A Class of Rank-Score Tests in Factorial Designs

Abbreviated Title: Rank-Score Tests in Factorial Designs

By
EDGAR BRUNNER
Department of Medical Statistics
Humboldt Alle 32
D-37073 Göttingen
Germany

and

MADAN L. PURI[1]
Department of Mathematics
Indiana University
Rawles Hall 232
Bloomington, IN, 47405
U.S.A.

AMS 1991 *subject classifications.* Primary 62G10; secondary 62G20, 62K15.

Key words and phrases: Rank tests, score functions, rank transform, nonparametric hypotheses, factorial designs

[1] The research of Madan L. Puri was supported by the Office of Naval Research Contract N00014-91-J-1020.

Abstract

The analysis of factorial designs in a nonparametric setup has been restricted mainly to the one-way layout. Procedures for higher-way layouts are either restricted to semiparametric models or to special designs. Moreover, the continuity of the underlying distribution functions is assumed in general.

The aim of this paper is to provide a general theory for the analysis of nonparametric factorial designs with fixed factors. Rank procedures for nonparametric hypotheses based on the distribution functions are proposed and the results are derived for score functions with bounded second derivatives. Unlike in most of the literature, we do not assume the continuity of the underlying distribution functions or the equality of the sample sizes. This means that data from continuous distributions as well as discrete ordinal data are covered by our approach. The results obtained are applied to some special factorial designs. For small sample sizes, the Box-approximation is applied to compute approximate p-values for the statistics. Within this framework, also the question of the *rank transform property* of rank statistics is briefly addressed. The application of the proposed tests is demonstrated by the analysis of real data for a two-way layout.

1 Introduction

When analyzing data from a factorial design, usually a linear model is assumed and the hypotheses are formulated by the parameters of this model. If no specific distribution functions are assumed, then there are no parameters by which the treatment effects can be described or by which the hypotheses can be formulated. Thus, artificial parameters (e.g. shift parameters) may be introduced to express the hypotheses. However, the analysis of such so-called semiparametric models is restricted to the artificially introduced parameters and the results are in general not invariant under strictly monotone transformations of the data. Moreover, a special semiparametric model may not be appropriate to describe the data. Note that a shift model is not justified e.g. when the observations are discrete values within a small range like quality scales in psychology and medicine.

The most general model for data from a factorial design uses only the distribution functions of the observations to define treatment effects and to formulate the hypotheses. For the simple case of two independent samples with distribution functions F_1 and F_2 respectively, the nonparametric treatment effect $p = \int F_1 dF_2$ was considered by Mann & Whitney (1947). Under the nonparametric hypothesis $H_0^F : F_1 = F_2$, they derived the asymptotic distribution of the natural estimator $\widehat{p} = \int \widehat{F}_1 d\widehat{F}_2$ of p where $\widehat{F}_1$ and $\widehat{F}_2$ are the empirical distribution functions of the two samples. It is well known that $\widehat{p}$ can easily be computed from the ranks of the observations.

Kruskal and Wallis (1952, 1953) generalized this idea to the one-way layout for a independent samples under the assumption that the observations have continuous distribution functions $F_1, \ldots, F_a$. They considered the nonparametric relative treatment effects $p_i = \int H dF_i$ where H denotes the weighted mean of the distribution functions F_i, $i = 1, \ldots, a$. Under the nonparametric hypothesis $H_0^F : F_1 = \cdots = F_a$, they derived the asymptotic distribution of the vector $\widehat{\mathbf{p}} = (\widehat{p}_1, \ldots, \widehat{p}_a)'$ where the estimators $\widehat{p}_i = \int \widehat{H} d\widehat{F}_i$ of p_i are computed from the ranks of all observations.

The natural generalization of the nonparametric hypothesis in the one-way layout to higher-way layouts proceeded in several steps: Lemmer & Stoker (1967) considered rank tests in a two-way layout with $i = 1, \ldots, a$ levels of the factor A and $j = 1, \ldots, b$ levels of the factor B and assumed continuous distribution functions F_{ij} for the observations in the factor level combination (i, j). The hypothesis was stated as $H_0 : F_{ij} = F$ for all i and j. Test statistics for main effects and interactions were considered under this hypothesis where no definition of a main effect or an interactions was given. Rinaman (1983) and Hora & Conover (1984) considered a two-way layout 'without interaction' and stated the hypothesis of no treatment effect as $H_0 : F_{ij} = F_i$ for all $j = 1, \ldots, b$; $i = 1, \ldots, a$. The same nonparametric hypothesis was used by Thompson (1991) to test the nested factor in a two-fold nested model. In a two-way layout 'with interaction', Akritas (1991) and Thompson (1991) formulated the hypothesis of no treatment effect in the same way and they noted that the hypothesis $H_0^F : F_{ij} = F_i$ is a 'joint hypothesis', i.e. the treatment effect and the interaction are amalgamated.

In the papers quoted above, the terminology 'treatment effect' and 'interaction' is used in the sense of a linear model. Patel and Hoel (1973) measured a nonparametric interaction through the differences of the Wilcoxon–Mann–Whitney statistic in a 2×2-design.

This idea has been used by Brunner, Puri and Sun (1995) to define a treatment effect, an interaction, and a simple treatment effect (where the main effect and the interaction are amalgamated) for a two-sample design with b strata where rank statistics for testing nonparametric hypotheses related to these effects are derived. Marden and Muyot (1995) generalized the idea of Patel and Hoel (1973) to factorial designs and studied rank tests based on linear functions of Wilcoxon–Mann–Whitney statistics. In this approach, main effects have been defined in the sense of simple main effects. However, no systematic approach for higher-way layouts or a generalization to score functions has been considered in this paper. The idea to formulate nonparametric hypotheses in factorial designs by contrasts of the distribution functions (Akritas & Arnold (1994)) seems to be a breakthrough towards a purely nonparametric formulation of hypotheses in higher-way layouts and this formulation of the hypotheses is a straightforward generalization of the nonparametric hypothesis in the one-way layout. Several important points and useful consequences of this simple idea should be noted. First, the hypotheses in the linear model are implied by these nonparametric hypotheses. Second, the generalization to higher-way layouts is straightforward in the same way as in the theory of linear models. Furthermore, the formulation of the hypotheses is not restricted to continuous distribution functions and models with discrete observations are included in this setup. The main point, however, is that under these hypotheses, the asymptotic covariance matrix of a vector of linear contrasts of the rank means has a very simple form and can be estimated by the ranks of the observations. These results give a new insight into the rank transform method (Conover & Iman, 1976, 1981 and Lemmer, 1980) and it can easily be seen when the rank transform fails. Based on the idea of the nonparametric hypotheses, Akritas, Arnold & Brunner (1997) developed rank tests for nonparametric main effects and interactions in an unbalanced two-factorial design without assuming continuous distribution functions. However, only the special case of Wilcoxon scores is considered in this paper.

The aim of the present paper is to generalize these results to score functions with a bounded second derivative and to relax the assumption that the sample sizes n_i converge to some constants λ_i. Also the problem of unbounded score functions is briefly addressed. Moreover, an elementary method of proving the results is given where the case of ties is automatically included. The same method has been used by Brunner & Denker (1994) for mixed models, however under the assumption of continuous distribution functions. The results are presented in a general form such that statistics for nonparametric hypotheses in any factorial design can be derived easily from this unified approach. Many rank (score) statistics given in the literature are special cases of the statistics derived in this paper.

The paper is organized as follows. In Section 2, the models, hypotheses and statistics are defined and the general notations used throughout the paper are given. General asymptotic results are derived in Section 3 and an approximation for small samples is considered in Section 4. Applications of the general results to one- and two-way layouts are given in Section 5 where explicit statistics for the nonparametric hypotheses are provided. The rank transform is shortly discussed in Section 5.1. The interpretation of the nonparametric hypotheses and some relations to the parametric hypotheses in the linear model are given in Section 5.2. Some Lemmas needed to prove the Theorems stated in the body of the paper, are provided in the Appendix.

2 Models, Hypotheses and Notations

Notations. For a convenient formulation of hypotheses and 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 $\mathbf{M}_{r\times d}$ is called a *contrast matrix* if $\mathbf{M}_{r\times d}\mathbf{1}_d = \mathbf{0}_{r\times 1}$ where $\mathbf{1}_d = (1, \ldots, 1)'$ denotes the d-dimensional vector of 1's. In particular, we use the contrast matrices

$$\mathbf{C}_d = (\mathbf{1}_{d-1} \vdots -\mathbf{I}_{d-1}) \in I\!R^{(d-1)\times d} \quad \text{and} \quad \mathbf{P}_d = \mathbf{I}_d - \tfrac{1}{d}\mathbf{J}_d \tag{2.1}$$

where $\mathbf{I}_d$ is the d-dimensional unit matrix and $\mathbf{J}_d = \mathbf{1}_d\mathbf{1}_d'$ is the $d \times d$ matrix of 1's. $\mathbf{C}_d$ is a partitioned matrix of order $(d-1) \times d$ where $\vdots$ denotes the partition. Note that $rank(\mathbf{C}_d) = d - 1$ and thus, $\mathbf{C}_d\mathbf{C}_d'$ is nonsingular while the matrix $\mathbf{P}_d$ is a d-dimensional projection matrix of rank $d - 1$.

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

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

is defined as

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

and the Kronecker-product of the matrices $\mathbf{A}_i$, $i = 1, \ldots, a$ is written as $\bigotimes_{i=1}^{a} \mathbf{A}_i$.

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

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

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

Note that

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

if the matrices $\mathbf{A}_i$ and $\mathbf{B}_i$ 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 $B(A)$.

The normalized-version of the distribution function of a random variable X_i is denoted by

$$F_i(x) = \frac{1}{2}\left[F_i^+(x) + F_i^-(x)\right] \tag{2.2}$$

where $F_i^+(x) = P(X_i \leq x)$ is the right continuous version and $F_i^-(x) = P(X_i < x)$ is the left continuous version of the distribution function. This definition of the distribution function includes the case of ties and, moreover discrete ordinal data are included in this setup. We exclude only the trivial case when $F_i(x)$ is a one-point distribution function. The terminology 'normalized-version' is taken over from functional analysis where in the context of Fourier series, it is used to denote the mean of the left- and right-continuous version of a function (see e.g. Lang (1993)). 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 if not stated otherwise. If a random variable X_i is distributed according to a distribution function $F_i(x)$, this is written as $X_i \sim F_i(x)$, and the asymptotic equivalence of two sequences of random variables X_N and Y_N is denoted by $X_N \doteq Y_N$.

Models. We consider independent random variables $X_{ij} \sim F_i$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$. To describe treatment effects in this general setup, we consider the relative treatment effects $p_i = \int H(x)dF_i(x)$ where $H(x) = N^{-1}\sum_{i=1}^{d} n_i F_i(x)$ is the weighted average distribution function. The p_i's can be regarded as 'relative effects' with respect to $H(x)$. If $H(x) \equiv x$, then $p_i = \mu_i = \int x dF_i(x)$ is the expectation. In this sense, $p_i = \int H(x)dF_i(x)$ is a generalized expectation. Denote by $\mathbf{p} = (p_1, \ldots, p_d)' = \int Hd\mathbf{F}$ the vector of the relative treatment effects where $\mathbf{F} = (F_1, \ldots, F_d)'$ is the vector of the distribution functions.

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 relative effects $p_1, \ldots, p_d$ are a lexicographic ordering of the higher-way layout relative effects, e.g., $p_{11}, \ldots, p_{d_1 d_2}$ such that the last index i_2 is running faster than the first index i_1.

The relative effects p_i are generalized by introducing a score function $J(u)$, $u \in (0,1) \to \mathbb{R}^1$ and it is assumed that the function $J(u)$ has a bounded second derivative, i.e. $\|J''\|_\infty = \sup_{0 \le u \le 1} |J''(u)| < \infty$ in order to avoid unnecessary technical complications. Thus, we consider the weighted relative effects

$$\mathbf{p}(J) \;=\; (p_1(J), \ldots, p_d(J))' = \int J[H(x)] d\mathbf{F}(x) \; . \tag{2.3}$$

Hypotheses. The simplest hypothesis is the hypothesis that there is no treatment effect at all. In a nonparametric setup, this hypothesis is formulated as $H_0^F : F_1 = \cdots = F_d$ which can formally be written as $H_0^F : \mathbf{P}_d \mathbf{F} = \mathbf{0}$, where $\mathbf{P}_d$ is given in (2.1) and $\mathbf{0}$ denotes a $d \times 1$ vector of functions mapping into $\mathbf{0}$. More complex hypotheses or hypotheses in higher-way layouts may be formulated by a suitable contrast matrix $\mathbf{C}$ as $H_0^F(\mathbf{C}) : \mathbf{CF} = \mathbf{0}$. This formulation of a hypothesis in a nonparametric setup is analogous to the formulation of a hypothesis in the theory of linear models where the hypotheses are formulated in terms of the expectations μ_i, i.e. $H_0^\mu(\mathbf{C}) : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)'$. In a nonparametric setup, the hypotheses must be formulated in terms of the distribution functions since no parameters are available. The interpretation of these nonparametric hypotheses and the relation to some parametric hypotheses will be discussed in Section 5 where the general results to be derived in the next section are applied to special designs. Note that in general, $H_0^F(\mathbf{C}) : \mathbf{CF} = \mathbf{0} \Rightarrow H_0^\mu(\mathbf{C}) : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ since $\mathbf{C}\boldsymbol{\mu} = \mathbf{C} \int x d\mathbf{F} = \int x d(\mathbf{CF})$.

Statistics. The quantities $p_i(J)$ are estimated by replacing the distribution functions $F_i(x)$ by their empirical counterparts

$$\widehat{F}_i(x) = \frac{1}{2}\left[\widehat{F}_i^+(x) + \widehat{F}_i^-(x)\right] = \frac{1}{n_i}\sum_{j=1}^{n_i} c(x - X_{ij}) \tag{2.4}$$

where $c(u) = \frac{1}{2}[c^+(u) + c^-(u)]$ 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 $\widehat{\mathbf{F}}(x) = (\widehat{F}_1(x), \ldots, \widehat{F}_d(x))'$ or shortly by $\widehat{\mathbf{F}} = (\widehat{F}_1, \ldots, \widehat{F}_d)'$. 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

$$\widehat{H}(x) = \frac{1}{N}\sum_{i=1}^{d} n_i \widehat{F}_i(x) = \frac{1}{N}\sum_{i=1}^{d}\sum_{j=1}^{n_i} c(x - X_{ij}) \tag{2.5}$$

and an estimator for $\mathbf{p}(J)$ is given by $\widehat{\mathbf{p}}(J) = (\widehat{p}_1(J), \ldots, \widehat{p}_d(J))'$ where

$$\widehat{p}_i(J) = \int J[\widehat{H}]d\widehat{F}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} J\left[\frac{1}{N}\left(R_{ij} - \frac{1}{2}\right)\right] = \frac{1}{n_i}\sum_{j=1}^{n_i}\phi_{ij} = \overline{\phi}_{i\cdot} \tag{2.6}$$

and

$$R_{ij} = \frac{1}{2} + N\widehat{H}(X_{ij}) = \frac{1}{2} + \sum_{r=1}^{d}\sum_{s=1}^{n_r} c(X_{ij} - X_{rs}) \tag{2.7}$$

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. The quantities $\phi_{ij} = J\left[\frac{1}{N}\left(R_{ij} - \frac{1}{2}\right)\right]$ are called *rank scores.*

REMARK 2.1 *To generate the rank scores ϕ_{ij}, the combined empirical distribution function $\widehat{H}(x)$ is commonly multiplied by $\frac{N}{N+1}$ in order to avoid an infinite value of $J(u)$ at* *the point $u = 1$ if the score function $J(\cdot)$ is unbounded. Moreover, $\frac{N}{N+1}\widehat{H}(x) \in (0,1)$ and $\widehat{H}(x) \in (0,1]$ if the counting function $c^+(u)$ is used to generate the empirical distribution function given in (2.4) which is commonly the case when the distribution functions $F_i(x)$ are assumed to be continuous. In this case, the definition of the scores is not symmetric with respect to the limits of the interval $[0,1]$. However, we consider only bounded score functions $J(\cdot)$ and moreover, $\widehat{H}(x) \in \left[\frac{1}{2N}, 1 - \frac{1}{2N}\right]$ if $F_i(x)$ and $\widehat{F}_i(x)$ are defined as given in (2.2) and (2.4) respectively. Thus, the normalized-version of the empirical distribution function leads automatically to a symmetric definition of the scores and there is no need to consider $J\left[\frac{N}{N+1}\widehat{H}(x)\right]$ instead of $J[\widehat{H}(x)]$.*

3 Asymptotic Results

3.1 Score Functions with Bounded Derivatives

In this Section, some asymptotic properties of the statistic $\widehat{\mathbf{p}}(J)$ are given and the asymptotic normality of $\sqrt{N}\mathbf{C}\widehat{\mathbf{p}}(J)$ is derived under the hypothesis $H_0^F(\mathbf{C}) : \mathbf{CF} = \mathbf{0}$ where $\mathbf{C}$ is a suitable contrast matrix.

To derive the asymptotic results, the following regularity conditions are needed.

ASSUMPTIONS

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

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

Let $J(u)$, $u \in (0,1) \to I\!R^1$, be a score function with

(c1) bounded first derivative, i.e. $\|J'\|_\infty = \sup_{0<u<1} |J'(u)| < \infty$.

(c2) bounded second derivative, i.e. $\|J''\|_\infty = \sup_{0<u<1} |J''(u)| < \infty$.

We note that (c2) $\Rightarrow$ (c1) $\Rightarrow$ $\|J\|_\infty = \sup_{0<u<1} |J(u)| < \infty$. Below, it will be stated separately for each theorem, proposition or lemma which of these assumptions are needed to prove the results.

First, conditions for the consistency of the estimators $\widehat{p}_i(J)$ are given.

PROPOSITION 3.1 *Let $X_{ij} \sim F_i(x)$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$ be independent random variables and let $p_i(J)$ and $\widehat{p}_i(J)$ be as given in (2.3) and (2.6) respectively. Then, under the assumptions (a), (b) and (c1),*

$$\widehat{p}_i(J) - p_i(J) \xrightarrow{p} 0 \ .$$

PROOF: It suffices to show that $E(\widehat{p}_i(J) - p_i(J))^2 \to 0$.

First consider

$$
\begin{aligned}
&(\widehat{p}_i(J) - p_i(J))^2 \\
&\quad = \left(\int J[\widehat{H}] d\widehat{F}_i - \int J[H] dF_i \right)^2 \\
&\quad = \left(\int \left(J[\widehat{H}] - J[H] \right) d\widehat{F}_i + \int J[H] d \left[\widehat{F}_i - F_i \right] \right)^2 \\
&\quad \le \frac{2}{n_i} \sum_{j=1}^{n_i} \left(J[\widehat{H}(X_{ij})] - J[H(X_{ij})] \right)^2 \\
&\qquad\qquad + \frac{2}{n_i^2} \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} \left(J[H(X_{ij})] - \int J[H] dF_i \right) \left(J[H(X_{ik})] - \int J[H] dF_i \right)
\end{aligned}
$$

using Jensen's inequality. Taking expectation and using independence and Lemma A.2, (5) of the Appendix, we obtain

$$
\begin{aligned}
E\left[\widehat{p}_i(J) - p_i(J)\right]^2 &\leq \frac{2}{N}\|J'\|_\infty^2 + \frac{2}{n_i^2}\sum_{j=1}^{n_i} E\left(J[H(X_{ij})] - \int J[H]dF_i\right)^2 \\
&\leq \frac{2}{n_i}\left(\|J'\|_\infty^2 + \|J\|_\infty^2\right) = O\left(\frac{1}{n_i}\right)
\end{aligned}
$$

and the proof follows. □

Next, we state the basic asymptotic equivalence.

THEOREM 3.2 *Let* $X_{ij} \sim F_i(x)$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$ *be independent random variables. Then under the assumptions* $(a), (b)$ *and* $(c2)$,

$$
\sqrt{N}\int J[\widehat{H}]d\left(\widehat{\mathbf{F}} - \mathbf{F}\right) \quad \doteq \quad \sqrt{N}\int J[H]d\left(\widehat{\mathbf{F}} - \mathbf{F}\right) \ .
$$

PROOF: It suffices to consider the i-th component of $\widehat{\mathbf{F}} - \mathbf{F}$. We note that

$$
\begin{aligned}
\sqrt{N}\int J[\widehat{H}]d\left(\widehat{F}_i - F_i\right) &= \sqrt{N}\int J[H]d\left(\widehat{F}_i - F_i\right) \\
&\quad + \sqrt{N}\int \left(J[\widehat{H}] - J[H]\right) d\left(\widehat{F}_i - F_i\right), \quad i = 1, \ldots, d \ .
\end{aligned}
$$

Using Taylor's expansion, we obtain

$$
J[\widehat{H}] - J[H] \quad = \quad J'[H]\left[\widehat{H} - H\right] + \frac{1}{2}J''\left(\widehat{\theta}_N\right)\left[\widehat{H} - H\right]^2
$$

where $\widehat{\theta}_N$ is between $\widehat{H}$ and H. Thus,

$$
\sqrt{N}\int J[\widehat{H}]d\left(\widehat{F}_i - F_i\right) \quad = \quad \sqrt{N}\int J[H]d\left(\widehat{F}_i - F_i\right) + \sqrt{N}\left(B_1 + B_2\right)
$$

where

$$B_1 = \int J'[H]\left[\widehat{H} - H\right] d(\widehat{F}_i - F_i) \quad \text{and} \quad B_2 = \frac{1}{2}\int J''[\widehat{\theta}_N]\left[\widehat{H} - H\right]^2 d(\widehat{F}_i - F_i)\ .$$

It follows from the Appendix, Lemma A.4 that $E(NB_j^2) \to 0$, $j = 1, 2$ which completes the proof. □

We note that $\sqrt{N}\int J[H]d\widehat{\mathbf{F}} = \sqrt{N}\ \overline{\mathbf{Y}}.(J)$ is a vector of independent (unobservable) random variables $\sqrt{N}\ \overline{Y}_{i\cdot}(J) = \sqrt{N} n_i^{-1}\sum_{j=1}^{n_i} Y_{ij}(J)$, $i = 1, \ldots, d$, where $Y_{ij}(J) = J\left[H(X_{ij})\right]$ is called *asymptotic rank (score) transform (ART)* because $Y_{ij}(J)$ is asymptotically equivalent to $\widehat{Y}_{ij}(J) = J[\widehat{H}(X_{ij})]$. Thus,

$$\mathbf{V}_N \;=\; Cov\left(\sqrt{N}\ \overline{\mathbf{Y}}.(J)\right) = N\bigoplus_{i=1}^{d}\frac{1}{n_i}\sigma_i^2(J) \tag{3.8}$$

where $\sigma_i^2(J) = Var\left(J\left[H(X_{ij})\right]\right)$, $j = 1, \ldots, n_i$. The unknown variances $\sigma_i^2(J)$ can be estimated from the rank scores $\phi_{ij} = \widehat{Y}_{ij}(J) = J\left[\frac{1}{N}(R_{ij} - \frac{1}{2})\right]$.

THEOREM 3.3 *Let $X_{ij} \sim F_i(x)$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$ be independent random variables and assume that $\sigma_i^2(J) \geq \sigma_0^2(J) > 0$ where $\sigma_i^2(J)$ is given in (3.8). Then, under the assumptions (a), (b) and (c1), $\widehat{\sigma}_i^2(J)/\sigma_i^2(J) \xrightarrow{p} 1$, where*

$$\widehat{\sigma}_i^2(J) \;=\; \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(\phi_{ij} - \overline{\phi}_{i\cdot}\right)^2, \quad \overline{\phi}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i}\phi_{ij}\ , \quad i = 1, \ldots, d \tag{3.9}$$

where $\phi_{ij} = J\left[\frac{1}{N}(R_{ij} - \frac{1}{2})\right]$. Moreover, $\widehat{\mathbf{V}}_N\mathbf{V}_N^{-1} \xrightarrow{p} \mathbf{I}_d$ where $\widehat{\mathbf{V}}_N = N\bigoplus_{i=1}^{d}\frac{1}{n_i}\widehat{\sigma}_i^2(J)$.

PROOF: Since $\sigma_i^2(J) > 0$ by assumption, it suffices to show that $\widehat{\sigma}_i^2(J) - \sigma_i^2(J) \xrightarrow{p} 0$. First we note that

$$\widehat{\sigma}_i^2(J) \;=\; \frac{1}{n_i - 1}\left[\sum_{j=1}^{n_i} J^2\left[\widehat{H}(X_{ij})\right] - n_i\left(\frac{1}{n_i}\sum_{j=1}^{n_i} J\left[\widehat{H}(X_{ij})\right]\right)^2\right]$$

$$= \frac{1}{n_i}\sum_{j=1}^{n_i} J^2\left[\widehat{H}(X_{ij})\right] - \left(\frac{1}{n_i}\sum_{j=1}^{n_i} J\left[\widehat{H}(X_{ij})\right]\right)^2 + \frac{1}{n_i}\widehat{\sigma}_i^2(J)$$

$$= \int J^2[\widehat{H}]d\widehat{F}_i - \left(\int J[\widehat{H}]d\widehat{F}_i\right)^2 + O\left(\frac{\|J\|_\infty^2}{n_i}\right)$$

and $\sigma_i^2(J) = \int J^2[H]dF_i - (\int J[H]dF_i)^2$. Thus,

$$\begin{aligned}
&\widehat{\sigma}_i^2(J) - \sigma_i^2(J) \\
&= \int J^2[\widehat{H}]d\widehat{F}_i - \int J^2[H]dF_i - \left(\int J[\widehat{H}]d\widehat{F}_i\right)^2 + \left(\int J[H]dF_i\right)^2 + O\left(\frac{\|J\|_\infty^2}{n_i}\right) \\
&= \int (J[\widehat{H}] + J[H])(J[\widehat{H}] - J[H])d\widehat{F}_i + \int J^2[H]d(\widehat{F}_i - F_i) \\
&\quad - \left(\int J[\widehat{H}]d\widehat{F}_i - \int J[H]dF_i\right)\left(\int J[\widehat{H}]d\widehat{F}_i + \int J[H]dF_i\right) + O\left(\frac{\|J\|_\infty^2}{n_i}\right) \\
&= A_1 + A_2 - A_3 + O\left(\frac{\|J\|_\infty^2}{n_i}\right) .
\end{aligned}$$

We consider the three terms A_j separately and show that $E(A_j^2) \to 0$, $j = 1, 2, 3$. For the first term, we have

$$A_1^2 \leq 4\|J\|_\infty^2 \int (J[\widehat{H}] - J[H])^2 d\widehat{F}_i = \frac{4\|J\|_\infty^2}{n_i}\sum_{j=1}^{n_i}\left(J\left[\widehat{H}(X_{ij})\right] - J\left[H(X_{ij})\right]\right)^2$$

by Jensen's inequality and $E(A_1^2) \leq 4\|J\|_\infty^2 \cdot \|J'\|_\infty^2/N \to 0$ by Lemma A.2, (6.).

Using independence, it follows for the second term that

$$\begin{aligned}
E\left(A_2^2\right) &= \frac{1}{n_i^2}\sum_{j=1}^{n_i} E\left(J^2[H(X_{ij})] - \int J^2[H]dF_i\right)^2 \\
&= \frac{1}{n_i^2}\sum_{j=1}^{n_i} Var\left(J^2[H(X_{ij})]\right) \leq \frac{\|J\|_\infty^4}{n_i} \to 0 .
\end{aligned}$$

For the third term, it follows that

$$A_3^2 = \left(\int J[\widehat{H}]d\widehat{F}_i - \int J[H]dF_i\right)^2\left(\int J[\widehat{H}]d\widehat{F}_i + \int J[H]dF_i\right)^2$$

$$\leq \quad 4\|J\|_\infty^2 \left(\int J[\widehat{H}] d\widehat{F}_i - \int J[H] dF_i \right)^2 \ .$$

Finally,

$$E\left(A_3^2\right) \quad \leq \quad \frac{8\|J\|_\infty^2}{n_i} \left(\|J\|_\infty^2 + \|J'\|_\infty^2\right) \to 0$$

using the result of Proposition 3.1. The proof is completed by noting that $\mathbf{V}_N$ is a diagonal matrix and thus, $\widehat{\mathbf{V}}_N \mathbf{V}_N^{-1} = \bigoplus_{i=1}^{d} \frac{\widehat{\sigma}_i^2(J)}{\sigma_i^2(J)} \xrightarrow{p} \mathbf{I}_d$. □

It should be pointed out that the statement of Theorem 3.2 is that $\sqrt{N}\left[\overline{\mathbf{Y}}.(J) - \mathbf{p}(J)\right]$ is asymptotically equivalent to the random vector $\sqrt{N}\left[\widehat{\mathbf{p}}(J) - \int J[\widehat{H}] d\mathbf{F}\right]$ where $\int J[\widehat{H}] d\mathbf{F}$ is unobservable. However, under the hypothesis $H_0^F : \mathbf{CF} = \mathbf{0}$ it follows that

$$\sqrt{N}\mathbf{C}\left[\widehat{\mathbf{p}}(J) - \int J[\widehat{H}] d\mathbf{F}\right] \quad = \quad \sqrt{N}\mathbf{C}\widehat{\mathbf{p}}(J) - \sqrt{N}\int J[\widehat{H}] d(\mathbf{CF}) = \sqrt{N}\mathbf{C}\widehat{\mathbf{p}}(J)$$

and $\sqrt{N}\mathbf{C}\widehat{\mathbf{p}}(J)$ is a vector of linear rank score statistics.

Now we give the asymptotic distribution of $\sqrt{N}\mathbf{C}\widehat{\mathbf{p}}(J)$ under $H_0^F : \mathbf{CF} = \mathbf{0}$.

THEOREM 3.4 *Let $X_{ij} \sim F_i(x)$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$ be independent random variables and assume that $\sigma_i^2(J) \geq \sigma_0^2(J) > 0$ where $\sigma_i^2(J)$ is given in (3.8). Let $\mathbf{V}_N$ as given in (3.8) and let $\widehat{\mathbf{V}}_N$ as given in Theorem 3.3. Then, under the assumptions (a), (b), (c2) and under the hypothesis $H_0^F : \mathbf{CF} = \mathbf{0}$,*

1. *the statistic $\sqrt{N}\mathbf{C}\widehat{\mathbf{p}}(J) = \sqrt{N}\mathbf{C}\int J[\widehat{H}] d\widehat{\mathbf{F}}$ has asymptotically a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{C}\mathbf{V}_N\mathbf{C}'$,*

2. *the quadratic form* $Q_N(\mathbf{C}) = N\widehat{\mathbf{p}}'(J)\mathbf{C}'\left[\mathbf{C}\mathbf{V}_N\mathbf{C}'\right]^{-}\mathbf{C}\widehat{\mathbf{p}}(J)$ *has asymptotically a central* χ^2_f*-distribution with* $f = rank(\mathbf{C})$ *where* $[\mathbf{C}\mathbf{V}_N\mathbf{C}']^{-}$ *denotes a generalized inverse of* $[\mathbf{C}\mathbf{V}_N\mathbf{C}']$.

3. *If* $\mathbf{C}$ *is of full row rank, then* $Q_N(\mathbf{C}) = N\widehat{\mathbf{p}}'(J)\mathbf{C}'\left[\mathbf{C}\widehat{\mathbf{V}}_N\mathbf{C}'\right]^{-1}\mathbf{C}\widehat{\mathbf{p}}(J)$ *has asymptotically a central* χ^2_f*-distribution with* $f = rank(\mathbf{C})$.

PROOF: Statement (1) follows immediately by the Theorems 3.2 and 3.3 by applying the central limit Theorem to $\sqrt{N}\ \overline{\mathbf{Y}}.(J)$ since the random variables $Y_{ij}(J)$ are uniformly bounded and $\sigma_i^2(J) > 0$, $i = 1, \ldots, d$ by assumption.

To prove (2), set $\mathbf{X} = \sqrt{N}\mathbf{C}\widehat{\mathbf{p}}(J)$. Then, under H_0^F, from (1), $\mathbf{X}$ has asymptotically a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{S} = \mathbf{C}\mathbf{V}_N\mathbf{C}'$. Note that, if $\mathbf{X} \sim N(\mathbf{0}, \mathbf{S})$, then $\mathbf{X}'\mathbf{S}^{-}\mathbf{X} \sim \chi^2_f$ with $f = rank(\mathbf{S})$ where $\mathbf{S}^{-}$ is a symmetric reflexive generalized inverse of $\mathbf{S}$ (see e.g. Rao & Mitra (1971), Theorem 9.2.3.). The result follows by noting that the matrix product $\mathbf{C}'\left[\mathbf{C}\mathbf{V}_N\mathbf{C}'\right]^{-}\mathbf{C}$ is invariant for any choice of the generalized inverse (see e.g. Rao & Mitra (1971), Lemma 2.2.6) and $rank\left(\mathbf{C}'\left[\mathbf{C}\mathbf{V}_N\mathbf{C}'\right]^{-}\mathbf{C}\right) = rank(\mathbf{C})$ if $rank\left(\mathbf{C}\mathbf{V}_N\mathbf{C}'\right) = rank(\mathbf{C})$ which follows by the fact that $\mathbf{V}_N$ is of full rank since $\sigma_i^2(J) > 0$, $i = 1, \ldots, d$ by assumption. Therefore, the result is true for any choice of the generalized inverse $[\mathbf{C}\mathbf{V}\mathbf{C}']^{-}$.

(3) follows immediately from (2) and Theorem 3.3. □

We shall now consider the asymptotic distribution of the statistic $Q_N(\mathbf{C})$ under the sequence of alternatives

$$\mathbf{F}_N = (F_{N,1}, \ldots, F_{N,d})' = \left(1 - \frac{1}{\sqrt{N}}\right) \mathbf{F} + \frac{1}{\sqrt{N}} \mathbf{K} \tag{3.10}$$

contiguous to the nonparametric null hypothesis $H_0^F : \mathbf{CF} = \mathbf{0}$ where $\mathbf{K} = (K_1, \ldots, K_d)'$ is some vector of distribution functions.

THEOREM 3.5 *Let $X_{ij} \sim F_{N,i}(x) = (1 - N^{-1/2})F_i(x) + N^{-1/2}K_i(x)$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$ be independent random variables and let $\widehat{F}_i(x)$ denote the empirical distribution function of $X_{i1}, \ldots, X_{in_i}$ as defined in (2.4). Let $\mathbf{F} = (F_1, \ldots, F_d)'$ in (3.10) and let $\mathbf{C}$ be some contrast matrix of full row rank such that $\mathbf{CF} = \mathbf{0}$. Further let $H(x) = N^{-1} \sum_{i=1}^d n_i F_i(x)$ denote the weighted average distribution function of $F_1, \ldots, F_d$. Let*

$$H^*(x) = \frac{1}{N} \sum_{i=1}^d n_i F_{N,i}(x) = H(x) - \frac{1}{\sqrt{N}} \sum_{i=1}^d \frac{n_i}{N} \left[F_i(x) - K_i(x)\right]$$

denote the weighted average distribution function of $F_{N,1}, \ldots, F_{N,d}$ and let $\widehat{H}(x)$ denote the average empirical distribution function as defined in (2.5). Assume that $\sigma_i^2(J) = \int J^2(H)dF_i - \left(\int J(H)dF_i\right)^2 \geq \sigma_0^2(J) > 0$ and let $\mathbf{V}_N$ as given in (3.8). Then, under the assumptions $(a), (b)$ and $(c2)$ and under the sequence of alternatives (3.10),

1. *the statistics*

   $$\widehat{\boldsymbol{\nu}}(J) = \sqrt{N}\mathbf{C}\widehat{\mathbf{p}}(J) = \sqrt{N}\mathbf{C} \int J[\widehat{H}]d\widehat{\mathbf{F}} \quad \text{and} \quad \sqrt{N}\mathbf{C}\overline{\mathbf{Y}}_{\cdot} = \sqrt{N}\mathbf{C} \int J(H)d\widehat{\mathbf{F}}$$

   are asymptotically equivalent,

2. *$\widehat{\boldsymbol{\nu}}(J)$ has asymptotically a multivariate normal distribution $N(\boldsymbol{\nu}(J), \mathbf{C}\mathbf{V}_N\mathbf{C}')$, where $\boldsymbol{\nu}(J) = \int J(H)d(\mathbf{CK})$,*

3. *the quadratic form* $Q_N(\mathbf{C}) = N\widehat{\mathbf{p}}'(J)\mathbf{C}'\left[\mathbf{C}\widehat{\mathbf{V}}_N\mathbf{C}'\right]^{-1}\mathbf{C}\widehat{\mathbf{p}}(J)$ *has asymptotically a noncentral* $\chi^2_f(\gamma(J))$*-distribution with* $f = rank(\mathbf{C})$ *and with noncentrality parameter* $\gamma(J) = \boldsymbol{\nu}'(J)\left[\mathbf{C}\mathbf{V}_N\mathbf{C}'\right]^{-1}\boldsymbol{\nu}(J)$.

PROOF: Proceeding as in the proof of Theorem 3.2, we decompose

$$\begin{aligned}
\widehat{\boldsymbol{\nu}}(J) &= \sqrt{N}\mathbf{C}\int J[\widehat{H}]d\widehat{\mathbf{F}} \\
&= \sqrt{N}\mathbf{C}\Big(\int J(H)d\widehat{\mathbf{F}} + \int \left[J[\widehat{H}] - J(H)\right] d\mathbf{F}_N \\
&\qquad + \int \left[J[\widehat{H}] - J(H) + J(H^*) - J(H^*)\right] d(\widehat{\mathbf{F}} - \mathbf{F}_N)\Big) \\
&= \sqrt{N}\mathbf{C}\int J(H)d\widehat{\mathbf{F}} + \mathbf{a}_1 + \mathbf{a}_2 - \mathbf{a}_3
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{a}_1 &= \mathbf{C}\int \left[J[\widehat{H}] - J(H)\right] d\mathbf{K} \\
\mathbf{a}_2 &= \sqrt{N}\mathbf{C}\int \left[J[\widehat{H}] - J(H^*)\right] d(\widehat{\mathbf{F}} - \mathbf{F}_N) \\
\mathbf{a}_3 &= \sqrt{N}\mathbf{C}\int \left[J(H) - J(H^*)\right] d(\widehat{\mathbf{F}} - \mathbf{F}_N)\ .
\end{aligned}$$

To prove the statement in (1), it suffices to consider the i-th components of $\mathbf{a}_j$, $j = 1, 2, 3$ and it will be shown that

$$\begin{aligned}
(i) &\quad \int [J[\widehat{H}] - J(H)]\ dK_i = o(1) \quad \text{in } L^2, \\
(ii) &\quad \sqrt{N}\int [J[\widehat{H}] - J(H^*)]\ d(\widehat{F}_i - F_{N,i}) = o(1) \quad \text{in } L^2, \\
(iii) &\quad \sqrt{N}\int [J(H) - J(H^*)]\ d(\widehat{F}_i - F_{N,i}) = o(1) \quad \text{in } L^2\ .
\end{aligned}$$

Statement (i) follows easily by Jensen's inequality,

$$E\Big(\int \left[J[\widehat{H}] - J(H)\right] dK_i\Big)^2 \ \le\ \int E\left[J[\widehat{H}] - J(H)\right]^2 dK_i$$

$$\leq \frac{4}{N}\|J'\|_\infty^2$$

and using Lemma A.3, (2) in the Appendix.

To prove (ii), we note that under the sequence of alternatives (3.10),

$$E\left(\widehat{F}_i(x)\right) = F_{N,i}(x) \quad \text{and} \quad E\left(\widehat{H}(x)\right) = H^*(x) \tag{3.11}$$

and the result follows in the same way as in the proof of Theorem 3.2 using Lemma A.4 in the Appendix.

To prove (iii), we note that

$$\begin{aligned}
&E\left(\sqrt{N}\int [J[H(x)] - J[H^*(x)]]\, d(\widehat{F}_i(x) - F_{N,i}(x))\right)^2 \\
&= \frac{N}{n_i^2}\sum_{j=1}^{n_i} E\left(J\left[H(X_{ij})\right] - J\left[H^*(X_{ij})\right] - \int \left(J\left[H(x)\right] - J\left[H^*(x)\right]\right) dF_{N,i}(x)\right)^2 \\
&\leq \frac{N}{n_i^2}\sum_{j=1}^{n_i}\left[2E\left(J\left[H(X_{ij})\right] - J\left[H^*(X_{ij})\right]\right)^2 + 2\int \left(J\left[H(x)\right] - J\left[H^*(x)\right]\right)^2 dF_{N,i}(x)\right]
\end{aligned}$$

by independence, (3.11) and Jensen's inequality. The result follows from Lemma A.3, (1) and (3) in the Appendix. This completes the proof of statement (1). The result (2) follows from Theorem 3.4, (1) by noting that $E\left(\sqrt{N}\mathbf{C}\int J(H)d\widehat{\mathbf{F}}\right) = \int J(H)d(\mathbf{CK})$ and that $\sigma_{N,i}^2(J)/\sigma_i^2(J) \to 1$ as $\min n_i \to \infty$ where $\sigma_{N,i}^2(J) = Var\left(J[H(X_{i1})]\right) = \sigma_i^2(J) + O\left(N^{-1/2}\right)$. Finally, statement (3) follows immediately from (2) which completes the proof of the Theorem. □

REMARK 3.1 *It follows from Theorem 3.5 that the set of alternatives for which the rank score tests are consistent is the set of all vectors* $\mathbf{p}(J)$ *for which* $\mathbf{Cp}(J) \neq \mathbf{0}$ *holds. Let*

$H_0^p(\mathbf{C}) : \mathbf{Cp}(J) = \mathbf{0}$ *denote the nonparametric hypotheses related to the relative treatment effects* $p_i(J)$, $i = 1, \ldots, d$. *Figure 1 shows the relations between the hypotheses* H_0^F, H_0^μ *and* H_0^p *in general.*

About here insert Figure 1.

Consider the case where X_{ij} has a Bernoulli distribution, $X_{ij} \sim B(q_i)$ where $q_i = P(X_{i1} = 1)$, $i = 1, \ldots, d$ and let $\mathbf{q} = (q_1, \ldots, q_d)'$. One possibility to formulate parametric hypotheses concerning $\mathbf{q}$ is to consider contrasts in $\mathbf{q}$ and write the hypotheses as $H_0^q : \mathbf{Cq} = \mathbf{0}$ where $\mathbf{C}$ is a suitable contrast matrix. Obviously, $H_0^q \iff H_0^F$ since F_i is completely determined by q_i.

To give some more insight into the meaning and interpretation of the nonparametric hypotheses in linear models, some relations to parametric hypotheses in the two-way layout are given in Section 5.2.

3.2 Unbounded Score Functions

The case of unbounded score functions is briefly discussed in this subsection. The main motivation is to include also the normal scores which are known to be locally optimal in the case of shift alternatives if the underlying distribution functions are normal. Assuming continuous distribution functions, the asymptotic normality of linear rank-score statistics for unbounded (square integrable) score functions has been shown by Hajek (1968). A similar result for score functions with bounded δ-norm was proved by Denker and Rösler

(1985). Regarding the technical details, we refer to above quoted papers. These general results can be applied to derive rank statistics in factorial designs for the case of continuous distribution functions in the same way as worked out in the proceding theorems. There is no motivation, however, to consider unbounded score functions in the case of discontinuous distribution functions and thus, we shall not consider this problem here.

4 Small Sample Approximations

The procedures considered in the previous sections are valid for large sample sizes. However, the speed of approximation to the asymptotic χ^2-distributions of the proposed statistics is rather slow, especially if the number of factor levels is large. Thus it is necessary to use some small sample approximation for the statistics. For completeness, we quote a simple modified Box-approximation (Box, 1954) for heteroscedastic factorial designs given by Brunner, Dette & Munk (1997). For details, we refer to this paper.

The hypothesis $H_0^\mu : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ can be formulated equivalently as $H_0^\mu : \mathbf{M}\boldsymbol{\mu} = \mathbf{0}$ where $\mathbf{M} = \mathbf{C}'(\mathbf{C}\mathbf{C}')^-\mathbf{C}$ is a projection matrix. Note that all elements of $\mathbf{M}$ are known constants and $\mathbf{M}$ does not depend on the special choice of the generalized inverse $(\mathbf{C}\mathbf{C}')^-$. In many cases, the contrast matrix $\mathbf{C}$ can be chosen such that all diagonal elements of $\mathbf{M}$ are identical to m, say (c.f. the applications given in Section 5). This leads to the simplified form of the Approximation Procedure given below. To test the hypothesis $H_0^F : \mathbf{CF} = \mathbf{0}$, consider the quadratic form $Q_N^*(\mathbf{M}) = N \cdot \widehat{\mathbf{p}}'(J)\ \mathbf{M}\ \widehat{\mathbf{p}}(J)$ instead of the quadratic forms given in Theorem 3.4 (Wald-type statistics). The asymptotic distribution of Q_N^* under the hypothesis is given in Theorem 4.1 and a small sample approximation is given in Approximation Procedure 4.2.

THEOREM 4.1 *Let* $\mathbf{M} = \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-}\mathbf{C}$ *and let* $\mathbf{V}_N$ *be as given in (3.8). Then, under the assumptions of Theorem 3.4 and under the hypothesis* $H_0^F : \mathbf{CF} = \mathbf{0}$*, the quadratic form* $Q_N^*(\mathbf{M}) = N \cdot \widehat{\mathbf{p}}'(J)\ \mathbf{M}\ \widehat{\mathbf{p}}(J)$ *has asymptotically the weighted* χ^2*-distribution* $\sum_{i=1}^{d} \lambda_i U_i$ *where the* U_i *are independent random variables each having a* χ_1^2*-distribution and the* λ_i *are the eigenvalues of* $\mathbf{M}\mathbf{V}_N\mathbf{M}$*.*

PROOF: First note that $\mathbf{MF} = \mathbf{0} \Longleftrightarrow \mathbf{CF} = \mathbf{0}$ since $\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-}$ is a generalized inverse of $\mathbf{C}$. Thus, under H_0^F, by Theorem 3.4, $\sqrt{N}\ \mathbf{M}\widehat{\mathbf{p}}(J)$ has asymptotically a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{M}\mathbf{V}_N\mathbf{M}$. From this, it follows that $Q_N^*(\mathbf{M}) = N \cdot (\mathbf{M}\widehat{\mathbf{p}}(J))'\ \mathbf{M}\ \mathbf{M}\ \widehat{\mathbf{p}}(J) = N \cdot \widehat{\mathbf{p}}'(J)\ \mathbf{M}\ \widehat{\mathbf{p}}(J)$ has asymptotically the weighted χ^2-distribution $\sum_{i=1}^{d} \lambda_i U_i$ where the U_i are independent random variables each having a χ_1^2-distribution and the λ_i are the eigenvalues of $\mathbf{M}\ \mathbf{M}\mathbf{V}_N\mathbf{M} = \mathbf{M}\mathbf{V}_N\mathbf{M}$, (see e.g. Graybill, 1976, p.136). □

Approximation Procedure 4.2 *Let* $\mathbf{M}$ *be as defined in Theorem 4.1 and assume that the diagonal elements* m_{ii} *of* $\mathbf{M}$ *are identical to* m*, say, i.e.* $m_{ii} \equiv m$*. Further let* $\mathbf{\Lambda}_d = diag\{n_1, \ldots, n_d\}$*. Then, under the assumptions of Theorem 4.1, the distribution of the statistic*

$$T_N(\mathbf{M}) \ = \ \frac{N}{m \cdot tr(\widehat{\mathbf{V}}_N)} \cdot \widehat{\mathbf{p}}'(J)\mathbf{M}\widehat{\mathbf{p}}(J) = \frac{Q_N^*(\mathbf{M})}{m \cdot tr(\widehat{\mathbf{V}}_N)} \qquad (4.12)$$

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

$$\widehat{f}_1 = m^2 \cdot \frac{\left[tr(\widehat{\mathbf{V}}_N)\right]^2}{tr\left((\mathbf{M}\widehat{\mathbf{V}}_N\mathbf{M}\widehat{\mathbf{V}}_N)\right)} = (Nm)^2 \cdot \frac{\left(\sum_{i=1}^d \widehat{\sigma}_i^2(J)/n_i\right)^2}{tr(\mathbf{M}\widehat{\mathbf{V}}_N\mathbf{M}\widehat{\mathbf{V}}_N)} \quad and \tag{4.13}$$

$$\widehat{f}_0 = \frac{\left[tr(\widehat{\mathbf{V}}_N)\right]^2}{tr\left(\widehat{\mathbf{V}}_N^2(\mathbf{\Lambda}_d - \mathbf{I}_d)^{-1}\right)} = \frac{\left(\sum_{i=1}^d \widehat{\sigma}_i^2(J)/n_i\right)^2}{\sum_{i=1}^d \widehat{\sigma}_i^4(J)/[n_i^2(n_i-1)]}\ , \tag{4.14}$$

where $\widehat{\sigma}_i^2(J)$ is given in (3.9) and $tr(\cdot)$ denotes the trace of a square matrix.

Derivation: See Brunner, Dette and Munk (1997) where also the the more general case is considered where $\mathbf{M}$ does not have identical diagonal elements and the accuracy of the approximation is verified by some simulation studies.

5 Application to Special Factorial Designs

In this section, we apply the general theory derived in Section 3 to some special factorial designs. On the one hand, we wish to give some explicit statistics; on the other hand, we will show that several known rank statistics which have been proposed for some simple designs, follow as special cases from our general approach. Particularly, we consider the one-factor design where the Kruskal-Wallis statistic (1952, 1953) and the rank-transform statistic (Conover & Iman, 1981) come out as special cases.

Cross-classified and nested models are considered as examples for two-way layouts where the statistics derived by Akritas, Arnold & Brunner (1997) are special cases if Wilcoxon scores are considered. The relations to some rank statistics proposed by Hora & Conover (1984), Akritas (1990) and Thompson (1991) are also discussed.

The extension to higher-way layouts is straightforward.

5.1 One-Factor Designs

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 $\mathbf{P}_a = \mathbf{I}_a - \frac{1}{a}\mathbf{J}_a$ be the contrast matrix defined in (2.1). Then the hypothesis for the one-way layout is written as $H_0^F : F_1 = \cdots = F_a$ or equivalently as $\mathbf{P}_a\mathbf{F} = \mathbf{0}$ where $\mathbf{F} = (F_1, \ldots, F_a)'$ denotes the vector of the distribution functions and $\mathbf{0}$ denotes a vector of functions mapping into 0. Let $\mathbf{p}(J) = (p_1(J), \ldots, p_a(J))' = \int J(H)d\mathbf{F}$ be the vector of the relative treatment effects $p_i(J) = \int J(H)dF_i$ where $H(x) = \frac{1}{N}\sum_{i=1}^a n_i F_i(x)$ is the weighted mean distribution function. The vector $\mathbf{p}(J)$ is estimated consistently by $\widehat{\mathbf{p}}(J) = \int J(\widehat{H})d\widehat{\mathbf{F}} = (\widehat{p}_1(J), \ldots, \widehat{p}_a(J))'$ where $\widehat{p}_i(J) = \overline{\phi}_{i\cdot}$ is given in (2.6).

Under H_0^F, the statistics $\sqrt{N}\mathbf{P}_a\widehat{\mathbf{p}}(J)$ and $\sqrt{N}\mathbf{P}_a\overline{\mathbf{Y}}_\cdot(J)$ are asymptotically equivalent (see Theorem 3.2) where $\overline{\mathbf{Y}}_\cdot(J) = (\overline{Y}_{1\cdot}(J), \ldots, \overline{Y}_{a\cdot}(J))'$, $\overline{Y}_{i\cdot}(J) = n_i^{-1}\sum_{j=1}^{n_i} J(H(X_{ij}))$. Let $\mathbf{\Lambda} = diag\{n_1, \ldots, n_a\}$ denote the diagonal matrix of the sample sizes. Then under $H_0^F : \mathbf{P}_a\mathbf{F} = \mathbf{0}$, it follows that $\sigma_1^2(J) = \cdots = \sigma_a^2(J) = \sigma^2(J)$ and $\mathbf{V}_N = Cov\left(\sqrt{N}\ \overline{\mathbf{Y}}_\cdot(J)\right) = N\sigma^2(J)\mathbf{\Lambda}^{-1}$. A consistent estimate of $\sigma^2(J)$ follows immediately from Theorem 3.3 by pooling the estimators $\widehat{\sigma}_i^2(J)$, viz.

$$\widehat{\sigma}_N^2(J) = \frac{1}{N-a}\sum_{i=1}^a\sum_{j=1}^{n_i}\left(\phi_{ij} - \overline{\phi}_{i\cdot}\right)^2 \qquad (5.1)$$

where ϕ_{ij} and $\overline{\phi}_{i\cdot}$ are given in (3.9). To test the hypothesis H_0^F, consider the quadratic form Q_N given in Theorem 3.4, (2) and let $\widehat{\mathbf{W}}_N = \widehat{\mathbf{V}}_N^{-1}(\mathbf{I}_a - \mathbf{J}_a\widehat{\mathbf{V}}_N^{-1}/tr(\widehat{\mathbf{V}}_N^{-1}))$ where $\widehat{\mathbf{V}}_N^{-1} = \mathbf{\Lambda}/[N\widehat{\sigma}_N^2(J)]$. Note that $\widehat{\mathbf{W}}_N$ is a generalized inverse of $\mathbf{P}_a\widehat{\mathbf{V}}_N\mathbf{P}_a$ and that $\mathbf{P}_a\widehat{\mathbf{W}}_N\mathbf{P}_a = \widehat{\mathbf{W}}_N$. Then it follows from Theorem 3.4 that

$$Q_N = N\widehat{\mathbf{p}}'(J)\mathbf{P}_a[\mathbf{P}_a\widehat{\mathbf{V}}_N\mathbf{P}_a]^-\mathbf{P}_a\widehat{\mathbf{p}}(J) = N\widehat{\mathbf{p}}'(J)\widehat{\mathbf{W}}_N\widehat{\mathbf{p}}(J)$$

$$
\begin{aligned}
&= \frac{1}{\widehat{\sigma}_N^2(J)} \widehat{\mathbf{p}}'(J) \left(\mathbf{\Lambda} - \frac{1}{N} \mathbf{\Lambda} \mathbf{J}_a \mathbf{\Lambda} \right) \widehat{\mathbf{p}}(J) \\
&= \frac{N-a}{\sum_{i=1}^{a} \sum_{j=1}^{n_i} \left(\phi_{ij} - \overline{\phi}_{i\cdot} \right)^2} \sum_{i=1}^{a} n_i \left(\overline{\phi}_{i\cdot} - \overline{\phi}_{\cdot\cdot} \right)^2 \qquad (5.2)
\end{aligned}
$$

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 $\sigma_1^2(J) = \cdots = \sigma_a^2(J)$ under H_0^F and thus, it is not necessary to apply the small sample approximation for heteroscedastic factorial designs stated in Section 4. A comprehensive simulation study for Wilcoxon scores showed that the approximation is quite accurate if $a \geq 3$ and $n_i \geq 6$.

REMARK 5.1 *Q_N given in (5.2) has the so called rank transform (RT) property, i.e. if the scores ϕ_{ij} are replaced by independent normally distributed random variables, then the corresponding normal theory statistic has asymptotically the same distribution as Q_N. For a further discussion of the RT-property of a rank statistic, see Section 5.3.*

If $\widehat{\sigma}_N^2(J)$ given in (5.1) is replaced by

$$
\widehat{\sigma}_0^2(J) = \frac{1}{N-1} \sum_{i=1}^{a} \sum_{j=1}^{n_i} (\phi_{ij} - \mu(J))^2 \qquad (5.3)
$$

where $\mu(J) = \int J(H)dH$, then the quadratic form Q_N given in (5.2) becomes the Kruskal-Wallis statistic for general scores (see e.g. Puri, 1964). Note that $\mu(J)$ may be replaced by the consistent estimator $\widehat{\mu}(J) = \int J(\widehat{H})d\widehat{H} = \overline{\phi}_{\cdot\cdot}$ and for Wilcoxon scores, the centering constant becomes $\overline{\phi}_{\cdot\cdot} = (N+1)/2N$. Moreover, if the distribution functions are assumed to be continuous and if Wilcoxon scores are used, then

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

and Q_N given in (5.2) becomes the Kruskal-Wallis H-statistic (Kruskal & Wallis, 1952, 1953). Note that both the variance estimators $\widehat{\sigma}_N^2(J)$ and $\widehat{\sigma}_0^2(J)$ are consistent for $\sigma^2(J)$.

Patterned Alternatives

We use the method of Page (1963) and Hettmansperger & Norton (1987) for constructing test statistics which are especially sensitive against a conjectured patterned alternative. The estimated treatment effects are weighted by a set of constants $w_1, \ldots, w_a$ reproducing the conjectured pattern of the alternative which has to be specified in advance. By renumbering the levels of the treatment, one can always establish that the pattern is related to an increasing treatment effect. Thus, without loss of generality, we take the weights $w_1 \leq w_2 \leq \cdots \leq w_a$, and test the hypothesis $H_0^F : F_1 = \cdots = F_a$ against the alternative $H_1^F : F_1 \geq F_2 \geq \cdots \geq F_a$ with at least one strict inequality.

Let $\mathbf{w} = (w_1, \ldots, w_a)'$ denote the vector of the weights w_i. Then under H_0^F, the linear rank score statistic

$$L_N(J) \;=\; \sqrt{N}\;\mathbf{w}'\mathbf{P}_a\widehat{\mathbf{p}}(J) = \sqrt{N}\sum_{i=1}^{a}(w_i - \overline{w}_\cdot)\overline{\phi}_{i\cdot} \qquad (5.4)$$

has asymptotically a normal distribution with mean 0 and variance

$$\sigma_w^2(J) \;=\; \mathbf{w}'\mathbf{P}_a\mathbf{V}_N\mathbf{P}_a\mathbf{w} = N\sigma^2(J)\sum_{i=1}^{a}\frac{1}{n_i}(w_i - \overline{w}_\cdot)^2\ .$$

Here, $\overline{w}_{\cdot} = a^{-1} \sum_{i=1}^{a} w_i$ and the variance $\sigma^2(J)$ is estimated by $\widehat{\sigma}_N^2(J)$ given in (5.1). Let $\widehat{\sigma}_w^2(J) = N\widehat{\sigma}_N^2(J) \sum_{i=1}^{a} n_i^{-1}(w_i - \overline{w}_{\cdot})^2$. Then the statistic $T_N(J) = L_N(J)/\widehat{\sigma}_w(J)$ has asymptotically a standard normal distribution under H_0^F. For small samples, the null distribution of $T_N(J)$ may be approximated by the central t_ν-distribution with $\nu = N - a - 1$. For Wilcoxon scores, the approximation is quite accurate for $a \geq 3$ and $n_i \geq 6$.

5.2 Two-Factor Designs

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

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

denote the vector of the distribution functions where the second index j is running faster than the first index i. Let $\mathbf{C}_A = \mathbf{P}_a \otimes \frac{1}{b}\mathbf{1}_b'$, $\mathbf{C}_B = \frac{1}{a}\mathbf{1}_a' \otimes \mathbf{P}_b$ and $\mathbf{C}_{AB} = \mathbf{P}_a \otimes \mathbf{P}_b$ where $\mathbf{P}_a$ and $\mathbf{P}_b$ are given in (2.1). Then the nonparametric hypotheses of 'no main effect A', 'no main effect B' or 'no interaction AB' are formulated as

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

The hypotheses of no *simple factor effects* $A|B$ and $B|A$ are typically analyzed in nested designs. However, such hypotheses have also been considered in cross-classified designs (Koch (1969, 1970), Akritas (1990), Thompson (1991) and Brunner, Puri & Sun (1995) where sometimes the terminology *joint hypothesis* is used. The hypothesis $H_0^F(A|B)$ of no simple factor effect A means that there is no effect of factor A within each level of factor B. Let $\mathbf{C}_{A|B} = \mathbf{P}_a \otimes \mathbf{I}_b$ and $\mathbf{C}_{B|A} = \mathbf{I}_a \otimes \mathbf{P}_b$. Then the hypotheses of no simple factor A effect and of no simple factor B effect are formulated as

$$H_0^F(A|B) : \mathbf{C}_{A|B}\mathbf{F} = \mathbf{0}, \quad H_0^F(B|A) : \mathbf{C}_{B|A}\mathbf{F} = \mathbf{0}\ .$$

The nonparametric hypotheses shall be compared with the usual parametric hypotheses in the linear model to give some insight into the interpretation and meaning of the more restrictive nonparametric hypotheses.

In the linear model, it is assumed that $X_{ijk} \sim F_{ij}(x) = F(x - \mu_{ij})$ where $\mu_{ij} = \alpha_i + \beta_j + (\alpha\beta)_{ij}$, $i = 1, \ldots, a$, $j = 1, \ldots, b$. Let $\boldsymbol{\mu} = (\mu_{11}, \ldots, \mu_{ab})'$. Then the hypotheses of no main effects A and B and of no interaction AB are expressed as

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

The hypotheses of no simple factor A or of no simple factor B effect are expressed as

$$H_0^\mu(A|B) : \mathbf{C}_{A|B}\ \boldsymbol{\mu} = \mathbf{0}, \quad H_0^\mu(B|A) : \mathbf{C}_{B|A}\ \boldsymbol{\mu} = \mathbf{0}\ .$$

In what follows, we quote some simple relations between the nonparametric and the parametric hypotheses in the linear model for the two-way cross-classification.

PROPOSITION 5.1

1. *In the nonparametric model,*

   (a) $H_0^F(A|B) \Rightarrow H_0^F(A)$ *and* $H_0^F(AB)$*,*

(b) $H_0^F(B|A) \Rightarrow H_0^F(B)$ *and* $H_0^F(AB)$.

2. *In the linear model where* $\mu_{ij} = \int x dF_{ij}$,

   (a) $H_0^F(A) \Rightarrow H_0^\mu(A)$, $\quad H_0^F(B) \Rightarrow H_0^\mu(B)$, $\quad H_0^F(AB) \Rightarrow H_0^\mu(AB)$,

   (b) $H_0^\mu(A|B) \Rightarrow H_0^\mu(A)$, $\quad H_0^\mu(B|A) \Rightarrow H_0^\mu(B)$,

   (c) $H_0^F(A|B) \Longleftrightarrow H_0^\mu(A|B)$, $\quad H_0^F(B|A) \Longleftrightarrow H_0^\mu(B|A)$,

   (d) *If* $\mathbf{C}_{AB}\boldsymbol{\mu} = \mathbf{0}$, *then* $H_0^F(A) \Longleftrightarrow H_0^\mu(A)$ *and* $H_0^F(B) \Longleftrightarrow H_0^\mu(B)$.

PROOF: For the proofs of the statements 1(a), (b) and 2(a)-(c), see Lemma 2.9 in Brunner and Puri (1996). Statement 2(d) follows by noting that $H_0^\mu(A) \Rightarrow H_0^\mu(A|B)$ if $\mathbf{C}_{AB}\ \boldsymbol{\mu} = \mathbf{0}$ and the result follows from 2(c) and 2(a). □

REMARK 5.2 *In a linear model without interaction (i.e. where the main effects are well defined), the hypotheses of no parametric main effect* A *or* B*, respectively are equivalent to the parametric hypotheses of no (linear) main effect* A *or* B*, respectively. For a further discussion of nonparametric hypotheses, see Akritas & Arnold (1994), Akritas, Arnold & Brunner (1997), Brunner, Puri & Sun (1995) and Brunner & Puri (1996).*

Statistics for the cross-classified model. Let $\widehat{\mathbf{F}}(x) = (\widehat{F}_{11}(x), \ldots, \widehat{F}_{ab}(x))'$ 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 $\widetilde{\phi}_{i\cdot\cdot} = b^{-1} \sum_{j=1}^{b} \overline{\phi}_{ij\cdot}$, $i = 1, \ldots, a$, denote the unweighted means of the cell means $\overline{\phi}_{ij\cdot} = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} \phi_{ijk}$ where $\phi_{ijk} = J[\frac{1}{N}(R_{ijk} - \frac{1}{2})]$ and 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 $\widehat{\mathbf{p}}(J) = \int J[\widehat{H}]d\widehat{\mathbf{F}} = (\overline{\phi}_{11\cdot}, \ldots, \overline{\phi}_{ab\cdot})'$ under the hypothesis $H_0^F : \mathbf{CF} = \mathbf{0}$ using the contrast matrices $\mathbf{C}_A$, $\mathbf{C}_B$, $\mathbf{C}_{AB}$, $\mathbf{C}_{A|B}$, and $\mathbf{C}_{B|A}$. Let

$$\widehat{\sigma}_{ij}^2(J) = \frac{1}{n_{ij}-1}\sum_{k=1}^{n_{ij}}(\phi_{ijk} - \overline{\phi}_{ij\cdot})^2, \qquad \widehat{\mathbf{V}}_N = N\bigoplus_{i=1}^a\bigoplus_{j=1}^b \frac{\widehat{\sigma}_{ij}^2(J)}{n_{ij}}, \tag{5.5}$$
$$\widehat{\tau}_i^2(J) = \frac{1}{b^2}\sum_{j=1}^b \frac{\widehat{\sigma}_{ij}^2(J)}{n_{ij}}, \qquad \widehat{\mathbf{\Sigma}}_a = \bigoplus_{i=1}^a \widehat{\tau}_i^2(J).$$

Let $\widehat{\mathbf{W}}_a = N^{-1}\widehat{\mathbf{\Sigma}}_a^{-1}(\mathbf{I}_a - \mathbf{J}_a\widehat{\mathbf{\Sigma}}_a^{-1}/\mathbf{1}_a'\widehat{\mathbf{\Sigma}}_a^{-1}\mathbf{1}_a)$ and note that $\widehat{\mathbf{W}}_a$ is a generalized inverse of $\mathbf{C}_A\widehat{\mathbf{V}}_N\mathbf{C}_A' = N\mathbf{P}_a\widehat{\mathbf{\Sigma}}_a\mathbf{P}_a$ and that $\mathbf{P}_a\widehat{\mathbf{W}}_a\mathbf{P}_a = \widehat{\mathbf{W}}_a$. Then, under $H_0^F(A)$, it follows from Theorem 3.4 that the quadratic form

$$\begin{aligned} Q_N(\mathbf{C}_A) &= N\widehat{\mathbf{p}}'(J)\mathbf{C}_A'(\mathbf{C}_A\widehat{\mathbf{V}}_N\mathbf{C}_A')^-\mathbf{C}_A\widehat{\mathbf{p}}(J) = N\widehat{\mathbf{p}}'(J)[\widehat{\mathbf{W}}_a \otimes \tfrac{1}{b}\mathbf{J}_b]\widehat{\mathbf{p}}(J) \\ &= \sum_{i=1}^a \frac{1}{\widehat{\tau}_i^2(J)}\left(\widetilde{\phi}_{i\cdot\cdot} - \frac{1}{\sum_{r=1}^a(1/\widehat{\tau}_r^2(J))}\sum_{r=1}^a\frac{\widetilde{\phi}_{r\cdot\cdot}}{\widehat{\tau}_r^2(J)}\right)^2, \end{aligned} \tag{5.6}$$

has asymptotically a central χ_f^2-distribution with $f = a - 1$.

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

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

is also derived from Theorem 3.4 and $Q_N(\mathbf{C}_{AB})$ has asymptotically a central χ_f^2-distribution with $f = (a-1)\times(b-1)$ under $H_0^F(AB)$.

Finally let $N_i = \sum_{j=1}^b n_{ij}$ denote the total sample size within level i of factor A and let $S_i^2(J) = \sum_{j=1}^b\sum_{k=1}^{n_{ij}}(\phi_{ijk} - \overline{\phi}_{ij\cdot})^2$. Then $\widehat{\sigma}_i^2(J) = (N_i - b)^{-1}S_i^2(J)$ is a consistent estimator for $\sigma_i^2(J) = \sigma_{ij}^2(J) = Var\,(J[H(X_{ij1})])$, $j = 1, \ldots, b$. (Note that $\sigma_{ij}^2(J) = \sigma_i^2(J)$ for all

$j = 1, \ldots, b$ under $H_0^F(B|A)$). The statistic for testing the hypothesis $H_0^F(B|A)$ of no simple factor B effect,

$$\begin{aligned} Q_N(\mathbf{C}_{B|A}) &= N\widehat{\mathbf{p}}'(J)\mathbf{C}'_{B|A}(\mathbf{C}_{B|A}\widehat{\mathbf{V}}_N\mathbf{C}'_{B|A})^{-}\mathbf{C}_{B|A}\widehat{\mathbf{p}}(J) \\ &= \sum_{i=1}^{a} \frac{N_i - b}{S_i^2(J)} \sum_{j=1}^{b} n_{ij}\left(\overline{\phi}_{ij\cdot} - \overline{\phi}_{i\cdot\cdot}\right)^2 \ , \quad \overline{\phi}_{i\cdot\cdot} = \frac{1}{N_i}\sum_{s=1}^{b}\sum_{k=1}^{n_{is}} \phi_{isk} \end{aligned} \qquad (5.7)$$

is derived in the same way and under $H_0^F(B|A)$, it follows that $Q(\mathbf{C}_{B|A})$ has asymptotically a central χ_f^2-distribution with $f = a(b-1)$.

Because rows and columns are interchangeable in this design, the quadratic forms $Q_N(\mathbf{C}_B)$ for testing $H_0^F(B)$ and $Q_N(\mathbf{C}_{A|B})$ for testing $H_0^F(A|B)$ are obtained from $Q_N(\mathbf{C}_A)$ and $Q_N(\mathbf{C}_{B|A})$ respectively, by interchanging rows and columns.

Nested model. The test statistics for the two-fold nested model with fixed factors are briefly given below in order to consider the two-way layout completely for models with fixed factors. Analogous results for Wilcoxon scores have been derived by Akritas, Arnold & Brunner (1997).

Let $X_{ijk} \sim F_{ij}(x) = \frac{1}{2}[F_{ij}^+ + F_{ij}^-]$, $i = 1, \ldots, a$, $j = 1, \ldots, b_i$, $k = 1, \ldots, n_{ij}$ be independent random variables and let $\mathbf{F} = (F_{11}, \ldots, F_{ab})'$ denote the vector of the distribution functions F_{ij} in the sub-categories $j = 1, \ldots, b_i$ within the categories $i = 1, \ldots, a$. The hypothesis of no sub-category effect in the nonparametric setup means that F_{ij} does not depend on j, i.e. $F_{i1} = \cdots = F_{ib_i}$. This can be written equivalently as $H_0^F(B(A)) : \mathbf{C}_{B(A)}\mathbf{F} = \mathbf{0}$ where

$$\mathbf{C}_{B(A)} = \bigoplus_{i=1}^{a} \mathbf{P}_{b_i}, \text{ and } \mathbf{P}_{b_i} = \mathbf{I}_{b_i} - \frac{1}{b_i}\mathbf{J}_{b_i}.$$

For testing the nonparametric hypothesis of no category effect, it may be appropriate to weight the distribution functions F_{ij} in the sub-categories by some weights q_{ij} which have to be chosen by the experimenter according to the meaning of the different sub-categories. Let $\mathbf{q}_i = (q_{i1}, \ldots, q_{ib_i})'$ denote the vector of the weights in category $i = 1, \ldots, a$ where $\sum_{j=1}^{b_i} q_{ij} = 1$. Then we consider the weighted means $\widetilde{F}_{i\cdot}^{q} = \sum_{j=1}^{b_i} q_{ij} F_{ij}$ and express the hypothesis of no category effect as $H_0^F(A) : \mathbf{P}_a \mathbf{Q}\mathbf{F} = 0$ where $\mathbf{Q} = \bigoplus_{i=1}^{a} \mathbf{q}_i'$.

Test statistics for these nonparametric hypotheses are derived by arguments similar to those in the previous section. Thus for testing $H_0^F(A)$ and $H_0^F(B(A))$, the test statistics are based on $\mathbf{P}_a \mathbf{Q} \widehat{\mathbf{p}}(J) = \mathbf{P}_a \mathbf{Q} \int J(\widehat{H}) d\widehat{\mathbf{F}}$, and $\mathbf{C}_{B(A)} \widehat{\mathbf{p}}(J) = \mathbf{C}_{B(A)} \int J(\widehat{H}) d\widehat{\mathbf{F}}$. The asymptotic distribution of these vectors under the hypotheses are obtained from Theorem 3.4.

Let $S_i^2(J) = \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} (\phi_{ijk} - \overline{\phi}_{ij\cdot})^2$, $N_B = \sum_{i=1}^{a} b_i$, $N = \sum_{i=1}^{a} \sum_{j=1}^{b_i} n_{ij}$ and $N_i = \sum_{j=1}^{b_i} n_{ij}$. Note that $\sigma_{ij}^2(J) = \sigma_i^2(J)$ for all j under $H_0(B(A))$. Then, under $H_0^F(B(A))$, the quadratic form

$$Q_N(\mathbf{C}_{B(A)}) = \sum_{i=1}^{a} \frac{N_i - b_i}{S_i^2(J)} \sum_{j=1}^{b_i} n_{ij} \left(\overline{\phi}_{ij\cdot} - \overline{\phi}_{i\cdot\cdot} \right)^2 , \quad \overline{\phi}_{i\cdot\cdot} = \frac{1}{N_i} \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} \phi_{ijk} \qquad (5.8)$$

has asymptotically a central χ_f^2-distribution with $f = (N_B - a)$. Note that the statistic $Q_N(\mathbf{C}_{B(A)})$ has the RT-property (see Remark 5.1) with respect to a parametric statistic with heteroscedastic errors across the levels of factor A.

Let $\widehat{\tau}_i^2(J) = \sum_{j=1}^{b_i} q_{ij}^2 \widehat{\sigma}_{ij}^2 / n_{ij}$ and $\widetilde{\phi}_{i\cdot\cdot} = \sum_{j=1}^{b_i} q_{ij} \overline{\phi}_{ij\cdot}$. Then, under the null hypothesis $H_0^F(A)$, the quadratic form

$$Q_N(\mathbf{P}_a \mathbf{Q}) = \sum_{i=1}^{a} \frac{1}{\widehat{\tau}_i^2(J)} \left(\widetilde{\phi}_{i\cdot\cdot} - \frac{1}{\sum_{r=1}^{a} (1/\widehat{\tau}_r^2(J))} \sum_{r=1}^{a} \frac{\widetilde{\phi}_{r\cdot\cdot}}{\widehat{\tau}_r^2(J)} \right)^2 ,$$

has a central χ^2_f-distribution with $f = a - 1$ and $Q_N(\mathbf{P}_a\mathbf{Q})$ has the RT-property with respect to a parametric statistic with heteroscedastic errors.

Approximations for small samples. The application of the method described in Section 4 is briefly shown in this paragraph. The hypothesis $H_0^F(A)$ in the cross-classification is equivalently restated as $H_0^F(A) : \mathbf{M}_A\mathbf{F} = \mathbf{0}$ where $\mathbf{M}_A = \mathbf{P}_a \otimes \frac{1}{b}\mathbf{J}_b$ is a projection matrix with constant diagonal elements $m_a = (a-1)/(ab)$. Let $\widetilde{\phi}_{i\cdot\cdot} = b^{-1}\sum_{j=1}^b \overline{\phi}_{ij\cdot}$ and $\widetilde{\phi}_{\cdot\cdot\cdot} = a^{-1}\sum_{i=1}^a \widetilde{\phi}_{i\cdot\cdot}$. Then, under $H_0^F(A)$, the statistic

$$T_N(\mathbf{M}_A) \;=\; \frac{Nab^2}{(a-1)\; tr(\widehat{\mathbf{V}}_N)} \cdot \sum_{i=1}^a \left(\widetilde{\phi}_{i\cdot\cdot} - \widetilde{\phi}_{\cdot\cdot\cdot}\right)^2$$

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 (4.13) and (4.14), respectively by replacing $\mathbf{M}$ with $\mathbf{M}_A$ and $\widehat{\mathbf{V}}_N$ is given in (5.5). Note that under $H_0^F(B|A)$, the variances $\sigma^2_{ij}(J)$ are the same for all j, i.e. $\sigma^2_i(J) = \sigma^2_{ij}(J)$, for all $j = 1, \ldots, b$, $i = 1, \ldots, a$. Thus, $\widehat{\sigma}^2_i(J) = \sum_{j=1}^b (n_{ij} - 1)\widehat{\sigma}^2_{ij}(J)/(N_i - b)$ as used for the derivation of (5.7).

The approximations for testing the other hypotheses are derived in the same way.

Comparison with other statistics. For equal sample sizes $n_{ij} \equiv n$ and under the assumption of continuous distribution functions, Thompson (1991) derived a rank tests for some hypotheses in different two-way layouts. In a model with interaction, the statistic for testing the hypothesis $H_0^F(B|A)$ given by Thompson (1991, p. 415, variance estimator $\widehat{\sigma}^2_3(i)$) is a special case of the statistic $Q_N(\mathbf{C}_{B|A})$ given in (5.7). In a model without interaction, the hypotheses $H_0^F(B)$ and $H_0^F(B|A)$ are equivalent. Thus, the statistic given by Thompson (1991, p. 415, variance estimator $\hat{\sigma}^2$) and the statistic F_N given by Hora & Conover (1984) are special cases of the statistic $Q_N(\mathbf{C}_A)$ given in (5.6) if rows and columns of the design are interchanged. Note that the assumptions of no interaction and of equal sample sizes are crucial for the derivation of the statistics given by Thompson and by Hora & Conover.

Thompson (1991, p. 414, Section 7.2) considered also a nested design where an equal number for the levels of the nested factor as well as an equal number of replications and continuous distribution functions are assumed. Thus the hypothesis and the statistic are formally identical to the case where a simple factor effect is considered in a cross-classified design and the statistic derived by Thompson is a special case of the statistic $Q_N(\mathbf{C}_{B(A)})$ given in (5.8).

Akritas (1990, Section 3, formula (3.2)) considers unequal sample sizes and derives a rank test (using only the Wilcoxon scores) for $H_0^\mu(B|A)$ in a nested linear model assuming continuous distribution functions. The statistic given there is also a special case of the statistic $Q_N(\mathbf{C}_{B(A)})$ given in (5.8).

5.3 The Rank Transform Property of the Statistic $Q_N(\mathbf{C})$

In this subsection, we pay special attention to the RT-technique suggested by Conover & Iman (1976, 1981) which has been criticized by Brunner & Neumann (1986), Blair, Sawilowski & Higgens (1987), Akritas (1990, 1991), Thompson & Ammann (1990) and Thompson (1991) and possibly others. The following considerations shall clarify the question when the method of the RT-technique works and when it fails. Moreover, we wish to distinguish the procedures derived in this paper from the usual RT-statistics and we suggest to replace the expression *RT-technique* by the terminology *RT-property of a rank (score) statistic.*

First we note that in the statistic $\widehat{\mathbf{p}}(J)$, the original observations are replaced by the scores of their (mid)-ranks among all observations. Recall that

$$\overline{\mathbf{Y}}_{\cdot}(J) \;=\; (\overline{Y}_{1\cdot}(J), \ldots, \overline{Y}_{d\cdot}(J))'$$

is the mean vector of the ART $Y_{ij}(J) = J[H(X_{ij})]$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$. It follows from Theorem 3.2 that $\sqrt{N}\mathbf{C}\widehat{\mathbf{p}}(J)$ is asymptotically equivalent to $\sqrt{N}\mathbf{C}\overline{\mathbf{Y}}_{\cdot}(J)$ if $\mathbf{CF} = \mathbf{0}$, i.e. if the hypotheses are formulated in terms of the distribution functions. Some cases where the hypotheses in the linear model are equivalent to the corresponding nonparametric hypotheses are given in Proposition 5.1. Note that most counterexamples for the RT-technique use linear hypotheses $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ such that $\mathbf{CF} \neq \mathbf{0}$ where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)'$ is the vector of the expectations. We are discussing here the problem in a general frame work and not in the special cases of one- or two-way layout problems.

Next, consider the covariance matrix

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

Note that in general the diagonal elements $\sigma_i^2(J) = Var(Y_{i1}(J))$ of $\mathbf{V}_N$ 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)). In particular, some of the diagonal elements in $\mathbf{V}_N$ may be equal under the hypothesis $H_0 : \mathbf{CF} = \mathbf{0}$ and the corresponding estimators given in (3.9) can be pooled to estimate $\mathbf{V}_N$ consistently.

Now let $U_{ij} \sim N(\mu_i(J), \sigma_i^2(J))$, $i = 1, \ldots, d$ be independent normally distributed random variables where $\mu_i(J) = E(Y_{i1}(J))$ and $\sigma_i^2(J) = Var(Y_{i1}(J))$. Let $\overline{\mathbf{U}}. = (\overline{U}_{1\cdot}, \ldots, \overline{U}_{d\cdot})'$ denote the mean vector of the U_{ij}'s. Then, by definition, the statistics $\sqrt{N}\ \overline{\mathbf{U}}.$ and $\sqrt{N}\ \overline{\mathbf{Y}}.(J)$ have asymptotically the same multivariate normal distribution. Furthermore define $\widetilde{\sigma}_i^2(J) = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (U_{ij} - \overline{U}_{i\cdot})^2$ and let $\widetilde{\mathbf{V}}_N$ denote the matrix $\mathbf{V}_N$ with $\sigma_i^2(J)$ replaced by $\widetilde{\sigma}_i^2(J)$. Then $\widetilde{\mathbf{V}}_N$ is consistent for $\mathbf{V}_N$ and from Theorem 3.9, $\widehat{\mathbf{V}}_N$ is consistent for $\mathbf{V}_N$ in the sense that $\widetilde{\mathbf{V}}_N \mathbf{V}_N^{-1} \stackrel{p}{\longrightarrow} 1$ and $\widehat{\mathbf{V}}_N \mathbf{V}_N^{-1} \stackrel{p}{\longrightarrow} 1$, respectively. Note that $\widetilde{\sigma}_i^2(J)$ is derived from $\widehat{\sigma}_i^2(J)$ by replacing the rank scores ϕ_{ij} by the corresponding normally distributed random variables U_{ij}. Thus, the statistic $\widehat{\mathbf{p}}(J)$ is a 'rank transform' of the statistic $\overline{\mathbf{U}}.$ and it follows from the above considerations that under $H_0 : \mathbf{CF} = \mathbf{0}$, the statistics $\sqrt{N}\mathbf{C}\widehat{\mathbf{p}}(J)$ and $\sqrt{N}\mathbf{C}\overline{\mathbf{U}}.$ have asymptotically a multivariate normal distribution $N(\mathbf{0}, \mathbf{C}\mathbf{V}_N\mathbf{C}')$. This property of the rank statistic $\sqrt{N}\mathbf{C}\widehat{\mathbf{p}}(J)$ shall be called *rank transform (RT) property with respect to the normal theory statistic* $\sqrt{N}\mathbf{C}\overline{\mathbf{U}}.$. Note that the expectations $\mu_i(J)$ and the variances $\sigma_i^2(J)$ of the corresponding normally distributed random variables U_{ij} are determined from the ART under $H_0 : \mathbf{CF} = \mathbf{0}$.

For testing $H_0 : \mathbf{CF} = \mathbf{0}$, there are two useful ways to define a statistic from $\sqrt{N}\mathbf{C}\widehat{\mathbf{p}}(J)$. One possibility is to define the quadratic form $Q_N(\mathbf{C}) = N\widehat{\mathbf{p}}(J)'(\mathbf{C}\widehat{\mathbf{V}}_N\mathbf{C}')^{-1}\widehat{\mathbf{p}}(J)$ which is the rank version of the Wald-type statistic. It follows from Theorem 3.4 that under $H_0 : \mathbf{CF} = \mathbf{0}$, the quadratic form $Q_N(\mathbf{C})$ has asymptotically a central χ_f^2-distribution with $f = rank(\mathbf{C})$ degrees of freedom. The other possibility is to define the quadratic form $T_N(\mathbf{M}) = N\widehat{\mathbf{p}}(J)'\mathbf{M}\widehat{\mathbf{p}}(J)'$ which has the RT-property with respect to the normal theory statistic $N\overline{\mathbf{U}}_\cdot'\mathbf{M}\overline{\mathbf{U}}_\cdot$ where $\mathbf{M} = \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\mathbf{C}$ is a projection matrix which is taken from the balanced ANOVA models. Under $H_0 : \mathbf{CF} = \mathbf{0}$, the asymptotic distribution of $N\overline{\mathbf{U}}_\cdot'\mathbf{M}\overline{\mathbf{U}}_\cdot$ is a weighted χ^2-distribution because the variances $\sigma_i^2(J)$ are not necessarily equal, in general. For small sample sizes, this weighted χ^2-distribution is approximated by a scaled F-distribution with estimated degrees of freedom (see Section 4).

6 Numerical Example: Two-Way Layout for Counting Data

In this Section, the statistics given in Section 5 and the approximations given in Section 4 are applied to a data set with counting data. The authors are grateful to Dr. Beuscher (Schaper & Brümmer, Inc., Salzgitter, Germany) for making available the data.

In order to detect adverse reactions concerning the fertility, three groups of female Wistar rats were treated with two different dosages (factor A) of a drug (groups 2 and 3) and with a placebo (group 1). 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.

About here insert Table 1

The intention of the experiment is to detect adverse reactions of the drug. Therefore it is not only necessary to investigate a change of the location but also a change of the dispersion under the treatment is of importance. The Wilcoxon-scores $a_{ijk}^W = \frac{1}{N}(R_{ijk} - \frac{1}{2})$ are known to be sensitive to location alternatives while the Mood-scores $a_{ijk}^M = (a_{ijk}^W - \frac{1}{2})^2$ are known to be sensitive to changes in dispersion. Note that a statistic based on the Wilcoxon-scores may be less sensitive to location alternatives in case of large dispersion effects while the Mood-scores may be less sensitive to dispersion alternatives in case of large location effects. Thus, either the Wilcoxon-scores or the Mood-scores should detect any meaningful effect in such a trial and both scores should be used when adverse reactions are analysed.

The results of an analysis using Wilcoxon-scores as well as Mood-scores are displayed in Tables 2 and 3. Table 2 contains the score means $\overline{a}_{ij\cdot}^W$ and $\overline{a}_{ij\cdot}^M$ together with the unweighted score means $\widetilde{a}_{i\cdot\cdot}^W$ and $\widetilde{a}_{i\cdot\cdot}^M$ for the treatment levels and $\widetilde{a}_{\cdot j\cdot}^W$ and $\widetilde{a}_{\cdot j\cdot}^M$ for the two years.

About here insert Table 2

About here insert Table 3

For the Mood-scores it turns out that neither an iteraction ($p = 0.260$), nor an effect of the year ($p = 0.254$), nor an effect of the treatment ($p = 0.244$) can be detected. This means that regarding dispersion, no effect can be detected in the experiment.

For the Wilcoxon-scores, the large p-value 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 ($p = 0.263$).

7 Conclusions and Discussion

In this paper, we have generalized the ideas given in Akritas, Arnold & Brunner (1997) to the case of score functions with a bounded second derivative and moreover, we did not assume that the relative sample sizes converge to some constants. A simple method of proving the asymptotic results is given which is based on elementary counting arguments and on some moment inequalities for empirical processes (see Lemma A.2). The question of unbounded score functions, e.g. normal scores, is briefly discussed. The details are ommited because we did not intend to overburden the paper with too many technical details. The asymptotic distribution of the proposed rank score statistics is considered under the nonparametric hypotheses based on the distribution functions. The general results given in Section 3 are applied to various experimental designs in Section 5. Rank tests for some special hypotheses in these designs are given in the literature under different more or less restrictive assumptions. In some cases, the results are derived only for Wilcoxon-scores and in other cases the continuity of the parent distribution functions or an equal number of replications or factor levels is assumed. In particular, either models without interactions are considered in the literature or only statistics for testing the simple factor effects are given. From our general results provided in Section 3, however, rank statistics for all nonparametric hypotheses in two- or higher-way layouts can be derived without assuming the continuity of the parent distribution functions. Thus, the proposed statistics are applicable to analyze data for balanced and unbalanced designs, data with continuous distribution functions or data with ties. This means that especially pure ordinal data (psychological test data, quality scales in ecology or medicine) can be analyzed by the proposed statistics. Also the extreme case of data having a Bernoulli distribution is included in this setup.

All proposed statistics are invariant under strictly monotone transformations of the data and moreover, they are robust against outliers. The set of alternatives for which the proposed tests are consistent is characterized by $\mathbf{Cp} \neq \mathbf{0}$. The quadratic forms $Q_N(\mathbf{C})$ given in Theorem 3.4 can easily be computed by the SAS procedure PROC MIXED applied to the rank data when the option CHISQ is used in the model statement and the structure of the covariance matrix $\mathbf{V}_N$ is taken care of in the REPEATED statement.

A Appendix

The following Lemma is needed to handle the ties.

LEMMA A.1 *Let $X_{ij} \sim F_i$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$ be independent random variables, let $F_i = \frac{1}{2}[F_i^+ + F_i^-]$ as given in (2.2) and let $c(u) = \frac{1}{2}\left[c^+(u) + c^-(u)\right]$ as given in (2.4). Then*

1. $E\left[c(x - X_{ij}) - F_i(x)\right] = 0,$
2. $\int F_i dF_i = \frac{1}{2},$

3. $E\left[c(X_{ij}-X_{is})-F_i(X_{ij})\right]=0.$

PROOF: (1) By definition,

$$\begin{aligned} E\left[c(x-X_{ij})-F_i(x)\right] &= P(X_{ij}<x)+\tfrac{1}{2}P(X_{ij}=x)-F_i(x) \\ &= F_i^-(x)+\tfrac{1}{2}\left[F_i^+(x)-F_i^-(x)\right]-F_i(x) \\ &= \tfrac{1}{2}\left[F_i^+(x)+F_i^-(x)\right]-F_i(x)=0. \end{aligned}$$

(2) The result follows using integration by parts (see Hewitt & Stromberg (1969), p. 419).

(3) follows in the same way as (1) using (2). □

To prove the asymptotic results, we first give some moment inequalities for empirical processes which are needed in the body of the paper.

LEMMA A.2 *Let $X_{ij}\sim F_i$, $i=1,\ldots,d$, $j=1,\ldots,n_i$ be independent random variables, and let F_i be as given in (2.2). Let $H(x)=N^{-1}\sum_{i=1}^d n_iF_i(x)$ and let $\widehat{F}_i(x)$ and $\widehat{H}(x)$ be as defined in (2.4) and (2.5) respectively. Then, under the assumption (c2) as stated in Section 3,*

1. $E\left[\widehat{F}_i(x)-F_i(x)\right]^2\le 1/n_i\ ,\quad i=1,\ldots,d\ ,$
2. $E\left[\widehat{F}_i(X_{rj})-F_i(X_{rj})\right]^2\le 1/n_i\ ,\quad i,r=1,\ldots,d\ ,\quad j=1,\ldots,n_i\ ,$
3. $E\left[\widehat{H}(x)-H(x)\right]^2\le 1/N\ ,$
4. $E\left[\widehat{H}(X_{ij})-H(X_{ij})\right]^2\le 1/N\ ,\quad i=1,\ldots,d\ ,\quad j=1,\ldots,n_i\ ,$

5. $E\left(J\left[\widehat{H}(x)\right] - J\left[H(x)\right]\right)^2 \leq \frac{1}{N}\|J'\|_\infty^2$,

6. $E\left(J\left[\widehat{H}(X_{ij})\right] - J\left[H(X_{ij})\right]\right)^2 \leq \frac{1}{N}\|J'\|_\infty^2$, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$,

7. $E\left[\widehat{H}(x) - H(x)\right]^4 = O\left(N^{-2}\right)$,

8. $E\left[\widehat{H}(X_{ij}) - H(X_{ij})\right]^4 = O\left(N^{-2}\right)$, $\quad i = 1, \ldots, d$, $\quad j = 1, \ldots, n_i$.

PROOF: 1. By independence, and by Lemma A.1,

$$\begin{aligned} E\left[\widehat{F}_i(x) - F_i(x)\right]^2 &= \frac{1}{n_i^2}\sum_{s=1}^{n_i}\sum_{t=1}^{n_i} E\left(\left[c(x - X_{is}) - F_i(x)\right]\left[c(x - X_{it}) - F_i(x)\right]\right) \\ &= \frac{1}{n_i^2}\sum_{s=1}^{n_i} E\left[c(x - X_{is}) - F_i(x)\right]^2 \leq \frac{1}{n_i} . \end{aligned}$$

2. The proof of (2) follows by conditioning on X_{rj} and proceeding as in (1).

$$\begin{aligned} E\left[\widehat{F}_i(X_{rj}) - F_i(X_{rj})\right]^2 &= \frac{1}{n_i^2}\sum_{s=1}^{n_i}\sum_{t=1}^{n_i} E\left(\left[c(X_{rj} - X_{is}) - F_i(X_{rj})\right]\left[c(X_{rj} - X_{it}) - F_i(X_{rj})\right]\right) \\ &= \frac{1}{n_i^2}\sum_{s=1}^{n_i} E\left[c(X_{rj} - X_{is}) - F_i(X_{rj})\right]^2 \leq \frac{1}{n_i} . \end{aligned}$$

3. It follows by independence and using (1.) that

$$\begin{aligned} E\left[\widehat{H}(x) - H(x)\right]^2 &= \frac{1}{N^2}\sum_{i=1}^{d}\sum_{r=1}^{d} E\left[n_i\left(\widehat{F}_i(x) - F_i(x)\right) n_r\left(\widehat{F}_r(x) - F_r(x)\right)\right] \\ &= \frac{1}{N^2}\sum_{i=1}^{d} n_i^2 E\left[\widehat{F}_i(x) - F_i(x)\right]^2 \leq \frac{1}{N^2}\sum_{i=1}^{d} n_i^2\frac{1}{n_i} = \frac{1}{N} . \end{aligned}$$

4. The result follows in the same way as (3.) using (2.).

5. Using the mean value Theorem, it follows that $\left(J[\widehat{H}] - J[H]\right)^2 \leq \|J'\|_\infty^2 \left[\widehat{H} - H\right]^2$ and the result follows using (3.).

6. The proof follows in the same way as for (5.) using (4.) instead of (3.).

7. To prove (7.), we collapse the two indices i and j into one index, r say, and the random variables $X_{ij},\ i = 1, \dots, d,\ j = 1, \dots, n_i$ are renumbered to $X_r,\ r = 1, \dots, N$ where N is the total number of observations and $X_r \sim G_r(x)$. Note that $G_1 = \cdots = G_{n_1} \equiv F_1$. Consider,

$$E\left[\widehat{H}(x) - H(x)\right]^4 \;=\; \frac{1}{N^4} \sum_{r=1}^{N} \sum_{s=1}^{N} \sum_{t=1}^{N} \sum_{u=1}^{N} E\left[\varphi(X_r)\varphi(X_s)\varphi(X_t)\varphi(X_u)\right]$$

where $\varphi(X_r) = c(x - X_r) - G_r(x)$. If one of the indices r, s, t, u is different from all the other indices, then it follows by independence that the expectation in the right hand side is 0. The number of cases where not one index is different from all the others is of order N^2. Thus,

$$E\left[\widehat{H}(x) - H(x)\right]^4 \;=\; O\left(\frac{1}{N^4} \cdot N^2\right) = O\left(\frac{1}{N^2}\right) \; .$$

8. The proof follows in a similar way as for (7.) by the arguments used to prove (2.). □

To prove the results for contiguous alternatives, we need the following Lemma.

LEMMA A.3 *Under the assumptions of Theorem 3.5,*

1. $\left(J\left[H(x)\right] - J\left[H^*(x)\right]\right)^2 \le \frac{1}{N} \|J'\|_\infty^2,$

2. $E\left(J\left[\widehat{H}(x)\right] - J\left[H(x)\right]\right)^2 \le \frac{4}{N} \|J'\|_\infty^2,$

3. $E\left(J\left[H(X_{ij})\right] - J\left[H^*(X_{ij})\right]\right)^2 \le \frac{1}{N} \|J'\|_\infty^2 \; .$

PROOF: To prove (1), we note that

$$|H(x) - H^*(x)| \le \frac{1}{\sqrt{N}} \sum_{i=1}^{d} \frac{n_i}{N} |F_i(x) - K_i(x)| \le \frac{1}{\sqrt{N}}$$

and that

$$J\left[H(x)\right]-J\left[H^*(x)\right]=\int_{H^*(x)}^{H(x)}dJ(s)=J'(\theta_1)\left[H(x)-H^*(x)\right]$$

where θ_1 is between $H^*(x)$ and $H(x)$ and the result follows.

Statement (2) follows from

$$\begin{aligned}J\left[\widehat{H}(x)\right]-J\left[H(x)\right] &= \int_{H(x)}^{\widehat{H}(x)}dJ(s)\\ &= J'(\theta_2)\left[\widehat{H}(x)-H(x)\right]=J'(\theta_2)\left[\widehat{H}(x)-H^*(x)+H^*(x)-H(x)\right]\end{aligned}$$

where θ_2 is between $H(x)$ and $\widehat{H}(x)$. Thus,

$$\left(J\left[\widehat{H}(x)\right]-J\left[H(x)\right]\right)^2 \leq \|J'\|_\infty^2\left(2\left[\widehat{H}(x)-H^*(x)\right]^2+2\left[H^*(x)-H(x)\right]^2\right) .$$

Since $E\left(\widehat{H}(x)\right)=H^*(x)$, it follows from Lemma A.2, (3) that $E\left(\widehat{H}(x)-H^*(x)\right)^2\leq 1/N$. Together with statement (1), it follows that

$$E\left(J\left[\widehat{H}(x)\right]-J\left[H(x)\right]\right)^2\leq 2\|J'\|_\infty^2\left(\frac{1}{N}+\frac{1}{N}\right)=\frac{4}{N}\|J'\|_\infty^2 .$$

Statement (3) follows analogously using (1). □

To prove Theorem 3.2, we need the following Lemma.

LEMMA A.4 *Let*

$$\begin{aligned}B_1 &= \int J'[H]\left[\widehat{H}-H\right]d(\widehat{F}_i-F_i),\\ B_2 &= \frac{1}{2}\int J''[\widehat{\theta}_N]\left[\widehat{H}-H\right]^2 d(\widehat{F}_i-F_i).\end{aligned}$$

Then, under the assumptions $(a),(b)$ *and* $(c2)$,

1. $E(NB_1^2) = NE\left(\int J'[H]\left[\widehat{H} - H\right] d(\widehat{F}_i - F_i)\right)^2 \to 0,$

2. $E(NB_2^2) = NE\left(\frac{1}{2}\int J''[\widehat{\theta}_N]\left[\widehat{H} - H\right]^2 d(\widehat{F}_i - F_i)\right)^2 \to 0.$

PROOF: To show that $E(NB_1^2) \to 0$, note that

$$B_1 = \frac{1}{Nn_i}\sum_{r=1}^{d}\sum_{s=1}^{n_r}\sum_{j=1}^{n_i}\left[\varphi_1(X_{ij}, X_{rs}) - \varphi_2(X_{rs})\right],$$

where

$$\begin{aligned}\varphi_1(X_{ij}, X_{rs}) &= J'\left[H(X_{ij})\right]\left[c(X_{ij} - X_{rs}) - F_r(X_{ij})\right] \ , \\ \varphi_2(X_{rs}) &= \int J'[H(x)]\left[c(x - X_{rs}) - F_r(x)\right] dF_i(x) \ .\end{aligned}$$

Then,

$$\begin{aligned}&NE(B_1^2)\\ &= \frac{N}{N^2 n_i^2}\sum_{r=1}^{d}\sum_{t=1}^{d}\sum_{s=1}^{n_r}\sum_{u=1}^{n_t}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i} E\left(\left[\varphi_1(X_{ij}, X_{rs}) - \varphi_2(X_{rs})\right]\left[\varphi_1(X_{ik}, X_{tu}) - \varphi_2(X_{tu})\right]\right)\\ &\ll \frac{1}{Nn_i^2}\sum_{r=1}^{d}\sum_{s=1}^{n_r}\sum_{j=1}^{n_i} E\left(\left[\varphi_1(X_{ij}, X_{rs}) - \varphi_2(X_{rs})\right]^2\right) \ll \frac{1}{n_i}\|J'\|_\infty^2\end{aligned}$$

by conditioning arguments (similar to those used in the proof of Lemma A.2,8.) where the Vinogradov symbol $' \ll '$ is used instead of the $O(\cdot)$-notation.

The second statement $E(NB_2^2) \to 0$ is easily proved by decomposing, $B_2 = B_{21} + B_{22}$ where

$$B_{21} = \frac{1}{2}\int J''(\widehat{\theta}_N)\left[\widehat{H} - H\right]^2 d\widehat{F}_i \quad \text{and} \quad B_{22} = -\frac{1}{2}\int J''(\widehat{\theta}_N)\left[\widehat{H} - H\right]^2 dF_i \ .$$

The two terms are estimated separately.

$$E(NB_{21}^2) \ \le \ \frac{N\|J''\|_\infty^2}{4n_i}\sum_{j=1}^{n_i} E\left[\widehat{H}(X_{ij}) - H(X_{ij})\right]^4 = O\left(\frac{\|J''\|_\infty^2}{N}\right)$$

by Lemma A.2, 8.

$$E(NB_{22}^2) \leq \frac{N\|J''\|_\infty^2}{4} E\left[\widehat{H}(x) - H(x)\right]^4 = O\left(\frac{\|J''\|_\infty^2}{N}\right)$$

by Lemma A.2, 7. This completes the proof. □

Acknowledgments. The authors would like to express their sincere gratitude to the referees for making extremely valuable suggestions which resulted in significant improvements in the presentation of the results. They are also grateful to Professor Manfred Denker for many helpful discussions and valuable comments during the preparation of the paper. Part of this paper was prepared during the first author's visit to the Department of Mathematics, Indiana University. He would like to thank the Department of Mathematics for its hospitality and support. The paper was revised during the second author's visit to the Center for Mathematics and Applications, Australian National University. He would like to thank Professor Peter Hall for providing an extremely congenial and stimulating environment for research during his stay at the Center for Mathematics and its Applications, Australian National University, Canberra in the summer of 1999.

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.

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.

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 Neumann, N. (1986). Rank tests in 2×2 designs. *Statist. Neerlandica* **40** 251–271.

Brunner, E. and Puri, M. L. (1996). Nonparametric methods in design and analysis of experiments. *Handbook of Statistics* Vol. **13**. (Eds.: S. Ghosh and C.R. Rao), 631–703.

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). *The American Statistician* **35** 124–133.

Denker, M. and Rösler, U. (1985). Some contributions to Chernoff–Savage theorems. *Statist. Dec.* **3** 49–75.

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

Hajek, J. (1968). Asymptotic normality of simple linear ranks statistics under alternatives. *Ann. Math. Statist.* **39** 325–346.

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

Hewitt, E. and Stromberg, K. (1969). *Real and Abstract Analysis.* Springer, Berlin, Heidelberg, New York, 2^{nd} edition.

Hora, S. C. and Conover, W. J. (1984). The F–statistic in the two-way layout with rank score transformed data. *J. Amer. Statist. Assoc.* **79** 668–673.

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

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.

LANG, S. (1993). *Real and Functional Analysis.* Springer, Berlin, Heidelberg, New York, 3^{rd} edition.

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

LEMMER, H. H. AND STOKER, D. J. (1967). A distribution-free analysis of variance for the two-way classification. *S. African Statist. Journal* **1** 67–74.

MANN, H. B. AND WHITNEY, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.* **18** 50–60.

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.

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.

Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and Its Applications.* Wiley, New York.

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

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. (1990). Efficiencies of interblock rank statistics for repeated measures designs. *J. Amer. Statist. Assoc.* **85** 519–528.

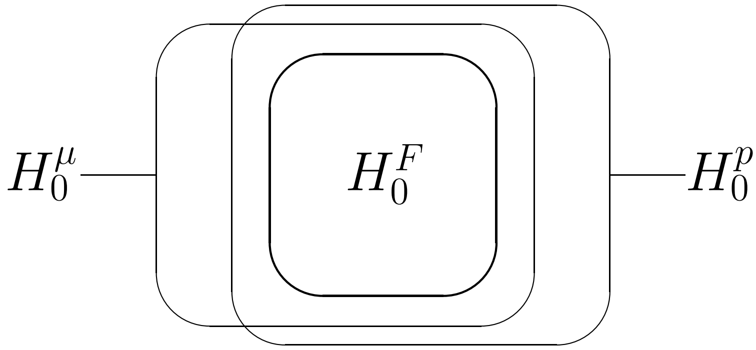

Figure 1. Implication of the hypotheses H_0^μ, H_0^p and H_0^F.

Table 1. *Number of corpora lutea from Wistar rats in a fertility trial to detect adverse reactions of a drug.*

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

Table 2. *Rank-score means (Wilcoxon-scores and Mood-scores) for the number of the corpora lutea of the Wistar rats in the fertility trial. The results for the Wilcoxon-scores* $a_{ijk}^W = \frac{1}{N}(R_{ijk} - \frac{1}{2})$ *are displayed in the left hand part of the table while the results for the Mood-scores* $a_{ijk}^M = (a_{ijk}^W - \frac{1}{2})^2$ *are displayed in the right hand part of the table.*

	Wilcoxon-Scores			Mood-Scores		
Group	Year 1	Year 2	$\widetilde{a}_{i\cdot\cdot}^W$	Year 1	Year 2	$\widetilde{a}_{i\cdot\cdot}^M$
Placebo	0.459	0.467	0.463	0.044	0.076	0.060
Dosage 1	0.501	0.280	0.390	0.070	0.123	0.097
Dosage 2	0.649	0.613	0.631	0.100	0.080	0.090
unweighted means	0.536	0.453		0.071	0.093	

Table 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 Wilcoxon-scores are given in the left part and the results obtained by the Mood-scores are given in the right part of the table.*

	Wilcoxon-Scores				Mood-Scores			
	asymptotic χ^2-distribution		modified Box-approximation		asymptotic χ^2-distribution		modified Box-approximation	
Hypothesis	$Q_N(\mathbf{C})$	p-value	$T_N(\mathbf{M})$	p-value	$Q_N(\mathbf{C})$	p-value	$T_N(\mathbf{M})$	p-value
$H_0^F(A)$	6.91	0.032	3.80	0.031	3.54	0.170	1.45	0.244
$H_0^F(B)$	1.28	0.257	1.28	0.263	1.33	0.248	1.33	0.254
$H_0^F(AB)$	1.78	0.412	0.92	0.403	2.35	0.308	1.39	0.260