Nonparametric Models for ANOVA and ANCOVA: A Review

Michael G. Akritas[a] and Edgar Brunner[b*]

[a]Department of Statistics, Penn State University, 326 Thomas Bldg., University Park, PA 16802, USA, email: mga@stat.psu.edu

[b]Department of Medical Statistics, University of Göttingen, Humboldtallee 32, D-37073 Göttingen, Germany, email: brunner@ams.med.uni-goettingen.de

Recently developed nonparametric models, hypotheses and test statistics for ANOVA and ANCOVA designs with independent and dependent ordinal data (continuous or not) are reviewed and discussed.

*Supported in part by DFG/Br-655/12

1. INTRODUCTION

Univariate and multivariate analysis of variance (ANOVA and MANOVA), as well as analysis of covariance (ANCOVA) form cornerstones of applied statistics. They are used in many disciplines (biological, medical, social or psychological studies) for addressing complex questions about the effects of factors and covariates and about the interaction between them. The data from such studies can be either independent or dependent (longitudinal studies). In two critical ways, the logic of the classical approach requires that the response variable has units of measurement that are equal (or have comparable meaning) across its entire range. First, the parameters of the model that constitute the effects being estimated or tested are defined in terms of mean values on that metric. Second, the desirable statistical properties of least squares estimators depend on assumptions about the distribution of the residuals (normality, homogeneity, and independence), and those residuals are defined by differences in this metric.

Unfortunately, measurements in many applied sciences rarely justify the assumption of an equal interval scale. For example, measures in the social sciences generally have arbitrary metrics because the theoretical processes being studied are specified without regard to any particular scale of measurement. As a consequence, applied scientists increasingly turn to parametric and semi-parametric alternatives to classical least squares statistics, such as generalized linear models, generalized additive models, and frailty models. These techniques avoid those assumptions of the classical model that are overly inconsistent with the nature of the dependent variable. For instance, ordinal regression with a logit link function (Winship & Mare [44]) provides a model that is plausible for Likert-type rating scales, improving on the classical model by assuming a discrete distribution for the response variable in place of a continuous error distribution and by using the logit link to avoid the possibility of fitted values falling outside the range of the response scale.

It must not be forgotten, however, that the parametric and semi-parametric alternatives also depend on assumptions, and those assumptions may or may not be correct for any given application. In the case of the ordinal model with a logit link, the logistic function is assumed to characterize the relationship between the linear form of the model and the response probabilities, and relationships to the independent variables are assumed to be homogeneous across response categories and consistent with the response thresholds corresponding to the logit link. There are several examples in the published literature where different models have produced conflicting results.

In this article we review recent developments in statistical modelling that are completely nonparametric (also nonlinear and nonadditive). Because this approach defines the effects of interest nonparametrically, in a manner general across possible transformation of the response scale, it is well suited to the ambiguity of scaling typical in many applied sciences. It further avoids assumptions about the form and consistency of distributions of the response variable, which applied scientists are rarely in a position to specify in advance.

The basic nonparametric models for ANOVA and ANCOVA with independent data are reviewed in Section 2. Section 3 presents the extension of the ANOVA model to repeated measures designs. Related work is discussed in Section 5, while other extensions are briefly outlined in Section 4.

2. THE BASIC NONPARAMETRIC MODELS

In this section we present the nonparametric models for two-way ANOVA and one-way ANCOVA. These are the simplest designs where all features of the nonparametric modelling can be appreciated. Following the idea of Ruymgaart [37], all distribution functions (so also conditional ones) in this paper are taken as the average of their left- and right-continuous version, $F(x) = \frac{1}{2}[F^{+}(x) + F^{-}(x)]$. This permits a unified presentation of the model and test procedures for all types of data. For the two-way ANOVA design, the observations will be denoted by Y_{ijk}, $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, n_{ij}$. For the one-way ANCOVA design, the observations will be denoted by (Y_{ij}, X_{ij}), $i = 1, \ldots, I$, $j = 1, \ldots, n_i$, where X_{ij} is the covariate, Y_{ij} is the response, i enumerates the factor levels and j the observations within each factor level.

The nonparametric model for the two-way ANOVA design specifies only that

$$Y_{ijk} \sim F_{ij} \ , \tag{2.1}$$

for some distribution function F_{ij} (Akritas & Arnold [4], Akritas, Arnold & Brunner [6]).

The nonparametric model for the one-way ANCOVA design specifies only

$$Y_{ij}|X_{ij} = x \sim F_{ix} \ , \tag{2.2}$$

i.e., that conditionally on $X_{ij} = x$, Y_{ij} has distribution function that depends on i and x (Akritas, Arnold & Du [7]). Note that models (2.1) and (2.2) do not specify how the response distribution changes when the levels, or covariate value changes. Thus they are completely nonparametic (also nonlinear and nonadditive).

For the two-way ANOVA design set $\overline{F}_{i\cdot}(y) = J^{-1}\sum_j F_{ij}(y)$, and $\overline{F}_{\cdot j}(y) = I^{-1}\sum_i F_{ij}(y)$. For the one-way ANCOVA design, choose a distribution function $G(x)$ and let

$$\overline{F}_{i\cdot}^{G}(y) = \int F_{ix}(y)\, dG(x) \, , \ \text{and } \overline{F}_{\cdot x}(y) = \frac{1}{k}\sum_{i=1}^{k} F_{ix}(y) \ . \tag{2.3}$$

If the X_{ij} are a random sample, we can choose $G(x)$ to be the overall distribution function of the covariate. Thus, if the covariate has the same distribution in all groups, $\overline{F}_{i\cdot}(y)$ is the marginal distribution function of Y_{ij}. The hypotheses of interest in model (2.1) or (2.2) are:

1. The $\overline{F}_{i\cdot}(y)$ do not depend on i, or $\overline{F}_{i\cdot}^{G}$ do not depend on i (no main effect);
2. The $\overline{F}_{\cdot j}(y)$ do not depend on j, or $\overline{F}_{\cdot x}(y)$ do not depend on x (no main effect);
3. The $F_{ij}(y) = \overline{F}_{i\cdot}(y) + K_j(y)$, or $F_{ix}(y) = \overline{F}_{i\cdot}^{G}(y) + K_x(y)$ (no interaction);
4. $F_{ij}(y)$ is independent of i, or $F_{ix}(y)$ is independent of i (no simple effect);
5. $F_{ij}(y)$ is independent of j, or $F_{ix}(y)$ is independent of x (no simple effect).

Note that for the ANCOVA setting, the first hypothesis is sensible even when the model is not additive (i.e., even when the slopes are not equal in the classical case), while the third hypothesis is the generalization of testing the equality of slopes in the classical model above. An important advantage of the non-parametric models is that these hypotheses and the procedures we suggest for analyzing them are unchanged by monotone transformations in the response. In the classical model, it is often necessary to find an appropriate transformation which simultaneously linearizes the expectation and equalizes the variances. In the non-parametric model such a transformation is not necessary.

The above hypotheses can also be described in terms of corresponding nonparametric effects being zero. These nonparametric effects are defined from decompositions of F_{ij} and F_{ix}. The decomposition of F_{ij} (Akritas & Arnold [4]) is

$$F_{ij}(y) = M(y) + A_i(y) + B_j(y) + C_{ij}(y), \tag{2.4}$$

where $M = \overline{F}_{\cdot\cdot}$, $A_i = \overline{F}_{i\cdot} - M$, $B_j = \overline{F}_{\cdot j} - M$, and $C_{ij} = F_{ij} - \overline{F}_{i\cdot} - \overline{F}_{\cdot j} + M$. A_i, B_j, and C_{ij} are, respectively, the nonparametric main row, main column, and interaction effects. The decomposition of F_{ix} (Akritas, Arnold & Du [7]) is

$$F_{ix}(y) = M^G(y) + A_i^G(y) + B_x^G(y) + C_{ix}^G(y) \ , \tag{2.5}$$

where $M^G(y) = I^{-1}\sum_{i=1}^I \int F_{ix}(y)dG(x)$, $A_i^G(y) = \overline{F}_{i\cdot}^G(y) - M(y)$, $B_x^G(y) = \overline{F}_{\cdot x}(y) - M^G(y)$, and $C_{ix}^G(y) = F_{ix}(y) - M^G(y) - A_i^G(y) - B_x^G(y)$ are, respectively, the nonparametric main factor, main covariate, and interaction effects.

Because decomposition (2.4) bears close resemblance to the decomposition of a rectangular array of means, μ_{ij}, it is straight forward to see how nonparametric effects are defined in higher-way layouts. A decomposition for two-way ANCOVA that displays the covariate adjusted main effects and interactions of the two factors is easily obtained from (2.5) by replacing the single index i by (i,j) and further decomposing $A_{(ij)}^G(y)$. This yields

$$F_{ijx}(y) = M^G(y) + A_i^G(y) + B_j^G(y) + (AB)_{ij}^G(y) + D_x^G(y) + C_{ijx}^G(y) \ , \tag{2.6}$$

where $M^G(y) = I^{-1}J^{-1}\sum_{i=1}^I\sum_{j=1}^J \int F_{ijx}(y)dG(x)$, $A_i^G(y) = J^{-1}\sum_{j=1}^J \int F_{ijx}(y)dG(x) - M^G(y)$, $B_j^G(y) = I^{-1}\sum_{i=1}^I \int F_{ijx}(y)dG(x) - M^G(y)$, $(AB)_{ij}^G = \int F_{ijx}(y)dG(x) - A_i^G(y) - B_j^G(y) + M^G(y)$, $D_x^G(y) = I^{-1}J^{-1}\sum_{i=1}^I\sum_{j=1}^J F_{ijx}(y) - M^G(y)$, and $C_{ijx}^G(y) = F_{ijx}(y) - M^G(y) - A_i^G(y) - B_j^G(y) - (AB)_{ij}^G(y) - D_x^G(y)$.

Estimates of the nonparametric effects provide useful graphical quantification of the effects; see Du, Akritas, Arnold & Osgood [27] for some such plots. Akritas & Arnold [4], Akritas, Arnold & Brunner [6], and Akritas, Arnold & Du [7] establish relations between the nonparametric hypotheses and the corresponding parametric ones. The basic result is that the nonparametric hypotheses are stronger than their parametric counterparts in the sense that they imply but are not implied by them. For example, Akritas, Arnold & Brunner [6] show the following

PROPOSITION 2.1 *The nonparametric hypothesis of no interaction is equivalent to the statement that, for any monotone transformation t, $Et(Y_{ijk}) = \mu_t + \alpha_{t,i} + \beta_{t,j}$.*

Thus the nonparametric hypothesis of no interaction is equivalent to the statement that the mean of any transformation of the response can be decomposed in an additive fashion. This strong form of additivity captures the *substantive* meaning of no interaction between factors as a scientist might think of it.

Because the nonparametric hypotheses are invariant under monotone transformations of the response, it is natural to use (mid-)rank test statistics. In addition, rank statistics are robust against outliers and have good power properties. A general theory of testing hypotheses in two- and higher-way ANOVA designs is possible from the observation that all hypotheses can be expressed as $\boldsymbol{C}\boldsymbol{F} = \mathbf{0}$, where $\boldsymbol{C}$ is a contrast matrix and $\boldsymbol{F}$ the column vector of cell distribution functions. For example in two-way ANOVA, $\boldsymbol{F} = (F_{11}, \ldots, F_{1J}, \ldots, F_{I1}, \ldots, F_{IJ})'$. An easy way to generate the appropriate contrast matrix for each hypothesis is given in Akritas, Arnold & Brunner [6]. The statistic for testing nonparametric hypothesis $\boldsymbol{C}\boldsymbol{F} = \mathbf{0}$ is based on the asymptotic distribution of

$$\widehat{\mathbf{T}}_{\boldsymbol{C}} = \boldsymbol{C} \int \widehat{H} d\widehat{\mathbf{F}}, \tag{2.7}$$

where $\widehat{\boldsymbol{F}}$ is the empirical estimator $\boldsymbol{F}$, i.e. the vector made up of the empirical distribution functions, and $\widehat{H}$ is the combined empirical distribution function. Note that the empirical distribution functions are also defined here as the average of their left- and right-continuous versions. Because the overall rank of Y_{ijk} among all N observations is given by $R_{ijk} = \frac{1}{2} + N\widehat{H}(Y_{ijk})$,

$$\boldsymbol{C} \int \widehat{H} d\widehat{\boldsymbol{F}} = N^{-1}\boldsymbol{C}\left(R_{11\cdot} - \frac{1}{2}, \ldots, R_{IJ\cdot} - \frac{1}{2}\right)' = N^{-1}\boldsymbol{C}\left(R_{11\cdot}, \ldots, R_{IJ\cdot}\right)'.$$

Thus (2.7) is a vector of linear (mid-)rank statistics. Let $\boldsymbol{V} = \mathrm{Diag}(\lambda_{11}^{-1}\sigma_{11}^2, \ldots, \lambda_{ab}^{-1}\sigma_{ab}^2)$, with $\lambda_{ij} = n_{ij}/N$, $\sigma_{ij}^2 = \mathrm{Var}[H(Y_{ijk})]$ for all $k = 1, \ldots, n_{ij}$, and $\widehat{\boldsymbol{V}}$ denote the matrix $\boldsymbol{V}$ with σ_{ij}^2 replaced by

$$\widehat{\sigma}_{ij}^2 = N^{-2}(n_{ij} - 1)^{-1} \sum_{k=1}^{n_{ij}} (R_{ijk} - R_{ij\cdot})^2. \tag{2.8}$$

Then, Akritas, Arnold & Brunner [6] show

THEOREM 2.2 *Under the hypothesis $\boldsymbol{C}\mathbf{F} = \mathbf{0}$, where $\boldsymbol{C}$ has full row rank, the test statistic*

$$Q(\boldsymbol{C}) = N\widehat{\boldsymbol{T}}_{\boldsymbol{C}}'(\boldsymbol{C}\widehat{\boldsymbol{V}}\boldsymbol{C}')^{-1}\widehat{\boldsymbol{T}}_{\boldsymbol{C}} \tag{2.9}$$

has, as $N \to \infty$, a χ_r^2 distribution, where χ_r^2 denotes the chi-square distribution with r degrees of freedom, and r is the number of rows of $\boldsymbol{C}$.

It should be noted that the above quadratic form can be computed with SAS, S-plus, Minitab, or any other higher level software system with matrix operation capabilities. Finally, Akritas, Arnold & Brunner [6] describe a finite-sample correction procedure that works remarkably well with small sample sizes. Such a correction is decribed in the more general context of Section 3.

For ANCOVA, we will only discuss testing the first type of hypotheses given below (2.3). This includes also testing for covariate adjusted main effects and interactions between factors in higher-way ANCOVA designs with one covariate. Testing procedures for the other type of hypotheses have been developed much more recently and will be briefly described in Section 4.

To describe the test statistics set $\overline{\boldsymbol{F}}_{\cdot}^{G} = (\overline{F}_{1\cdot}^{G}, \ldots, \overline{F}_{I\cdot}^{G})'$. Then any of the nonparametric hypotheses in one- two- and higher-way ANCOVA designs is of the form

$$H_0 : \boldsymbol{C}\overline{\boldsymbol{F}}_{\cdot}^{G} = \mathbf{0} \ , \tag{2.10}$$

for some full-rank contrast matrix $\boldsymbol{C}$. The statistic for testing such a hypothesis is based on

$$\widehat{T}_{\boldsymbol{C}}^{G} = \boldsymbol{C} \int \widehat{H} d\widehat{\overline{\boldsymbol{F}}^{G}}_{\cdot} \ , \tag{2.11}$$

where $\widehat{H}$ is the empirical distribution function of all Y_{ij}, $\widehat{\overline{\boldsymbol{F}}^{G}}_{\cdot} = (\widehat{\overline{F}^{G}}_{1\cdot}(y), \ldots, \widehat{\overline{F}^{G}}_{I\cdot}(y))'$, and

$$\widehat{\overline{F}^{G}}_{i\cdot}(y) = \int \widehat{F}_{ix}(y) d\widehat{G}(x), \quad \text{with} \quad \widehat{F}_{ix}(y) \ = \ \sum_{j=1}^{n_i} \frac{K\left(\frac{x-X_{ij}}{a_{n_i}}\right)}{\sum_{j'} K\left(\frac{x-X_{ij'}}{a_{n_i}}\right)} I(Y_{ij} \leq y) \ , \tag{2.12}$$

where $\widehat{G}$ is the empirical distribution function of all X_{ij}, and K is a known probability density function (kernel). Note that (2.11) is a vector of weighted mid-rank statistics. Let $\widehat{\boldsymbol{V}}^{G}$ denote the estimate of the asymptotic covariance matrix of $\int \widehat{H} d\widehat{\overline{\boldsymbol{F}}^{G}}_{\cdot}$ given in Akritas, Arnold & Du [7]. Then in the aforementioned paper it is shown that under suitable smoothness assumptions

THEOREM 2.3 *Let r denote the rank of $\boldsymbol{C}$. Under the null hypothesis (2.10),*

$$N\left(\widehat{T}_{\boldsymbol{C}}^{G}\right)'\left(\boldsymbol{C}\widehat{\boldsymbol{V}}^{G}\boldsymbol{C}'\right)^{-1}\widehat{T}_{\boldsymbol{C}}^{G} \to \chi_r^2, \quad \textit{in distribution.}$$

Note that even though the kernel estimator of the conditional distribution that enters (2.11) has a slower rate of convergence, the test statistic does have the usual square root of N rate of convergence.

We finish this subsection with a biological illustration (taken from Akritas, Arnold and Brunner (1996)) of the (mid-)rank procedures for testing the nonparametric hypotheses with independent data.

Example 1.

Two inhalable test substances, drug 1 and drug 2 (factor A), are to be compared with regard to their irritative activity in the respiratory tract of the rat after subchronic inhalative exposure. Reserve cell hyperplasia in the respiratory epithelium of the nose after exposure to 2, 5, and 10[*ppm*] of the test substances served as a criterion for irritation. The result was histopathologically evaluated by the grading scales: 0 = 'no changes', 1 = 'slight changes', 2 = 'distinct changes', 3 = 'severe changes'. The results for the two drugs and three concentration groups (factor B) with 20 rats each, are given in Table 1.

Mid-rank statistics of the form described in (2.9) below where used to analyze this data set. The test for interaction gave p-value 0.77 indicating that the results are quite homogeneous within the three concentrations (no interaction). A significant treatment effect for the drug is proved at the 5% level ($p = 0.028$). For the concentration a highly significant effect is proved ($p < 10^{-7}$). We remark that in this example, both the asymptotic and small sample approximations give quite similar p-values.

Table 1
Data and rank means for the reserve cell hyperplasia trial.

	Test substance								Results		
	drug 1				drug 2						
Concentration	no. of rats with scale				no. of rats with scale				Rank Means $R_{ij\cdot}$		$\widetilde{R}_{\cdot j\cdot}$
	0	1	2	3	0	1	2	3	drug 1	drug 2	
2	18	2	0	0	16	3	1	0	33.95	39.68	36.81
5	12	6	2	0	8	8	3	1	49.85	62.08	55.96
10	3	7	6	4	1	5	8	6	83.18	94.28	88.73
									55.66	65.34	$\widetilde{R}_{i\cdot\cdot}$

3. REPEATED MEASURES

As mentioned in the Section 1, in many biological experiments and medical or psychological studies, randomly chosen subjects are observed repeatedly under the same or under different treatments. Such designs include growth curves, longitudinal data or repeated measures designs. For the analysis of such designs there exists a variety of classical linear models assuming the multivariate normality of the observed random vectors; a special structure for the dependencies of the multivariate observations, e.g. compound symmetry, may or may not be assumed. In this section we describe nonparametric generalizations of the classical ANOVA and MANOVA procedures where not only the assumption of normality of the error terms is relaxed but also treatment effects and hypotheses are defined in a nonparametric framework. Based on the ideas explained in Section 2, we formulate nonparametric hypotheses in various designs and derive (mid-)rank statistics for testing these hypotheses. We re-emphasize that the procedures to be described are applicable to all ordinal data including discrete such as scores in psychological tests, and quality scales in order to describe the degree of the damage of plants or trees in ecological or environmental studies.

Akritas and Brunner [8] consider a general class of repeated measures designs with some factors having subjects nested within their levels (whole-plot factors) and other factors being crossed with the subjects (sub-plot factors). They also allow cluster sampling, i.e. each subject may receive more than one measurement at each occasion (sub-plot factor level combination). For simplicity we consider here only designs where each subject is observed at all occasions but receives only one observation per occasion. As before, we will use a single index to enumerate the factor-level combinations of possibly several whole-plot factors and one index for the level combinations of sub-plot factors. In such designs we observe random vectors $\boldsymbol{X}_{ik} = (X_{ik1}, \ldots, X_{ikd})'$, where i enumerates the cells of the whole-plot factors, k enumerates the subjects within cell i, and the third index, which we will call s, enumerates the d cells of the sub-plot factors. Thus $\boldsymbol{X}_{ik}$ is the vector of repeated observations or measurements taken on subject k within cell i. The nonparametric mixed model assumes only that the vectors $\boldsymbol{X}_{ik}$, $i = 1, \ldots, r$, $k = 1, \ldots, n_i$, are independent and

$$\boldsymbol{X}_{ik} = (X_{ik1}, \ldots, X_{ikd})' \sim F_i(\boldsymbol{x}),\ i = 1, \ldots, r,\ k = 1, \ldots, n_i, \tag{3.13}$$

so that the marginal distribution functions $F_{is}(x) = \frac{1}{2}\left[F_{is}^{+}(x) + F_{is}^{-}(x)\right]$, $i = 1, \ldots, r$, $s = 1, \ldots, d$, of X_{iks} do not depend on k (Akritas and Brunner [8]). In particular, no special structure is assumed for the dependencies between the components of the vectors $\boldsymbol{X}_{ik}$.

A difference between the marginal distributions F_{is} is quantified by the so-called relative treatment effects $p_{is} = \int H(x)dF_{is}(x)$, where $H(x) = \frac{1}{N}\sum_{i=1}^{r}\sum_{s=1}^{d} n_i F_{is}(x)$ is the weighted average of the $N = d \cdot \sum_{i=1}^{r} n_i$ distribution functions.

Let $\boldsymbol{F} = (F_{11}, \ldots, F_{1d}, \ldots, F_{r1}, \ldots, F_{rd})'$ denote the vector of the marginal distributions and let $\boldsymbol{p} = \int H d\boldsymbol{F} = (p_{11}, \ldots, p_{1d}, \ldots, p_{r1}, \ldots, p_{rd})'$, denote the vector of the relative treatment effects. Then nonparametric hypotheses are formulated by means of the marginal distributions $F_{is}(x)$ in the same way as for independent observations. Let $\boldsymbol{C}$ denote a suitable contrast matrix to formulate a hypothesis. Then, $H_0^F : \boldsymbol{C}\boldsymbol{F} = \mathbf{0}$ is the most general form of a nonparametric hypothesis in the mixed model (3.13).

The relative treatment effects $p_{is} = \int H(x)dF_{is}(x)$ are estimated by replacing $F_{is}(x)$ and $H(x)$ by the obvious empirical estimators $\widehat{F}_{is}(x)$, $\widehat{H}(x) = \frac{1}{N}\sum_{i=1}^{r}\sum_{s=1}^{d} n_i \widehat{F}_{is}(x)$. An asymptotically unbiased and L_2-consistent estimator of p_{is} is then given by

$$\widehat{p}_{is} = \int \widehat{H} d\widehat{F}_{is} = \frac{1}{n_i}\sum_{k=1}^{n_i} \widehat{H}(X_{iks}) = \frac{1}{n_i}\sum_{k=1}^{n_i} \frac{1}{N}\left(R_{iks} - \frac{1}{2}\right) = \frac{1}{N}\left(\overline{R}_{i\cdot s} - \frac{1}{2}\right), \tag{3.14}$$

where R_{iks} denotes the mid-rank of X_{iks} among all N observations. Thus, $\boldsymbol{p} = \int H d\boldsymbol{F}$ is estimated by

$$\begin{aligned}
\widehat{\boldsymbol{p}} &= \int \widehat{H} d\widehat{\boldsymbol{F}} = (\widehat{p}_{11}, \ldots, \widehat{p}_{rd})' \\
&= \frac{1}{N}\left(\overline{R}_{1\cdot 1} - \tfrac{1}{2}, \ldots, \overline{R}_{1\cdot d} - \tfrac{1}{2}, \ldots, \overline{R}_{r\cdot 1} - \tfrac{1}{2}, \ldots, \overline{R}_{r\cdot d} - \tfrac{1}{2}\right). \qquad (3.15)
\end{aligned}$$

Let now $Y_{iks} = H(X_{iks})$, $s = 1, \ldots, d$, be the so-called asymptotic rank transform of X_{iks} (Akritas [1]), and set $\boldsymbol{Y}_{ik} = (Y_{ik1}, \ldots, Y_{ikd})'$, and $\overline{\boldsymbol{Y}}_{\cdot} = \left(\overline{\boldsymbol{Y}}_{1\cdot}', \ldots, \overline{\boldsymbol{Y}}_{r\cdot}'\right)'$, where $\overline{\boldsymbol{Y}}_{i\cdot} = \left(\overline{Y}_{i\cdot 1}, \ldots, \overline{Y}_{i\cdot d}\right)' = \frac{1}{n_i}\sum_{k=1}^{n_i} \boldsymbol{Y}_{ik}$. Finally, let $\boldsymbol{V}_i = \mathrm{Cov}(\boldsymbol{Y}_{i1})$, $k = 1, \ldots, n_i$. By the independence of the $\overline{\boldsymbol{Y}}_{i\cdot}$, the covariance matrix $\boldsymbol{V}_n = \mathrm{Cov}(\sqrt{n}\,\overline{\boldsymbol{Y}}_{\cdot})$ is block diagonal, i.e.

$$\boldsymbol{V}_n = \bigoplus_{i=1}^{r} \frac{n}{n_i}\boldsymbol{V}_i\,. \tag{3.16}$$

Akritas and Brunner [8] show

THEOREM 3.1 *Let $\boldsymbol{X}_{ik} = (X_{ik1}, \ldots, X_{ikd})'$ be independent and identically distributed random vectors and let $\boldsymbol{V}_n$ be the covariance matrix given in (3.16). Then, under suitable assumptions and under $H_0^F : \boldsymbol{C}\boldsymbol{F} = \mathbf{0}$, the statistic $\sqrt{n}\boldsymbol{C}\widehat{\boldsymbol{p}} = \sqrt{n}\boldsymbol{C}\int \widehat{H} d\widehat{\boldsymbol{F}}$ has, asymptotically, a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{C}\boldsymbol{V}_n\boldsymbol{C}'$.*

Since the covariance matrix $\boldsymbol{V}_n$ is unknown, it must be estimated from the data. A consistent estimator is given in the next theorem (Akritas and Brunner [8]).

THEOREM 3.2 *For $i = 1, \ldots, r$, let $\overline{\boldsymbol{R}}_{i\cdot} = n_i^{-1} \sum_{k=1}^{n_i} \boldsymbol{R}_{ik}$ denote the mean of the vectors $\boldsymbol{R}_{ik} = (R_{ik1}, \ldots, R_{ikd})'$, and set*

$$\widehat{\boldsymbol{V}}_i = \frac{1}{N^2(n_i - 1)} \sum_{k=1}^{n_i} \left(\boldsymbol{R}_{ik} - \overline{\boldsymbol{R}}_{i\cdot}\right) \left(\boldsymbol{R}_{ik} - \overline{\boldsymbol{R}}_{i\cdot}\right)' \ , \quad \widehat{\boldsymbol{V}}_n = \bigoplus_{i=1}^{r} \frac{n}{n_i} \widehat{\boldsymbol{V}}_i \ . \tag{3.17}$$

Then, under the same assumptions as in Theorem 3.1,$\|\widehat{\boldsymbol{V}}_i - \boldsymbol{V}_i\| \xrightarrow{p} 0, \quad i = 1, \ldots, r$, and $\|\widehat{\boldsymbol{V}}_n - \boldsymbol{V}_n\| \xrightarrow{p} 0$, where $\boldsymbol{V}_i = \mathrm{Cov}(\boldsymbol{Y}_{i1})$, $\boldsymbol{V}_n$ is defined in (3.16), and $\|\cdot\|$ denotes the Euclidean norm of a matrix.

To test the nonparametric hypothesis $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$ explained in Section 2, we consider quadratic forms of the random vectors $\sqrt{n}\boldsymbol{C}\widehat{\boldsymbol{p}}$. Other statistics used in multivariate analysis, like Wilk's Λ, for example, are not discussed here since they require the equality of the covariance matrices. In nonparametric models, however, this assumption is not reasonable since, in general, the assumption of homoscedastic distribution functions is not transferred to the asymptotic rank transform vectors $\boldsymbol{Y}_{ik} = (Y_{ik1}, \ldots, Y_{ikd})'$, $i = 1, \ldots, r$, $k = 1, \ldots, n_i$, where $Y_{iks} = H(X_{iks})$, $s = 1, \ldots, d$. This is easily seen by the fact that $H(\cdot)$ is a non-linear transformation in general as pointed out already by Akritas [1]. We consider two different quadratic forms.

Wald-Type Statistics (WTS)

The simplest way to derive a quadratic form for testing the hypothesis $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$ is to use a generalized inverse of the covariance matrix $\boldsymbol{CV}_n\boldsymbol{C}$' under H_0^F as the generating matrix of the quadratic form where the unknown covariance matrix $\boldsymbol{V}_n$ is replaced by the consistent estimator $\widehat{\boldsymbol{V}}_n$ given in (3.17). Let $[\boldsymbol{C}\widehat{\boldsymbol{V}}_n\boldsymbol{C}']^+$ denote the Moore-Penrose inverse of $\boldsymbol{C}\widehat{\boldsymbol{V}}_n\boldsymbol{C}'$. Assume that $\boldsymbol{V}_n \to \boldsymbol{V} \neq \boldsymbol{0}$ such that $r(\boldsymbol{CV}_n) = r(\boldsymbol{CV})$, where $r(\cdot)$ denotes the rank of a matrix. Then under $H_0^F : \boldsymbol{CF} = \boldsymbol{0}$, it follows from Theorem 3.1 and Theorem 3.2 that

$$Q_n^W(\boldsymbol{C}) \ = \ n\, \widehat{\boldsymbol{p}}'\boldsymbol{C}' \left[\boldsymbol{C}\widehat{\boldsymbol{V}}_n\boldsymbol{C}'\right]^+ \boldsymbol{C}\widehat{\boldsymbol{p}} \tag{3.18}$$

has, asymptotically, a central χ^2-distribution with $f = r(\boldsymbol{CV})$ degrees of freedom. This statistic is called the rank version of the WTS. However, extremely large sample sizes are needed for a satisfactory approximation.

ANOVA-Type Statistics (ATS)

The estimation of the covariance matrix $\boldsymbol{V}_n$ in (3.18) requires large sample sizes. Therefore, Brunner, Munzel and Puri [19] suggested to leave out the estimator $\widehat{\boldsymbol{V}}_n$ in the generating matrix of the quadratic form and to consider the asymptotic distribution of

$$Q_n^A(\boldsymbol{C}) \ = \ n\widehat{\boldsymbol{p}}'\boldsymbol{C}'[\boldsymbol{CC}']^-\boldsymbol{C}\widehat{\boldsymbol{p}} \ = \ n\widehat{\boldsymbol{p}}'\boldsymbol{T}\widehat{\boldsymbol{p}}. \tag{3.19}$$

We note that $\boldsymbol{T} = \boldsymbol{C}'[\boldsymbol{CC}']^-\boldsymbol{C}$ is a projection matrix where $[\boldsymbol{CC}']^-$ denotes some g-inverse of $\boldsymbol{CC}'$. The matrix $\boldsymbol{T}$ provides a standard formulation of the hypothesis $\boldsymbol{CF} =$

$\mathbf{0}$ since two hypotheses $\boldsymbol{C}_1\boldsymbol{F} = \mathbf{0}$ and $\boldsymbol{C}_2\boldsymbol{F} = \mathbf{0}$ are equivalent if and only if $\boldsymbol{T}_1 = \boldsymbol{C}_1'[\boldsymbol{C}_1\boldsymbol{C}_1']^-\boldsymbol{C}_1 = \boldsymbol{T}_2 = \boldsymbol{C}_2'[\boldsymbol{C}_2\boldsymbol{C}_2']^-\boldsymbol{C}_2$. To see this, note that $\boldsymbol{T}\boldsymbol{F} = \mathbf{0} \Longleftrightarrow \boldsymbol{C}\boldsymbol{F} = \mathbf{0}$ because $\boldsymbol{C}'[\boldsymbol{C}\boldsymbol{C}']^-$ is a generalized inverse of $\boldsymbol{C}$. The asymptotic distribution of $Q_n^A(\boldsymbol{C})$ is given in the next Theorem.

THEOREM 3.3 *Let $\boldsymbol{T} = \boldsymbol{C}'[\boldsymbol{C}\boldsymbol{C}']^-\boldsymbol{C}$ and let $\boldsymbol{V}_n$ be as in (3.16). Then, under the the same assumptions as in Theorem 3.1, and under the hypothesis $H_0^F : \boldsymbol{C}\boldsymbol{F} = \mathbf{0}$, it follows that*

$$Q_n^A(\boldsymbol{C}) \;=\; n\widehat{\boldsymbol{p}}'\boldsymbol{T}\widehat{\boldsymbol{p}} \sim \sum_{i=1}^{r}\sum_{s=1}^{d} \lambda_{is}Z_{is}, \quad \text{as } n \to \infty, \tag{3.20}$$

where the λ_{is} are the eigenvalues of $\boldsymbol{T}\boldsymbol{V}_n\boldsymbol{T}$ and the $Z_{is} \sim \chi_1^2$ are independent random variables which are χ_1^2 distributed.

Brunner, Munzel and Puri [19] suggested to approximate the distribution of the random variable $\sum_{i=1}^{r}\sum_{s=1}^{d}\lambda_{is}Z_{is}$ by a scaled χ^2-distribution such that the first two moments coincide. This approximation dates back to Box [18] and is quite accurate. The result is given in the following approximation procedure.

APPROXIMATION PROCEDURE 3.4 *If $tr(\boldsymbol{T}\boldsymbol{V}_n) \geq t_0 > 0$ then, under $H_0^F : \boldsymbol{C}\boldsymbol{F} = \mathbf{0}$, the first two moments of the asymptotic distribution of $tr(\boldsymbol{T}\boldsymbol{V}_n) \cdot Q_n^A(\boldsymbol{C})/tr(\boldsymbol{T}\boldsymbol{V}_n\boldsymbol{T}\boldsymbol{V}_n)$ and of the χ_f^2-distribution coincide for $f = [tr(\boldsymbol{T}\boldsymbol{V}_n)]^2/tr(\boldsymbol{T}\boldsymbol{V}_n\boldsymbol{T}\boldsymbol{V}_n)$, where $tr(\cdot)$ denotes the trace of a square matrix.*

The unknown traces $tr(\boldsymbol{T}\boldsymbol{V}_n)$ and $tr(\boldsymbol{T}\boldsymbol{V}_n\boldsymbol{T}\boldsymbol{V}_n)$ can be estimated consistently by replacing $\boldsymbol{V}_n$ with $\widehat{\boldsymbol{V}}_n$ given in (3.17) and it follows under $H_0^F : \boldsymbol{C}\boldsymbol{F} = \mathbf{0}$ that the statistic

$$F_n(\boldsymbol{C}) \;=\; \frac{n \cdot tr(\boldsymbol{T}\widehat{\boldsymbol{V}}_n)}{tr(\boldsymbol{T}\widehat{\boldsymbol{V}}_n\boldsymbol{T}\widehat{\boldsymbol{V}}_n)}\, \widehat{\boldsymbol{p}}'\, \boldsymbol{T}\, \widehat{\boldsymbol{p}} \quad \dot{\sim} \quad \chi_f^2 \tag{3.21}$$

has approximately a central χ_f^2-distribution where f is estimated by

$$\widehat{f} \;=\; \frac{[tr(\boldsymbol{T}\widehat{\boldsymbol{V}}_n)]^2}{tr(\boldsymbol{T}\widehat{\boldsymbol{V}}_n\boldsymbol{T}\widehat{\boldsymbol{V}}_n)}. \tag{3.22}$$

We note that for very small sample sizes the estimator $\widehat{f}$ in (3.22) may be slightly biased.

We note that $Q_n^W(\boldsymbol{C}) = F_n(\boldsymbol{C})/f$ if $r(\boldsymbol{C}) = 1$ which follows from simple algebraic arguments. See Brunner, Munzel and Puri [19] for details regarding the consistency of the tests based on $Q_n^W(\boldsymbol{C})$ or $F_n(\boldsymbol{C})/f$.

In some special cases the so-called compound symmetry of the covariance matrix can be assumed under the hypothesis. In particular, in repeated measures designs with one homogeneous group of subjects and d repeated measures, compound symmetry can be assumed under the hypothesis $H_0^F : F_1 = \cdots = F_d$ if the subjects are blocks which can be split into homogeneous parts and each part is treated separately. In this case, only two quantities have to be estimated: the common variance and the common covariance. For more details, we refer to Brunner, Munzel and Puri [19].

The recent book Brunner, Domhof and Langer [20] presents many examples and discusses software for the computation of the statistics $Q_n^W(\boldsymbol{C})$ and $F_n(\boldsymbol{C})/f$.

4. EXTENSIONS AND FURTHER DEVELOPMENTS

The nonparametric ANOVA methodology has been extended to designs with independent and dependent censored data (Akritas & Brunner [9], O'Gorman & Akritas [35]), to data missing completely at random (Brunner, Munzel and Puri [19]), and to data missing at random (Akritas, Kuha & Osgood [10], Antoniou & Akritas [14]). An interesting generalization to multivariate ANOVA using spatial ranks is developed in Choi & Marden [22]. The paper by Brunner, Munzel and Puri [19] deals also with general rank-scores.

The nonparametric ANCOVA methodology described in Section 2 has been extended to up to three covariates also for repeated measures designs (Tsangari & Akritas [39], Tsangari & Akritas [40]). For four or more covariates the "curse of dimensionality" effects take over and this methodology, which requires consistent estimation of the conditional distributions, does not work.

As mentioned in Section 2 testing hypotheses involving the covariate in the nonparametric ANCOVA model is much more recent. The theory for testing such hypotheses is closely connected with the theory of ANOVA when the number of factor levels is large. See Boos and Brownie (1995), Akritas & Arnold [5], Bathke [15], Akritas & Papadatos [11], Wang & Akritas [42], Wang & Akritas [41] are representative publications in this new area; the last three deal also with heteroscedastic designs, while the last considers rank statistics for this problem. To see the connection with the nonparametric ANCOVA problem, think of the covariate as a factor with many levels. Then, hypotheses regarding the factor "covariate" in this hypothetical ANOVA design approximate the corresponding hypotheses in the ANCOVA design, and coincide with them asymptotically. The theory of ANOVA when the number of factor levels is large is not directly applicable because it requires at least two observations per factor level combination, whereas in typical ANCOVA designs there is only one observation per covariate value. To remedy this we use smoothness assumptions to augment the hypothetical ANOVA design by considering windows around each covariate value. This induces dependence (since different cells may share observations) and thus the aforementioned theory for ANOVA when the number of factor levels is large is still not directly applicable. However a new theory can be developed; see Wang & Akritas [43]. This approach also yields an alternative approach (and test statistic) to the procedure of Akritas, Arnold & Du [7] for the covariate adjusted main effects and interactions of factors. The interesting aspect of this theory is that it does not require consistent estimation of the conditional distribution functions, as that of Akritas, Arnold and Du [7] does, since the window sizes need not tend to infinity. Thus, there is hope that this new methodology will allow the extension of the nonparametric methodology to ANCOVA designs with more than three covariates.

5. RELATED WORK

There is a plethora of related work for ANOVA designs, with both independent and dependent observations, which emphasize invariance under monotone transformations. See Patil and Hoel (1973), Govindarajulu [28], Conover and Iman [24], Brunner & Neumann [21], Kepner & Robinson [31], Akritas [1], Akritas [2], Alvo & Cabilio [12], Thompson [38], Akritas [3], Marden and Muyot (1995), Cliff (1996), Alvo & Cabilio [13], Handcock & Janssen [30]. These, however, do not make reference to the nonparametric model and hypotheses (2.1), (2.4) and thus their applicability is limited. Some of these approaches are discussed in the recent book Brunner, Domhof and Langer [20] (e.g. Section 5.8) which, however, emphasizes testing the nonparametric hypotheses.

There is a much richer bibliography dealing with rank-related methods for testing the usual parametric hypotheses, such as aligned rank tests. Such statistics cannot be invariant under monotone transformations since they test hypotheses that are not. Thus they belong in the area of robust statistics.

There is also a lot of work in nonparametric ANCOVA. See Hall & Hart [29], King, Hart & Wehrly [32], Young & Bowman [45], Kulasekera [33], Bowman & Azzalini [17], p.80, Dette & Munk [25], Dette & Neumeyer [26], to mention a few. Again, this literature does not make use of the nonparametric model (2.2), (2.5) and, as a consequence, its scope is limited. In particular, these papers deal only with the so-called problem of curve comparison, which corresponds to the fourth of the hypotheses listed under (2.3), i.e. no simple factor effect, and their methods are only applicable to one-way ANCOVA with continuous response. Different from the above, Bathke & Brunner [16] contains an interesting alternative approach to testing for covariate-adjusted factor effects.

REFERENCES

1. Akritas, M. G. (1990). The Rank Transform Method in Some Two-Factor Designs. *Journal of the American Statistical Association* 85, 73–78.
2. Akritas, M. G. (1991). Limitations of the Rank Transform Drocedure: A Study of Repeated Measures Designs, Part I. *Journal of the American Statistical Association* 85, 73–78.
3. Akritas, M. G. (1993). Limitations of the Rank Transform Drocedure: A Study of Repeated Measures Designs, Part II, *Statistics & Probability Letters* 17, 149–156.
4. Akritas, M.G. and S.F. Arnold (1994). Fully Nonparametric Hypotheses for Factorial Designs I: Multivariate Repeated Measures Designs. *Journal of the American Statistical Association*, 89, 336–343.
5. Akritas, M. G. and Arnold, S. F. (2000). Asymptotics for ANOVA when the number of levels is large. *Journal of the American Statistical Association* 95, 212-226.
6. Akritas, M.G., Arnold, S.F. and Brunner, E. (1997). Nonparametric Hypotheses and Rank Statistics for Unbalanced Factorial Designs. *Journal of the American Statistical Association* 92, 258–265.
7. Akritas, M. G., Arnold, S. F. and Du, Y. (2000). Nonparametric models and methods for nonlinear analysis of covariance. *Biometrika* 87, 507-526.
8. Akritas, M.G. and Brunner , E. (1997a). A Unified Approach to Rank Tests in Mixed Models, *Journal of Statistical Planning and Inference* 61, 249–277.
9. Akritas, M.G. and Brunner, E. (1997b). Nonparametric methods for designs with censored data. *Journal of the American Statistical Association* 92, 568-576.
10. Akritas, M.G, Kuha, J. and Osgood (2002). A nonparametric approach to matched pairs with censored data. *Sociological Methods & Research* 30, 425-462. (With Discussion)
11. Akritas, M. G. and Papadatos, N. (2002). Heteroskedastic One-Way ANOVA and Lack-of-Fit Tests. *Journal of the American Statistical Association*, tentatively accepted.
12. Alvo, M. and Cabilio, P. (1991). On the balanced incomplete block design for rankings. *Annals of Statistics* 19,1597-1613.
13. Alvo, M. and Cabilio, P. (1999). A general rank based approach to the analysis of block data. *Commun. Stat.-Theor. M.* 28, 197-215.
14. Antoniou, E. and Akritas, M.G. (2003). Nonparametric methods for designs with data missing at random. Submitted.
15. Bathke, A. (2002). ANOVA for a large number of treatments. *Mathematical Methods of Statistics* 11, 118-132.
16. Bathke, A. and Brunner, E. (2002). A nonparametric alternative to analysis of covariance. Preprint.
17. Bowman, A. W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis.* Oxford: Oxford University Press.
18. 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. *Annals of Mathematical Statistics* 25, 290–302.
19. Brunner, E., Munzel, U. and Puri, M.L. (1999). Rank-Score Tests in Factorial Designs

With Repeated Measures. *Journal of Multivariate Analysis* 70, 286–317.
20. Brunner, E., Domhof, S. and Langer, F. (2002). *Nonparametric Analysis of Longitudinal Data in Factorial Designs*, Wiley, New York.
21. Brunner, E. and Neumann, N. (1982). Rank Tests for Correlated Random Variables. *Biometrical Journal*, 24, 373–389.
22. Choi, K. and Marden, J. (2002). Multivariate analysis of variance using spatial ranks. *Sociological Methods & Research* 30, 341-366.
23. Cliff, N. (1996). Answering ordingal questions with ordinal data using ordinal statistics. *Multivariate Behavioral Research* 31, 331-350.
24. Conover, W. J. and Iman, R. L. (1981), Rank transformations as a bridge between parametric and nonparametric statistics (with discussion). *American Statistician* 35, 124–133.
25. Dette, D. and Munk, A. (1998). Nonparametric comparison of several regression functions: exact and asymptotic theory. *Annals of Statistics* 26, 2339-2368.
26. Dette, D. and Neumeyer, N. (2001). Nonparametric analysis of covariance. *Annals of Statistics* 29, 1361-1400.
27. Du, Y., Akritas, M. G., Arnold, S. F. and Osgood, D. W. (2002). Analysis of teenage deviant behavior data. *Sociological Methods & Research* 30, 309-340.
28. Govindarajulu, Z. (1975), Robustness of Mann-Whitney-Wilcoxon test to dependence in the variables, *Studia Scientiarum Mathematicarum Hungarica* 10, 39–45.
29. Hall, P. and Hart, J.D. (1990). Bootstrap test for defference between means in nonparametric regression. *Journal of the American Statistical Association* 85, 1039-1049.
30. Handcock, M. S., and Janssen, P. (2002). Statistical inference for the relative density. *Sociological Methods & Research* **30**, 394-424.
31. Kepner, J. L. and Robinson, D. H. (1988). Nonparametric Methods for Detecting Treatment Effects in Repeated Measures Designs. *Journal of the American Statistical Association* 83, 456–461.
32. King, E.C., Hart, J.D. and Wehrly, T.E. (1991). Testing the equality of regression curves using linear smoothers. *Statistics & Probability Letters* 12, 239-247.
33. Kulasekera, K.B. (1995). Comparison of regression curves using quadi residuals. *Journal of the American Statistical Association* 90, 1085-1093.
34. Marden, J. I. and Muyot, M. E. T. (1995). Rank tests for main and interaction effects in Analysis of Variance. *Journal of the American Statistical Association* 90, 1388-1398.
35. O'Gorman, J.T. and Akritas, M.G. (2001). Nonparametric models and methods for designs with dependent censored data. *Biometrics* 57, 88-95
36. Patil, K. M. and Hoel, D. G. (1973). A nonparametric test for interaction in factorial experiments. *Journal of the American Statistical Association* 68, 615-620.
37. Ruymgaart, F.H. (1980). A Unified Approach to the Asymptotic Distribution Theorey of Certain Midrank Statistics, in *Statistique non Parametrique Asymptotique*, 1–18, J.P. Raoult (Ed.), Lecture Notes on Mathematics, No. 821, Springer, Berlin.
38. Thompson, G. L. (1991). A unified approach to rank tests for multivariate and repeated measures designs. *Journal of the American Statistical Association* 86, 410–419.
39. Tsangari, H. and Akritas, M.G. (2003). Nonparametric ANCOVA with two and three covariates. *Journal of Multivariate Analysis*. In press.

40. Tsangari, H. and Akritas, M.G. (2003). Nonparametric models and methods for ANCOVA with dependent data. *Journal of Nonparametric Statistics.* In press.
41. Wang, H. and Akritas, M.G. (2003). Rank tests for ANOVA with large number of factor levels. *Journal of Nonparametric Statistics.* In press.
42. Wang, L. and Akritas, M.G. (2003). Two-way Heteroscedastic ANOVA when the Number of Levels is Large. Submitted.
43. Wang, L. and Akritas, M.G. (2003). Testing for the covariate effect in nonparametric ANCOVA. Submitted.
44. Winship, C. and Mare, R. D. (1984). Regression models with ordinal variables. *American Sociological Review* 49, 512-525.
45. Young, S. G. and Bowman, A. W. (1995). Non-parametric analysis of covariance. *Biometrics* 51, 920-31.