of the equality of the distribution functions of the covariates over the treatment groups may be justified in designs where the experimental units are randomly assigned to the treatments (or to the treatment combinations). Thus, by randomization, this assumption is reasonable.

2. In case of 'unhappy randomization' and in observational studies, however, the relative treatment effects $p_i^{(0)} = \int H dF_i^{(0)}$ may depend on the relative effects $p_i^{(\ell)} = \int H^{(\ell)} dF_i^{(\ell)}$, $\ell = 1, \ldots, m$; $i = 1, \ldots, d$, of the covariates within the treatment groups. In this case, adjustment for covariates is not only intended for a possible reduction of the variances but also for the correction of a potential bias caused by the dependence on the covariates.

The first case was considered in literature by a more or less heuristically motivated procedure (Quade, 1967) and by a procedure which is based on the asymptotic multivariate normality of a vector of linear rank statistics in the one-way layout for shift models with continuous distribution functions (Puri and Sen, 1969). This procedure was recently generalized by Langer (1998) to factorial designs and possibly discontinuous distribution functions where the concept of formulating hypotheses by the distribution functions was used. The second case causes more problems and is still under research. First encouraging results have been derived by Siemer (1999) but they have to be developed further.

3.2 Unweighted Treatment Effects

As briefly indicated in subsection 1.3.1, the relative treatment effects $p_i = \int H dF_i$ depend on the sample sizes $n_1, \ldots, n_d$, through the weighted mean distribution functions $H(x)$. In case of unequal sample sizes n_i, $i = 1, \ldots, d$, the unweighted relative effects $\pi_i = \int H^* dF_i$ as given in (1.4) may be used to formulate the hypotheses and to derive meaningful confidence intervals for relative treatment effects. Some recent results (Kulle, 1999) need to be improved regarding the approximation of the null distribution of the statistics in case of small sample sizes.

3.3 Multivariate Designs

The methods presented in this paper can easily be developed for multivariate observations (Munzel, 1996; Munzel and Brunner, 2000). The application of these results to other multivariate problems like principal components analysis or studies with multiple endpoints have yet to be developed.

REFERENCES

AKRITAS, M. G. (1990). The rank transform method in some two-factor designs. *J. Amer. Statist. Assoc.* **85**, 73–78.

AKRITAS, M. G. (1991). Limitations on the rank transform procedure: A study of repeated measures designs, Part I. *J. Amer. Statist Assoc.* **86**, 457–460.

AKRITAS, M. G., AND ARNOLD, S. F. (1994). Fully nonparametric hypotheses for factorial designs I: Multivariate repeated measures designs. *J. Amer. Statist. Assoc.* **89**, 336–343.

AKRITAS, M. G., ARNOLD, S. F. AND BRUNNER, E. (1997). Nonparametric hypotheses and rank statistics for unbalanced factorial designs. *J. Amer. Statist. Assoc.* **92**, 258–265.

AKRITAS, M. G. AND BRUNNER, E. (1996). Rank Tests for Patterned Alternatives in Factorial Designs with Interactions. *Festschrift on the Occasion of the* 65^{th} *birthday of Madan L. Puri*, VSP-International Science Publishers, Utrecht, The Netherlands, 277–288.

AKRITAS, M. G. AND BRUNNER, E. (1997). A unified approach to ranks tests in mixed models. *J. Statist. Plann. Inference* **61**, 249–277.

BLAIR, R. C., SAWILOWSKI, S. S. AND HIGGENS, J. J. (1987). Limitations of the rank transform statistic in tests for interactions. *Comm. in Statist., Part B — Simulation and Computation* **16**, 1133–1145.

BOOS, D. D. AND BROWNIE, C. (1992). Rank based mixed model approach to multisite clinical trials. *Biometrics* **48**, 61–72.

BOX, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *Ann. Math. Statist.* **25**, 290–302.

BRUNNER, E. AND DENKER, M. (1994). Rank statistics under dependent observations and applications to factorial designs. *J. Statist. Plann. Inference* **42**, 353–378.

BRUNNER, E., DETTE, H. AND MUNK, A. (1997). Box–type approximations in nonparametric factorial designs. *J. Amer. Statist. Assoc.* **92**, 1494–1502.

BRUNNER, E. AND LANGER, F. (1999). *Nichtparametrische Analyse longitudinaler Daten*, Oldenbourg, München.

BRUNNER, E., MUNZEL, U. AND PURI, M. L. (1999). Rank-Score Tests in Factorial Designs with Repeated Measures. *J. Mult. Analysis* **70**, 286–317.

BRUNNER, E. AND NEUMANN, N. (1982). Rank Tests for Correlated Random Variables. *Biometr. J.* **24**, 373–389.

BRUNNER, E. AND NEUMANN, N. (1986*a*). Rank tests in 2×2 designs. *Statist. Neerlandica* **40**, 251–271.

BRUNNER, E. AND NEUMANN, N. (1986*b*). Two-sample rank tests in general models. *Biometr. J.* **28**, 395–402.

BRUNNER, E. AND PURI, M. L. (1996). Nonparametric methods in design and analysis of experiments. *Handbook of Statistics* **13**, *(S. Ghosh and C.R. Rao, Eds.)*, 631-703.

BRUNNER, E. AND PURI, M. L. (2000). A class of rank-score tests in factorial designs. *J. Statist. Plann. Inference* (to appear).

BRUNNER, E., PURI, M. L. AND SUN, S. (1995). Nonparametric methods for stratified two-sample designs with application to multi clinic trials. *J. Amer. Statist. Assoc.* **90**, 1004–1014.

CONOVER, W. J. AND IMAN, R. L. (1976). On some alternative procedures using ranks for the analysis of experimental designs. *Commun. Statist.* **A5, 14**, 1349–1368.

CONOVER, W. J. AND IMAN, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics (with discussion). *Amer. Statist.* **35** 124–133.

GOVINDARAJULU, Z. (1975). Robustness of Mann-Whitney-Wilcoxon test to dependence in the variables. *Studia Scientiarum Mathematicarum Hungarica* **10**, 39–45.

GRAYBILL, F. A. (1976). *Theory and Applications of the Linear Model.* Duxbury Press, North Scituate, Mass.

GROGGLE, D. J. AND SKILLINGS, J. H. (1986). Distribution-free tests for main effects in multifactor designs. *Amer. Statist.* **40**, 99–102.

HETTMANSPERGER, T. P. AND NORTON, R. M. (1987). Tests for patterned alternatives in *k*–sample problems. *J. Amer. Statist. Assoc.* **82**, 292–299.

HORA, S. C. AND IMAN, R. L. (1988). Asymptotic relative efficiencies of the rank-transformation procedure in randomized complete block designs. *J. Amer. Statist. Assoc.* **83**, 462–470.

HOLLANDER, M., PLEDGER, G. AND LIN, P. E. (1974). Robustness of the Wilcoxon test to a certain dependency between samples. *Ann. Statist.* **2**, 177–181.

KEPNER, J. L. AND ROBINSON, D. H. (1988). Nonparametric Methods for Detecting Treatment Effects in Repeated Measures Designs. *J. Amer. Statist. Assoc.*, **83**, 456–461.

KOCH, G. G. (1969). Some aspects of the statistical analysis of 'split-plot' experiments in completely randomized layouts. *J. Amer. Statist. Assoc.* **64**, 485–506.

KOCH, G. G. (1970). The use of nonparametric methods in the statistical analysis of a complex split plot experiment. *Biometrics* **26**, 105–128.

KOCH, G. G. AND SEN, P. K. (1968). Some aspects of the statistical analysis of the 'mixed model'. *Biometrics*, **24**, 27–48.

KRUSKAL, W. H. (1952). A nonparametric test for the several sample problem. *Ann. Math. Statist.* **23**, 525–540.

KRUSKAL, W. H. AND WALLIS, W. A. (1952). The use of ranks in one-criterion variance analysis. *J. Amer. Statist. Assoc.* **47**, 583–621.

KRUSKAL, W. H. AND WALLIS, W. A. (1953). Errata in: The use of ranks in one-criterion variance analysis. *J. Amer. Statist. Assoc.* **48**, 907–911.

KULLE, B. (1999). Nichtparametrisches Behrens-Fisher-Problem im Mehr-Stichprobenfall. *Diplomarbeit, Universität Göttingen, Inst. für Mathematische Stochastik.*

LANGER, F. (1998). Berücksichtigung von Kovariablen im nichtparametrischen gemischten Modell. *Dissertation, Universität Göttingen, Inst. für Mathematische Stochastik.*

LEMMER, H. H. (1980). Some empirical results on the two-way analysis of variance by ranks. *Comm. Statist. Theory and Methods* **14**, 1427–1438.

LÉVY, P. (1925). *Calcul des Probabilitées.* Gauthiers-Villars, Éditeurs, Paris.

MACK, G. A. AND SKILLINGS, J. H. (1980). A Friedman-type rank test for main effects in a two-factor ANOVA. *J. Amer. Statist. Assoc.* **75**, 947–951.

MARDEN, J. I. AND MUYOT, M. E. (1995). Rank tests for main and interaction effects in analysis of variance. *J. Amer. Statist. Assoc.* **90**, 1388–1398.

MATHAI, A. M. AND PROVOST, S. B. (1992). *Quadratic Forms in Random Variables.* Marcel Dekker, New York.

MEHRA, K. L. AND PURI, M. L. (1967). Multi-sample analogues of some one-sample tests. *Ann. Math. Statist.* **38**, 523–549.

MUNZEL, U. (1996). Multivariate nichtparametrische Verfahren für feste Faktoren in mehrfaktoriellen Versuchsanlagen. *Dissertation, Universität Göttingen, Inst. für Mathematische Stochastik.*

MUNZEL, U. (1999*a*). Linear rank score statistics when ties are present. *Statist. Probab. Lett.* **41**, 389–395.

MUNZEL, U. (1999*b*). Nonparametric methods for paired samples. *Statistica Neerlandica* **53**, 277–286.

MUNZEL, U. AND BRUNNER, E. (2000). Nonparametric methods in multivariate factorial designs. *J. Statist. Plann. Inference* (to appear).

PAGE, E. B. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *J. Amer. Statist. Assoc.* **58**, 216–230.

PATEL, K. M. AND HOEL, D. G. (1973). A nonparametric test for interaction in factorial experiments. *J. Amer. Statist. Assoc.* **68**, 615–620.

PURI, M. L. (1964). Asymptotic efficiency of a class of c-sample tests. *Ann. Math. Statist.* **35**, 102–121.

PURI, M. L. AND SEN, P. K. (1969). Analysis of covariance based on general rank scores. *Ann. Math. Statist.* **40**, 610–618.

QUADE, D. (1967). Rank analysis of covariance. *J. Amer. Statist. Assoc.* **62**, 1187–1200.

RAVIV, A. (1978). A Non-parametric Test for Comparing Two Non-independent Distributions. *J. R. Statist. Soc.* B **40**, 253–261.

RINAMAN, W. C., JR. (1983). On distribution-free rank tests for two-way layouts. *J. Amer. Statist. Assoc.* **78**, 655–659.

RUYMGAART, F.H. (1980). A unified approach to the asymptotic distribution theory of certain midrank statistics. In: *Statistique non Parametrique Asymptotique*, 1–18, J.P. Raoult (Ed.), Lecture Notes on Mathematics, No. 821, Springer, Berlin.

SEN, P. K. (1967). On some non-parametric generalizations of Wilk's test for H_M, H_{VC} and H_{MVC}. *Ann. Inst. Statist. Math.* **19**, 541–571.

SIEMER, A. (1999). Berücksichtigung von heterogen verteilten Kovariablen in einem nichtparametrischen Modell. *Diplomarbeit, Universität Göttingen, Inst. für Mathematische Stochastik.*

THOMPSON, G. L. (1990). Asymptotic Distribution of Rank Statistics Under Dependencies with Multivariate Applications. *J. Mult. Analysis* **33**, 183–211.

THOMPSON, G. L. (1991). A unified approach to rank tests for multivariate and repeated measures designs. *J. Amer. Statist. Assoc.* **86**, 410–419.

THOMPSON, G. L. AND AMMANN, L. P. (1989). Efficacies of rank-transform statistics in two-way models with no interaction. *J. Amer. Statist. Assoc.* **84**, 325–330.

THOMPSON, G. L. AND AMMANN, L. P. (1990). Efficiencies of interblock rank statistics for repeated measures designs. *J. Amer. Statist. Assoc.* **85**, 519–528.

Edgar Brunner
Abt. Medizinische Statistik
Universität Göttingen
Humboldtallee 32
37073 Göttingen, Germany

Madan L. Puri
Department of Mathematics
Indiana University
Bloomington, IN 47405, USA