

This article presents a nonparametric approach to analyzing research designs that include repeated observations on the same set of individuals or units, such as longitudinal panel studies. The data collected from different individuals are generally assumed to be independent, while several observations from the same individual or unit may be dependent. The approach suggested in this study is nonparametric in the sense that only the distribution functions are used to define the treatment effects, and suggested procedures are extensions of well-established rank order methods. Missing observations are especially common in studies with repeated measures, and three strategies for addressing this problem are compared: complete case analysis, last observation carried forward, and complete set analysis. The authors demonstrate these methods with an analysis of age trends and sex differences in alcohol consumption in a sample of U.S. adolescents who completed questionnaires four times in the 7th through 10th grades.

Rank Procedures for Repeated Measures With Missing Values

SEBASTIAN DOMHOF
EDGAR BRUNNER
University of Göttingen, Germany

D. WAYNE OSGOOD
Pennsylvania State University

1. INTRODUCTION

In many sociological, psychological, or medical studies, the subjects are observed repeatedly under different conditions, called *treatments* in the terminology of experimental design. Such designs can be described by several more or less complicated models that come under the framework of *mixed models*. They include

AUTHORS' NOTE: *Wayne Osgood's contribution to this research was supported under Award 94-IJ-CX-0058 (Finn Esbensen, Principal Investigator) from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Justice.*

growth curves, longitudinal data, or repeated-measures designs. The experimental units—often referred to as *subjects* or *individuals*—are typically people in sociological research, but they might instead be groups or organizations. The data collected from different experimental units are generally assumed to be independent while that from the same unit may be, to some degree, dependent. As a result of this dependence, the analysis and interpretation of such studies become considerably more difficult than for experimental designs in which all observations are independent. Moreover, the special problem of *missing values* is an additional difficulty arising in studies with repeated measures.

The theory for the analysis of such designs using parametric or semiparametric mixed models is well developed and is well described in some recent textbooks (Diggle, Liang, and Zeger 1994; Lindsey 1993). For the analysis of general nonnormal data—particularly for discrete data or data that are measured on an ordinal scale—the theory is less developed, particularly when several factors are present in the experiment or study. Such data occur in many sociological or psychological experiments and in numerous areas in medicine, where ratings or gradings of any type are observed. The designs of these trials are more or less complicated, and appropriate procedures for the analysis are required.

Here, we suggest an approach that is nonparametric in the sense that only the distribution functions are used to define and estimate the treatment effects and to formulate the hypotheses. The suggested procedures are extensions of the Wilcoxon-Mann-Whitney and the Kruskal-Wallis test to multidimensional factorial research designs and to repeated-measures designs. These extensions are based on (1) the idea of Akritas and Arnold (1994) to formulate nonparametric hypotheses in factorial designs by means of the distribution functions and (2) on an extension of this idea to unbalanced repeated-measures designs in which the distribution functions may be discontinuous (Akritas and Brunner 1997; Brunner, Munzel, and Puri 1999). To motivate the procedures and to present the ideas, we consider the alcohol consumption study described in the following subsection.

1.1. EXAMPLE: ALCOHOL CONSUMPTION STUDY

Our empirical example is an analysis of the relationship of gender and age to alcohol consumption in a sample of American students, assessed once each year in the 7th through 10th grades. Prior research leads us to expect that we will find alcohol use to be associated with age and that this age trend will differ between boys and girls. National data sources, such as the *Monitoring the Future Study*, report that the rate of alcohol use increases sharply over the age span we are studying (Johnston, O'Malley, and Bachman 2000:95, 99-100). These data also indicate that in early adolescence (e.g., 8th grade), there is negligible gender difference in the prevalence of alcohol consumption. By the end of high school, however, males use alcohol at substantially higher rates than females (Johnston et al. 2000:95, 99-100), thus matching the gender difference that is typical of most forms of deviant or illegal behavior (Gottfredson and Hirschi 1990). The relatively early initiation of alcohol use by girls is consistent with Caspi, Moffitt, and Silva's (1993) portrayal of early pubertal maturation of girls prompting attention from older boys, who encourage them to join in activities such as substance use and delinquency.

The example is well suited to the goals of this study because there is a sharp decline in the rate of participation in the study over the four waves of data. Of the 321 students who completed the questionnaire in the 7th grade, 211 (66 percent) did so in the 8th grade, 184 (57 percent) in the 9th grade, and 169 (53 percent) in the 10th grade. Furthermore, there are likely to be systemic differences between the students who participated for different lengths of time.

The largest decline in participation came between the first and second years of the study, and most of that attrition was due to a change from accepting passive parental consent to requiring active consent. Students could participate in the first wave as long as parents did not disapprove, but a signed parental consent form was necessary to participate in later waves. Even after extensive efforts (including letters, phone calls, and inducements to students and teachers), researchers were unable to obtain completed forms from many families. Later attrition primarily resulted from families relocating away from the target schools and not providing a new address,

which became more common as students entered high school in the 9th or 10th grades. Overall, it is likely that students who provided fewer waves of data came from families that lived less stable lives and were less cooperative with authorities. Thus, they may well be at higher risk for alcohol use.

These data come from the national evaluation of the G.R.E.A.T. gang prevention program (Esbensen et al. 2001). This evaluation included more than 3,000 students from 153 classrooms in 22 schools, located in six different cities. Although this sample is not representative in any systematic sense, it does capture much of the diversity of the United States. The research sites include many regions of the country, small to very large cities, very poor to solidly middle-class areas, and areas with very different racial and ethnic compositions. In the total sample, 51 percent of respondents were female and 49 percent male. The ethnic/racial composition of the sample was 46 percent white, 17 percent African American, 20 percent Latino, 4 percent Native American, and 4 percent Asian, with the remainder in other groups or of mixed ethnicity.

We report analyses based on a 12 percent random subsample from the larger study. The total sample is large enough that even trivial effects might reach statistical significance, so we preferred this smaller sample for illustrating the method we are presenting. Also, there is considerably less clustering by classroom and school in this smaller sample. Parametric analyses indicate only weak clustering effects in the total sample (Esbensen et al. 2001), and they should be of no consequence in the reduced sample.

The data come from self-administered questionnaires, completed under the supervision of the research team. Teachers were not present. The measure of alcohol use is a single item asking, "How many times in the last six months have you used alcohol?" Respondents responded with a specific number rather than choosing from response categories. The few highest responses were recoded to a maximum of 100. Although respondents are reporting a count of events, it would be unwise to assume that responses will follow formal statistical distributions that typify event counts, such as the Poisson or negative binomial. Almost all respondents who indicated using alcohol more than five times chose round numbers, such as multiples of 10. Limitations of memory and of time avail-

able for the task mean that these numbers are not literal counts but rather rough approximations. Thus, it is prudent to analyze these data as ordinal rather than metric. Therefore, we will use ranking methods to analyze this example.

1.2. STATISTICAL MODEL

Nonparametric models concerning rank methods for repeated measures were first investigated for the paired two-sample design in which the hypothesis was formulated through the marginal distributions of both samples. In the context of longitudinal data, the paired two-sample design corresponds to measurements on the same experimental unit at two different occasions. The marginal distributions are considered to investigate whether the observed variable has changed with time.

The first ideas to use the so-called *marginal model* to define a treatment effect and to formulate a hypothesis in a nonparametric mixed model date back to Mehra and Puri (1967); Hollander, Pledger, and Lin (1974); and Govindarajulu (1975). This approach was later extended and studied in more detail by Brunner and Neumann (1982), Thompson (1990, 1991), and Brunner and Denker (1994).

The idea underlying the nonparametric marginal model is to make use of solely the independence structure of the observations and to determine from the design which marginal distributions of the observed random vectors are identical. The questions of interest are then investigated using quantities that can be derived from the marginal distributions. Thus, one completely forgoes a parameterization and modeling of the distributions. The simplest way of modeling the independence structure is by associating independent replications with independent random variables while allowing random variables not corresponding to independent replications an arbitrary dependence structure.

To introduce the ideas and to keep the notation simple, we restrict ourselves to the case of two groups of subjects, which corresponds to the repeated-measures design underlying the example in the previous subsection. Here, two groups (boys: $i = 1$, girls: $i = 2$), each consisting of n_i experimental units (subjects), are examined on t dif-

ferent occasions (in the example, $t = 4$), where the measurements X_{ik1}, \dots, X_{ikt} (alcohol consumption) are observed on the k th subject in group i . These t repeated measures are written in vector form as $\mathbf{X}_{ik} = (X_{ik1}, \dots, X_{ikt})'$, $k = 1, \dots, n_i$. The $n = n_1 + n_2$ vectors \mathbf{X}_{ik} are assumed to be independent for $(i, k) \neq (i', k')$, with marginals

$$X_{iks} \sim F_{is}(x) = \frac{1}{2} [F_{is}^-(x) + F_{is}^+(x)], \quad k = 1, \dots, n_i, \quad (1.1)$$

where $F_{is}^-(x) = P(X_{iks} < x)$ denotes the left-continuous version, and $F_{is}^+(x) = P(X_{iks} \leq x)$ denotes the right-continuous version of the distribution function of the random variable X_{iks} . The distribution function $F_{is}(x)$, defined in (1.1), is called the *normalized version* of the distribution function and is useful to describe models for random variables with continuous as well as with discontinuous distribution functions in a unified form (Ruymgaart 1980). Clearly, if the distribution function is continuous, then all three versions are identical; that is, $F_{is} = F_{is}^- = F_{is}^+$.

The components of each vector \mathbf{X}_{ik} can be arbitrarily dependent on one another. Since the observations of different subjects within one group are considered as replications of the experiment, it is reasonable to assume that the common distribution functions $G_{ik}(\mathbf{x})$ of the vectors \mathbf{X}_{ik} are identical (i.e., they do not depend on the index k). Additional dependence structures or assumptions on the distributions will not be needed within this framework. This nonparametric model is very general and makes use exclusively of the distributions and the independence of the random vectors. Thus, the notion of group and time effects and of the interaction between them, as well as the corresponding hypotheses, must be defined by the distributions. This will be provided in the next section.

2. RELATIVE EFFECTS AND HYPOTHESES

2.1. RELATIVE EFFECTS

To describe treatment effects in the general marginal model (1.1), we generalize Birnbaum's (1956) idea of describing a nonparametric treatment effect for two independent random variables $Y_1 \sim F_1$

and $Y_2 \sim F_2$ with continuous distribution functions by the probability $p = P(Y_1 \leq Y_2)$, applying it to the case of several samples, to data with ties, as well as to the case of repeated measures.

Let $X_{ik} = (X_{ik1}, \dots, X_{ikt})'$ denote independent random vectors with marginal distribution functions $F_{is}^+(x) = P(X_{iks} \leq x)$, $i = 1, 2$, $k = 1, \dots, n_i$, $s = 1, \dots, t$. Then a treatment effect for group i at time s with respect to group r at time ℓ ($i, r = 1, 2$ and $s, \ell = 1, \dots, t$) is defined as the probability that a randomly selected observation from group r at time ℓ is less or equal than a randomly selected observation from group i at time s (if $r = i$, then the observations from different subjects k and k' must be selected); that is, $P(X_{r\ell} \leq X_{ik's})$, $(r, k) \neq (i, k')$.

We note that $P(X_{r\ell} \leq X_{ik's}) = P(X_{r1\ell} \leq X_{i2s})$, $(r, k) \neq (i, k')$, since the observations $X_{r\ell} \sim F_{r\ell}^+(x)$ and $X_{ik's} \sim F_{is}^+(x)$ are independent and identically distributed replications of the experiment. Let $N = t \cdot (n_1 + n_2)$ denote the total number of observations. Then the weighted mean

$$p_{is} = \frac{1}{N} \sum_{r=1}^2 \sum_{\ell=1}^t n_r P(X_{r1\ell} \leq X_{i2s}), \quad i = 1, 2, \quad s = 1, \dots, t, \quad (2.1)$$

is called a *relative treatment effect* of F_{is}^+ with respect to the total experiment.

Some explanation is required as to why a weighted sum of all pairwise probabilities is used in this definition. First, it should be noted that it may be difficult to draw consistent conclusions only from a subset of pairwise comparisons since they are not transitive in general. This is easily seen from the following counterexample.

Let $X_i \sim F_i^+(x)$, $i = 1, 2, 3$, where

$$F_1^+(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{3}, & 1 \leq x < 4 \\ 1, & x \geq 4 \end{cases}, \quad F_2^+(x) = \begin{cases} 0, & x < 3 \\ 1, & x \geq 3 \end{cases},$$

$$F_3^+(x) = \begin{cases} 0, & x < 2 \\ \frac{2}{3}, & 2 \leq x < 5 \\ 1, & x \geq 5 \end{cases}$$

and assume that larger values correspond to a better outcome. Then it follows that $P(X_1 \leq X_2) = \frac{1}{3}$ (i.e., F_1^+ is "better" than F_2^+),

$P(X_2 \leq X_3) = \frac{1}{3}$ (i.e., F_2^+ is “better” than F_3^+), and $P(X_3 \leq X_1) = \frac{1}{3}$ (i.e., F_3^+ is “better” than F_1^+), which appears to be a paradox result (for details, see Gardner 1970). Such inconsistent results can be avoided if all distributions are compared with the same reference distribution $H^+(x)$ of a random variable X_0 . Obviously, $H^+(x)$ must cover a wide range of possible observations if different distributions $F_i^+(x)$ should correspond to different relative effects $p_i = P(X_0 \leq X_i)$. If, for example, $H^+(x)$ is chosen as

$$H^+(x) = \begin{cases} 0, & x < -2 \\ \frac{1}{3}, & -2 \leq x < -1 \\ 1, & x \geq -1 \end{cases}$$

then $p_1 = P(X_0 \leq X_1) = p_2 = P(X_0 \leq X_2) = p_3 = P(X_0 \leq X_3) = 1$, although the pairwise relative effects for the three distributions are different from $\frac{1}{2}$. Therefore, a reasonable reference distribution should include all distributions of the experiment. One useful choice of a suitable reference distribution is the mean of all distributions in the experiment

$$H^+(x) = \frac{1}{N} \sum_{r=1}^2 \sum_{\ell=1}^t n_r F_{r\ell}^+(x) = \frac{1}{N} \sum_{r=1}^2 \sum_{\ell=1}^t n_r P(X_{r1\ell} \leq x),$$

where the distribution functions $F_{r\ell}^+(x)$ are weighted by the sample sizes n_1 and n_2 in the two experimental groups of subjects. The relative treatment effect p_{is} , defined in (2.1), can be considered a numerical quantification of the comparison between the independent random variables $X_{iks} \sim F_{is}^+(x)$ and $X_0 \sim H^+(x)$. We note that the procedures considered by Brunner and Neumann (1982), Kepner and Robinson (1988), Thompson (1990, 1991), and Akritas and Arnold (1994) are based on the relative treatment effects p_{is} , defined in (2.1).

The case of distribution functions with common points of discontinuity is taken into account by using the normalized version $F_{is}(x) = \frac{1}{2}[F_{is}^-(x) + F_{is}^+(x)]$ of the distribution function, which leads to the normalized version of the mean distribution function

$$H(x) = \frac{1}{2} [H^+(x) + H^-(x)] \quad (2.2)$$

$$\begin{aligned}
&= \frac{1}{N} \sum_{r=1}^2 \sum_{\ell=1}^t \frac{n_r}{2} [F_{r\ell}^+(x) + F_{r\ell}^-(x)] \\
&= \frac{1}{N} \sum_{r=1}^2 \sum_{\ell=1}^t n_r \left[P(X_{r1\ell} < x) + \frac{1}{2} P(X_{r1\ell} = x) \right]. \quad (2.3)
\end{aligned}$$

By using the normalized version $H(x)$, the relative effect p_{is} ($i = 1, 2, s = 1, \dots, t$) is generalized to possibly discontinuous distribution functions; that is,

$$p_{is} = \frac{1}{N} \sum_{r=1}^2 \sum_{\ell=1}^t n_r \left[P(X_{r1\ell} < X_{i2s}) + \frac{1}{2} P(X_{r1\ell} = X_{i2s}) \right]. \quad (2.4)$$

If the distribution functions $F_{is}(x)$ are continuous, then $P(X_{r1\ell} < X_{i2s}) = P(X_{r1\ell} \leq X_{i2s})$ and $P(X_{r1\ell} = X_{i2s}) = 0$, which shows that p_{is} , defined in (2.1), is just a special case of p_{is} , defined in (2.4). The procedures considered by Akritas and Brunner (1997) and by Brunner et al. (1999) are based on this generalization of the relative treatment effect p_{is} .

2.2. HYPOTHESES

To conveniently formulate the hypotheses in this design with t repeated measures and two independent groups of subjects, we use vector and matrix notation. Let $\mathbf{F} = (F_{11}, \dots, F_{1t}, F_{21}, \dots, F_{2t})'$ denote the vector of the marginal distributions, and let $\mathbf{p} = (p_{11}, \dots, p_{1t}, p_{21}, \dots, p_{2t})'$ denote the vector of the relative treatment effects. In the sequel, the situations of *no group effect*, *no time effect*, and *no interaction* between group and time are characterized by means of the distribution functions.

In the two-sample design with independent observations, the hypothesis of no treatment effect is commonly formulated as $H_0 : F_1 = F_2$ or equivalently as $H_0 : F_1 - F_2 = 0$, where 0 denotes a function that is identical 0. Let $\mathbf{F} = (F_1, F_2)'$ denote the vector of the distribution functions. Then this hypothesis can be written in vector notation as $(1, -1)\mathbf{F} = 0$.

This concept to formulate nonparametric hypotheses by means of the distribution functions has been extended to factorial designs by Akritas and Arnold (1994) and is applied to the nonparametric

marginal model in (1.1). The fixed factor group with the two levels M and F (gender) is denoted by A , and the fixed factor time with the four levels $1, \dots, 4$ (time points) is denoted by T and the interaction between these two factors by AT . The following hypotheses are considered:

1. No effect of the factor A (group): $H_0(A) : \bar{F}_{1.} = \bar{F}_{2.}$, where $\bar{F}_{i.} = \frac{1}{t} \sum_{s=1}^t F_{is}$, $i = 1, 2$.
2. No effect of the factor T (time): $H_0(T) : \bar{F}_{.1} = \dots = \bar{F}_{.t}$, where $\bar{F}_{.s} = \frac{1}{2}(F_{1s} + F_{2s})$, $s = 1, \dots, t$.
3. No interaction between the factors A and T : $H_0(AT) : F_{is} = \bar{F}_{i.} + \bar{F}_{.s} - \bar{F}_{..}$, $i = 1, 2$, $s = 1, \dots, t$, where $\bar{F}_{..} = \frac{1}{2t} \sum_{s=1}^t (F_{1s} + F_{2s})$.

To write these hypotheses equivalently in matrix notation, let $\mathbf{1}_t = (1, \dots, 1)' \in \mathbb{R}^t$ denote the t -dimensional vector with all components equal to 1, let $\mathbf{J}_t = \mathbf{1}_t \mathbf{1}_t'$ denote the $t \times t$ matrix with all elements equal to 1, let \mathbf{I}_t denote the t -dimensional identity matrix, and let $\mathbf{p}_t = \mathbf{I}_t - \frac{1}{t} \mathbf{J}_t$ denote the so-called t -dimensional centering matrix. Finally, let $\mathbf{A} \otimes \mathbf{B}$ denote the Kronecker product of the two matrices \mathbf{A} and \mathbf{B} , which means that each element of the matrix \mathbf{A} is multiplied by the matrix \mathbf{B} . Then the three hypotheses are written in matrix notation as

$$\begin{aligned} H_0(A) : ((1, -1) \otimes \frac{1}{t} \mathbf{1}_t') \mathbf{F} &= \mathbf{C}_A \mathbf{F} = 0, \\ H_0(T) : (\frac{1}{2}, \frac{1}{2} \otimes \mathbf{P}_t) \mathbf{F} &= \mathbf{C}_T \mathbf{F} = 0, \\ H_0(AT) : (\mathbf{P}_2 \otimes \mathbf{P}_t) \mathbf{F} &= \mathbf{C}_{AT} \mathbf{F} = 0. \end{aligned} \quad (2.5)$$

Similar relations follow for the relative treatment effects under these nonparametric hypotheses since from $\mathbf{C}\mathbf{F} = \mathbf{0}$, it follows that also $\mathbf{C}\mathbf{p} = \mathbf{0}$ (a formal derivation of this implication is given in Appendix 5.1). In particular, we have

$$H_0(A) : \bar{F}_{1.} = \bar{F}_{2.}, \Rightarrow \bar{p}_{1.} = \bar{p}_{2.},$$

where $\bar{p}_{i.} = \frac{1}{t} \sum_{s=1}^t p_{is}$, $i = 1, 2$.

$$H_0(T) : \bar{F}_{.1} = \dots = \bar{F}_{.t} \Rightarrow \bar{p}_{.1} = \dots = \bar{p}_{.t},$$

where $\bar{p}_{.s} = \frac{1}{2}(p_{1s} + p_{2s})$, $s = 1, \dots, t$.

$$H_0(AT) : F_{is} = \bar{F}_{i.} + \bar{F}_{.s} - \bar{F}_{..} \Rightarrow$$

$$p_{is} = \bar{p}_{i.} + \bar{p}_{.s} - \bar{p}_{..}, \quad i = 1, 2, \quad s = 1, \dots, t,$$

where $\bar{p}_{..} = \frac{1}{2t} \sum_{s=1}^t (p_{1s} + p_{2s})$.

If the random variables X_{iks} are observed on a metric scale and if the expectations $\mu_{is} = E(X_{iks})$, $i = 1, 2$, $s = 1, \dots, t$ exist, then, under the nonparametric hypotheses, similar relations hold for the expectations μ_{is} since $\mathbf{CF} = \mathbf{0} \Rightarrow \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$, where $\boldsymbol{\mu} = (\mu_{11}, \dots, \mu_{1t}, \mu_{21}, \dots, \mu_{2t})'$ denotes the vector of the expectations (for a formal derivation, see also Appendix 5.1). In particular, the implications for the three hypotheses of no group effect, no time effect, and no interaction between group and time are

$$H_0(A) : \bar{F}_{1.} = \bar{F}_{2.}, \Rightarrow \bar{\mu}_{1.} = \bar{\mu}_{2.},$$

where $\bar{\mu}_{i.} = \frac{1}{t} \sum_{s=1}^t \mu_{is}$, $i = 1, 2$.

$$H_0(T) : \bar{F}_{.1} = \dots = \bar{F}_{.t} \Rightarrow \bar{\mu}_{.1} = \dots = \bar{\mu}_{.t},$$

where $\bar{\mu}_{.s} = \frac{1}{2}(\mu_{1s} + \mu_{2s})$, $s = 1, \dots, t$.

$$H_0(AT) : F_{is} = \bar{F}_{i.} + \bar{F}_{.s} - \bar{F}_{..} \Rightarrow \\ \mu_{is} = \bar{\mu}_{i.} + \bar{\mu}_{.s} - \bar{\mu}_{..}, \quad i = 1, 2, \quad s = 1, \dots, t,$$

where $\bar{\mu}_{..} = \frac{1}{2t} \sum_{s=1}^t (\mu_{1s} + \mu_{2s})$.

This means that under the nonparametric hypothesis $H_0 : \mathbf{CF} = \mathbf{0}$, the corresponding linear contrasts \mathbf{Cp} of the relative treatment effects, as well as the corresponding linear contrasts $\mathbf{C}\boldsymbol{\mu}$ of the expectations, are equal to 0. Therefore, the hypotheses in (2.5) can be used as reasonable characterizations to describe the situations of no group effect, no time effect, or no interaction in the nonparametric model.

3. MISSING VALUES, ESTIMATORS, AND STATISTICS

3.1. STRATEGIES FOR ANALYZING DATA WITH MISSING VALUES

First we discuss some strategies for analyzing longitudinal data with missing values. Each of these can be applied in combination with the estimators and statistics described in the next two subsections. The simplest way is to leave out all subjects where at least one

observation is missing. This strategy is referred to as *complete case analysis* (CCA). The advantage of CCA is that there are no longer missing values in the data set that is analyzed, and all formulas for complete observations can be used. The disadvantages are that a lot of information is thrown away (in the worst case, all available data) and that the subjects for which all observations are available must be considered a random sample of all subjects in the study. In the terminology of Rubin (1976), this means that the observations are missing completely at random. Because this assumption may be at least questionable in many studies, several strategies are in use to replace the unobserved values by some reasonable values. Such more or less sophisticated estimation procedures take into account that the fact whether a measurement could be observed or not may be related to the outcome of the study. Such *imputation* methods have been developed for parametric as well as semiparametric models during the past three decades. It seems to be difficult to apply these ideas to pure nonparametric models as considered in this study, and the development of appropriate procedures has to be seen in the future.

A rather simple but nevertheless popular method of imputing unobserved values is the so-called *last observation carried forward* (LOCF) method, where the missing values are replaced with the last observed measurement. The advantage of LOCF is also that there are no longer missing values in the data set, and all formulas for complete observations can be used. The obvious disadvantage is that missing data are replaced with “guessed” data, which means that there is no real gain of information; instead, a bias may be caused by the virtual data. Moreover, the assumption of identically distributed random vectors \mathbf{X}_{ik} within group i is no longer true since observations carried forward are completely dependent on the last observed value.

A third method, the so-called *complete set analysis* (CSA), uses all data that have been observed. This leads to technically difficult statistical procedures. Moreover, the estimated covariance matrix may no longer meet the basic properties of a covariance matrix. In particular, it may happen that the variances and covariances do not fulfill the Cauchy-Schwarz inequality and/or that the covariance matrix is not positive semidefinite. This requires rather complicated

correction procedures (see the remark in Section 3.3). The advantage is that all information that is available is used. It may be noted, however, that even CSA requires that the observations are missing completely at random.

Currently, the experimenter is left with only one practical recourse when faced with the problem of missing values in the framework of a nonparametric model: One must carry out different analyses and then compare and discuss the results obtained. A careful subgroup analysis may improve the interpretation of these results. Such subgroups can be obtained from a stratification of the data, where the strata are defined by those subjects who drop out after the first time point, after the second time point, and so on. The data are then analyzed for those earlier time points when the groups are complete. This technique will be briefly demonstrated in the next subsection.

3.2. ESTIMATORS FOR THE RELATIVE EFFECTS

Statistics for testing the nonparametric hypotheses considered in Section 2 are based on consistent estimators of the relative marginal effects p_{is} . If there are no missing values in the data set, the notation introduced in Section 1.2 will be sufficient for the derivation of the estimators and statistics. This is also true if there are missing values and the CCA or the LOCF strategy is used to create a complete data set. If, however, the CSA strategy is used, we need further notation. To indicate whether an observation is missing or not, indicators λ_{iks} are introduced; that is,

$$\lambda_{iks} = \begin{cases} 1, & \text{if } X_{iks} \text{ is observed,} \\ 0, & \text{if } X_{iks} \text{ is missing.} \end{cases}$$

The indicators λ_{iks} can be considered as random variables. If, however, the observations are *missing completely at random* (Rubin 1976), the indicators are independent of the observations. We can therefore carry out the analysis conditionally on the indicators and treat them as fixed.

Furthermore, note that the new notation with the indicators λ_{iks} is a generalization of the notation in Section 1.2. If there are no missing values in the original data set or in the data set created by

the CCA or LOCF strategy, all indicators are equal to 1. Therefore, we can use the more general notation, regardless of whether there are missing values or which strategy for handling missing values is used.

Denote by $\lambda_{i,s} = \sum_{k=1}^{n_i} \lambda_{iks}$ the number of available observations at time s in group i , and let $N = \sum_{i=1}^2 \sum_{k=1}^{n_i} \sum_{s=1}^t \lambda_{iks}$ denote the number of all observations X_{iks} . Then the relative treatment effect p_{is} , defined in (2.4), can be estimated from the ranks of the observations by

$$\hat{p}_{is} = \frac{1}{N\lambda_{i,s}} \sum_{k=1}^{n_i} \lambda_{iks} \left(R_{iks} - \frac{1}{2} \right) = \frac{1}{N} \left(\bar{R}_{i,s} - \frac{1}{2} \right),$$

where $\bar{R}_{i,s} = \lambda_{i,s}^{-1} \sum_{k=1}^{n_i} \lambda_{iks} R_{iks}$, $i = 1, 2$, $s = 1, \dots, t$, and R_{iks} denotes the rank of X_{iks} among all N observed values. If X_{iks} is missing, then $R_{iks} = 0$, for convenience. Thus, $\bar{R}_{i,s}$ is the mean in group i at time s of the ranks of the N observed X_{iks} . For a derivation of the estimators \hat{p}_{is} , see Appendix 5.2. The estimator \hat{p}_{is} is consistent for the relative effect p_{is} if the minimal number of subjects at time s in group i for which observations are available tends to infinity (for details, see Appendix 5.3 and Proposition 3.1 of Brunner et al. 1999).

The vector of the estimators \hat{p}_{is} for study group i is denoted by $\hat{\mathbf{p}}_i = (\hat{p}_{i1}, \dots, \hat{p}_{it})'$, and $\hat{\mathbf{p}} = (\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2)'$ is the vector of the estimators in both study groups. Let $\bar{\mathbf{R}}_i = (\bar{R}_{i,1}, \dots, \bar{R}_{i,t})'$ denote the vector of the rank means $\bar{R}_{i,s}$ for the t repeated measures within group i . Then, the rank representation of $\hat{\mathbf{p}}_i$ is given by

$$\hat{\mathbf{p}}_i = \frac{1}{N} \left(\bar{R}_{i,1} - \frac{1}{2}, \dots, \bar{R}_{i,t} - \frac{1}{2} \right)' = \frac{1}{N} \left(\bar{\mathbf{R}}_i - \frac{1}{2} \mathbf{1}_t \right). \quad (3.1)$$

The estimators \hat{p}_{is} of the relative marginal effects p_{is} can be used to describe the results of the study. The quantity \hat{p}_{is} estimates the probability that a randomly selected observation from the total data set (both groups, all subjects, and all time points) is smaller than a randomly selected observation at time point s within group i .

For the alcohol consumption study, this means that the lower the value of \hat{p}_{is} , the lower the alcohol consumption at time point s in group i compared with the consumption at all time points in the total

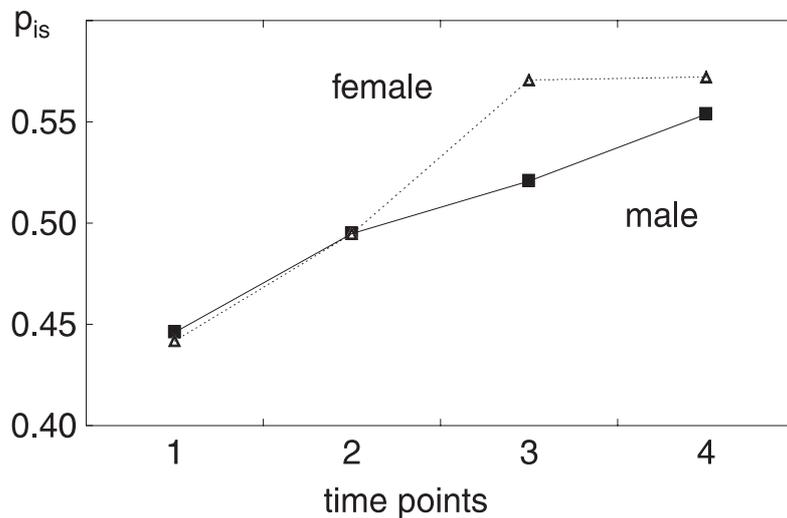


Figure 1: Time Profiles of the Relative Marginal Effects p_{is} for Alcohol Consumption in the Two Study Groups

study. If we apply the CSA strategy to the example, the rank means for the groups M ($i = 1$) and F ($i = 2$) are

$$\bar{\mathbf{R}}_1 = (395.20, 438.42, 461.39, 490.63)',$$

$$\bar{\mathbf{R}}_2 = (391.49, 438.30, 505.55, 506.92)',$$

respectively. By (3.1), this leads to the estimators

$$\hat{\mathbf{p}}_1 = (0.46, 0.49, 0.52, 0.55)', \quad (\text{group } M),$$

$$\hat{\mathbf{p}}_2 = (0.44, 0.49, 0.57, 0.57)', \quad (\text{group } F)$$

the relative marginal effects. The time curves of these estimators \hat{p}_{is} for the two study groups M and F are displayed in Figure 1.

One draws from Figure 1 that the relative marginal effects seem to increase during time in both study groups, while no uniform effect of the gender of the subjects is to be seen. Rather, there seems to be an interaction between gender and time (i.e., differing age trends for boys and girls). To evaluate this descriptive impression by an

inferential analysis, the distributional aspects of the estimators $\hat{\mathbf{p}}_i$ shall be considered below.

3.3. TEST STATISTICS

For testing the general nonparametric hypothesis $H_0 : \mathbf{CF} = \mathbf{0}$, we first consider the large sample distribution of the contrast vector of the relative marginal effects $n^{1/2} \mathbf{C}\hat{\mathbf{p}}$, where $n = n_1 + n_2$ denotes the total number of subjects. In the case that no observations are missing, Akritas and Brunner (1997) have shown that for large sample sizes, the distribution of $n^{1/2} \mathbf{C}\hat{\mathbf{p}}$ has a multivariate normal distribution with expectation $\mathbf{0}$ and covariance matrix $\mathbf{C}\mathbf{V}_n\mathbf{C}'$ under the hypothesis $H_0 : \mathbf{CF} = \mathbf{0}$. The matrix \mathbf{V}_n is a covariance matrix that has a block-diagonal structure and is given by

$$\mathbf{V}_n = \begin{pmatrix} \frac{n}{n_1} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \frac{n}{n_2} \mathbf{V}_2 \end{pmatrix}. \quad (3.2)$$

The unknown covariance matrices \mathbf{V}_i , $i = 1, 2$, can be consistently estimated by the sample covariance matrices of the ranks R_{iks} . The only assumption needed to show this result is that the ratios $n/n_i \leq N_0 < \infty$, $i = 1, 2$ of the total number of subjects n and number n_i of the subjects in the study group i remain bounded when $n \rightarrow \infty$. In the case of completely at random missing values, a similar result has been derived by Brunner et al. (1999) under the assumption that $n/\lambda_0 \leq N_0 < \infty$, where λ_0 denotes the minimal number of subjects at time s in group i for which observations are available. For details, we refer to Appendix 5.3.

Remark. The estimation of the covariance matrix $\hat{\mathbf{V}}_i$ can yield a matrix that is not positive semidefinite, if the CSA strategy is applied in case of missing values. Numerous suggestions on the modification of the matrix ensuring positive semidefiniteness can be found in the literature (see, e.g., Rousseeuw and Molenberghs 1993). In any case, great caution is necessary in the analysis and interpretation when many observations are missing.

To test the global hypotheses of no group or time effect or of no interaction, we shall use the quadratic forms $Q_n(\mathbf{C})$ and $F_n(\mathbf{C})$ given

below in (3.3) and (3.4), respectively, while hypotheses concerning a conjectured trend or paired comparisons are tested by the linear form $T_n(\mathbf{w})$, given in 3.5.

For testing the global hypothesis $H_0 : \mathbf{C}\mathbf{F} = \mathbf{0}$, the Wald-type statistic (WTS) $Q_n(\mathbf{C})$ is commonly used. This statistic is a quadratic form in the contrasts $n^{1/2} \mathbf{C}\hat{\mathbf{p}}$, where a generalized inverse $[\mathbf{C}\hat{\mathbf{V}}_n\mathbf{C}']^-$ of the estimated covariance matrix $\mathbf{C}\hat{\mathbf{V}}_n\mathbf{C}'$ is used to generate the quadratic form. Note that it is not possible to replace the generalized inverse by a simple inverse in general because the matrix \mathbf{C} might be singular. It follows from Theorem 5.1 in Appendix 5.3 that the statistic

$$Q_n(\mathbf{C}) = n \cdot \hat{\mathbf{p}}' \mathbf{C}' [\mathbf{C}\hat{\mathbf{V}}_n\mathbf{C}']^- \mathbf{C}\hat{\mathbf{p}} \quad (3.3)$$

has, asymptotically, a central $\chi^2_{\text{rank}(\mathbf{C})}$ -distribution under the hypothesis H_0 if \mathbf{V}_n is of full rank. However, extremely large sample sizes are needed to achieve an acceptable approximation by this distribution. Therefore, we will not use the WTS but the so-called ANOVA-type statistic. This statistic is described below.

The problems appearing with Wald-type statistics are mainly caused by estimating the covariance matrix \mathbf{V}_n , in particular, when all elements of \mathbf{V}_n must be estimated. The idea, therefore, is to leave out the estimated matrix $\hat{\mathbf{V}}_n$ and consider the quadric form $Q_n^A(\mathbf{C}) = n \cdot \hat{\mathbf{p}}' \mathbf{M} \hat{\mathbf{p}}$, where $\mathbf{M} = \mathbf{C}'[\mathbf{C}\mathbf{C}']^- \mathbf{C}$, and $[\mathbf{C}\mathbf{C}']^-$ denotes some generalized inverse of $\mathbf{C}\mathbf{C}'$. Note that \mathbf{M} is a projection matrix and that $\mathbf{M}\mathbf{F} = \mathbf{0} \Leftrightarrow \mathbf{C}\mathbf{F} = \mathbf{0}$ because $\mathbf{C}'[\mathbf{C}\mathbf{C}']^-$ is a generalized inverse of \mathbf{C} . Thus, it is reasonable to use the quadratic form $Q_n^A(\mathbf{C})$ as a statistic for testing the hypothesis $H_0 : \mathbf{C}\mathbf{F} = \mathbf{0}$. The large sample distribution of $Q_n^A(\mathbf{C})$ can be approximated by a scaled χ^2 -distribution. The same approximation technique is used for estimating the degrees of freedom for t -test with unequal variances. The results are summarized below.

Under $H_0 : \mathbf{C}\mathbf{F} = \mathbf{0}$, the first two moments of the ANOVA-type statistic (ATS),

$$F_n(\mathbf{C}) = \frac{n}{\text{tr}(\mathbf{M}\mathbf{V}_n)} \hat{\mathbf{p}}' \mathbf{M} \hat{\mathbf{p}} \quad (3.4)$$

and of the $F(f, \infty)$ -distribution coincide asymptotically for

$$f = \frac{[\text{tr}(\mathbf{M}\mathbf{V}_n)]^2}{\text{tr}(\mathbf{M}\mathbf{V}_n\mathbf{M}\mathbf{V}_n)},$$

where the unknown traces $\text{tr}(\mathbf{M}\mathbf{V}_n)$ and $\text{tr}(\mathbf{M}\mathbf{V}_n\mathbf{M}\mathbf{V}_n)$ can be estimated consistently by replacing \mathbf{V}_n with $\hat{\mathbf{V}}_n$, given in (5.3) in Appendix 5.3. For the derivation of these results see, for example, Brunner et al. (1999).

Remark. This approximation dates back to Box (1954) and turns out to be quite accurate for independent observations (see Brunner, Dette, and Munk 1997). For repeated measures, the estimator \hat{f} may be slightly biased for small sample sizes.

To derive test statistics that are especially sensitive to a conjectured patterned alternative, the method of Page (1963) and Hettmansperger and Norton (1987) was extended to repeated-measures designs by Akritas and Brunner (1996). The idea is to weight the estimated treatment effects by a set of constants w_{11}, \dots, w_{2t} reproducing the conjectured pattern of the alternative, which has to be specified in advance. Let $\mathbf{w} = (w_{11}, \dots, w_{2t})$ denote the vector of the weights w_{is} . Then, under $H_0 : \mathbf{C}\mathbf{F} = \mathbf{0}$, it follows from Theorem 5.1 in Appendix 5.3 that the linear rank statistic $L_n(\mathbf{w}) = n^{1/2} \mathbf{w}' \mathbf{C} \hat{\mathbf{p}}$ has, asymptotically, a normal distribution with mean 0 and variance $\sigma_w^2 = \mathbf{w}' \mathbf{C} \mathbf{V}_n \mathbf{C}' \mathbf{w}$, which can be estimated consistently by $\hat{\sigma}_w^2 = \mathbf{w}' \mathbf{C} \hat{\mathbf{V}}_n \mathbf{C}' \mathbf{w}$, where $\hat{\mathbf{V}}_n$ is given in (5.3) in Appendix 5.3. Finally, it follows that the linear form

$$T_n(\mathbf{w}) = L_n(\mathbf{w}) / \sigma_w = \frac{n^{1/2}}{\hat{\sigma}_w} \mathbf{w}' \mathbf{C} \hat{\mathbf{p}} \quad (3.5)$$

has, asymptotically, a standard normal distribution under $H_0 : \mathbf{C}\mathbf{F} = \mathbf{0}$.

In the next section, we apply these statistics to the repeated-measures design underlying the alcohol consumption study.

TABLE 1: Number of Completed Questionnaires During the Course of the Study Stratified With Respect to Gender

<i>Gender</i>	<i>Year</i>			
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
Male	165	104	93	80
Female	156	107	91	89
Total	321	211	184	169

TABLE 2: Relative Marginal Effects of Alcohol Consumption Over the Four Time Points in the Two Study Groups of Boys and Girls Computed With the Three Different Strategies: Complete Case Analysis (CCA), Last Observation Carried Forward (LOCF), and Complete Set Analysis (CSA)

<i>Time</i>	<i>Strategy</i>					
	<i>CCA</i>		<i>LOCF</i>		<i>CSA</i>	
	<i>M</i>	<i>F</i>	<i>M</i>	<i>F</i>	<i>M</i>	<i>F</i>
1	0.45	0.44	0.46	0.45	0.45	0.44
2	0.48	0.45	0.48	0.49	0.49	0.49
3	0.52	0.54	0.50	0.54	0.52	0.57
4	0.55	0.57	0.52	0.56	0.55	0.57

4. ANALYSIS OF THE EXAMPLE

4.1. RESULTS

In the example of the alcohol study, we have two groups of subjects (boys and girls) who report their alcohol consumption at four time points (7th through 10th grades). There is a considerable amount of missing data and dropouts. Only a total of 60 male and 60 female participants reported data for all four time points. The number of observations available at each time point is listed in 1, while the relative marginal effects for the three strategies are displayed in Table 2. The analysis of the data is performed by the nonparametric methods described in the previous sections.

First, a potential interaction between gender and time is investigated where the hypothesis $H(AT)$ in (2.5) is tested by the ATS

TABLE 3: ANOVA-Type Statistics and p -Values for the Alcohol Study Using the Three Strategies: Complete Case Analysis (CCA), Last Observation Carried Forward (LOCF), and Complete Set Analysis (CSA)

<i>Strategy</i>	<i>Factor</i>		<i>Statistic</i>	<i>d.f.</i>	<i>p-value</i>
CCA	Gender	A	0.011	1	0.9169
	Time	T	12.013	2.7	$<10^{-4}$
	Interaction	AT	0.443	2.7	0.7043
LOCF	Gender	A	0.733	1	0.3919
	Time	T	23.120	2.5	$<10^{-4}$
	Interaction	AT	1.768	2.5	0.1617
CSA	Gender	A	0.441	1	0.5065
	Time	T	14.320	2.7	$<10^{-4}$
	Interaction	AT	0.752	2.7	0.5101

$F_n(\mathbf{C})$ in (3.4). The results for all three strategies are quite similar (see Table 3).

Apparently, no interaction can be detected, and the questions of a potential gender effect or time effect can be investigated by the ATS for the hypotheses $H_0(A)$ and $H_0(T)$ in (2.5). All results are listed in Table 3. Obviously, the decisions agree for all three strategies. No effect of gender can be detected, while there is a clear time effect. The conjecture was that alcohol consumption increases with time. To assess an increasing trend, the pattern $\mathbf{w} = (1, 2, 3, 4)'$ is appropriate, and the statistic $T_n(\mathbf{w})$ given in (3.5) is computed. For all three strategies, a highly significant increasing trend is detected. The results are $T_n(\mathbf{w}) = 4.96$ (p -value $< 10^{-4}$) for CCA, $T_n(\mathbf{w}) = 6.47$ (p -value $< 10^{-4}$) for LOCF, and $T_n(\mathbf{w}) = 5.52$ (p -value $< 10^{-4}$) for CSA.

Some caution is necessary for the interpretation of the results since about 50 percent of the data are missing at the end of the study. To assess the question of a potential selection process causing the missingness of the data, which was already discussed in Section 1.1, a subgroup analysis is performed. This investigation, however, should not be considered a confirmatory analysis but rather an exploratory data inspection.

To investigate the question of whether those participants who drop out after the third time point have different time profiles than

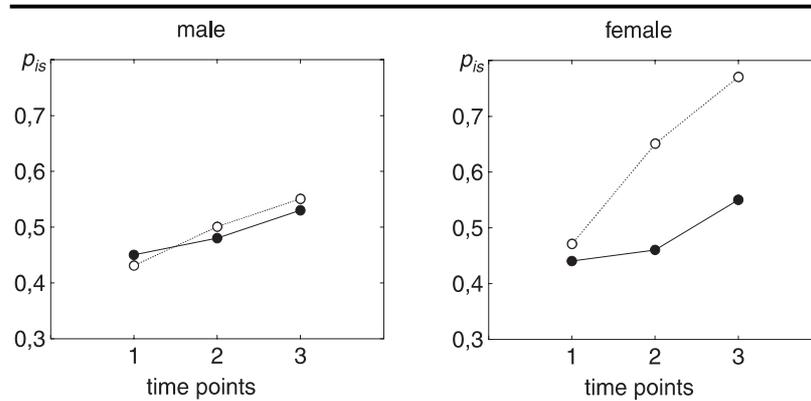


Figure 2: Relative Marginal Effects for the Male (left) and Female (right) Participants With Complete Data (solid line) and for Those Dropping Out After the Third Time Point (dashed line)

those who completed the questionnaire at all four time points, the relative marginal effects are computed for the male and female participants separately for the first three time points. The results are displayed in Figure 2.

Obviously, all time profiles are nearly identical—except that of the female participants who drop out after the third time point. This is a strong indication that the observations might not be missing completely at random. This conjecture should be confirmed by an appropriate analysis for the remaining part of the total study.

5. APPENDIX

5.1. INTEGRAL REPRESENTATION OF RELATIVE EFFECTS

To show the general relations $\mathbf{CF} = \mathbf{0} \Rightarrow \mathbf{Cp} = \mathbf{0}$ and $\mathbf{CF} = \mathbf{0} \Rightarrow \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ as stated in Section 2.2, we use the integral representation of the relative effect. Let $H(x) = N^{-1} \sum_{i=1}^2 \sum_{s=1}^t n_i F_{is}(x)$ denote the weighted average of all marginal distribution functions, where $N = t \cdot (n_1 + n_2)$ denotes the total number of observations. If the random variables $X_{r1\ell}$ and X_{i2s} are independent, then

$$P(X_{r1\ell} < X_{i2s}) + \frac{1}{2}P(X_{r1\ell} = X_{i2s}) = \int F_{rl} dF_{is}$$

and by the representation of $H(x)$ in (2.2), it follows that

$$p_{is} = \int H dF_{is}, \quad i = 1, 2, \quad s = 1, \dots, t. \quad (5.1)$$

Thus, the vector of the relative treatment effects $\mathbf{p} = (p_{11}, \dots, p_{1t}, p_{21}, \dots, p_{2t})'$ can be formally written as $\mathbf{p} = \int H d\mathbf{F}$.

To see the relation of the hypotheses in (2.5) to the relative marginal effects p_{is} , we note that for any matrix \mathbf{C} with $2 \cdot t$ columns, $\mathbf{C}\mathbf{p} = \mathbf{C} \int H d\mathbf{F} = \int H d(\mathbf{C}\mathbf{F})$, which proves the statement that $\mathbf{C}\mathbf{F} = \mathbf{0} \Rightarrow \mathbf{C}\mathbf{p} = \mathbf{0}$.

To prove a similar relation for the expectations $\boldsymbol{\mu} = (\mu_{11}, \dots, \mu_{1t}, \mu_{21}, \dots, \mu_{2t})'$, assume that the random variables X_{iks} are defined on a metric scale and that $\mu_{is} = E(X_{i1s} < \infty)$. Then $\boldsymbol{\mu}$ can be formally written as $\boldsymbol{\mu} = \int x d\mathbf{F}$, and it follows that $\mathbf{C}\boldsymbol{\mu} = \int x d(\mathbf{C}\mathbf{F})$ and $\mathbf{C}\mathbf{F} = \mathbf{0} \Rightarrow \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$.

5.2. DERIVATION OF THE ESTIMATORS

To estimate the relative effect $p_{is} = \int H dF_{is}$, the functions $H(x)$ and $F_{is}(x)$ are replaced by their empirical counterparts. Let

$$\hat{F}_{is}(x) = \frac{1}{\lambda_{i.s}} \sum_{k=1}^{n_i} \lambda_{iks} \frac{1}{2} [\mathbb{I}\{x > X_{iks}\} + \mathbb{I}\{x \geq X_{iks}\}], \quad (5.2)$$

$$i = 1, 2, \quad s = 1, \dots, t,$$

denote the empirical distribution function at time s in group i , where $\mathbb{I}\{\cdot\}$ denotes the indicator function. The mean distribution function $H(x)$ and the corresponding empirical distribution function $\hat{H}(x)$ are thus given by

$$H(x) = \frac{1}{N} \sum_{i=1}^2 \sum_{s=1}^t \lambda_{i.s} F_{is}(x),$$

$$\hat{H}(x) = \frac{1}{N} \sum_{i=1}^2 \sum_{s=1}^t \lambda_{i.s} \hat{F}_{is}(x),$$

where $N = \sum_{i=1}^2 \sum_{k=1}^{n_i} \sum_{s=1}^t \lambda_{iks}$ denotes the number of all observed values X_{iks} . Thus, one obtains the “plug-in” estimator

$$\begin{aligned} \hat{p}_{is} &= \int \hat{H}d\hat{F}_{is} = \frac{1}{\lambda_{i.s}} \sum_{k=1}^{n_i} \lambda_{iks} \hat{H}(X_{iks}) \\ &= \frac{1}{N\lambda_{i.s}} \sum_{k=1}^{n_i} \lambda_{iks} \left(R_{iks} - \frac{1}{2} \right) = \frac{1}{N} \left(\bar{R}_{i.s} - \frac{1}{2} \right), \end{aligned}$$

for $p_{is} = \int HdF_{is}$, where $\bar{R}_{i.s} = \lambda_{i.s}^{-1} \sum_{k=1}^{n_i} \lambda_{iks} R_{iks}$, $i = 1, 2$, $s = 1, \dots, t$. Here, R_{iks} denotes the rank of X_{iks} among all N observed values since $R_{iks} = N \cdot \hat{H}(X_{iks}) + \frac{1}{2}$.

5.3. ASYMPTOTIC RESULTS

The estimator \hat{p}_{is} is asymptotically unbiased and consistent for the relative effect p_{is} in the sense that $E(\hat{p}_{is} - p_{is})^2 \rightarrow 0$ if $\min_{i,s} \lambda_{i.s} \rightarrow \infty$ (see Proposition 3.1 of Brunner et al. 1999). The asymptotic normality of the contrast $n^{1/2} \mathbf{C}\hat{\mathbf{p}}$ under $H_0 : \mathbf{C}\mathbf{F} = \mathbf{0}$ is stated in the following theorem.

Theorem 5.1. Let $\mathbf{X}_{ik} = (X_{ik1}, \dots, X_{ikt})'$, $i = 1, 2$, $k = 1, \dots, n_i$, be independent random vectors, and let $n = n_1 + n_2$. Let $F_{is}(x)$, $\hat{F}_{is}(x)$, λ_{iks} , $H(x)$, $\hat{H}(x)$, p_{is} , and \hat{p}_{is} be as defined earlier. Furthermore, let $\lambda_0 = \min_{i,s} \lambda_{i.s}$ denote the minimal number of the observed data in group i at time s . If $n/\lambda_0 \leq n_0 < \infty$, then, under the hypothesis $H_0 : \mathbf{C}\mathbf{F} = \mathbf{0}$, the statistic $n^{1/2} \mathbf{C}\hat{\mathbf{p}}$ has, asymptotically, a multivariate normal distribution with expectation $\mathbf{0}$ and covariance matrix $\mathbf{C}\mathbf{V}_n\mathbf{C}'$. The unknown covariance matrix \mathbf{V}_n is consistently estimated by

$$\hat{\mathbf{V}}_n = \begin{pmatrix} \frac{n}{n_1} \hat{\mathbf{V}}_1 & \mathbf{0} \\ \mathbf{0} & \frac{n}{n_2} \hat{\mathbf{V}}_2 \end{pmatrix},$$

where for $i = 1, 2$, the elements of $\hat{\mathbf{V}}_i = (\hat{v}_i(s, s'))_{s, s'=1, \dots, t}$ are estimated by means of the ranks R_{iks} ; that is,

$$\hat{v}_i(s, s) = \frac{n_i}{N^2 \lambda_{i.s} (\lambda_{i.s} - 1)} \sum_{k=1}^{n_i} \lambda_{iks} (R_{iks} - \bar{R}_{i.s})^2,$$

$$\hat{v}_i(s, s') = \frac{n_i}{N^2 K_i(s, s')} \sum_{k=1}^{n_i} \lambda_{iks} \lambda_{iks'} (R_{iks} - \bar{R}_{i.s})(R_{iks'} - \bar{R}_{i.s'}),$$

where $K_i(s, s') = (\lambda_{i.s} - 1)(\lambda_{i.s'} - 1) + \Lambda_i(s, s') - 1$ and where $\Lambda_i(s, s') = \sum_{k=1}^{n_i} n_i \lambda_{iks} \lambda_{iks'}$.

Proof. See Brunner et al. (1999).

5.4. SPECIAL RESULTS FOR THE DESIGN OF THE EXAMPLE

Below, we give the explicit quadratic forms for the ATS to test the hypotheses of no group effect, no time effect, and no interaction. Moreover, a linear form that is particularly sensitive to an increasing trend over the t time points is derived from (3.5).

Let $\bar{R}_{i.} = \frac{1}{t} \sum_{s=1}^t \bar{R}_{i.s}$, $i = 1, 2$ denote the unweighted mean of the rank means $\bar{R}_{i.s}$ in group i . Then, in the case of two study groups, straightforward computation shows that for the group effect A , the estimated degree of freedom is $\hat{f} = 1$, and the statistic $F_n(\mathbf{C})$ in (3.4) reduces to

$$F_n(\mathbf{C}_A) = \frac{n}{\text{tr}(\mathbf{M}_A \hat{\mathbf{V}}_n)} \hat{\mathbf{p}}' \mathbf{M}_A \hat{\mathbf{p}} = \frac{n \cdot t}{2N^2 \cdot S_n} (\bar{R}_{1.} - \bar{R}_{2.})^2, \quad (5.3)$$

where, $S_n = \text{tr}(\mathbf{M}_A \hat{\mathbf{V}}_n)$, $\mathbf{M}_A = \mathbf{C}'_A [\mathbf{C}_A \mathbf{C}'_A]^{-1} \mathbf{C}_A$ and $\mathbf{C}_A = (1, -1) \otimes \frac{1}{t} \mathbf{1}'_t$, as given in (2.5). Under $H_0(A) : \bar{F}_{1.} = \bar{F}_{2.}$, the statistic $F_n(\mathbf{C}_A)$ has, approximately, a central χ^2_1 -distribution.

In the same way, the statistic for testing the time effect

$$\begin{aligned} F_n(\mathbf{C}_T) &= \frac{n}{\text{tr}(\mathbf{M}_T \hat{\mathbf{V}}_n)} \hat{\mathbf{p}}' \mathbf{M}_T \hat{\mathbf{p}} \\ &= \frac{n}{2N^2 \cdot S_n} \sum_{s=1}^t [\bar{R}_{1.s} + \bar{R}_{2.s} - (\bar{R}_{1.} + \bar{R}_{2.})]^2 \end{aligned} \quad (5.4)$$

is derived from (3.4) by letting $\mathbf{C}_T = (\frac{1}{2}, \frac{1}{2}) \otimes \mathbf{P}_t$, where $S_n = \text{tr}(\mathbf{M}_T \hat{\mathbf{V}}_n)$ and $\mathbf{M}_T = \mathbf{C}'_T [\mathbf{C}_T \mathbf{C}'_T]^{-1} \mathbf{C}_T$. Under $H_0(T) : \mathbf{C}_T \mathbf{F} = \mathbf{0} \Leftrightarrow \bar{F}_{.1} = \dots = \bar{F}_{.t}$, the statistic $F_n(\mathbf{C}_T)$ has, approximately, a central $F(\hat{f}_T, \infty)$ -distribution where

$$\hat{f}_T = \frac{[\text{tr}(\mathbf{M}_T \hat{\mathbf{V}}_n)]^2}{\text{tr}(\mathbf{M}_T \hat{\mathbf{V}}_n \mathbf{M}_T \hat{\mathbf{V}}_n)}.$$

For the interaction AT , let $\tilde{R}_{..s} = \frac{1}{2} \sum_{i=1}^2 \bar{R}_{i,s}$, $\tilde{R}_{...} = \frac{1}{t} \sum_{s=1}^t \tilde{R}_{..s}$ and $\mathbf{C}_{AT} = \mathbf{P}_2 \otimes \mathbf{P}_t$ denote the contrast matrix to test the hypothesis of no interaction $H_0(AT)$, as given in (2.5). Further note that $\mathbf{M}_{AT} = \mathbf{C}_{AT}$. Then, under the hypothesis $H_0(AT) : \mathbf{C}_{AT}\mathbf{F} = \mathbf{0} \Leftrightarrow F_{is} = \bar{F}_{i..} + \bar{F}_{.s} - \bar{F}_{...}$, $i = 1, 2$, $s = 1, \dots, t$, the statistic

$$\begin{aligned} F_n(\mathbf{C}_{AT}) &= \frac{n}{\text{tr}(\mathbf{M}_T \hat{\mathbf{V}}_n)} \hat{\mathbf{p}}' \mathbf{M}_{AT} \hat{\mathbf{p}} & (5.5) \\ &= \frac{n}{N^2 \cdot \text{tr}(\mathbf{M}_T \hat{\mathbf{V}}_n)} \sum_{i=1}^2 \sum_{s=1}^t (\bar{R}_{i,s} - \bar{R}_{i..} - \tilde{R}_{..s} + \tilde{R}_{...})^2 \end{aligned}$$

has, approximately, a central $F(\hat{f}_T, \infty)$ -distribution where

$$\hat{f}_T = \frac{[\text{tr}(\mathbf{M}_{AT} \hat{\mathbf{V}}_n)]^2}{\text{tr}(\mathbf{M}_{AT} \hat{\mathbf{V}}_n \mathbf{M}_{AT} \hat{\mathbf{V}}_n)}.$$

Finally, we derive a linear form $T_n(\mathbf{w})$, which is sensitive to an increasing trend of the relative marginal effects $\bar{p}_{.1}, \dots, \bar{p}_{.t}$, where $\bar{p}_{.s} = \frac{1}{2}(p_{1s} + p_{2s})$, $s = 1, \dots, t$. Conjectured increasing trend is reflected by a set of increasing constants $w_1 < w_2 < \dots < w_t$ for the statistic $T_n(\mathbf{w})$ in (3.5). Following Hettmansperger and Norton (1987) and Akritas and Brunner (1996), we take $w_s = s$, $s = 1, \dots, t$, that is, $\mathbf{w} = (1, 2, \dots, t)'$. Then, $\mathbf{w}'\mathbf{C}_T = \frac{1}{2}(\mathbf{w}'\mathbf{P}_t, \mathbf{w}'\mathbf{P}_t)$ and $\mathbf{w}'\mathbf{P}_t = (1, 2, \dots, t) - \frac{t+1}{2}\mathbf{1}'_t$. Let $\hat{\mathbf{p}} = (\hat{p}_{.1}, \dots, \hat{p}_{.t})'$, where $\hat{p}_{.s} = \frac{1}{2N}(\bar{R}_{1,s} + \bar{R}_{2,s} - 1)$. Then, under $H_0 : \mathbf{C}_T\mathbf{F} = \mathbf{0} \Leftrightarrow \bar{F}_{.1} = \dots = \bar{F}_{.t}$, the linear form

$$T_n(\mathbf{w}) = \frac{n^{1/2}}{N\hat{\sigma}_w} \sum_{s=1}^t \left(s - \frac{t+1}{2} \right) (\bar{R}_{1,s} + \bar{R}_{2,s}) \quad (5.6)$$

has, asymptotically, a standard normal distribution where

$$\sigma_w^2 = \sum_{i=1}^2 \frac{n}{n_i} \mathbf{w}'\mathbf{P}_t \hat{\mathbf{V}}_i \mathbf{P}_t \mathbf{w},$$

and $\hat{\mathbf{V}}_i$ is given in (5.3).

REFERENCES

- Akritis, Michael G. and S. F. Arnold. 1994. "Fully Nonparametric Hypotheses for Factorial Designs: I. Multivariate Repeated Measures Designs." *Journal of the American Statistical Association* 89:336-43.
- Akritis, Michael G. and E. Brunner. 1996. "Rank Tests for Patterned Alternatives in Factorial Designs With Interactions." Pp. 277-88 in *Festschrift on the Occasion of the 65th Birthday of Madan L. Puri*, edited by E. Brunner and M. Denker. Utrecht, the Netherlands: VSP.
- . 1997. "A Unified Approach to Rank Tests in Mixed Models." *Journal of Statistical Planning and Inference* 61:249-77.
- Birnbaum, Z. W. 1956. "On a Use of the Mann-Whitney Statistic." Pp. 13-17 in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, edited by J. Neyman. Berkeley: University of California Press.
- Box, George 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.
- Brunner, Edgar and M. Denker. 1994. "Rank Statistics Under Dependent Observations and Applications to Factorial Designs." *Journal of Statistical Planning and Inference* 42:353-78.
- Brunner, Edgar, H. Dette, and A. Munk. 1997. "Box-Type Approximations in Nonparametric Factorial Designs." *Journal of the American Statistical Association* 92:1494-502.
- Brunner, Edgar, U. Munzel, and M. L. Puri. 1999. "Rank-Score Tests in Factorial Designs With Repeated Measures." *Journal of Multivariate Analysis* 70:286-317.
- Caspi, Avshalom D., Terrie E. Moffitt, and Phil A. Silva. 1993. "Unraveling Girls' Delinquency: Biological, Dispositional, and Contextual Contributions to Adolescent Misbehavior." *Developmental Psychology* 29:19-30.
- Diggle, Peter J., K.-Y. Liang, and S. L. Zeger. 1994. *Analysis of Longitudinal Data*. Oxford, UK: Oxford University Press.
- Esbensen, Finn-Aage, Dwayne W. Osgood, Terrance J. Taylor, Dana P. Lynskey, and Adrienne Freng. 2001. "How Great Is G.R.E.A.T.?: Results From a Quasi-Experimental Design." *Criminology and Public Policy* #1:87-115.
- Gardner, Martin. 1970. "The Paradox of the Nontransitive Dice and the Elusive Principle of Indifference." *Scientific American* 223:110-14.
- Gottfredson, Michael R. and Travis Hirschi. 1990. *A General Theory of Crime*. Stanford, CA: Stanford University Press.
- Govindarajulu, Zakula. 1975. "Robustness of Mann-Whitney-Wilcoxon Test to Dependence in the Variables." *Studia Scientiarum Mathematicarum Hungarica* 10:39-45.
- Hettmansperger, Thomas P. and R. M. Norton. 1987. "Tests for Patterned Alternatives in *k*-Sample Problems." *Journal of the American Statistical Association* 82:292-99.
- Hollander, Myles, G. Pledger, and P. E. Lin. 1974. "Robustness of the Wilcoxon Test to a Certain Dependency Between Samples." *Annals of Statistics* 2:177-81.
- Johnston, Lloyd D., Patrick M. O'Malley, and Jerald G. Bachman. 2000. *Monitoring the Future: National Survey Results on Drug Use, 1975-1999*. Bethesda, MD: National Institute on Drug Abuse.
- Kepner, James L. and D. H. Robinson. 1988. "Nonparametric Methods for Detecting Treatment Effects in Repeated Measures Designs." *Journal of the American Statistical Association* 83:456-61.

- Lindsey, James K. 1993. *Models for Repeated Measurements*. Oxford, UK: Oxford University Press.
- Mehra, K. L. and M. L. Puri. 1967. "Multi-Sample Analogues of Some One-Sample Tests." *Annals of Mathematical Statistics* 38:523-49.
- Page, E. B. 1963. "Ordered Hypotheses for Multiple Treatments: A Significance Test for Linear Ranks." *Journal of the American Statistical Association* 58:216-30.
- Rousseeuw, Peter J. and G. Molenberghs. 1993. "Transformation of Non Positive Semidefinite Correlation Matrices." *Communications in Statistics A* 22:965-84.
- Rubin, Donald B. 1976. "Inference and Missing Data." *Biometrika* 63:581-92.
- Ruymgaart, Frits H. 1980. "A Unified Approach to the Asymptotic Distribution Theory of Certain Midrank Statistics." Pp. 1-18 in *Statistique non Parametrique Asymptotique*, edited by J. P. Raoult. Berlin, Germany: Springer.
- Thompson, G. L. 1990. "Asymptotic Distribution of Rank Statistics Under Dependencies With Multivariate Applications." *Journal of Multivariate Analysis* 33:183-211.
- . 1991. "A Unified Approach to Rank Tests for Multivariate and Repeated Measures Designs." *Journal of the American Statistical Association* 86:410-19.

Sebastian Domhof is an assistant professor in the Department of Medical Statistics at the University of Göttingen, Germany. His current research interests are in the field of nonparametric statistics. His recent publications include the forthcoming Nonparametric Analysis of Longitudinal Data in Factorial Experiments.

Edgar Brunner is a professor in the Department of Medical Statistics at the University of Göttingen, Germany. His recent publications include the forthcoming Nonparametric Analysis of Longitudinal Data in Factorial Experiments.

D. Wayne Osgood is a professor of crime, law and justice, and Sociology at Pennsylvania State University. His research interests include quantitative methods, the evaluation of programs in juvenile justice, the transition to adulthood, and a variety of topics concerning crime and deviance such as age and crime, time use and deviance, and peer influence.