ADDIS ABABA UNIVERSITY

DEPARTMENT OF STATISTICS

DOCTORAL DISSERTATION

Rank-Based Directional Test in k-Sample Multivariate Problems

---

By:

Taddesse Kassahun Melese

Supervisors:

Prof. Eshetu Wencheko

Prof. Arne C. Bathke

A dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics (Biostatistics)

Addis Ababa University
Graduate Programs

Dissertation Title: *Rank-Based Directional Test in k-Sample Multivariate Problems*

By: Taddesse Kassahun Melese
Department: Statistics

A dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics (Biostatistics)

The dissertation was reviewed and approved by the following:[1]

Prof. Eshetu Wencheko *Supervisor*	_____ Signature
Prof. Arne C. Bathke *Supervisor*	_____ Signature
Prof. Asheber Abebe *External Examiner*	_____ Signature
Dr. Shibru Temesgen *Internal Examiner*	_____ Signature
Dr. Bedilu Alamirie *Chairman*	_____ Signature

Date Approved: July 1, 2021

[1] Signatures are on file in the Department of Statistics.

Declaration of Authorship[2]

I, Taddesse Kassahun Melese, declare that this dissertation entitled:"*Rank-Based Directional Test in k-Sample Multivariate Problems*", and the work presented in it is my own. I confirm that the use of all material from other sources has been properly and fully acknowledged, and any part of this dissertation has not been submitted for another School or degree of this University, or for a qualification at any other institution.

Signature: _____

Date of submission: July 1, 2021

[2]Signature is on file in the Department of Statistics.

"A dream doesn’t become reality through magic; it takes sweat, determination and hard work."

Colin Powell

I dedidacte this dissertation to my late father Kassahun Melese, who taught me the value of hard work and education. I can only imagine how you would have been very proud if you lived to see this day.

Addis Ababa University
College of Natural and Computational Sciences
Department of Statistics

Doctor of Philosophy

Rank-Based Directional Test in k-sample Multivariate Problems

By: Taddesse Kassahun Melese

Supervisors: Prof. Eshetu Wencheko and
Prof. Arne C. Bathke

Abstract

Data from several response variables, potentially measured on different scales, occur naturally in various practical settings such as clinical trials. Treatment differences with respect to these response variables are usually analyzed using parametric methods under the assumptions of multivariate normality and covariance homogeneity. In many situations, however, these assumptions are not fulfilled, in particular when the response variables under consideration are ordered categorical. Thus, a rank-based (nonparametric) approach which disregards the above assumptions is desirable. More importantly, an investigator may be interested to test a global hypothesis of no treatment difference versus the alternative that treatment effects are monotonically increasing (decreasing) with respect to all responses. A number of studies have been conducted to address the issue of testing directional alternatives in two or more multivariate samples both in the parametric as well as nonparametric framework. Given that the response variables are measured in a mix of metric and ordered categorical scales, the parametric methods are not suitable. In turn, most of the contributions in the area of nonparametric statistics base their inferences on several pairwise comparisons of treatments in such a way that the ranks are being computed pairwise only, that is, only between those two levels that are compared at each step. This reduces the amount of available information and is well known to potentially lead to paradoxical situations. In order to incorporate more information from multivariate data for testing directional hypotheses which involve variables measured

both in metric as well as ordered categorical scales in a unified manner we propose a new rank-based test statistic. The statistic we have derived is a multivariate generalization based on a coordinate-wise approach of a univariate test statistic proposed by Bathke (2009) for alternative patterns within a nonparametric framework. Separate ranking for different variables is employed in order to ensure invariance under monotone transformations of the responses as well as the weights describing alternative patters. Unlike most methods available in the literature, the newly introduced test handles data with ties, in particular, ordered categorical data as the underlying distribution is not required to be continuous. The test statistic introduced in this dissertation is proved to be accurate in detecting pre-specified equi-directional alternative patterns across two or more multivariate samples trough extensive simulation studies. A comparison is also made with that of the rank-sum type test for directional multivariate problems proposed by O'Brien (1984) in which the newly developed test is in par and sometimes better than the test by O'Brien. Applications to several datasets obtained from clinical trials are presented and potential extensions in different directions are discussed.

The other more interesting practical issue in directional multivariate problems is to test conjectured alternative patterns in which treatments effects are monotonically increasing for some of the responses and monotonically decreasing for others. So long as treatment effects can be specified on a priori basis, we suggest interchanging the signs of different responses and making the anticipated direction of treatment effects similar. Following this, we employ the newly developed test statistic to test treatment effects in opposite directions in two or more multivariate samples. Furthermore, we employ the closed testing principle in conjunction with the test we have proposed in order to identify on which specific responses or sets of responses the effects are actually observed. An application of this procedure is demonstrated by re-analyzing a dose-response dataset. In summary, the test developed in this dissertation can handle monotone trends based on a complete case multivariate data. Developing a test statistic which can handle umbrella alternatives, and/or incomplete cases is differed to future research.

Acknowledgments

Primarily, thanks God for all the countless blessings you have given me in every step of my life.

A very special gratitude goes to my supervisors Prof. Eshetu Wencheko and Prof. Arne C. Bathke for their consistent and constructive guidance. I owe a great debt to Prof. Eshetu for all that I learned including but not limited to punctuality, dedication, and reverence. Prof. Arne, thank you for igniting my interest as well as inspirational discussions to work on rank-based procedures. The direction and encouragement I got from both of you made the completion of this work possible.

My thanks are due to colleagues at the Department of Statistics, Addis Ababa University for the pleasant atmosphere they have provoked through times of stress. Special thanks go to Mr. Mekonnen Tadesse (Associate Prof.) for his moral support starting from the first day of beginning the PhD project, and to Mr. Zelalem Tazu for his acquiescence to be my guarantor so that the PhD study leave is get approved.

I would like to thank Prof. Asheber Abebe and Dr. Shibru Temesgen for accepting to review my dissertation as referees. Their many useful comments and suggestions helped to improve this dissertation.

I would never have reached this stage of my life without the support and motivation my sister Ejeyensu Kassahun provided me with. Thank you, and I wish you good health.

I am very grateful to have the most supportive and compassionate wife Melkam Muluken. Her endless support and patience during the entire PhD journey has been unbelievable. I also wish to thank my children Aaron, Leul and Nahom for their perpetual love and understanding when I could not treat them properly during the busy time of my PhD research.

Taddesse Kassahun Melese

Addis Ababa, Ethiopia
July 1, 2021

Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA	Analysis of Variance
cm	Centimeter
CC-BY-NC	Creative Commons Attribution Non-commercial
d.f.	Degrees of freedom
EEG	Electroencephalogram
FWE	Family wise error
g	Gram
iid	Identically independently distributed
kg	Kilogram
max	Maximum
m	Meter
min	Minimum
MANOVA	**M**ultivariate Analysis of Variance
S.No.	Serial number
vs.	Versus

List of Symbols

$\mathbb{R}$	All real numbers.
$\dot{\sim}$	Approximately distributed as.
$\in$	Belongs to.
$\mathcal{O}$	Big O.
$\xrightarrow{a.s.}$	Converges almost surely.
$\xrightarrow{p}$	Converges in probability.
$\sim$	Distributed as, distributed according to.
$\varnothing$	Empty set.
$\forall$	for all.
$\bigcap$	Intersection operator.
$\bigotimes$	Kronecker product.
$\bigoplus$	Kronecker (direct) sum.
$c^{-}(x)$	Left-continuous version of the count function.
H_0^F	Nonparametric null hypothesis defined in terms of distribution function F.
$c(x)$	Normalised version of the count function.
Pr(X)	Probability of a random variable X.
$c^{+}(x)$	Right-continuous version of the count function.
$\subseteq$	Subset
$\ni$	Such that.
$\mathbf{X}^T$	T denotes the transpose of a vector $\mathbf{X}$.
$Corr(\cdot)$	The correlation function.
$Cov(\cdot)$	The covariance function.
$\mathbb{E}(\cdot)$	The expected value function.
$\exists$	There exists.
$\therefore$	Therefore.
$tr(\cdot)$	The trace of a square matrix.

Publications

This dissertation is based on the following publications:

1) **Kassahun T.**, Wencheko E., and Bathke A. C. (2021). A rank-based test for monotone trends in k-sample multivariate problems. Communications in Statistics - Theory and Methods, *https://doi:10.1080/03610926.2021.1931333*

2) **Kassahun T.**, Wencheko E., and Bathke A. C. (2021). Nonparametric directional testing for multivariate problems based on a closed testing principle. submitted to *Metrika*

CHAPTER 1

Overview of the Dissertation

1.1 Motivation

Scientific explanations for occurrences of real life phenomena are based on empirical evidences. In an attempt to find avenues for investigating such phenomena, researchers usually collect data from two or more variables on a subject-wise basis. The need to compare different treatment groups of subjects with regard to these variables leads to a multivariate problem. Multivariate problems appear in many areas of research such as in the life sciences, engineering, business and the social sciences. For example, patients suffering from back pain may be randomly assigned into three back pain therapies (T_1, T_2, T_3), or treatment groups. Suppose that each patient received the treatment for t weeks, and at the end an investigator interviewed them about their psychological (X_1), emotional (X_2) and cognitive (X_3) pain status. The investigator is interested in finding out whether the pain status is significantly different between the three therapies which calls for multivariate analysis. Let us consider another example where a researcher is interested in determining if geographic region (*Region A*, *Region B*, *Region C* and *Region D*) has an effect on consumers' taste preferences (X_1), purchase intentions (X_2), and attitude towards product (X_3). This problems is again analysed by multivariate methods.

In each of the above examples, one can alternatively apply univariate methods

to separately assess each variable across groups although the problem of multiple tests would arise. To address this issue, a number of studies have employed the classical Multivariate Analysis of Variance (MANOVA) models assuming multivariate normality and homogeneity of the variance-covariance matrices. Some of the significant contributers for inference about mean vectors using MANOVA model include, Wilks (1946); Hotelling (1951); Pillai (1955); and Dempster (1960). The assumptions of multivariate normality and/or homogeneity of the variance-covariance matrices are often not met in practice. It may also happen that the different responses (endpoints) under consideration are measured on different scales in such a way that test statistics based on classical MANOVA models are not invariant under change of units of these variables. In addition, these classical methods of analysis can only be applied if the response variables are measured on interval or ratio scale. However, practical datasets such as the one considered by Bathke and Harrar (2008) involve a combination of response variables potentially measured on metric, ordinal or binary scales, i.e., body mass index (metric), exercise activity (binary) and education (ordinal).

As an alternative to the parametric methodology based on MANOVA, nonparametric approaches have been considered to analyze k-sample multivariate data without assuming either normality or equal covariance matrices or both by Puri and Sen (1966); Akritas and Arnold (1994); Brunner et al. (1999); Munzel and Brunner (2000b); Bathke and Harrar (2008); Bathke et al. (2008); Liu et al. (2011); Brunner et al. (2016), among others.

Each of the multivariate methods proposed by the authors discussed above, be it in a parametric or nonparametric framework, can only be applied to test whether one or more of the treatments differ. On the other hand, directional decisions involving several distinct response variables are often desired in a number of applied settings such as in clinical trials. For example, in a dose-response trial, the response might monotonically increase with dose, so that the vector of successive differences of the mean responses would be expected to be positive. This leads to a multivariate test problem with alternatives hypothesized to be in an a priori specified direction.

The topic of directional multivariate tests has been addressed by several authors (O'Brien (1984); Tsai and Koziol (1994); Silvapulle (1995); Follmann (1996);

Reitmeir and Wassmer (1996); Bregenzer and Lehmacher (1998); Minhajuddin et al. (2007); Nadar (2013)). In cases where the null hypothesis of no treatment effect is tested against equidirected alternatives of all response variables (endpoints), a rank sum type test proposed by O'Brien, is widely applicable. Despite its versatility, however, the O'Brien (1984), rank-sum test considers continuous distribution functions only although practical situations are expected to have a mixed set of continuous and discrete outcomes. Moreover, the rank-sum type test has low power when correlations are unequal (Karrison and O'Brien (2004)). To this end, there are tests proposed by, for example, Dietz (1989), Tsai and Koziol (1994), Hyun and Song (2009), and Nadar (2013). Although each of these tests can handle two or more multivariate samples, their derivation is based on pairwise comparison which potentially leads to paradoxical results as a pairwise significance may not imply an overall significance. It is thus appealing to develop a test statistic which can handle both metric and ordered categorical data in two or more multivariate samples. In a univariate setup, Hettmansperger and Norton (1987) developed nonparametric methods of testing patterned alternatives in k-sample problems involving continuous distribution function only. Later on, Bathke (2009) proposed an alternative test statistic to that of Hettmansperger and Norton (1987) statistic. This test was fully nonparametric in the sense that it analyzes both quantitative and ordered categorical data. Moreover, the test statistic developed by Bathke (2009), was invariant under monotone transformation of the weights describing alternative patterns. Therefore, the aim of the present dissertation is to develop a directional test statistic which can handle both metric and ordered categorical data in two or more multivariate samples by generalizing the test proposed by Bathke (2009).

Additionally, a more practical situation of directional multivariate problems, that is, treatment effects are a priori specified to monotonically increase for some of the responses and monotonically decrease for others is a point of discussion. This procedure is combined with the closed testing principle in order to further identify treatment effects in lower dimensional responses. The motivation to consider a closed testing principle emanates from the fact that global superiority (inferiority) can be demonstrated by using a global test statistic such as the one proposed in this thesis, but such statistics cannot show us on which of the response(s) the treatment effects

are actually significant.

1.2 Organization of the dissertation

This dissertation is composed of three Parts: Part I - theoretical introduction, Part II - procedures to test directional hypotheses, and Part III - conclusion and future research.

The first part (Chapter 2) presents an extensive introduction to rank-based tests in the context of multivariate datasets especially for a reader who is unfamiliar with these statistical methods.

The second part which encompasses Chapter 3 and Chapter 4 discusses the problem of testing multivariate directional hypotheses. In Chapter 3, the newly developed directional test statistic which can handle metric as well as ordered categorical data in a unified form from two or more multivariate samples is described. Details of mathematical derivations, simulation studies and real data applications are presented. The focus is on detection of equidirected treatment differences, that is, to test a monotonically increasing or decreasing trend with respect to all responses in two or more groups. Chapter 4 is devoted to present the method of testing a more general directional problem in two or more multivariate samples, that is, to test whether treatment effect is positive with respect to some of the responses while it is not (inferior) with respect to the remaining responses based on an a priori conjectured patters for each response. Furthermore, a closure principle in conjunction with the newly developed statistic is discussed when interest lies to identify on which of the response(s) treatment effects are actually observed.

In the last part of the dissertation, Part III, we present conclusion and potential extensions in several directions on the issues which are not addressed in this dissertation. Sample R codes (simulation as well as testing based on real data) are given in the Appendix.

Part I

Theoretical Introduction

CHAPTER 2

Basics of Rank-based Tests

In this chapter, a theoretical introduction to rank-based tests is discussed based on the books by Brunner et al. (2019), Hollander et al. (2014), Gibbons and Chakraborti (2011), Brunner et al. (2001) as well as other articles cited herein.

2.1 Introduction

Conventional methods of testing hypotheses involve parametric procedures. These hypotheses are formulated based on unknown population parameters such as mean, variance etc. which characterize theoretical probability distributions, for example normal, Poisson, etc. If the data under consideration are ordinal or ordered categorical, then the classical parametric models are not appropriate as they are not quantitative and therefore, the mean and variance are not defined. In addition, parametric tests are sensitive to violations of model assumptions which are pertinent in the derivation and construction of the tests. Suppose that the distribution under assumption for a given dataset is not true and/or the parametric assumptions are not fulfilled, then other methods, which do not require knowledge/assumption of the true distribution of data, are needed. In such cases, nonparametric or distribution-free methods are the viable alternatives to test hypotheses.

Currently, nonparametric methods are widely used in a number of applied settings because of their simplicity, requirement of minimal/no assumptions and robustness to outliers in the data. In general, nonparametric methods are popular due to the following advantages (for more see, for example, Hollander et al. (2014)).

- Nonparametric methods require few/ or no assumptions regarding the the underlying populations from which the data are obtained.
- We can obtain exact P-values for a hypothesis testing, exact coverage probabilities for confidence intervals, and exact experiment-wise error rates for multiple comparisons although the assumption of normality for the population under consideration is not satisfied.
- Nonparametric techniques are often easier to understand and apply than parametric methods and robust to the presence of outliers in the data.
- The loss of efficiency due to applying nonparametric techniques is small even when the underlying population is normal.
- Unlike parametric procedures, which require actual values of observations, many nonparametric procedures require ranks of the observations and hence they can be applied to ordered categorical data.

In rank-based procedures which are categorized under the class of nonparamatric techniques, actual values (magnitudes) of observations are transformed into their ranks. That is, actual values of observations are used to determine relative positions (ranks) in the sample. Accordingly, the set of numbers obtained when each actual value of observation is replaced by the value of some order-preserving function is called rank-order statistics. Suppose that a random variable X takes real values $\{x_1, x_2, \ldots, x_n\}$. Then the rank of x_i, $i = 1, \ldots, n$ can be obtained by counting how many values x_j, j between 1 and n, are less than or equal to x_i. Counting the number of pairs (x_i, x_j) where $x_j \leq x_i$ can be facilitated by using the count function $(c(\cdot))$ in Definition 1 below.

Definition 1 The functions

$$c^-(x) = \begin{cases} 0, x \leq 0 \\ 1, x > 0 \end{cases}, c^+(x) = \begin{cases} 0, x < 0 \\ 1, x \geq 0 \end{cases} \text{ and } c(x) = \frac{1}{2}\left[c^-(x) + c^+(x)\right]$$

are called left-continuous,right-continuous, and normalized version of the counting function, respectively.

There is a possibility that two or more observations ($x_i's$) have identical magnitudes which we call them tied. Practically ties may occur either because the variable is discrete or due to limitations on the precision of measurement in continuous variables. In such a case, the rank of observations will not be unique. This leads an investigator to choose one of the three different ranks for the tied observations, that is, either the smallest possible (minimal) rank, or the largest possible (maximal) rank, or the simple average of the minimal and maximal rank called mid-rank. One can obtain these ranks for a specific value x_i among all numbers $(x_1, \ldots, x_n)$ from counting functions as:

$$\begin{aligned} Minimal\ rank(r_i^-) &= 1 + \sum_{j=1}^{n} c^-(x_i - x_j) \\ Maximal\ rank(r_i^+) &= \sum_{j=1}^{n} c^+(x_i - x_j) \\ Mid\ rank(r_i) &= \frac{1}{2} + \sum_{j=1}^{n} c(x_i - x_j) = \frac{1}{2}\left[r_i^+ + r_i^-\right] \end{aligned} \tag{2.1}$$

Suppose that each x_i, $i = 1, \ldots, n$, is different from the others, then r_i, the rank of x_i, equals 1 plus the number of comparisons between x_i and x_j for which $x_j < x_i$ $(i \neq j)$, i.e., $c^-(x)$. Note in this case that $x_{(1)}$ denotes the smallest of $\{x_1, x_2, \ldots, x_n\}$, $x_{(2)}$ denotes the second smallest of them, . . ., and $x_{(n)}$ denotes the largest, i.e., $\{x_{(1)} < x_{(2)} < \ldots < x_{(n)}\}$ then the rank of the i^{th} observation x_i, say r_i, is the place number of x_i in the ordered collection. Alternatively, the rank of x_i can be obtained by using the indicator function:

$$Rank(x_i) = \sum_{j=1}^{n} I(x_j \leq x_i) \tag{2.2}$$

Consider a vector of observations $\mathbf{X} = (x_1, \ldots, x_n)^T$ and the rank vector $\mathbf{R} = (r_1, \ldots, r_n)^T$ associated with it. Then a statistic $T = T(\mathbf{R})$ which is a function of $\mathbf{R}$ is called a rank statistic. Hypotheses tested by rank-based statistics are invariant under any monotone transformations of response variables.

This thesis deals with a novel rank-based procedure to test directional hypotheses regarding multivariate data from several samples. Multivariate dataset incorporates two or more response variables. In the context of clinical trials, two or more response variables are referred to as multiple endpoints. The values of these endpoints are recorded for each subject or unit under consideration. Multivariate datasets are usually displayed in a rectangular format similar to spreadsheets in which the rows constitute the subjects (units) and the columns include the endpoints. Suppose that x_{ij} denotes the value of the i^{th} subject based on the j^{th} endpoint. Then measurements obtained from n subjects on p endpoints can be displayed as follows.

Subject	**Response variable 1**	...	**Response variable p**
1	x_{11}	...	x_{1p}
⋮	⋮	⋮	⋮
n	x_{n1}	...	x_{np}

TABLE 2.1: Multivariate data from n subjects and p response variables in a rectangular form

One of the two methods of ranking, namely separate ranking or overall ranking of different responses are applied for multivariate problems. In the case of overall ranking, the values of all response variables are ranked together as if they are from one response variable. However, in separate ranking, the values of each response variable are ranked in a variable-wise fashion. In this thesis, we apply separate ranking due to the following advantages.

- Tests based on separate rankings are invariant under any monotone transformation of the response variables.
- Practical multivariate data entail a mixed scale of response variables, i.e., some are metric and others are ordinal for which separate ranking can only be feasible.

- Separate ranking preserves independence when the different response variables are actually independent.

The derivation of asymptotic distribution of the test statistic proposed in this thesis applies the method of asymptotic rank transform which is introduced by Akritas (1990). In this method, it is possible to show the asymptotic equivalence of the rank statistic to another statistic which is based on a non-random transformation of the data. It is evident that the asymptotic theory for the rank-based statistics will follow as the results for the statistics based on the original observations can be used for any non-random transformation of the data.

2.2 Distribution functions

The probability that the realization of a random variable is less than or equal to a certain value can be determined from a cumulative distribution function. Formally, a distribution function, F_X for a random variable X can be defined as follows.

Definition 2 Suppose that X is a random variable. Then the left continuous, right continuous and normalized versions of the the distribution function for X are, respectively, given by

$$F^-(x) = Pr(X < x)$$

$$F^+(x) = Pr(X \leq x)$$

$$F(x) = \frac{1}{2}\left[F^-(x) + F^+(x)\right].$$

When a random variable X is continuous, the probability at a specific point is zero, i.e., $Pr(X = x) = 0$ so that $Pr(X \leq x) = Pr(X < x)$. In this case all three versions of the distribution function are identical, that is, $F^-(x) = F^+(x) = F(x)$. On the other hand, if a random variable X is discrete or ordered categorical, then specific value(s) of a response variable may be observed more than once. Hence, the probability at a specific point can exceed zero in such a way that $Pr(X \leq x) >$

$Pr(X < x)$. As a result, the usual (right continuous version) of the distribution function is different from its left continuous version. Therefore, the normalized version (see Definition 2) of a distribution function can be used to analyze both continuous as well as discrete data in particular ordered categorical data in a unified manner.

Let $\mathbf{X}_{ij} = (X_{ij}^{(1)}, X_{ij}^{(2)}, \ldots, X_{ij}^{(p)})^T$ be a $p \times 1$ vector of observations on p response variables potentially measured on different scales for the j^{th} subject in treatment i ($j = 1, 2, \ldots, n_i$, $i = 1, \ldots, k$) and T stands for transpose. Then the $k-$sample multivariate dataset layout can be given in Table 2.2. For each of the p-dimensional vectors ($\mathbf{X}_{ij}$), the multivariate distribution function is denoted by $F_i(\mathbf{x})$, and the marginal distribution function for the d^{th} ($d = 1, \ldots, p$) response variable in treatment i is denoted as $F_i^{(d)}(x)$.

Sample 1			**Sample 2**			...	**Sample k**		
$X_{11}^{(1)}$	$X_{12}^{(1)}$	$\ldots X_{1n_1}^{(1)}$	$X_{21}^{(1)}$	$X_{22}^{(1)}$	$\ldots X_{2n_2}^{(1)}$	...	$X_{k1}^{(1)}$	$X_{k2}^{(1)}$	$\ldots X_{kn_k}^{(1)}$
$X_{11}^{(2)}$	$X_{12}^{(2)}$	$\ldots X_{1n_1}^{(2)}$	$X_{21}^{(2)}$	$X_{22}^{(2)}$	$\ldots X_{2n_2}^{(2)}$	...	$X_{k1}^{(2)}$	$X_{k2}^{(2)}$	$\ldots X_{kn_k}^{(2)}$
	$\vdots$			$\vdots$		$\vdots$		$\vdots$	
$X_{11}^{(p)}$	$X_{12}^{(p)}$	$\ldots X_{1n_1}^{(p)}$	$X_{21}^{(p)}$	$X_{22}^{(p)}$	$\ldots X_{2n_2}^{(p)}$	...	$X_{k1}^{(p)}$	$X_{k2}^{(p)}$	$\ldots X_{kn_k}^{(p)}$

TABLE 2.2: Data layout for p-dimensional multivariate observations from k-samples

The empirical distribution function for the d^{th} response variable denoted $\hat{F}_i^{(d)}(x)$ consistently estimates the theoretical distribution function $F_i^{(d)}(x)$. In turn, the counting functions defined in Definition 1 can be used to calculate the three versions of empirical distribution function in the following way. Let $X_{i1}^{(d)}, \ldots, X_{in_i}^{(d)}$ be a sample of observations from $X_{ij}^{(d)} \sim F_i^{(d)}(x)$, $j = 1, \ldots, n_i$, $i = 1, \ldots, k$, $d = 1, \ldots, p$. Then the left continuous, right continuous and normalized versions of the empirical distribution function are

$$\hat{F}_i^{-(d)}(x) = \frac{1}{n_i} \sum_{j=1}^{n_i} c^-(x - X_{ij})$$

$$\hat{F}_i^{+(d)}(x) = \frac{1}{n_i}\sum_{j=1}^{n_i} c^+(x - X_{ij})$$

$$\hat{F}_i^{(d)}(x) = \frac{1}{2}\left[\hat{F}_i^-(x) + \hat{F}_i^+(x)\right],$$

respectively.

For response variable d, the average of distribution functions from k samples is given by $H^{(d)}(x) = \frac{1}{N}\sum_{i=1}^{k} n_i F_i^{(d)}(x)$, where $N = \sum_{i=1}^{k} n_i$. It is consistently estimated by $\widehat{H}^{(d)}(x) = \frac{1}{N}\sum_{i=1}^{k} n_i \hat{F}_i^{(d)}(x)$.

Define $Y_{ij}^{(d)} = H^{(d)}(X_{ij}^{(d)})$, $\hat{Y}_{ij}^{(d)} = \hat{H}^{(d)}(X_{ij}^{(d)})$ and the $p \times N$ matrix of asymptotic rank transforms $\mathbf{Y}$ as

$$\mathbf{Y} = \begin{bmatrix} Y_{11}^{(1)} & Y_{12}^{(1)} & \dots & Y_{1n_1}^{(1)} & \dots & Y_{k1}^{(1)} & Y_{k2}^{(1)} & \dots & Y_{kn_k}^{(1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{11}^{(p)} & Y_{12}^{(p)} & \dots & Y_{1n_1}^{(p)} & \dots & Y_{k1}^{(p)} & Y_{k2}^{(p)} & \dots & Y_{kn_k}^{(p)} \end{bmatrix}$$

The $p \times N$ matrix of rank transforms ($\hat{\mathbf{Y}}$) can be defined in a similar way where the element at the i^{th} row and j^{th} column is $\hat{Y}_{ij}^{(d)}$. Applying the results in Equation 2.1, the elements of the rank transform matrix, $\hat{Y}_{ij}^{(d)}$ can be determined from mid-ranks, i.e., $\hat{Y}_{ij}^{(d)} = \frac{1}{N}\left(r_{ij}^{(d)} - \frac{1}{2}\right)$. The asymptotic rank transforms defined above are independent while the rank transforms are not (cf. Akritas and Arnold (1994)).

2.3 Nonparametric models, hypotheses and effects

Non-parametric models refer to statistical models that do not often make assumptions about the population distribution or sample size to generate a model. Put in other words, the model structure of nonparametric models is not specified a priori, but rather determined from data. When we talk about the term non-parametric we do

not mean that the procedure does not incorporate parameters, rather the parameters are flexible and defined by assuming an infinite dimension.

Consider an experimental design in k groups each containing n_i units (subjects) and data are recorded from each subject based on p response variables. The general nonparametric model can be written as

$$\mathbf{X}_{ij} \sim \mathcal{P}_i, \tag{2.3}$$

where,

$$\mathbf{X}_{ij} = \begin{pmatrix} X_{ij}^{(1)} \\ X_{ij}^{(2)} \\ \vdots \\ X_{ij}^{(p)} \end{pmatrix}$$

$i = 1, \ldots, k$ represents the treatment groups, $j = 1, \ldots, n_i$ the subjects in treatment group i and $\mathcal{P}_i$ is the multivariate distribution of $\mathbf{X}_{ij}$.

The vectors $\mathbf{X}_{ij}$ are assumed to be independent while the components of each one of these vectors, that is, $X_{ij}^{(d)}$, $d = 1, \ldots, p$ can be dependent since they are measurements from the same subject. The observations of different subjects within the i^{th} group based on a d^{th} response are considered as replications of the experiment. Thus, it is reasonable to assume that the distribution functions of these vector of observations are identical, that is, they do not depend on the index j, i.e., $X_{ij}^{(d)} \sim F_i^{(d)}(x)$. Here $F_i^{(d)}(x)$ denotes the normalized version of the distribution function. In this nonparametric model, one can consider continuous, discrete quantitative as well as ordered categorical data as the normalized version of the distribution function is used. The normalized version of distribution function leads to the use of mid-ranks.

In experimental data, statistical analysis is mainly performed with the objectives of describing and estimating of differences between treatments, evaluating whether observed differences could be explained by chance alone or there is some systematic effect, and test hypotheses regarding treatment differences. Cumulative distribution function can be used to describe treatment effects. In the parametric framework, the differences between distributions can be described by differences or ratios of the

parameter values. Consider for example, data from a normal distribution. Here one can formulate treatment effects using differences or ratios of expected values. On the other hand, if the distribution function from which the data came is not (assumed to be) known, it is not always possible to quantify differences between the distributions. However, one can use differences or ratios of quantiles (e.g., median) or moments (e.g., expectation, variance, or skewness). It is evident that moments can only be determined from quantitative data. In addition, moments are not invariant under monotone transformations of the data in such a way that the model assumption (e.g., symmetry) cannot be satisfied. In turn, the median is a crude measure of location especially when the data under analysis is discrete and the number of categories is few. Also it may remain the same even though other observations are included into the data. A nonparametric (relative) treatment effect (see Definition 3) can describe differences between distributions or treatment effects, in a fully nonparametric way.

Definition 3 Consider two independently distributed random variables $X_1 \sim F_1$ and $X_2 \sim F_2$. Then the probability

$$p = Pr(X_1 < X_2) + \frac{1}{2} Pr(X_1 = X_2)$$

is called nonparametric relative effect of X_2 with respect to X_1 or relative effect of F_2 with respect to F_1.

According to the above definition, the relative effect of X_1 with respect to X_2 is equal to one minus the relative effect of X_2 with respect to X_1. Considering again two independent random variables, $X_1 \sim F_1$ and $X_2 \sim F_2$, we can have the following interpretation of the nonparametric relative effect. X_1 is said to have a stochastic tendency to take greater values than X_2 if $p < \frac{1}{2}$. On the other hand, X_1 will have a stochastic tendency to take smaller values than X_2 if $p > \frac{1}{2}$. X_1 and X_2 are stochastically comparable when $p = \frac{1}{2}$, that is, X_1 has no stochastic tendency to take greater or smaller values than X_2. Furthermore, the nonparametric relative effect equals $1/2$ if the random variables X_1 and X_2 are independent and identically distributed. Invariance under arbitrary order-preserving transformations is the other interesting property of nonparametric relative effect.

In diagnostic trials, accuracy of a procedure being implemented is measured by the area under a receiver operating characteristic curve. This quantity is the same as the nonparametric relative effect. Therefore, the methods for estimation and drawing inference about the nonparametric relative effect can also serve the same purpose for area under a receiver operating characteristic curve and the vise-versa (we refer to the book by Brunner et al. (2019) for detail discussion on nonparametric relative effects).

The nonparametric relative effect (p) defined above can also be expressed using the Lebesgue–Stieltjes integral as $p = \int F_2 dF_1$.

In general, for N independent random variables $X_i \sim F_i$, the multi-distribution nonparametric relative effect of X_i with respect to $X_1, \ldots, X_N$ can be obtained by averaging the relative effect as

$$p_i = \frac{1}{N} \sum_{l=1}^{N} \left[Pr(X_l < X_i) + \frac{1}{2} Pr(X_l = X_i) \right].$$

Equivalently, the above multi-distribution nonparametric relative effect can be given in terms of the Lebesgue–Stieltjes integral as

$$p_i = \int H dF_i, \tag{2.4}$$

where $H = \frac{1}{N} \sum_{l=1}^{N} F_l$ is the average of the cumulative distribution functions $F_1, \ldots, F_N$.

In the case of multi-distribution nonparametric relative effect given that $X_i \sim F_i$, $i = 1, \ldots, N$ are independent random variables, the variable X_i has a stochastic tendency to take greater values than X_j if $p_i > p_j$. When $p_i < p_j$ the variable X_i has a stochastic tendency to take smaller values than X_j. The two variables are said to be stochastically comparable if $p_i = p_j$.

Practical phenomena dictate that the distributions of the N independent observations are not all different. Instead n_i of them have same distribution for which the $N = \sum_{i=1}^{k} n_i$ observations are divided into k groups (treatments). As a result, the N observations $X_1, \ldots, X_N$ can be given as:

$$X_{11}, \ldots X_{1n_1}, \ldots, X_{k1}, \ldots X_{kn_k}, \text{ where } X_{ij} \sim F_i, i = 1, \ldots, k, j = 1, \ldots, n_i$$

A nonparametric treatment effect for group i having observations $X_{i1}, \ldots, X_{in_i}$ with respect to all k groups can be defined as

$$p_i = \frac{1}{N} \sum_{l=1}^{k} \sum_{j=1}^{n_l} \left[Pr(X_{lj} < X_{i1}) + \frac{1}{2} Pr(X_{lj} = X_{i1}) \right]$$

We can use either of the two possibilities for the formulation of null hypotheses based on the nonparametric model discussed above. The first is using the distribution functions and the second one is using the relative treatment effect.

Let $\mathbf{F}$ denotes the vector of marginal distribution functions from p responses in k samples.

$$\mathbf{F} = \begin{pmatrix} F_1^{(1)} \\ \vdots \\ F_1^{(p)} \\ \vdots \\ F_k^{(1)} \\ \vdots \\ F_k^{(p)} \end{pmatrix} \tag{2.5}$$

It may be assumed that the covariance matrix of $\mathbf{X}_{ij}$ and $\mathbf{X}_{ij*}$ is equal since these two vectors denote the observations of subjects within the same sample. In such a case one can formulate the nonparametric hypothesis of no treatment effect using the multivariate hypothesis $H_0^P : \mathcal{P}_1 = \ldots = \mathcal{P}_k$ (see Munzel and Brunner (2000a)). However, this multivariate hypothesis is restrictive in practical situations as it means that the variance-covariance matrices of vectors of responses in different samples are all the same which is not practically feasible in most cases. Therefore, the hypothesis based on marginal distributions is preferred, and can be defined as follows.

$$H_0^F : F_1^{(d)} = \ldots = F_k^{(d)}, \quad d = 1, \ldots, p. \tag{2.6}$$

One can reformulate the marginal hypothesis given above (see equation 2.6) in a more convenient way using the vector of marginal distributions as:

$$\mathbf{M}_k\mathbf{F} = \mathbf{0}, \tag{2.7}$$

where $\mathbf{M}_k = \mathbf{C}_k \otimes \mathbf{I}_p$, $\mathbf{I}_p$ denotes the p-dimensional identity matrix, $\mathbf{C}_k$ is the centering matrix which leads to a symmetric formulation of hypothesis

$$\mathbf{C}_k = \mathbf{I}_k - k^{-1}\mathbf{J}_k = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} - \frac{1}{k}\begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix}.$$

Suppose that $p_i^{(d)} = \int H^{(d)} dF_i^{(d)}$ denote the relative treatment effect of group i for the d^{th} response variable, where $H^{(d)} = \frac{1}{N}\sum_{i=1}^{k} n_i F_i^{(d)}$ is the weighted average of all distribution functions of the d^{th} response variable for the $i = 1, \dots, k$ groups each with sample size n_i and $N = \sum_{i=1}^{k} n_i$ is the total number of subjects in the entire sample. This formulation of $p_i^{(d)}$ can be regarded as an extension of the expected value of a random variable X, i.e., $\int x dF_i^{(d)}$. The vector of relative treatment effects for p response variables in k samples is denoted by

$$\mathbf{p} = \begin{pmatrix} p_1^{(1)} \\ \vdots \\ p_1^{(p)} \\ \vdots \\ p_k^{(1)} \\ \vdots \\ p_k^{(p)} \end{pmatrix} \tag{2.8}$$

The integral representation of vector of relative effects is:

$$\mathbf{p} = \int \mathbf{H} d\mathbf{F} \tag{2.9}$$

where $\mathbf{H} = \left(H^{(1)}, \ldots, H^{(p)}\right)^T$ is the vector of average distribution functions and $\mathbf{F}$ is as in (2.5).

Therefore, we can state the nonparametric hypothesis of the relative effects via linear contrasts as:

$$H_0^p : \mathbf{M_k p} = \mathbf{0}, \tag{2.10}$$

where $\mathbf{M_k}$ and $\mathbf{p}$ are as defined above.

2.4 Test statistics for testing nonparametric hypotheses in one-way layout

In this section, a review of the commonly used rank-based test statistics to test nonparametric hypotheses are presented for the univariate as well as multivariate cases.

2.4.1 Univariate test statistics

Consider k-independent samples in which the random variables X_{ij}, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$ are independently distributed, that is, $X_{ij} \sim F_i$. Here F_i is the normalized version of the cumulative distribution function. We can formulate the hypothesis of no difference among the k treatment effects as

$$H_0^F : F_1 = \ldots = F_k = \overline{F}., \tag{2.11}$$

where $\overline{F}. = \frac{1}{k} \sum_{i=1}^{k} F_i$.

If we are interested in the location shift effects, the cumulative distribution functions from k samples $F_1, \ldots, F_k$ are expressed as $F_i(x) = F(x - \tau_i)$, $i = 1, \ldots, k$, $-\infty < x < \infty$ and τ_i is the unknown treatment effect for the i^{th} sample. The null hypothesis of no differences among the k treatment effects can be formulated as:

$$H_0 : \tau_1 = \ldots = \tau_k \tag{2.12}$$

Kruskal–Wallis test

The Kruskal-Wallis test is one of the classical nonparametric tests which is applied to test the hypothesis of no differences among two or more treatment effects against the alternative that at least two of the treatment effects are not equal. The test statistic is proposed by Kruskal (1952) and Kruskal and Wallis (1952). Because ranks are used instead of actual observations, the test is a substitute for parametric ANOVA as long as samples are from identical and continuous distributions. In order to apply the Kruskal–Wallis test, first combine all N observations from the k samples, i.e., $X_{11},\ldots,X_{1n_1},\ldots,X_{k1},\ldots,X_{kn_k}$ and order them from least to greatest. Suppose that r_{ij} denote the rank of the j^{th}, $j = 1,\ldots,n_i$, observation in the i^{th} sample. Then the sum of the ranks for all observations in the i^{th} sample and the corresponding average rank can be obtained, respectively, from:

$$R_i = \sum_{j=1}^{n_i} r_{ij} \quad and \quad \overline{R_i} = \frac{R_i}{n_i} \tag{2.13}$$

Consider testing the null hypothesis that all distribution functions are equal against the two sided alternative, i.e.,

$$H_0^F : \mathbf{P}_k\mathbf{F} = \mathbf{0},$$

where $\mathbf{0}$ is a k-dimensional vector of zeros, $\mathbf{F} = (F_1,\ldots,F_k)^T$, and $\mathbf{P}_k$ is the centering matrix so that $\mathbf{P}_k\mathbf{F}$ can be rewritten as

$$\mathbf{P}_k\mathbf{F} = \begin{pmatrix} F_1 - \overline{F.} \\ \vdots \\ F_k - \overline{F.} \end{pmatrix}$$

Note that $H_0^F : \mathbf{P}_k\mathbf{F} = \mathbf{0}$ is equivalent to $H_0^F : F_1 = \ldots = F_k = \overline{F.}$, where $\overline{F.} = \frac{\sum_{i=1}^k F_i}{k}$.

ASSUMPTIONS

A1. The $N = \sum_{i=1}^k n_i$ random variables, $X_{1i}, \ldots, X_{in_i} \sim F_i(x)$, $i = 1, \ldots, k$ are mutually independent.

A2. $\sigma_i^2 = Var(H(X_{i1})) \geq \sigma_0^2 > 0$, where $H(\cdot)$ is the weighted average distribution function.

A3. The sample sizes grow at the same rate, that is, $N/n_i \leq N_0 < \infty$ for $N \to \infty$.

Under assumptions **(A1)** - **(A3)** and the null hypothesis H_0^F is true, the distribution of test statistic (H) tends to a central χ^2 distribution with $k-1$ degrees of freedom when $N \to \infty$ (for more, see Brunner et al. (2019), pp. 195 - 202).

$$H = \frac{N-1}{\sum_{i=1}^k \sum_{j=1}^{n_i} (r_{ij} - \frac{N+1}{2})^2} \sum_{i=1}^k n_i \left(\overline{R_i} - \frac{N+1}{2}\right)^2, \tag{2.14}$$

where n_i, $i = 1, \ldots, k$ is the number of observations in the i^{th} sample, $N = \sum_{i=1}^k n_i$, r_{ij}, and $\overline{R_i}$ are as defined above.

When ties are not present, the Kruskal-Wallis test statistic simplifies to

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k n_i \overline{R_i}^2 - 3(N+1), \tag{2.15}$$

where n_i, N, and $\overline{R_i}$ are as defined above.

The null hypothesis of no difference in treatment effects is rejected at the α level of significance if the observed value H_{cal} of $H \geq h_\alpha$. Note that h_α is the upper α percentile for the null distribution of H.

Jonckheere–Terpstra test

The null hypothesis formulated in equation (2.12) can also be tested using a test developed by Terpstra (1952) and Jonckheere (1954). However, this test is specifically recommended if an experimenter expects any deviation from the null hypothesis to be in a particular direction, i.e., either an increasing or a decreasing treatment effect.

Consider testing a hypothesis given in (2.12) versus a priori ordered alternatives $H_1 : \tau_1 \leq \ldots \leq \tau_k$ with at least one strict inequality. A test in the opposite direction can be performed by reversing all inequalities. The treatments are labeled in line with the conjectured alternative patterns. The exact permutation test can be conducted based on the following Jonckheere–Terpstra statistic (J).

$$J = \sum_{u=1}^{v-1} \sum_{v=2}^{k} U_{uv} \tag{2.16}$$

where, $U_{uv} = \sum_{i=1}^{n_u} \sum_{j=1}^{n_v} \Phi(X_{iu}, X_{jv})$ is the Mann-Whitney count relevant to the u^{th} sample ($u = 1, \ldots, k-1$) and any sample v for which $u < v$, and

$$\Phi(a,b) = \begin{cases} 1 & if\ a < b \\ 0 & if\ a > b \end{cases}$$

In the presence of ties, we redefine Φ as:

$$\Phi(a,b) = \begin{cases} 1 & if\ a < b \\ \frac{1}{2} & if\ a = b \\ 0 & if\ a = b \end{cases}$$

The null hypothesis in (2.12) is rejected at level of significance α if the observed value J_{cal} of $J \geq j_\alpha$, where j_α is chosen to make the type I error probability equal to α.

Observe that the expectation and variance of J under the null hypothesis are

$\frac{1}{4}\left(N^2 - \sum_{i=1}^{k} n_i\right)$ and $\frac{1}{72}\left(N^2(2N+3) - \sum_{i=1}^{k} n_i^2(2n_i+3)\right)$, respectively. For large sample size and under H_0, the Jonckheere-Terpstra statistic (J) is asymptotically normally distributed. That is, when $min(n_1,\ldots,n_k) \to \infty$, and the null hypothesis is true, J^* defined below has asymptotically a standard normal distribution.

$$J^* = \frac{J - (\frac{N^2 - \sum_{i=1}^{k} n_i^2}{4})}{\sqrt{\frac{N^2(2N+3) - \sum_{i=1}^{k} n_i^2(2n_i+3)}{72}}}, \tag{2.17}$$

where J is as in (2.16) and n_i is as defined above.

Hettmansperger–Norton test

Hettmansperger and Norton (1987) proposed a rank-based test for patterned alternatives based on a linear combination of ranks. The alternative hypothesis to be considered can be formulated as $H_A : \tau_i = \tau_0 + \tau c_i (\tau > 0)$, where $c_1,\ldots,c_k$ are constants that specify the pattern to be detected. For example, an odered alternative can be specified by $c_1 < \ldots < c_k$. Equally spaced constants are recommended for such ordered alternatives.

Let r_{ij} be the rank of X_{ij}, $i = 1,\ldots,k$, $j = 1,\ldots,n_i$, among the combined observations and $\overline{r_i}$ be the average rank of the i^{th} sample. Then the following optimal linear rank test is proposed by Hettmansperger and Norton.

$$HN = \sum_{i=1}^{k} \lambda_i (c_i - \bar{c}_w)\overline{r_i}, \tag{2.18}$$

where, $\lambda_i = n_i/N$, $N = \sum_{i=1}^{k} n_i$ and $\bar{c}_w = \sum_{i=1}^{k} \lambda_i c_i$.

If the null hypothesis is true and the sample size is large enough, that is, $N \to \infty$, then the null expected value $E_0(HN) = 0$, and the null variance $Var_0(HN) = [(N+1)/12] \sum_{i=1}^{k} \lambda_i (c_i - \bar{c}_w)^2$ so that the test by Hettmansperger and Norton can

be approximated by a normal distribution, i.e.,

$$HN^* = \sqrt{\frac{12}{N+1}} \frac{\sum_{i=1}^{k} \lambda_i (c_i - \bar{c}_w)\overline{r_i}}{\sqrt{\sum_{i=1}^{k} \lambda_i (c_i - \bar{c}_w)^2}} \sim N(0,1), \tag{2.19}$$

where, λ_i, n_i, N, $\overline{r_i}$, c_i, and $\bar{c}_w$ are as defined above.

Test by Bathke (2009)

Bathke (2009) proposed a univariate k-sample test statistic to test trend alternatives in which the alternative pattern was considered as a covariate whose influence is being tested. The test statistic was derived based on a test for monotone influence of covariates proposed by Bathke (2005). This test is invariant under both monotone transformations of the response variable as well as the weights describing alternative patters. It thus suffices to choose any value of the covariate which leads to a strictly monotonic trend. He considered a repeated measures design with a fixed row factor A having a levels and a fixed column factor B having b levels. Factor A represents groups with subjects nested in them, while factor B represents groups crossed with subjects, i.e., for each subject at each level of factor B, a response variable is observed. A response variable, X_{imt} which follows a distribution F_{im}, $i = 1,\ldots,a$, $m = 1,\ldots b$, $t = 1,\ldots,n_i$ was considered. A deterministic constant, x_{imt} was taken to represent the corresponding weights (explanatory variables).

2.4.2 Other univariate tests

The problem of testing against trend and umbrella alternatives, with known and unknown peaks, in two-way layouts was studied by Callegari and Akritas (2004). Two types of statistics: one using weights similar to Hettmansperger and Norton (1987),

and the other with weights which maximize the asymptotic efficacy, were considered for testing main effects with known umbrella peak. However, only the Hettmansperger–Norton type weights were considered for testing the hypothesis of no main effects against umbrella alternatives with unknown peak. The tests can be applied to continuous data, data with ties and ordinal data.

Terpstra et al. (2011) proposed a nonparametric test for detecting ordered alternatives in a completely randomized design by generalizing a test developed by Terpstra and Magel (see Terpstra and Magel (2010)). The test statistic is equivalent to the sum of Spearman correlation coefficients, where each Spearman correlation coefficient was calculated based on one observation from each treatment group, i.e., $\{(1, X_{1i_1}), (2, X_{2i_2}), \ldots, (k, X_{ki_k})\}$; $i_j = 1, 2, \ldots, n_j$, $j = 1, 2, \ldots, k$. As a result, it is possible to provide a summary measure of association between the response variable and treatment groups unlike most existing tests for ordered alternatives.

Nonparametric tests for testing scale parameters against umbrella alternative, when the peak of the umbrella is known, were proposed by Gaur et al. (2012). Different test statistics based on linear combination of two-sample U-statistics and assuming continuous distributions were used. The asymptotic relative efficiency of these tests were compared through simulating data from uniform, exponential, Cauchy, Laplace and logistic distributions.

Rank difference between two observations from different groups was considered by Shan et al. (2014), in order to develop a nonparametric test statistic for monotonic ordered alternatives. It was assumed that the actual difference can be measured by the rank difference in the nonparametric setting so that the efficiency of the test can be improved. The asymptotic distribution of the test statistic was derived assuming that the underlying distributions are continuous. Hypotheses were formulated vis á vis location parameters. The proposed test is a generalization of the well known sign test and Wilcoxon rank-sum test.

2.4.3 Multivariate test statistics

Consider the nonparametric multivariate model stated in (2.3). Let $R_{ij}^{(d)}, i = 1,\ldots,k$, $j = 1,\ldots,n_i$, $d = 1,\ldots,p$ denote the (mid-)rank of $X_{ij}^{(d)}$ of all observations, $\mathbf{R}_{ij} = (R_{ij}^{(1)},\ldots,R_{ij}^{(p)})^T$, and $\overline{\mathbf{R}}_i = n_i^{-1}\sum_{j=1}^{n_i}\mathbf{R}_{ij}$ denote the vector of rank means.

When an analyst is interested in testing the hypotheses stated in Section 2.3 versus a global (two-sided) alternative, the Wald-type statistic, ANOVA-type statistic, Lawley-Hotelling-type statistic (cf. Munzel and Brunner (2000b)), among others, can be employed. On the other hand, if the alternative hypothesis of interest is directional, a number of test statistics such as the test by Dietz and Killeen (1981), and O'Brien (1984) have been proposed.

Wald-type statistic

Suppose that the hypothesis $\mathbf{M_k F} = \mathbf{0}$ is true. Let $\lambda_1,\ldots,\lambda_{kp}$ denote the eigenvalues of $\mathbf{V} = Cov(\sqrt{N}\bar{\mathbf{Y}})$ and $min(\lambda_1,\ldots,\lambda_{kp}) \geq \lambda_0 > 0$. Suppose further that for large sample size, $N/n_i \leq N_0 < \infty$. Then the statistic $\sqrt{N}\mathbf{M_k}\hat{\mathbf{p}}$ has asymptotically a multivariate normal distribution with covariance matrix $\mathbf{M_k V M_k^T}$.

In the above, the components of $\bar{\mathbf{Y}} = (\bar{Y}_1^{(1)},\ldots,\bar{Y}_1^{p},\ldots,\bar{Y}_k^{(1)},\ldots,\bar{Y}_k^{(p)})^T$ are the means of the asymptotic rank transforms, i.e., $\bar{Y}_i^{(d)} = n_i^{-1}\sum_{j=1}^{n_i} Y_{ij}^{(d)}$, $i = 1,\ldots,k$, $d = 1,\ldots,p$, $\hat{\mathbf{p}} = (\hat{\mathbf{p}}_1^T,\ldots,\hat{\mathbf{p}}_k^T)^T$ is the estimated matrix of relative effects with $\hat{\mathbf{p}}_i = (\hat{p}_i^{(1)},\ldots,\hat{p}_i^{(p)})^T$ and $\hat{p}_i^{(d)} = \frac{\overline{R_i}^{(d)} - 1/2}{N}$, where $\overline{R_i}^{(d)} = n_i^{-1}\sum_{j=1}^{n_i} r_{ij}^{(d)}$ is the average of ranks $r_{ij}^{(d)}$ in the i^{th} sample for response d.

The nonparametric multivariate Wald-type statistic is expressed in the quadratic form,

$$Q_N^W = N\hat{\mathbf{p}}^T\mathbf{M_k}^T(\mathbf{M_k}\hat{\mathbf{V}}\mathbf{M_k}^T)^{-1}\mathbf{M_k}\hat{\mathbf{p}}, \tag{2.20}$$

where, $\hat{\mathbf{V}} = \bigoplus_{i=1}^{k}\hat{\mathbf{V}}_i$ is the consistent estimator of the covariance matrix $\mathbf{V}$ in the

sense that $||\hat{\mathbf{V}} - \mathbf{V}||_2 \to 0$, $\hat{\mathbf{V}}_i = \frac{1}{Nn_i(n_i-1)} \sum_{j=1}^{n_i} (\mathbf{R}_{ij} - \bar{\mathbf{R}}_i)(\mathbf{R}_{ij} - \bar{\mathbf{R}}_i)^T$, $\hat{\mathbf{p}}$, $\mathbf{M_k}$ and N are as defined before.

This quadratic form of Wald-type statistic is asymptotically distributed as a central χ^2 distribution with $(k-1)p$ degrees of freedom. Note that this statistic leads to a poor approximation when sample sizes are small.

ANOVA-type statistic

In the Wald-Type statistic Q_N^W given in (2.20) large sample sizes are necessary to maintain a pre-assigned level α using quantiles of the limiting χ^2 distribution. Therefore, an alternative test statistic which satisfactorily maintains the preassigned level, called ANOVA-type statistic, can be used. In order to construct an ANOVA-type statistic the simple quadratic form of analysis of variance is employed.

Consider a quadratic form $Q_N^A = N\hat{\mathbf{p}}^T \mathbf{P}_k \hat{\mathbf{p}}$, where $\mathbf{P}_k = (\mathbf{C}_k^T [\mathbf{C}_k \mathbf{C}_k^T]^{-1} \mathbf{C}_k) \otimes \mathbf{I}_p$, $\mathbf{I}_p$ and $\mathbf{C}_k$ are as in Section 2.3, and $\otimes$ denotes the Kronecker product. If the null hypothesis $\mathbf{M_k F} = \mathbf{0}$ is true, and under the assumptions considered in the construction of Wald-type statistic, $\sqrt{N}\mathbf{P}_k\hat{\mathbf{p}}$ has asymptotically a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{P}_k\mathbf{V}\mathbf{P}_k$. Thus, the distribution of Q_N^A can be determined by considering a weighted sum of independent χ^2 distributed random variables with 1 degree of freedom. That is, the distribution of Q_N^A has asymptotically the same distribution as $Z = \sum_{i=1}^{k} \sum_{d=1}^{p} v_i^{(d)} Z_i^{(d)}$, where $v_i^{(d)}$ denote the eigenvalues of $\mathbf{P}_k\mathbf{V}\mathbf{P}_k$ and $Z_i^{(d)}$ are independently distributed random variables each having a χ^2 distribution with 1 degree of freedom.

In practical situations, the values of $v_i^{(d)}$'s are unknown; and hence the sampling distribution of Q_N^A needs to be approximated. One can apply the Box (1954) approximation which is based on a scaled χ^2-distribution. In this approximation, the distribution of a random variable $g.C_f$ is considered, where C_f has a χ^2 distribution with f degrees of freedom, and g and f are constants such that the first two moments of Q_N^A and $g.C_f$ coincide. Therefore, $Q_N^A \big/ (g \cdot f)$ has approximately a χ^2/f-distribution, where $gf = tr(\mathbf{P}_k\mathbf{V})$ and $f = \frac{[tr(\mathbf{P}_k\mathbf{V})]^2}{tr(\mathbf{P}_k\mathbf{V}\mathbf{P}_k\mathbf{V})}$. Note that $tr(\mathbf{P}_k\mathbf{V})$

and $tr(\mathbf{P}_k\mathbf{V}\mathbf{P}_k\mathbf{V})$ involve unknown constants which should be estimated by replacing $\mathbf{V}$ with its consistent estimator $\hat{\mathbf{V}}$. Under $H_0^F : \mathbf{M}_k\mathbf{F} = \mathbf{0}$, the statistic

$$F_N = \frac{N}{tr(\mathbf{P}_k\hat{\mathbf{V}})}\hat{\mathbf{p}}\mathbf{P}_k\hat{\mathbf{p}} \tag{2.21}$$

has approximately a central $F(\hat{f},\infty)$ distribution, where $\hat{f} = [tr(\mathbf{P}_k\hat{\mathbf{V}})]^2 \Big/ tr(\mathbf{P}_k\hat{\mathbf{V}}\mathbf{P}_k\hat{\mathbf{V}})$, N, $\hat{\mathbf{p}}$, $\mathbf{P}_k$ and $\hat{\mathbf{V}}$ are as defined above.

Lawley-Hotelling-type statistic

The Lawley-Hotelling trace, a generalization of the test statistic by Hotelling (1951), is one of the most commonly used parametric multivariate test statistics. Equivalently, its nonparametric version can be applied to test for equality of distribution functions in k samples. Consider $k \times p$ dimensional matrices of distribution functions $\mathbb{F}$, relative treatment effects $\mathbf{p}$, and estimated relative treatment effects $\hat{\mathbf{p}}$.

$$\mathbb{F} = \begin{pmatrix} F_1^{(1)} & \dots & F_1^{(p)} \\ \vdots & \ddots & \vdots \\ F_k^{(1)} & \dots & F_k^{(p)} \end{pmatrix}, \mathbf{p} = \begin{pmatrix} p_1^{(1)} & \dots & p_1^{(p)} \\ \vdots & \ddots & \vdots \\ p_k^{(1)} & \dots & p_k^{(p)} \end{pmatrix}, \hat{\mathbf{p}} = \begin{pmatrix} \hat{p}_1^{(1)} & \dots & \hat{p}_1^{(p)} \\ \vdots & \ddots & \vdots \\ \hat{p}_k^{(1)} & \dots & \hat{p}_k^{(p)} \end{pmatrix}$$

We can rewrite the null hypothesis $H_0^F : F_1^{(d)} = \dots = F_k^{(d)}$, $(d = 1,\dots,p)$ as $\mathbf{C}_k\mathbb{F} = \mathbf{0}$, where $\mathbf{C}_k$ is as given in Section 2.3. When this null hypothesis is true and N is large, the distribution of $\sqrt{N}\mathbf{C}_k\hat{\mathbf{p}}$ is asymptotically the same as that of $\sqrt{N}\mathbf{C}_k\bar{\mathbb{Y}}$, where $\bar{\mathbb{Y}}$ is given as:

$$\bar{\mathbb{Y}} = \begin{pmatrix} \bar{Y}_1^{(1)} & \dots & \bar{Y}_1^{(p)} \\ \vdots & \ddots & \vdots \\ \bar{Y}_k^{(1)} & \dots & \bar{Y}_k^{(p)} \end{pmatrix}, \quad \bar{Y}_i^{(d)} = \sum_{j=1}^{n_i} \frac{Y_{ij}^{(d)}}{n_i} \text{ and } Y_{ij}^{(d)} \text{ are as in Section 2.2.}$$

Note that under H_0^F, $\sqrt{N}\mathbf{C}_k\hat{\mathbf{p}}$ is asymptotically distributed as multivariate normal

in which the rows are independent while the columns have asymptotic covariance matrix which is the same as covariance matrix of the of the rows of $\sqrt{N}\mathbf{C}_k\bar{\mathbb{Y}}$.

Under the stronger hypothesis $H_0^{\mathcal{P}}$ and after some mathematical manipulations, it is shown in Munzel and Brunner (2000b) that the nonparametric Lawley-Hotelling type statistic

$$LH = tr(\hat{\mathbf{p}}^T\mathbf{T}\hat{\mathbf{p}}\hat{\mathbf{\Sigma}}^{-1}) \tag{2.22}$$

has asymptotically a central χ^2 distribution with $(k-1)p$ degrees of freedom, where $\mathbf{T} = \mathbf{C}_k^T[\mathbf{C}_k(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{C}_k^T]^{-1}\mathbf{C}_k$, $\mathbf{L} = \bigoplus_{i=1}^{k}\mathbf{1}_{n_i}$,

$\hat{\mathbf{\Sigma}} = \frac{1}{N-k}\sum_{i=1}^{k}\frac{n_i(n_i-1)}{N}\hat{\mathbf{V}}_i$, $\hat{\mathbf{p}}$, $\mathbf{C}_k$, and $\hat{\mathbf{V}}$ are as given above.

Dietz and Killeen test

Dietz and Killeen (1981) proposed a single sample multivariate test for monotone trend over time based on Kendall's τ statistic. Let $\mathbf{X}^{(d)} = (X_1^{(d)},\ldots,X_n^{(d)})^T$, $d = 1,\ldots,p$, denote a sequence of continuous vectors observed at times $1,\ldots,n$. The hypothesis to be tested is formulated as:

H_0 : The p vectors are randomly ordered, versus

H_1 : There is a monotone trend in one or more of the p variables.

Let $\mathbf{K} = (K^{(1)},\ldots,K^{(p)})^T$ be a vector of p Mann statistics, that is, $K^{(d)} = \sum_{i<j} sign(X_j^{(d)} - X_i^{(d)})$, $d = 1,\ldots,p$. Under H_0 and if $\mathbf{\Sigma}$ is of full rank, the statistic

$$\mathbf{K}^T\mathbf{\Sigma}^{-1}\mathbf{K}/\sigma^2 \tag{2.23}$$

is asymptotically χ^2 distributed with p degrees of freedom. If on the other hand, $\mathbf{\Sigma}$ is of rank $q < p$, then

$$\mathbf{K}^T\mathbf{\Sigma}^{-}\mathbf{K}/\sigma^2 \tag{2.24}$$

is asymptotically χ^2 distributed with q degrees of freedom, where $\mathbf{\Sigma}^-$ is a generalized inverse of $\mathbf{\Sigma}$.
In the above,

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{1p} & \dots & \sigma_{pp} \end{bmatrix},$$

$\sigma_{gh} = K^{(g)(h)}/3 + (n^3 - n)r^{(g)(h)}/9,\ g,h = 1,\dots,p,$
$K^{(g)(h)} = \sum_{i<j} sign[(X_j^{(g)} - X_i^{(g)})(X_j^{(h)} - X_i^{(h)})],$
$r^{(g)(h)} = \frac{3}{n^3-n} \sum_{i,j,k} sign[(X_j^{(g)} - X_i^{(g)})(X_j^{(h)} - X_k^{(h)})],$

$$sign(x) = \begin{cases} +1, & if\ x > 0 \\ 0, & if\ x = 0 \\ -1, & if\ x < 0 \end{cases},$$

$\sigma^2 = n(n-1)(2n+5)/18.$

Test statistic by O'Brien

O'Brien (1984) proposed a rank-sum type test statistic to detect directional treatment differences, i.e., an effect where a treatment shows some improvement over another with respect to all response variables. This statistic is more powerful than the conventional Hotelling's T^2 for detecting alternatives where the treatment effect has the same direction for all response variables. In applying this statistic, the response variables are considered in a way that large values, say, denote better outcomes than small values.

Let $X_{ij}^{(d)}$, $i = 1,2$, $j = 1,\dots,n_i$, $d = 1,\dots,p$, denote the d^{th} response variable for the j^{th} unit in (treatment) group i. Assume that each $X_{ij}^{(d)}$ has a continuous distribution $F_i^{(d)}$. Furthermore, the vectors $\mathbf{X}_{ij} = (X_{ij}^{(1)},\dots,X_{ij}^{(p)})^T$ are assumed to be independently distributed with distribution function $\mathbf{F}_i$. For two samples, the hypothesis to be tested can be formulated as:

$$H_0 : \mathbf{F}_1 = \mathbf{F}_2 \text{ versus } H_1 : \mathbf{F}_1 < \mathbf{F}_2,$$

where the $<$ sign is interpreted component-wise.

In order to apply the O'Brien's statistic first pool the observations from the two response variables and then replace each observation by the corresponding rank. Following this, find the sum of ranks across the p response variables for each unit. Finally, these rank-sums are compared among the two groups using standard methods such as two-sample t test or nonparametric two sample univariate tests.

Test by Tsai and Koziol

Let $\mathbf{X}_{ij} = (X_{ij}^{(1)}, \ldots, X_{ij}^{(p)})^T, i = 1, \ldots, k, j = 1, \ldots, n_i$ be a random sample of size n_i from population π_i, and the cumulative distribution function of $\mathbf{X}_{ij}$ is denoted by $F_i(\mathbf{X}) = F(\mathbf{X} - \boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i$ is a p-dimensional vector of unknown parameters.

The hypothesis to be tested is formulated as:

$$H_0 : \boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_k, \text{ versus}$$

$$H_1 : \boldsymbol{\theta}_1 \leq \ldots \leq \boldsymbol{\theta}_k \text{ with at least one strict inequality.}$$

The alternative hypothesis will be tested by taking a pair of samples, that is, $\boldsymbol{\theta}_i \leq \boldsymbol{\theta}_l, i \neq l = 1, \ldots, k$ if and only if $\theta_i^{(d)} \leq \theta_l^{(d)}$ for all $1 \leq d \leq p$ with at least one strict inequality.

Tsai and Koziol (1994) proposed the following multivariate test statistic by generalizing the Terpstra (1952) and Jonckheere (1954) trend statistic.

$$T_N = \mathbf{J}_N^T \boldsymbol{\Sigma}_N^{-1} \mathbf{J}_N \text{ with large values leading to rejection of } H_0, \tag{2.25}$$

where $\mathbf{J}_N = (J^{(1)}, \ldots, J^{(p)})^T, J^{(d)} = \left(\Pi_{i=1}^k n_i\right)^{-1} \sum \phi_d(X_{1j}, \ldots, X_{kj})$, the summation extends over all possible $1 \leq j \leq n_i, 1 \leq i \leq k$, and

$$\phi_d(X_{1j}, \ldots, X_{kj}) = \sum_{i=1}^{k-1} \sum_{l=i+1}^{k} \left[I(X_{ij}^{(d)} < X_{lj}^{(d)}) - I(X_{lj}^{(d)} < X_{ij}^{(d)}); \right] \quad \boldsymbol{\Sigma}_N = (\sigma_{dd'}, N),$$

$$\sigma_{dd',N} = Cov(J^{(d)}, J^{(d')}) = \left(\Pi_{i=1}^{k} n_i\right)^{-1} \sum_{r_1=0}^{1} \cdots \sum_{r_k=0}^{1} \Pi_{c=1}^{k} \binom{n_c - 1}{1 - r_c} \xi_{r_1,\ldots,r_k}^{(d,d')}$$

$\xi_{r_1,\ldots,r_k}^{(d,d')}$ is the covariance between ϕ_d and $\phi_{d'}$, when r_i of the $X_{ij}'s$ are common between the two sets of X_{ij}'s, one in each of ϕ_d and $\phi_{d'}$.

Note that $\xi_{0,\ldots,0}^{(d,d')} = 0$ for all d, d', and any other $\xi_{r_1,\ldots,r_k}^{(d,d')}$ is obtained from the average of $(\phi_d - E(\phi_d))(\phi_{d'} - E(\phi_{d'}))$ the average is taken over all possible terms for which r_i of the X_{ij}'s are common in the i^{th} component of ϕ_d and $\phi_{d'}$.

Under H_0, and $n_i/N \to \lambda_i$, $\sum_{i=1}^{k} \lambda_i = 1$, the asymptotic distribution of T_N is chi-square with p degrees of freedom.

2.4.4 Other nonparametric tests in multivariate framework

Koziol et al. (1981), proposed a distribution-free test appropriate for ordered alternatives which allows incomplete data. The hypothesis testing problem was formulated based on continuous distribution functions. A test statistic was devised by applying the rank permutation principle conditioned on the patterns of missing data. Asymptotic critical values of a quadratic form for the rank-sum statistic when the covariance matrix is nonsingular were also determined.

Bregenzer and Lehmacher (1998) proposed a two-sample multivariate directional test statistic to detect an overall (treatment) superiority allowing incomplete data. The ideas of O'Brien (1984) procedures were employed to construct rank tests which allow mixed observation vectors consisting of both quantitative and categorical data.

Hyun and Song (2009) developed a nonparametric multivariate test statistic for monotone trend among k-samples. The vectors of observations on $p(\geq 2)$ variables were assumed to be independent with continuous distribution functions. The multivariate trend statistic was derived based on coordinate-wise Jonckheere statistics, i.e., the Jonckheere trend statistics for each response variable in all k-samples are calculated and summarized into a single statistic.

Nadar (2013) proposed a multivariate generalization of k-sample rank test for umbrella alternatives to test for equality of medians. Mean-centered statistic for

testing umbrella alternatives when the peak is known was defined for each response variable by considering pairs of treatments and the count function. Monte Carlo simulations show that the performance of the test is good in terms of keeping nominal Type-I error rate for the normal, light-tailed and heavy-tailed distributions. The test also performs well in the presence of a large number of outliers.

Note that each of the tests by Dietz (1989), Tsai and Koziol (1994), Hyun and Song (2009) and Nadar (2013) was derived based on several pairwise comparisons in such a way that the ranks are being computed pairwise only, that is, only between those two levels that are compared at each step. This reduces the amount of available information and is well known to potentially lead to paradox situations. See Efron's paradox dice in Brunner et al. (2019). Moreover, the underlying distributions have to be continuous.

Part II

Rank-Based Multivariate Directional Testing Problems

CHAPTER 3

A rank-based test for monotone trends in k-sample multivariate problems

Summary

In many practical applications, a multivariate outcome is measured on two or more groups of subjects. In clinical trials, one refers to different endpoints, often classified as primary and secondary. When these variables are analyzed separately, the available information is not fully exhausted, as possible dependencies among the endpoints are not taken into consideration. Within parametric frameworks, a large amount of literature exists to test whether two or more groups differ with respect to multivariate outcomes. Also, a few methods have been proposed using a fully nonparametric approach. It is of particular interest to test the hypothesis of no treatment effect against conjectured alternative patterns that are associated with treatment efficacy. We propose a rank-based test for detecting pre-specified alternative patterns across all endpoints. Here, we do not assume continuous distributions. That is, ties are allowed, and the proposed method can specifically be applied to multivariate

ordered categorical data. The test statistic we have derived is a multivariate generalization of a univariate test proposed by Bathke (2009) for alternative patterns within a nonparametric framework. Since we do not require the continuity of the distribution functions, the newly proposed test statistic is applied to data with ties, in particular, to multivariate-ordered categorical data. Moreover, the test is invariant under monotone transformations of the responses and the weights describing the alternative pattern. Finite sample performance of the proposed test statistic is assessed through a simulation study and comparisons are made with the most popular test for one-sided multivariate problems. The application of the proposed test statistic is demonstrated with an electroencephalogram dataset. We also reanalyzed some datasets given in the literature.

3.1 Introduction

Treatment comparisons in many practical applications such as clinical trials, psychology or sociology often involve a multivariate outcome potentially measured on different scales. The distributions of these response variables may not be normal which calls for nonparametric methods. Although the problem of multiple testing arises, univariate methods for assessing each response variable individually have been widely used because of their flexibility and ease of interpretation. On the other hand, incomplete information is used when these variables are analyzed separately as the possible dependencies among them are not taken into consideration. Moreover, a single overall objective probability statement is usually needed. An overall probability statement can be obtained by taking the minimum of per-experiment error rates on each univariate test by using the Bonferroni inequality. Although this technique maintains the family-wise error rate, it becomes conservative when the number of response variables is large.

Under the assumption of multivariate normality and homogeneity of the variance-covariance matrices, Wilks (1946), Hotelling (1951), and Pillai (1955), among others employed the classical Multivariate Analysis of Variance (MANOVA) models. In real life situations, however, the assumptions of multivariate normality and/or

homogeneity of the variance-covariance matrices are often not fulfilled. More importantly, a dataset under analysis may incorporate response variables which are measured on different scales for which parametric test statistics are not invariant under scale transformation, e.g., $cm \to m$, $g \to kg$, etc. of such responses. To address these issues, nonparametric approaches have been proposed by, for example, Puri and Sen (1966), Akritas and Arnold (1994), Brunner et al. (1999), Bathke and Harrar (2008), and Brunner et al. (2016).

The methods developed by the above authors are of a quadratic nature and do not use information on the direction of treatment differences. As a result, they lack power if they are applied to test a hypothesis with an alternative of interest in a specified direction. For example, a directional alternative hypothesis is tested in a dose-response trial if one is interested to see whether levels of enzymes in the blood increase with increasing dose levels of a chemical. There can be two or more enzymes that are measured within a given volume of blood. The dose levels can be expressed as, for example, none, low, medium, and high. In such cases, directional multivariate testing problems will come into picture.

In this part of the dissertation, we propose a rank-based test statistic which handles ordered alternatives in the case of several sample multivariate problems. Apart from disregarding a degenerate distribution, the statistic allows for continuous as well as discrete distributions.

3.2 Nonparametric model and hypothesis

Let $\mathbf{X}_{ij} = (X_{ij}^{(1)}, X_{ij}^{(2)}, \ldots, X_{ij}^{(p)})'$ be a $p \times 1$ vector of observations on p response variables potentially measured on different scales, and x_{ij} be non-stochastic constants determining the alternative pattern for the j^{th} subject in treatment i, $j = 1, 2, \ldots, n_i$, $i = 1, \ldots, k$. Suppose that the $\mathbf{X}_{ij}$'s are independent random vectors with possibly dependent components $X_{ij}^{(d)}$, $d = 1, 2, \ldots, p$ and $X_{ij}^{(d)} \sim F_i^{(d)}$, where $F_i^{(d)}$ is understood as the normalised version of the distribution function, which is the average of its right continuous ($F_i^{(d)+} = Pr(X_{ij}^{(d)} \leq x)$) and left-continuous

version ($F_i^{(d)-} = Pr(X_{ij}^{(d)} < x)$), that is,

$$F_i^{(d)}(x) = \frac{1}{2}\left[Pr(X_{ij}^{(d)} \leq x) + Pr(X_{ij}^{(d)} < x)\right] \tag{3.1}$$

We use this normalised version of the distribution function in order to handle continuous as well as ordered categorical data in a unified form.

In this thesis, the hypothesis of no treatment effect is expressed in terms of distribution functions. Suppose that the multivariate cumulative distribution function corresponding to each of the p-dimensional vectors ($\mathbf{X}_{ij}$) is $F_i(\mathbf{x})$, i.e., $\mathbf{X}_{ij} \sim F_i(\mathbf{x})$. Then the nonparametric null hypothesis is formulated as:

$$H_0 : F_1(\mathbf{x}) = F_2(\mathbf{x}) = \ldots = F_k(\mathbf{x}) \quad \text{for all } \mathbf{x} \tag{3.2}$$

and the alternative hypothesis is

$$H_1 : F_1^{(d)}(x) \leq \ldots \leq F_k^{(d)}(x) \tag{3.3}$$

for all x and $d = 1, \ldots, p$ with at least one strict inequality for at least one d, that is, each of the p response variables is stochastically decreasing across the factor levels $1, \ldots, k$. A test against a conjectured stochastically increasing trend in each of the p responses can be formulated by reversing the direction of inequalities in equation (3.3), that is, $H_1 : F_1^{(d)}(x) \geq \ldots \geq F_k^{(d)}(x)$.

3.3 Test statistic

Let $R_{ij}^{(d)}$ denote the (mid-) rank of $X_{ij}^{(d)}$ among all N observations $(X_{11}^{(d)}, \ldots, X_{kn_k}^{(d)})$. Then the $p \times 1$ vector of (mid-) ranks for observations obtained from p response variables for the j^{th} subject in treatment i, is given by $\mathbf{R}_{ij} = (R_{ij}^{(1)}, \ldots, R_{ij}^{(p)})^T$. Accordingly, the (mid-) rank matrix ($\mathbf{R}$) corresponding to the matrix of actual observations can be expressed as:

Sample 1			Sample 2			...	Sample k		
$R_{11}^{(1)}$	$R_{12}^{(1)}$	$\dots R_{1n_1}^{(1)}$	$R_{21}^{(1)}$	$R_{22}^{(1)}$	$\dots R_{2n_2}^{(1)}$	...	$R_{k1}^{(1)}$	$R_{k2}^{(1)}$	$\dots R_{kn_k}^{(1)}$
$R_{11}^{(2)}$	$R_{12}^{(2)}$	$\dots R_{1n_1}^{(2)}$	$R_{21}^{(2)}$	$R_{22}^{(2)}$	$\dots R_{2n_2}^{(2)}$	...	$R_{k1}^{(2)}$	$R_{k2}^{(2)}$	$\dots R_{kn_k}^{(2)}$
	...			...		...		...	
$R_{11}^{(p)}$	$R_{12}^{(p)}$	$\dots R_{1n_1}^{(p)}$	$R_{21}^{(p)}$	$R_{22}^{(p)}$	$\dots R_{2n_2}^{(p)}$	...	$R_{k1}^{(p)}$	$R_{k2}^{(p)}$	$\dots R_{kn_k}^{(p)}$

The covariance and correlation matrices corresponding to the matrices of observations are respectively given by:

$$\mathbf{\Sigma}_i = \begin{pmatrix} \sigma_i^{(1)^2} & \sigma_i^{(1)(2)} & \dots & \sigma_i^{(1)(p)} \\ \sigma_i^{(2)(1)} & \sigma_i^{(2)^2} & \dots & \sigma_i^{(2)(p)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_i^{(p)(1)} & \sigma_i^{(p)(2)} & \dots & \sigma_i^{(p)^2} \end{pmatrix}; \text{ for } i = 1, \dots, k$$

$$\boldsymbol{\rho}_i = \begin{pmatrix} 1 & \rho_i^{(1)(2)} & \dots & \rho_i^{(1)(p)} \\ \rho_i^{(2)(1)} & 1 & \dots & \rho_i^{(2)(p)} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_i^{(p)(1)} & \rho_i^{(p)(2)} & \dots & 1 \end{pmatrix}; \text{ for } i = 1, \dots, k$$

Observe that when data are from ordinal variables, the correlation is calculated based on the Spearman's rank correlation coefficient.

The weighted average of all unobservable distribution functions ($H^{(d)}(x)$) and that of empirical distribution function ($\widehat{H}^{(d)}(x)$) for variable d are defined as:

$$H^{(d)}(x) = \frac{1}{N} \sum_{i=1}^{k} n_i F_i^{(d)}(x) \tag{3.4}$$

$$\widehat{H}^{(d)}(x) = \frac{1}{N} \sum_{i=1}^{k} n_i \widehat{F}_i^{(d)}(x), \tag{3.5}$$

where $N = \sum_{i=1}^{k} n_i$, $\widehat{F}_i^{(d)}(x)$ is the empirical distribution function, and $F_i^{(d)}(x)$ is as in (3.1).

The empirical distribution function ($\widehat{F}_i^{(d)}(x)$) which is used to estimate the distribution function $H^{(d)}(x)$ is expressed in terms of a counting function as follows

(see Brunner et al. (2001), pp. 45-47):

$$\widehat{F}_i^{(d)}(x) = n_i^{-1} \sum_{j=1}^{n_i} c(x - X_{ij}^{(d)}), \text{ where } c(u) = \begin{cases} 0, & u < 0 \\ 1/2, & u = 0 \\ 1, & u > 0 \end{cases}$$

Therefore, the weighted average of all empirical distribution functions for the d^{th} variable, $\widehat{H}^{(d)}(x) = \frac{1}{N}\sum_{i=1}^{k} n_i \widehat{F}_i^{(d)}(x)$ can be expressed as

$$\widehat{H}^{(d)}(x) = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} c(x - X_{ij}^{(d)}). \tag{3.6}$$

Furthermore, $\widehat{H}^{(d)}(X_{ij}^{(d)})$ is related to mid-ranks ($R_{ij}^{(d)}$) of the random variables $X_{ij}^{(d)}$ among all N observations, $X_{11}^{(d)}, \ldots, X_{kn_k}^{(d)}$ by

$$\widehat{H}^{(d)}(X_{ij}^{(d)}) = \frac{1}{N}\left(R_{ij}^{(d)} - \frac{1}{2}\right). \tag{3.7}$$

The unobservable weighted average distribution function for the deterministic constant, $G(x)$ coincides with its empirical version, namely,

$$\widehat{G}(x) = G(x) = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} c(x - x_{ij}) \tag{3.8}$$

Let $Y_{ij}^{(d)} = H^{(d)}(X_{ij}^{(d)})$ denote the asymptotic rank transform, $\widehat{Y}_{ij}^{(d)} = \widehat{H}^{(d)}(X_{ij}^{(d)})$ the rank transform of the d^{th} response variable and $y_{ij} = G(x_{ij})$ the rank transform of the constants. The term "rank transform" stems from the fact that $\widehat{Y}_{ij}^{(d)}$ and y_{ij} are related to (mid-)ranks as:

$$\widehat{Y}_{ij}^{(d)} = \frac{1}{N}\left(R_{ij}^{(d)} - \frac{1}{2}\right) \text{ and } y_{ij} = \frac{1}{N}\left(r_{ij} - \frac{1}{2}\right) \tag{3.9}$$

where, $R_{ij}^{(d)}$ is the (mid-) rank of the d^{th} response variable $X_{ij}^{(d)}$ among all $N =$

$\sum_{i=1}^{k} n_i$ observations, $X_{11}^{(d)}, \ldots, X_{kn_k}^{(d)}$ and r_{ij} is the (mid-) rank of x_{ij} among the N constants $x_{11}, \ldots, x_{kn_k}$.

The asymptotic rank transforms are used in the theoretical derivation of the test statistic while the actual test utilizes the (mid-)ranks of the response variable ($R_{ij}^{(d)}$) and of the non-stochastic constants (r_{ij}). If random variables X_{ij} are independent, then the corresponding asymptotic rank transforms are also independent. However, the rank transforms may not be independent although the random variables X_{ij} are independent. Thus, theoretical derivation using asymptotic rank transforms is amenable.

Let us consider the univariate trend test statistic proposed by Bathke (2009) which can handle dependent and independent samples, and factorial designs. In general, the test was developed based on a two-factorial nonparametric mixed model. Values of a response variable for the two-factor factorial design is given in Table 3.1.

TABLE 3.1: Values of a response variable observed at each level of factors AB combination.

		Factor B		
		1	...	b
Factor A	1	$X_{111}, X_{112}, \ldots, X_{11n_1}$	$\ldots$	$X_{1b1}, X_{1b2}, \ldots, X_{1bn_1}$
	2	$X_{211}, X_{212}, \ldots, X_{21n_2}$	$\ldots$	$X_{2b1}, X_{2b2}, \ldots, X_{2bn_2}$
	$\ldots$	$\ldots$	$\ldots$	$\ldots$
	a	$X_{a11}, X_{a12}, \ldots, X_{a1n_a}$	$\ldots$	$X_{ab1}, X_{ab2}, \ldots, X_{abn_a}$

Assuming that experimental units may differ in different levels of factor A, Bathke (2009) defined the weighted averages of distribution function and empirical distribution function as

$$H = \frac{\sum_{i=1}^{a} \sum_{m=1}^{b} n_i F_{im}}{N} \quad \text{and} \tag{3.10}$$

$$\widehat{H} = \frac{\sum_{i=1}^{a}\sum_{m=1}^{b} n_i \widehat{F}_{im}}{N}, \text{ respectively.} \tag{3.11}$$

The corresponding asymptotic rank-transforms and the rank transforms are respectively given by $Y_{imt} = H(X_{imt})$ and $\widehat{Y}_{imt} = \widehat{H}(X_{imt})$. Ranking was performed by combining all observations in the two factor combinations in such a way that R_{imt} and r_{imt} denote the (mid-) ranks of X_{imt} among the N observations $X_{111}, \ldots, X_{abn_a}$ and x_{imt} among the N constants $x_{111}, \ldots, x_{abn_a}$, respectively. Note that n_a denotes the number of observations at the last (a) level of factor A so that the total number of observations (N) is given by $N = b.\sum_{i=1}^{a} n_i$. Under the following four technical assumptions

1. $\min_{i=1,\ldots,k}(n_i) \to \infty$
2. $\exists \sigma_0^2 > 0 : \ \forall i,m,t : \ Var(Y_{imt}) = \sigma_{im}^2 \geq \sigma_0^2$.
3. $\forall i,i',m,m',t,t' : \ Cov(Y_{imt},\ Y_{i'm't'}) = \begin{cases} \sigma_{imm'}, & (i,t) = (i',t'),\ m \neq m' \\ 0, & (i,t) \neq (i',t') \end{cases}$
4. Let $Q_{it} = \sum_{m=1}^{b}\left(r_{imt} - \frac{N+1}{2}\right) Y_{imt}$. Then, $\exists \tau_0^2 > 0 : \ \forall i,t : \ Var(Q_{it}) \geq \tau_0^2$

Bathke (2009) proposed a test statistic

$$\widehat{v} = \widehat{s}^{-1} \sum_{i=1}^{a}\sum_{m=1}^{b}\sum_{t=1}^{n_i}\left(r_{imt} - \frac{N+1}{2}\right)\left(R_{imt} - \frac{N+1}{2}.\right) \tag{3.12}$$

In the above:
$\widehat{s}$ is a consistent estimator of standard deviation for the expression given in the numerator of equation (3.12), where:

$$\widehat{s}^2 = \sum_{i=1}^{a}\sum_{m=1}^{b}\widehat{\sigma}_{im}^2 \sum_{t=1}^{n_i}\left(r_{imt} - \frac{N+1}{2}\right)^2 + \sum_{i=1}^{a}\sum_{m\neq l}\widehat{\sigma}_{iml}\sum_{t=1}^{n_i}\left(r_{imt} - \frac{N+1}{2}\right)\left(r_{ilt} - \frac{N+1}{2}\right)$$

$$\widehat{\sigma}_{im}^2 = (n_i - 1)^{-1} \sum_{t=1}^{n_i} \left(R_{imt} - \frac{N+1}{2} \right)^2$$

$$\widehat{\sigma}_{iml} = (n_i - 1)^{-1} \sum_{t=1}^{n_i} \left(R_{imt} - \frac{N+1}{2} \right) \left(R_{ilt} - \frac{N+1}{2} \right)$$

As a special case for k independent samples with sizes n_i, $i = 1, \ldots, k$, and the total number of observations $N = \sum_{i=1}^{k} n_i$, the test by Bathke (2009) is reduced to

$$\widehat{v} = \hat{s}^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(r_{ij} - \frac{N+1}{2} \right) \left(R_{ij} - \frac{N+1}{2} \right), \tag{3.13}$$

where

r_{ij} is the (mid-) rank of the constants x_{ij} among all N constants $x_{11}, \ldots, x_{kn_k}$.

R_{ij} is the (mid-) rank of the response variable (X_{ij}) among all N observations $X_{11}, \ldots, X_{kn_k}$,

$$\hat{s}^2 = \sum_{i=1}^{k} (n_i - 1)^{-1} \sum_{j=1}^{n_i} \left(R_{ij} - \frac{N+1}{2} \right)^2 \sum_{j=1}^{n_i} \left(r_{ij} - \frac{N+1}{2} \right)^2$$

The sum of (mid-) ranks for the response variable in the i^{th} sample can be denoted by $R_{i\cdot} = \sum_{j=1}^{n_i} R_{ij}$. Suppose, without loss of generality, that the pattern to be detected is an increasing trend from sample $i = 1$ to $i = k$ in such a way that the constants x_{ij} take the value $x_{ij} = i$, $i = 1, \ldots, k$, and $j = 1, \ldots, n_i$. Let us now formulate a simplified form of the test statistic given in equation (3.13). Consider the expression given in the numerator, i.e.,

$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(r_{ij} - \frac{N+1}{2} \right) \left(R_{ij} - \frac{N+1}{2} \right)$, we can rewrite this as

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(r_{ij} - \frac{N+1}{2} \right) \left(R_{ij} - \frac{N+1}{2} \right) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(r_{ij} - \frac{N+1}{2} \right) R_{ij}$$

The expression on the right hand side can be expanded as:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(r_{ij}-\frac{N+1}{2}\right)R_{ij} = \left(r_{11}-\frac{N+1}{2}\right)R_{11}+\left(r_{12}-\frac{N+1}{2}\right)R_{12}+\ldots+\left(r_{1n_1}-\frac{N+1}{2}\right)R_{1n_1}+\ldots+\left(r_{k1}-\frac{N+1}{2}\right)R_{k1}+\ldots+\left(r_{kn_k}-\frac{N+1}{2}\right)R_{kn_k}$$

In order to detect an increasing trend from sample 1 to k, the constants x_{ij} can be chosen as $x_{ij} = i$ for $i = 1,\ldots,k$, $j = 1,\ldots,n_i$. Thus, $x_{11} = 1$, $x_{12} = 1$, …, $x_{1n_1} = 1$. The mid ranks corresponding to each of these observations (r_{1j}), $j = 1,\ldots,n_1$ can be obtained by finding the average of $\{1,2,\ldots,n_1\}$, that is, $\sum_{i=1}^{n_1}\frac{i}{n_1} = \frac{(1+n_1)}{2}$. Applying the same technique to $x_{2j},\ldots,x_{kj}$, gives the following ranks.

$r_{2j} = n_1 + \frac{n_2+1}{2}$, $j = 1,\ldots,n_2$

$r_{3j} = n_1 + n_2 + \frac{n_3+1}{2}$, $j = 1,\ldots,n_3$

$\vdots$

$r_{kj} = n_1 + n_2 + \ldots + n_{k-1} + \frac{n_k+1}{2}$, $j = 1,\ldots,n_k$.

Thus, $\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(r_{ij}-\frac{N+1}{2}\right)R_{ij}$ can be rewritten by taking $\frac{(1+n_1)}{2}$ in the place of r_{1j}, $j = 1,\ldots,n_1$, $(n_1 + \frac{n_2+1}{2})$ in the place of r_{2j}, $j = 1,\ldots,n_2$, and so on as:

$$\left(\frac{1+n_1}{2}-\frac{\sum_{i=1}^{k}n_i+1}{2}\right)R_{11}+\ldots+\left(\frac{1+n_1}{2}-\frac{\sum_{i=1}^{k}n_i+1}{2}\right)R_{1n_1}+\ldots+\left(\sum_{i=1}^{k-1}n_i+\frac{1+n_k}{2}-\frac{\sum_{i=1}^{k}n_i+1}{2}\right)R_{k1}+\ldots+\left(\sum_{i=1}^{k-1}n_i+\frac{1+n_k}{2}-\frac{\sum_{i=1}^{k}n_i+1}{2}\right)R_{kn_k}$$

$$= \left(-\frac{\sum_{l=2}^{k} n_l}{2} \right) R_{11} + \ldots + \left(-\frac{\sum_{l=2}^{k} n_l}{2} \right) R_{1n_1} + \ldots + \left(\frac{\sum_{l=1}^{k-1} n_l}{2} \right) R_{kn_k}$$

$$= \left(-\frac{\sum_{l=2}^{k} n_l}{2} \right) R_{1.} + \left(n_1 - \frac{\sum_{l=3}^{k} n_l}{2} \right) R_{2.} + \ldots + \left(\frac{\sum_{l=1}^{k-1} n_l}{2} \right) R_{k.}$$

$$= \sum_{i=1}^{k} \left(\sum_{l=1}^{i-1} n_l - \sum_{l=i+1}^{k} n_l \right) \frac{R_{i.}}{2}.$$

Let us now consider the expression in the denominator of equation (3.13), i.e.,

$$\hat{s}^2 = \sum_{i=1}^{k} (n_i - 1)^{-1} \sum_{j=1}^{n_i} (r_{ij} - \frac{N+1}{2})^2 \sum_{j=1}^{n_i} (R_{ij} - \frac{N+1}{2})^2. \tag{3.14}$$

When $i = 1$ and substituting r_{1j}, $j = 1, \ldots, n_1$ by $\frac{(1+n_1)}{2}$, the expression on the right hand side of (3.14) will be

$$\frac{1}{n_1 - 1} \left[(r_{11} - \frac{N+1}{2})^2 \sum_{j=1}^{n_1} (R_{1j} - \frac{N+1}{2})^2 + \ldots + (r_{1n_1} - \frac{N+1}{2})^2 \sum_{j=1}^{n_1} (R_{1j} - \frac{N+1}{2})^2 \right]$$

$$= \left(\frac{n_1}{n_1 - 1} \right) \left(\frac{-\sum_{l=2}^{k} n_l}{2} \right)^2 \sum_{j=1}^{n_1} \left(R_{1j} - \frac{N+1}{2} \right)^2$$

If we continue like this for $i = 2, \ldots, k$, we can obtain the following when $i = k$.

$$\left(\frac{n_k}{n_k - 1} \right) \left(\frac{\sum_{l=1}^{k-1} n_l}{2} \right)^2 \sum_{j=1}^{n_k} \left(R_{kj} - \frac{N+1}{2} \right)^2$$

Thus,

$$\hat{s}^2 = \left(\frac{n_1}{n_1-1}\right)\left(\frac{-\sum_{l=2}^{k} n_l}{2}\right)^2 \sum_{j=1}^{n_1}\left(R_{1j}-\frac{N+1}{2}\right)^2 + \ldots + \left(\frac{n_k}{n_k-1}\right)\left(\frac{\sum_{l=1}^{k-1} n_l}{2}\right)^2 \sum_{j=1}^{n_k}\left(R_{kj}-\frac{N+1}{2}\right)^2$$

$$= \sum_{i=1}^{k}\left(\frac{n_i}{n_i-1}\right)\left(\frac{\sum_{l=1}^{i-1} n_l - \sum_{l=i+1}^{k} n_l}{2}\right)^2 \sum_{j=1}^{n_k}\left(R_{ij}-\frac{N+1}{2}\right)^2.$$

Combining the simplified expressions for the numerator and denominator of $\hat{v}$ gives the following simplified formula.

$$\hat{v} = \frac{\sum_{i=1}^{k}\left(\sum_{l=1}^{i-1} n_l - \sum_{l=i+1}^{k} n_l\right) R_{i.}}{\sqrt{\sum_{i=1}^{k}\frac{n_i}{n_i-1}\left(\sum_{l=1}^{i-1} n_l - \sum_{l=i+1}^{k} n_l\right)^2 \sum_{j=1}^{n_i}(R_{ij}-\frac{N+1}{2})^2}} \tag{3.15}$$

Define a vector of test statistics $\hat{\mathbf{v}} = (\hat{v}^{(1)},\ldots,\hat{v}^{(p)})^T$ in order to derive a test statistic for testing the hypotheses formulated in Section 3.2. Note that each $\hat{v}^{(d)}$, $d = 1,\ldots,p$, is asymptotically distributed as a standard normal according to Bathke (2009).

We reconsider the following sets of assumptions to derive the asymptotic result of the test statistic which can handle multivariate data.

ASSUMPTIONS

Assumption 1 $\min_{i=1,\ldots,k}(n_i) \to \infty$.

Assumption 2 $\exists {\sigma^{(d)}}^2 > 0 : \forall i, j, Var(Y_{ij}^{(d)}) = {\sigma_i^{(d)}}^2 \geq {\sigma^{(d)}}^2$, where $Y_{ij}^{(d)}$ is the asymptotic rank transform for the d^{th} response.

Assumption 3 $Cov(Y_{ij}^{(d)}, Y_{i^*j^*}^{(d^*)}) = \begin{cases} \sigma_i^{(d)(d^*)}, & (i,j) = (i^*, j^*),\ d \neq d^* = 1, \ldots, p \\ 0, & (i,j) \neq (i^*, j^*) \end{cases}$

Assumption 4 $Q_{ij} = \sum_{d=1}^{p} (y_{ij} - \bar{y}_i) Y_{ij}^{(d)}$. $\exists \tau_0^2 > 0 : \forall i, j;\ Var(Q_{ij}) \geq \tau_0^2$, where y_{ij} is the rank transform of the constants, and $\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$.

Assumption 1 is about the minimum sample size required to apply the test statistic. The test becomes valid when the number of observations in each of the samples is large enough. The expected variation to apply central limit theorem is expressed under the Assumptions 2 and 4. Assumption 3 deals with the covariance structure of the asymptotic rank transforms. It is assumed that any two asymptotic rank transforms corresponding to two responses measured (obtained) from the same subject (unit) in the same sample are correlated. The correlation between these asymptotic rank transforms is assumed to zero if responses are considered from either different units or samples.

Theorem 1 Under the null hypothesis H_0 given in (3.2) and Assumptions 2 & 3, the covariance between $\widehat{v}^{(d)}$ and $\widehat{v}^{(d^*)}$ for any $d \neq d^* = 1, \ldots, p$, is

$$Cov\left(\widehat{v}^{(d)}, \widehat{v}^{(d^*)}\right) = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(r_{ij} - \frac{N+1}{2}\right)^2 \sigma_i^{(d)(d^*)}}{\sqrt{\sum_{i=1}^{k} {\sigma_i^{(d)}}^2 \sum_{j=1}^{n_i} \left(r_{ij}^{(d)} - \frac{N+1}{2}\right)^2} \sqrt{\sum_{l=1}^{k} {\sigma_l^{(d^*)}}^2 \sum_{m=1}^{n_l} \left(r_{lm}^{(d^*)} - \frac{N+1}{2}\right)^2}} \tag{3.16}$$

Proof.

Consider univariate test statistics for two response variables, say $\widehat{v}^{(1)}$ and $\widehat{v}^{(2)}$.

$$\widehat{v}^{(1)} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (r_{ij}^{(1)} - \frac{N+1}{2})(R_{ij}^{(1)} - \frac{N+1}{2})}{\hat{\sigma}^{(1)}}$$

$$\widehat{v}^{(2)} = \frac{\sum_{l=1}^{k}\sum_{m=1}^{n_l}(r_{lm}^{(2)} - \frac{N+1}{2})(R_{lm}^{(2)} - \frac{N+1}{2})}{\hat{\sigma}^{(2)}}$$

where,
$R_{ij}^{(1)}$ and $R_{lm}^{(2)}$ are the (mid-) ranks for the first and second response variables, respectively.
$r_{ij}^{(1)}$ and $r_{lm}^{(2)}$ are the (mid-) ranks for the covariate.
$\hat{\sigma}^{(1)}$ and $\hat{\sigma}^{(2)}$ are the estimated standard deviations of the expressions in the numerators of $\widehat{v}^{(1)}$ and $\widehat{v}^{(2)}$, respectively.
$N = \sum_{i=1}^{k} n_i$ is the total number of observations in the k-samples.

Let us first consider the expressions given in the numerators of the two univariate test statistics ($\widehat{v}^{(1)}$ and $\widehat{v}^{(2)}$). We then substitute the (mid-)ranks obtained from the two responses, that is, $R_{ij}^{(1)}$ and $R_{ij}^{(2)}$ by the asymptotic rank transforms $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ respectively. Define two expressions $Nu1$ and $Nu2$ which are functions of the asymptotic rank transforms $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ as:

$$Nu1 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(r_{ij}^{(1)} - \frac{N+1}{2})(Y_{ij}^{(1)} - \frac{N+1}{2}) \text{ and} \tag{3.17}$$

$$Nu2 = \sum_{l=1}^{k}\sum_{m=1}^{n_l}(r_{lm}^{(2)} - \frac{N+1}{2})(Y_{lm}^{(2)} - \frac{N+1}{2}) \tag{3.18}$$

where,
$Y_{ij}^{(1)}$ and $Y_{lm}^{(2)}$ are the asymptotic rank transforms corresponding to the first and the second response variables, respectively.
$R_{ij}^{(1)}$, $R_{lm}^{(2)}$, $r_{ij}^{(1)}$, $r_{lm}^{(2)}$, and N are as defined above.

$$Cor(Nu1, Nu2) = \frac{Cov(Nu1, Nu2)}{\sigma_{Nu1}\sigma_{Nu2}} = \frac{\mathbb{E}\{(Nu1 - \mathbb{E}(Nu1))(Nu2 - \mathbb{E}(Nu2))\}}{\sigma_{Nu1}\sigma_{Nu2}}$$

Assuming that H_0 is true, $\mathbb{E}(Y_{ij}^{(1)})$ is the same for each ij and $\mathbb{E}(Y_{lm}^{(2)})$ is the same for each lm so that $\mathbb{E}(Nu1) = \mathbb{E}(Nu2) = 0$.

$$\implies Corr(Nu1, Nu2) = \mathbb{E}\left\{\left(\frac{Nu1}{\sigma_{Nu1}}\right)\left(\frac{Nu2}{\sigma_{Nu2}}\right)\right\}$$
$$= Cov\left(\frac{Nu1}{\sigma_{Nu1}}, \frac{Nu2}{\sigma_{Nu2}}\right)$$

That is, the co-variance between $\frac{Nu1}{\sigma_{Nu1}}$ and $\frac{Nu2}{\sigma_{Nu2}}$ is the same as the correlation between $Nu1$ and $Nu2$. Applying the results of **Lemma A1** and **Lemma A2** of Bathke (2005), the asymptotic rank transforms can be replaced by the corresponding rank transforms and, $\hat{\sigma}_{Nu1}$ is a consistent estimator of σ_{Nu1}, and $\hat{\sigma}_{Nu2}$ is a consistent estimator of σ_{Nu2}. If we further apply Slutsky's theorem,

$$\left(\widehat{v}^{(1)} - \frac{Nu1}{\sigma_{Nu1}}\right) \xrightarrow{p} 0 \quad \text{and} \quad \left(\widehat{v}^{(2)} - \frac{Nu2}{\sigma_{Nu2}}\right) \xrightarrow{p} 0.$$

Thus, $\left(\sum \frac{\widehat{v}^{(1)}\widehat{v}^{(2)}}{n} - \frac{1}{n}\sum \frac{Nu1}{\sigma_{Nu1}}\frac{Nu2}{\sigma_{Nu2}}\right)$ converges in probability to zero.
In turn,

$$\frac{1}{n}\sum \frac{Nu1}{\sigma_{Nu1}}\frac{Nu2}{\sigma_{Nu2}} \xrightarrow{a.s.} \mathbb{E}\left(\frac{Nu1}{\sigma_{Nu1}}\frac{Nu2}{\sigma_{Nu2}}\right) = cov\left(\frac{Nu1}{\sigma_{Nu1}}, \frac{Nu2}{\sigma_{Nu2}}\right)$$

This convergence is achieved as per the strong law of large numbers and assumption **(A1)** since $\mathbb{E}(Nu1) = \mathbb{E}(Nu2) = 0$.
Considering $\widehat{v}^{(1)}$ and $\widehat{v}^{(2)}$ as any two random variables, their covariance can be given by:

$$Cov\left(\widehat{v}^{(1)}, \widehat{v}^{(2)}\right) = \frac{\sum_{l=1}^{n}\left(\widehat{v}^{(1)} - \mathbb{E}(\widehat{v}^{(1)})\right)\left(\widehat{v}^{(2)} - \mathbb{E}(\widehat{v}^{(2)})\right)}{n} \tag{3.19}$$

It was shown by Bathke (2005) that $\widehat{v}^{(1)} \dot{\sim} N(0,1)$ and $\widehat{v}^{(2)} \dot{\sim} N(0,1)$ (asymptotically) so that $\mathbb{E}(\widehat{v}^{(1)})$ and $\mathbb{E}(\widehat{v}^{(2)})$ both take a value 0. As a result, the expression given in equation (3.19) can be rewritten as:

$$Cov\left(\widehat{v}^{(1)}, \widehat{v}^{(2)}\right) = \frac{\sum_{l=1}^{n}\widehat{v}^{(1)}\widehat{v}^{(2)}}{n} \tag{3.20}$$

but, it is described above that $\frac{\sum_{l=1}^{n}\widehat{v}^{(1)}\widehat{v}^{(2)}}{n}$ is asymptotically equal to $Cov(\frac{Nu1}{\sigma_{Nu1}}, \frac{Nu2}{\sigma_{Nu2}})$.

$$i.e., Cov\left(\widehat{v}^{(1)}, \widehat{v}^{(2)}\right) \equiv Cov(\frac{Nu1}{\sigma_{Nu1}}, \frac{Nu2}{\sigma_{Nu2}})$$

Furthermore, it is shown above that $Cov(\frac{Nu1}{\sigma_{Nu1}}, \frac{Nu2}{\sigma_{Nu2}}) = Cor(Nu1, Nu2)$. The definition of correlation between two random variables $Nu1$ and $Nu2$ gives

$$Cor(Nu1, Nu2) = \frac{Cov(Nu1, Nu2)}{\sqrt{Var(Nu1)}\sqrt{Var(Nu2)}}$$

Combining these expressions yields:

$$Cov\left(\widehat{v}^{(1)}, \widehat{v}^{(2)}\right) \equiv \frac{Cov(Nu1, Nu2)}{\sqrt{Var(Nu1)}\sqrt{Var(Nu2)}}. \tag{3.21}$$

Let us now derive the covariance between $Nu1$ and $Nu2$. Using bilinearity as a function of the two variables ($Y_{ij}^{(1)}$ and $Y_{lm}^{(2)}$) :

$$Cov(Nu1, Nu2) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\sum_{l=1}^{k}\sum_{m=1}^{n_l}(r_{ij}^{(1)} - \frac{N+1}{2})(r_{lm}^{(2)} - \frac{N+1}{2})Cov(Y_{ij}^{(1)}, Y_{lm}^{(2)}).$$

The co-variance between $Y_{ij}^{(1)}$ and $Y_{lm}^{(2)}$ will be zero if $(i,j) \neq (l,m)$. Note also that $r_{ij}^{(1)} = r_{ij}^{(2)} = r_{ij}$. Accordingly, the covariance term will reduce to:

$$Cov(Nu1, Nu2) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(r_{ij} - \frac{N+1}{2}\right)\left(r_{ij} - \frac{N+1}{2}\right)Cov(Y_{ij}^{(1)}, Y_{ij}^{(2)})$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(r_{ij} - \frac{N+1}{2}\right)^2 Cov(Y_{ij}^{(1)}, Y_{ij}^{(2)})$$

We know that asymptotic rank transforms are independent (cf. Akritas and Arnold (1994)) and hence the variance of each of the expressions given in (3.17) and (3.18) does not involve co-variance terms. Thus, the variances of $Nu1$ and $Nu2$ can be expressed as:

$$\sigma_{Nu1}^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(r_{ij}^{(1)} - \frac{N+1}{2}\right)^2 Var(Y_{ij}^{(1)})$$

$$= \sum_{i=1}^{k} \sigma_i^{(1)^2} \sum_{j=1}^{n_i} \left(r_{ij}^{(1)} - \frac{N+1}{2} \right)^2$$

$$\sigma_{Nu2}^2 = \sum_{l=1}^{k} \sum_{m=1}^{n_l} \left(r_{lm}^{(2)} - \frac{N+1}{2} \right)^2 Var(Y_{lm}^{(2)})$$

$$= \sum_{l=1}^{k} \sigma_l^{(2)^2} \sum_{m=1}^{n_l} \left(r_{lm}^{(2)} - \frac{N+1}{2} \right)^2$$

Therefore,

$$Cov\left(\hat{v}^{(1)}, \hat{v}^{(2)}\right) = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(r_{ij} - \frac{N+1}{2} \right)^2 Cov(Y_{ij}^{(1)}, Y_{ij}^{(2)})}{\sqrt{\sum_{i=1}^{k} \sigma_i^{(1)^2} \sum_{j=1}^{n_i} \left(r_{ij}^{(1)} - \frac{N+1}{2} \right)^2} \sqrt{\sum_{l=1}^{k} \sigma_l^{(2)^2} \sum_{m=1}^{n_l} \left(r_{lm}^{(2)} - \frac{N+1}{2} \right)^2}} \tag{3.22}$$

The unknown variance and covariance terms, i.e, $Cov(Y_{ij}^{(1)}, Y_{ij}^{(2)})$, $\sigma_i^{(1)^2}$ and $\sigma_l^{(2)^2}$ can be consistently estimated by

$$\widehat{\sigma}_i^{(1)(2)} = (n_i - 1)^{-1} \sum_{j=1}^{n_i} \left(R_{ij}^{(1)} - \frac{N+1}{2} \right) \left(R_{ij}^{(2)} - \frac{N+1}{2} \right)$$

$$\widehat{\sigma}_i^{(1)^2} = (n_i - 1)^{-1} \sum_{j=1}^{n_i} \left(R_{ij}^{(1)} - \frac{N+1}{2} \right)^2$$

$$\widehat{\sigma}_l(2)^2 = (n_l - 1)^{-1} \sum_{m=1}^{n_i} \left(R_{lm}^{(2)} - \frac{N+1}{2} \right)^2$$

in the sense that

$$\widehat{\sigma}_i^{(1)(2)} - Cov(Y_{ij}^{(1)}, Y_{ij}^{(2)}) \xrightarrow{p} 0$$

$$\widehat{\sigma}_i^{(1)^2} - \sigma_i^{(1)^2} \xrightarrow{p} 0$$

$$\widehat{\sigma}_l^{(2)^2} - \sigma_l^{(2)^2} \xrightarrow{p} 0$$

The proof of consistently estimating the unknown variance and covariance terms for the above statistics can be established in a similar way as for **Theorem 2.3** in Akritas and Brunner (1997) and **Lemma A2** in Bathke (2005). □

Theorem 2 Let $R_{ij}^{(d)}$ be the (mid-) rank of $X_{ij}^{(d)}$ with $X_{ij}^{(d)} \sim F_i^{(d)}$, $d = 1,\ldots,p$, $i = 1,\ldots,k$, $j = 1,\ldots,n_i$, and r_{ij} be the (mid-) rank of the constants determining the alternative pattern among all N observations. Then under the assumptions 1 - 4, the test statistic

$$M_v = \frac{\mathbf{1}'\widehat{\mathbf{v}}}{(\mathbf{1}'\widehat{\mathbf{\Sigma}}_{\widehat{v}}\mathbf{1})^{1/2}} \tag{3.23}$$

is asymptotically distributed as standard normal.

This statistic will be employed to test the null hypothesis stated in equation (3.2) versus the alternative hypothesis given in equation (3.3). Thus, the null hypothesis is rejected if the observed value of the test statistic (M_v) exceeds the α-quantile of the standard normal distribution.
In the above,
$\mathbf{1}$ is a $p-$dimensional vector of 1's.
$\widehat{\mathbf{v}} = (\widehat{v}^{(1)},\ldots,\widehat{v}^{(p)})^T$
$\widehat{\mathbf{\Sigma}}_{\widehat{\mathbf{v}}}$ is a consistent estimator of the covariance matrix of $\widehat{\mathbf{v}}$, which is given by

$$\widehat{\mathbf{\Sigma}}_{\widehat{\mathbf{v}}} = \begin{bmatrix} 1 & \widehat{Cov\left(\hat{v}^{(1)},\hat{v}^{(2)}\right)} & \cdots & \widehat{Cov\left(\hat{v}^{(1)},\hat{v}^{(p)}\right)} \\ \widehat{Cov\left(\hat{v}^{(2)},\hat{v}^{(1)}\right)} & 1 & \ldots & \widehat{Cov\left(\hat{v}^{(2)},\hat{v}^{(p)}\right)} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{Cov\left(\hat{v}^{(p)},\hat{v}^{(1)}\right)} & \widehat{Cov\left(\hat{v}^{(p)},\hat{v}^{(2)}\right)} & \ldots & 1 \end{bmatrix} \tag{3.24}$$

For $d \neq d^* = 1,\ldots,p$, the components of this matrix (3.24) are given by

$$\widehat{Cov\left(\hat{v}^{(d)},\hat{v}^{(d^*)}\right)} = \frac{\sum_{i=1}^{k}(n_i-1)^{-1}\sum_{j=1}^{n_i}\left(R_{ij}^{(d)}-\frac{N+1}{2}\right)\left(R_{ij}^{(d^*)}-\frac{N+1}{2}\right)\sum_{j=1}^{n_i}\left(r_{ij}-\frac{N+1}{2}\right)^2}{\sqrt{\sum_{i=1}^{k}\frac{1}{n_i-1}\sum_{j=1}^{n_i}\left(R_{ij}^{(d)}-\frac{N+1}{2}\right)^2\sum_{j=1}^{n_i}\left(r_{ij}-\frac{N+1}{2}\right)^2}\sqrt{\sum_{l=1}^{k}\frac{1}{n_l-1}\sum_{m=1}^{n_i}\left(R_{lm}^{(d^*)}-\frac{N+1}{2}\right)^2\sum_{m=1}^{n_l}\left(r_{lm}-\frac{N+1}{2}\right)^2}} \tag{3.25}$$

Proof.

Consider a symmetric and idempotent matrix $(\mathbf{I}_N - \bigoplus_{i=1}^{k} n_i^{-1}\mathbf{J}_{n_i})$ in order to express vectors of adjusted asymptotic rank transforms and rank transforms as:

$$\mathbf{D}^{(d)} = \left(\mathbf{I}_N - \bigoplus_{i=1}^{k} n_i^{-1}\mathbf{J}_{n_i}\right)\mathbf{Y}^{(d)} \tag{3.26}$$

$$\widehat{\mathbf{D}}^{(d)} = \left(\mathbf{I}_N - \bigoplus_{i=1}^{k} n_i^{-1}\mathbf{J}_{n_i}\right)\widehat{\mathbf{Y}}^{(d)} \tag{3.27}$$

$$\mathbf{D} = \left(\mathbf{I}_N - \bigoplus_{i=1}^{k} n_i^{-1}\mathbf{J}_{n_i}\right)\mathbf{y} \tag{3.28}$$

In the above,
$\mathbf{D}^{(d)}$ is an $N \times 1$ vector of adjusted asymptotic rank transforms for the d^{th}response,
$\mathbf{I}_N$ is an identity matrix of dimension N,
$\mathbf{J}_{n_i}$ is an $n_i \times n_i$ matrix of ones,
$\bigoplus$ is the direct sum operator,
$\mathbf{Y}^{(d)}$ is an $N \times 1$ vector of asymptotic rank transforms for the d^{th} response,
$\widehat{\mathbf{D}}^{(d)}$ is an $N \times 1$ vector of adjusted rank transforms for the d^{th} response,
$\widehat{\mathbf{Y}}^{(d)}$ is an $N \times 1$ vector of rank transforms for the d^{th} response,
$\mathbf{D}$ is an $N \times 1$ vector of adjusted rank transforms for the constant term, and
$\mathbf{y}$ is an $N \times 1$ vector of rank transforms for the constant term.

Define independent random variables as in Assumption 4.

$$Q_{ij} = \sum_{d=1}^{p} D_{ij}Y_{ij}^{(d)}, i = 1,\ldots,k, j = 1,\ldots,n_i.$$

Let the means and variances of Q_{ij} be denoted by μ_{ij} and v_{ij}, respectively. Suppose that the overall (for all Q_{ij}) standard deviation is represented as $s = (\sum_{i=1}^{k}\sum_{j=1}^{n_i} v_{ij})^{1/2}$. Note that the overall variance does not involve co-variance terms since $Q_{ij}'s$ are independent. Under Assumption 1, $s^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Q_{ij} - \mu_{ij})$ is asymptotically distributed

as standard normal since the Q_{ij} are uniformly bounded and hence the Kolmogorov condition for Central Limit Theorem is applicable.

Let us now rewrite $\sum_{i=1}^{k}\sum_{j=1}^{n_i} Q_{ij}$ in vector form.

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} Q_{ij} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} D_{ij}\left(Y_{ij}^{(1)} + \ldots + Y_{ij}^{(p)}\right) = \mathbf{D}^T\left(\mathbf{Y}^{(1)} + \ldots + \mathbf{Y}^{(p)}\right)$$

We can employ equations (3.26) - (3.28) to rewrite $\mathbf{D}^T\left(\mathbf{Y}^{(1)} + \ldots + \mathbf{Y}^{(p)}\right)$ as:

$$\mathbf{D}^T\left[\mathbf{Y}^{(1)} + \ldots + \mathbf{Y}^{(p)}\right] = \mathbf{y}^T\left(\mathbf{I}_N - \bigoplus_{i=1}^{k} n_i^{-1}\mathbf{J}_{n_i}\right)^T\left[\mathbf{Y}^{(1)} + \ldots + \mathbf{Y}^{(p)}\right]$$

The next equation follows as $\left(\mathbf{I}_N - \bigoplus_{i=1}^{k} n_i^{-1}\mathbf{J}_{n_i}\right)$ is a symmetric and idempotent matrix.

$$\mathbf{D}^T\left[\mathbf{Y}^{(1)} + \ldots + \mathbf{Y}^{(p)}\right] = \mathbf{y}^T\left(\mathbf{I}_N - \bigoplus_{i=1}^{k} n_i^{-1}\mathbf{J}_{n_i}\right)^T\left(\mathbf{I}_N - \bigoplus_{i=1}^{k} n_i^{-1}\mathbf{J}_{n_i}\right)\left[\mathbf{Y}^{(1)} + \ldots + \mathbf{Y}^{(p)}\right]$$

$$i.e., \mathbf{D}^T\left[\mathbf{Y}^{(1)} + \ldots + \mathbf{Y}^{(p)}\right] = \mathbf{D}^T\left(\mathbf{I}_N - \bigoplus_{i=1}^{k} n_i^{-1}\mathbf{J}_{n_i}\right)\left[\mathbf{Y}^{(1)} + \ldots + \mathbf{Y}^{(p)}\right]$$

$$= \mathbf{D}^T\left(\mathbf{D}^{(1)} + \ldots + \mathbf{D}^{(p)}\right).$$

Let us now find the expected value of $\sum_{i=1}^{k}\sum_{j=1}^{n_i} Q_{ij} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \mathbb{E}(Q_{ij}) = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \mu_{ij}$.

$$But, \mathbb{E}(Q_{ij}) = \mathbb{E}\left[\mathbf{D}^T\left(\mathbf{D}^{(1)} + \ldots + \mathbf{D}^{(p)}\right)\right] = \mathbb{E}\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} D_{ij}\left(D_{ij}^{(1)} + \ldots + D_{ij}^{(p)}\right)\right]$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n_i} D_{ij}\mathbb{E}\left(D_{ij}^{(1)} + \ldots + D_{ij}^{(p)}\right)$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n_i} D_{ij}\mathbb{E}\left[\left(Y_{ij}^{(1)} - \overline{Y}_i^{(1)}\right) + \ldots + \left(Y_{ij}^{(p)} - \overline{Y}_i^{(p)}\right)\right]$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n_i} D_{ij}\left[\mathbb{E}\left(Y_{ij}^{(1)} - \overline{Y}_i^{(1)}\right) + \ldots + \mathbb{E}\left(Y_{ij}^{(p)} - \overline{Y}_i^{(p)}\right)\right] = 0$$

The above result follows since $\mathbb{E}\left(Y_{ij}^{(d)} - \overline{Y}_i^{(d)}\right) = 0 \ \forall d = 1, \ldots, p$ under H_0^F.

Finally, we can obtain that $\sum_{i=1}^{k}\sum_{j=1}^{n_i} \mu_{ij} = 0$.

We presented earlier that $s^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Q_{ij} - \mu_{ij}) = s^{-1}\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} Q_{ij} - \sum_{i=1}^{k}\sum_{j=1}^{n_i}\mu_{ij}\right) \dot{\sim} N(0,1)$.

As $\sum_{i=1}^{k}\sum_{j=1}^{n_i}\mu_{ij} = 0$ (shown above),

$$s^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Q_{ij} = s^{-1}\left[\mathbf{D}^T\left(\mathbf{D}^{(1)} + \ldots + \mathbf{D}^{(p)}\right)\right] \dot{\sim} N(0,1).$$

The above implies that $\mathbf{D}^T\left(\mathbf{D}^{(1)} + \ldots + \mathbf{D}^{(p)}\right) \dot{\sim} N(0, s^2)$. It is well known that the distribution will remain the same if we multiply and/or divide $\mathbf{D}^T\left(\mathbf{D}^{(1)} + \ldots + \mathbf{D}^{(p)}\right)$ by a constant, i.e.,

$$\frac{\mathbf{D}^T\mathbf{D}^{(1)}}{s^{(1)}} + \ldots + \frac{\mathbf{D}^T\mathbf{D}^{(p)}}{s^{(p)}} \dot{\sim} N(0, s^{*2})$$

where the constants $s^{(1)}, \ldots, s^{(p)}$ are assumed to be the standard deviations of $\mathbf{D}^T\mathbf{D}^{(1)}, \ldots, \mathbf{D}^T\mathbf{D}^{(p)}$, respectively, and s^{*2} is the variance of the expressions before $\dot{\sim}$.

The asymptotic rank transforms can be replaced by the corresponding rank transforms, i.e., $N^{-1/2}\mathbf{D}^T(\widehat{\mathbf{D}}^{(d)} - \mathbf{D}^{(d)}) \xrightarrow{p} 0$ and the standard deviations can be replaced by their consistent estimators, i.e., $\widehat{s^{(d)}}^2 - s^{(d)2} \xrightarrow{p} 0$ for each $d = 1, \ldots, p$

Lemma A1 and **Lemma A2** in Bathke (2005)). Furthermore, we apply a continuous mapping theorem in order to get the following result.

$$\frac{\mathbf{D}^T\widehat{\mathbf{D}}^{(1)}}{\widehat{s^{(1)}}}+\ldots+\frac{\mathbf{D}^T\widehat{\mathbf{D}}^{(p)}}{\widehat{s^{(p)}}}\dot{\sim}N(0,s^{*2}) \tag{3.29}$$

If we substitute $\frac{\mathbf{D}^T\widehat{\mathbf{D}}^{(d)}}{\widehat{s^{(d)}}}$ by $\hat{v}^{(d)}$, $d=1,\ldots,p$, and standardise the expression given on the left side of $\dot{\sim}$ in (3.29), we can assert

$$\frac{\hat{v}^{(1)}+\ldots+\hat{v}^{(p)}}{s^*}\dot{\sim}N(0,1) \quad \square$$

Let us now determine a consistent estimator of s^* as this quantity is unknown and needs to be estimated.
Consider the variances and covariances of the sum of test statistics $\{\hat{v}^{(1)},\ldots,\hat{v}^{(p)}\}$, that is,

$$\begin{aligned} s^{*2} = Cov(\hat{v}^{(1)}+\ldots+\hat{v}^{(p)}) = \mathbf{1}^T\Sigma_{\hat{\mathbf{v}}}\mathbf{1} = 1+Cov(\widehat{v}^{(2)},\widehat{v}^{(1)})+\ldots+ \\ Cov(\widehat{v}^{(p)},\widehat{v}^{(1)})+\ldots+Cov(\widehat{v}^{(1)},\widehat{v}^{(p)})+\ldots+Cov(\widehat{v}^{(p-1)},\widehat{v}^{(p)})+1 \end{aligned}$$

The formula to find the covariance between any two test statistics $\hat{v}^{(d)}$ and $\hat{v}^{(d^*)}$, $d\neq d^*=1,\ldots,p$ is derived in Theorem 1. Again, this involves unknown variance and covariance terms. When these unknown quantities are substituted by their consistent estimators by applying a similar technique as in **Theorem 2.3** of Akritas and Brunner (1997) and **Lemma A2** of Bathke (2005), we can have the following $\widehat{Cov(\hat{v}^{(d)},\hat{v}^{(d^*)})}-Cov(\hat{v}^{(d)},\hat{v}^{(d^*)})\xrightarrow{p}0$, where $\widehat{Cov(\hat{v}^{(d)},\hat{v}^{(d^*)})}$ is as given in equation (3.25). Furthermore, applying the continuous mapping theorem on the function of these covariances yields

$$\mathbf{1}^T\hat{\Sigma}_{\hat{\mathbf{v}}}\mathbf{1}-s^{*2}\xrightarrow{p}0, \text{ where } \hat{\mathbf{\Sigma}}_{\hat{\mathbf{v}}} \text{ is as in equation (3.24).}$$

3.4 Simulation and real data example

3.4.1 Simulation

We performed 10,000 simulation runs in order to evaluate the performance of the newly proposed test statistic for small to moderate sample sizes with regard to whether it maintains the preassigned type-I error rate ($\alpha = 5\%$) under the null hypothesis and its power to detect directional alternatives. According to this simulation setup, the 95% likely interval for type-I error rate is given by (4.573%, 5.432%). All simulations were carried out using R (version 3.4.4, R Core Team (2018)).

Samples of sizes of $n = 5, 10, 15, 20, 30,$ and 50 were taken in each setting for which level and power had been evaluated on the basis of one-sided tests. In order to compare the type-I error rates and powers of the newly developed statistic with the most popular procedure to test directional alternatives in multivariate problems, i.e., the test by O'Brien (1984), we first restricted our considerations to $k = 2$ independent treatment groups, and data were generated from normal, exponential, and lognormal distributions. We applied the Cholesky decomposition to draw multivariate samples from normal, exponential, and lognormal distributions. In addition, we generated multivariate ordered categorical data using the R package GenOrd version 1.4.0. Assuming the pattern to be detected is an increasing trend, the covariate determining the alternative pattern takes a value 1 in the first treatment group and 2 in the second one. However, the proposed statistic is invariant under any monotone transformations of the response variables as well as the covariate. Therefore, an increasing trend can be defined by (1, 2) or by (1, 5) or (0, 4), and so on.

We considered a wide variety of different correlation structures (strong positive, strong negative, weak positive, and weak negative) between response variables, but only correlations 0.90, 0.35, -0.25, -0.35, -0.90 and unequal correlation are reported here (see Tables 3.2 - 3.4). Note that in the case of data from ordinal variables, the Spearman's correlation was used.

The simulation results presented in Table 3.2 indicate that the newly developed test keeps the type-I error rate equally well, sometimes even better than that of

O'Brien (1984). It can be observed that the actual significance levels of the rank sum test by O'Brien exceed the nominal level when correlations between responses are strong positive, weak positive, weak negative and/or unequal (ranged from 5.01% to 5.38%). However, these simulated type-1 error values lie within (4.573%, 5.432%), i.e., the 95% likely interval for the nominal type-1 error rate. The results also reveal that both the test by O'Brien and the newly developed test are conservative (simulated type-I error rate less than 4.573%) when $n = 5$ and in some cases when $n = 10$. When data are generated from the normal distribution with unequal correlation between responses, the newly proposed test gives a type-I error rate which exceeds 5%, but this value is still within the 95% likely interval for the nominal type-I error rate. In general, the simulated type-I error rates when $\alpha = 5\%$ and n between 10 and 20 for M_v lie between 4.15% to 4.88%.

Distribution	n_i	p	Correlation	O'Brien	M_v	Correlation	O'Brien	M_v
Normal	5	2	0.9	0.0350	0.0338	-0.9	0.401	—
	10			0.0406	0.0416		0.0422	0.0445
	15			0.0445	0.0419		0.0502	0.0486
	20			0.0454	0.0437		0.0479	0.0455
	30			0.0486	0.0475		0.0505	0.0477
	50			0.0485	0.0492		0.0473	0.0489
Normal	5	2	0.35	0.0459	0.0422	-0.35	0.0442	0.0388
	10			0.0538	0.045		0.0504	0.0459
	15			0.0539	0.0452		0.0516	0.0461
	20			0.0479	0.0442		0.0497	0.0453
	30			0.0517	0.0459		0.0512	0.047
	50			0.0484	0.0499		0.0463	0.0499
Normal	5	5	0.9	0.0494	0.0409	-0.25	0.0453	0.04
	10			0.0509	0.0454		0.0524	0.0415
	15			0.0517	0.0488		0.0499	0.0484
	20			0.0491	0.0476		0.0484	0.0453
	30			0.0512	0.0473		0.0486	0.0506
	50			0.0478	0.0445		0.0532	0.0503
Normal	5	5	0.35	0.0483	0.0388	Unequal	0.0478	0.0357
	10			0.0501	0.0431		0.0471	0.0434
	15			0.0499	0.0483		0.0492	0.047
	20			0.0483	0.0495		0.0503	0.0477
	30			0.05	0.0493		0.0504	0.0515
	50			0.051	0.0487		0.0495	0.0525

TABLE 3.2: Simulated $\alpha-$ levels for O'Brien (1984) rank sum and the newly proposed test (M_v) under normal distribution and different correlations between responses for two samples

Table 3.3 reports the simulated type-I error rates of the statistics by O'Brien and M_v for two multivariate samples under exponential distribution. As in the case of normal distribution, M_v is conservative for small sample sizes ($n = 10$; $n = 5$).

When we compare the two tests and is slightly more conservative than O'Brien for smaller sample size ($n = 5$) with α values between 3.52% and 4.44%. On the other hand, the rank sum test by O'Brien is found to be liberal for $n = 10$, correlation = 0.35, and five response variables with $\alpha = 5.5\%$. In addition, the rank sum test by O'Brien is found to be liberal (simulated error rate = 5.55% which is above the upper confidence limit of the 95% confidence interval for α) when correlation is strong negative and sample size is 30.

Distribution	n_i	**p**	**Correlation**	**O'Brien**	M_v	**Correlation**	**O'Brien**	M_v
Exponential	5	2	0.9	0.0508	0.0405	-0.9	0.0363	—
	10			0.0482	0.0424		0.0453	0.0444
	15			0.0481	0.0447		0.475	0.0498
	20			0.0479	0.0463		0.0472	0.0508
	30			0.0492	0.0496		0.0555	0.0501
	50			0.0488	0.0464		0.0490	0.0478
Exponential	5	2	0.35	0.0478	0.0402	-0.35	0.0449	0.0352
	10			0.0494	0.0433		0.0493	0.0436
	15			0.0473	0.0481		0.0473	0.0436
	20			0.0478	0.0485		0.0484	0.0436
	30			0.0507	0.0521		0.0524	0.0507
	50			0.047	0.0487		0.049	0.0467
Exponential	5	5	0.9	0.0500	0.0380	-0.25	0.0441	0.0367
	10			0.0525	0.0495		0.0488	0.0467
	15			0.0524	0.0462		0.0508	0.049
	20			0.0478	0.0470		0.0521	0.0518
	30			0.0500	0.0512		0.0513	0.0482
	50			0.0516	0.0489		0.0473	0.0513
Exponential	5	5	0.35	0.0499	0.0388	Unequal	0.0481	0.037
	10			0.055	0.0487		0.0553	0.0484
	15			0.0519	0.0473		0.0497	0.0465
	20			0.0469	0.0455		0.0497	0.0466
	30			0.0512	0.0494		0.051	0.0491
	50			0.0528	0.0494		0.0507	0.0481

TABLE 3.3: Simulated $\alpha-$ levels for O'Brien (1984) rank sum and the newly proposed test (M_v) under exponential distribution and different correlations between responses for two samples

In general, the newly proposed test accurately controls the nominal type-I error rate when data are generated from lognormal distribution and sample sizes are 10 or more. There are in fact some situations where the test is slightly conservative, for

example when correlation is strong negative or unequal (see Table 3.4 for details).

Distribution	n_i	**p**	**Correlation**	**O'Brien**	M_v	**Correlation**	**O'Brien**	M_v
Lognormal	5	2	0.9	0.0459	0.0431	-0.9	0.0369	—
	10			0.0526	0.0453		0.0443	0.0430
	15			0.0542	0.0426		0.0496	0.0438
	20			0.0479	0.0429		0.0507	0.0441
	30			0.0516	0.0479		0.0495	0.0463
	50			0.0493	0.0516		0.0483	0.0513
Lognormal	5	2	0.35	0.0449	0.0433	-0.35	0.0438	0.0366
	10			0.0525	0.0454		0.0519	0.0442
	15			0.0535	0.0451		0.0524	0.0454
	20			0.0491	0.0449		0.0513	0.0442
	30			0.0521	0.047		0.0489	0.0455
	50			0.0481	0.0486		0.0485	0.0502
Lognormal	5	5	0.9	0.0509	0.0383	-0.25	0.0433	0.0391
	10			0.0499	0.0460		0.0485	0.0451
	15			0.0547	0.0478		0.0504	0.0459
	20			0.0488	0.0456		0.0493	0.0459
	30			0.0519	0.0493		0.0506	0.0476
	50			0.0484	0.0453		0.0478	0.0469
Lognormal	5	5	0.35	0.0494	0.0406	Unequal	0.0484	0.0381
	10			0.0503	0.0462		0.0522	0.0419
	15			0.0512	0.047		0.0515	0.0478
	20			0.0464	0.0471		0.0467	0.0461
	30			0.0499	0.0527		0.0513	0.0506
	50			0.052	0.0485		0.0514	0.049

TABLE 3.4: Simulated $\alpha-$ levels for O'Brien (1984) rank sum and the newly proposed test (M_v) under lognormal distribution and different correlations between responses for two samples

When data are generated from multivariate ordinal variable, the newly developed test accurately controls the type-I error rate except that correlation is weak positive

(see Table 3.5 for more).

n_i	**p**	**Correlation**	M_v	p	**Correlation**	M_v	p	**Correlation**	M_v
15			0.0458			0.0468			0.0477
20	2	0.60	0.0465	2	-0.53	0.0464	2	0.35	0.0461
30			0.048			0.0469			0.0444
50			0.0501			0.0524			0.0472
15			0.0471			0.0504			0.0475
20	2	-0.35	0.0573	5	0.60	0.0469	5	0.35	0.0445
30			0.0507			0.0486			0.0447
50			0.047			0.0497			0.05
15			0.048						
20	5	Unequal	0.0489						
30			0.047						
50			0.0525						

TABLE 3.5: Simulated $\alpha-$ levels for the newly proposed test (M_v) when data are simulated from multivariate ordinal variable with different correlations between responses for two samples

In summary, the performance of the newly proposed test is found to be on a par with, or sometimes even better than O'Brien (1984)'s rank-sum test when the correlation between responses is strong as well as unequal. A close look at of the above results indicates that it allows a fair control of the type-I error rate by the newly proposed statistic for small sample sizes ($10 \leq n \leq 20$) under all investigated distributions including discrete distributions. The new test is thus rather versatile in use.

Simulated α levels in the case of three samples are presented in the following graphs (Figure 3.1 - 3.3).

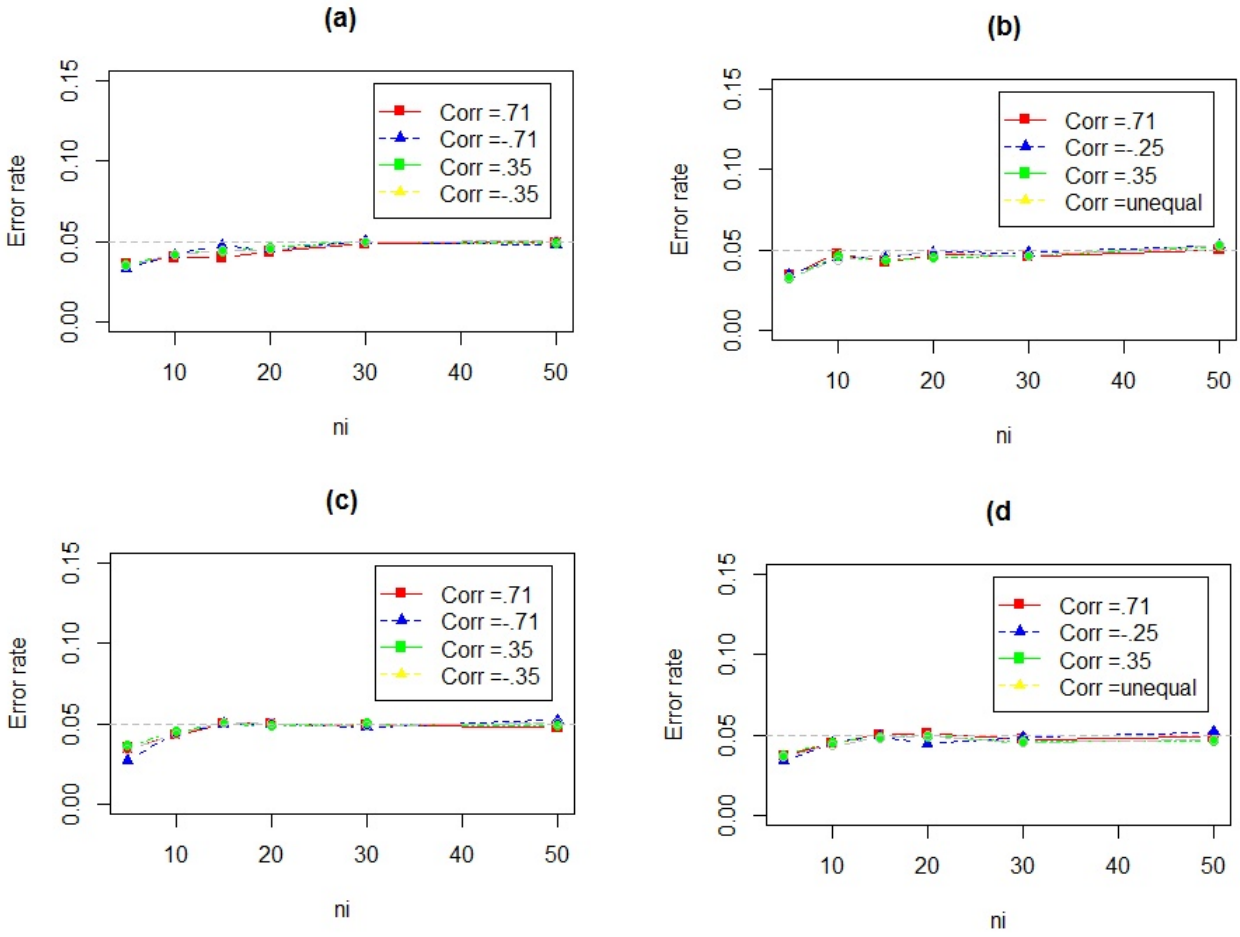

FIGURE 3.1: Simulated α levels for M_v in $k = 3$ from (a) normal distribution and $p = 2$, (b) normal distribution and $p = 5$, (c) exponential distribution and $p = 2$, and (d) exponential distribution and $p = 5$

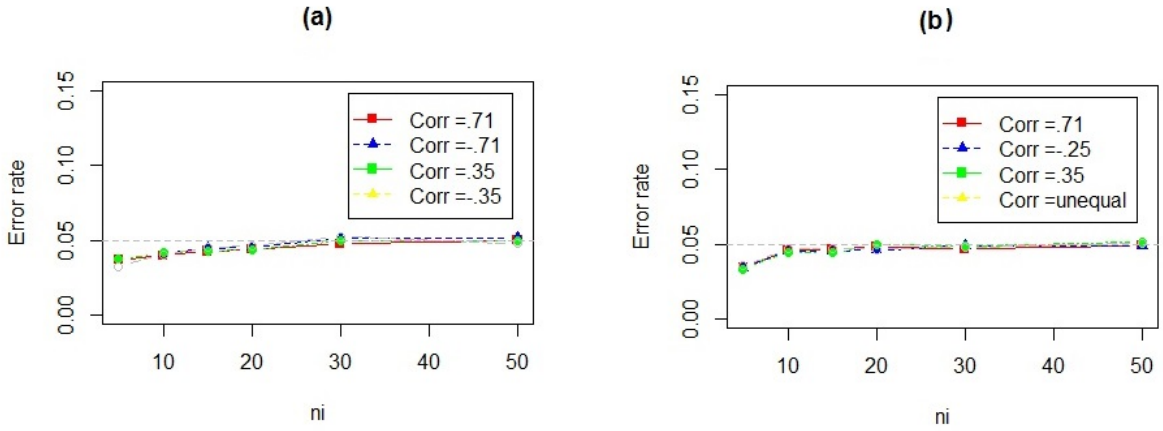

FIGURE 3.2: Simulated α levels for M_v in $k = 3$ from (a) Lognormal distribution and $p = 2$, and (b) Lognormal distribution and $p = 5$

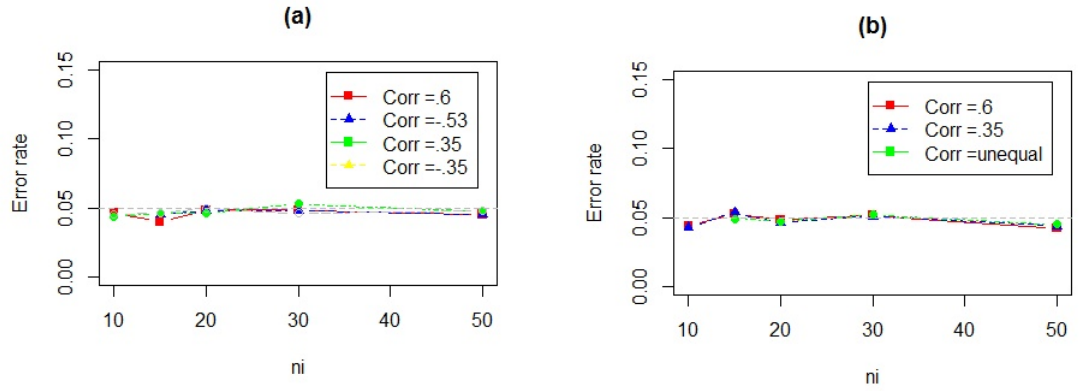

FIGURE 3.3: Simulated α levels for M_v in $k = 3$ from (a) discrete distribution and $p = 2$, (b) discrete distribution and $p = 5$

The performance of M_v to keep the nominal type-I error rate ($\alpha = 0.05$) in three samples ($k = 3$) is found to be similar to that of the case in two samples. Each of the different cases considered above reveal that the test can be reasonably applied for $n \geq 10$ (see Figures 3.1 - 3.3).

The simulated powers for M_v and O'Brien's rank sum tests for different distributions and correlation structures are presented in the following graphs.

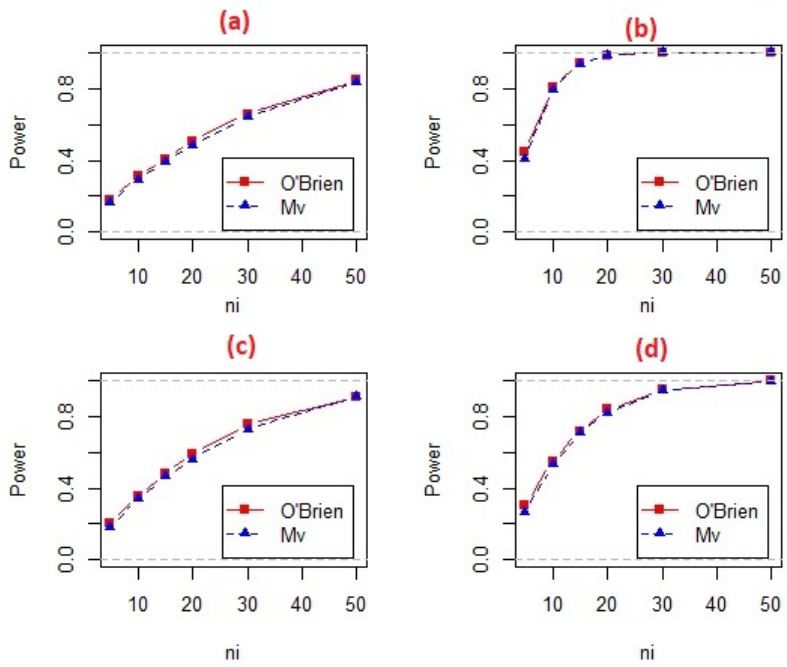

FIGURE 3.4: Simulated power for M_v and O'Brien with two response variables from normal distribution with (a) correlation = 0.71, (b) correlation = -0.71, (c) correlation = 0.35, and (d) correlation = -0.35

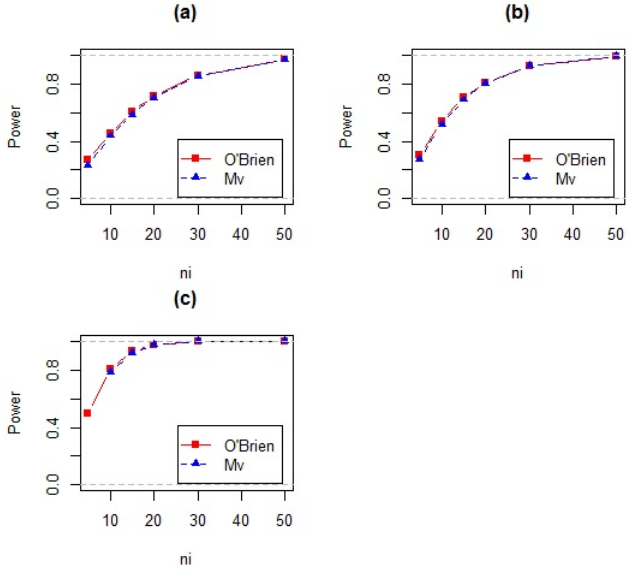

FIGURE 3.5: Simulated power for M_v and O'Brien with two response variables from exponential distribution with (a) correlation = 0.71, (b) correlation = 0.35, and (c) correlation = -0.35.

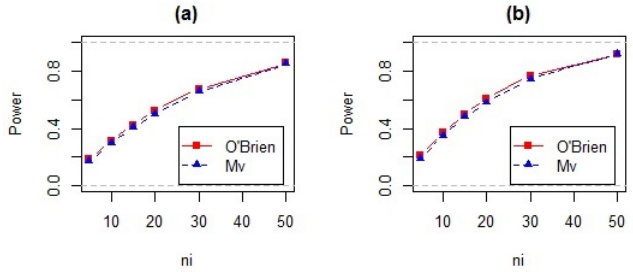

FIGURE 3.6: Simulated power for M_v and O'Brien with two response variables from (a) Lognormal distribution with correlation = 0.71, and (b) Lognormal distribution with correlation = 0.35

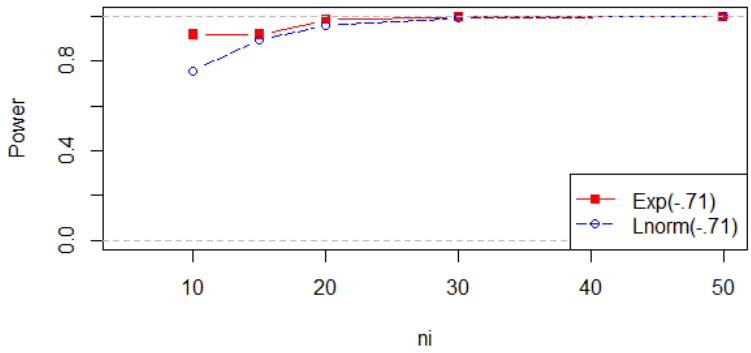

FIGURE 3.7: Simulated power for M_v with two response variables from exponential distribution with correlation = -0.71, and lognormal distribution with correlation = -0.71

Figures 3.4 - 3.7 report the simulated power of the statistics by O'Brien's rank sum and M_v. To evaluate the respective power, we considered a location shift alternative by adding 0.5 to each observation in the second sample as in O'Brien. The results indicate that both M_v and O'Brien's rank sum tests provide rather comparable power and can generally be recommended for $n \geq 15$. The rank sum test by O'Brien is slightly more powerful than M_v when the correlation between responses is negative and/or unequal. This stems from the fact that the rank sum test by O'Brien is is slightly more liberal than M_v when the correlation between responses is negative

and/or unequal. There is an indication of superiority of M_v over O'Brien's rank sum test for a moderately large sample size ($n = 50$).

The simulated power for M_v in three samples when data are from normal, exponential, lognormal and discrete distributions are provided below.

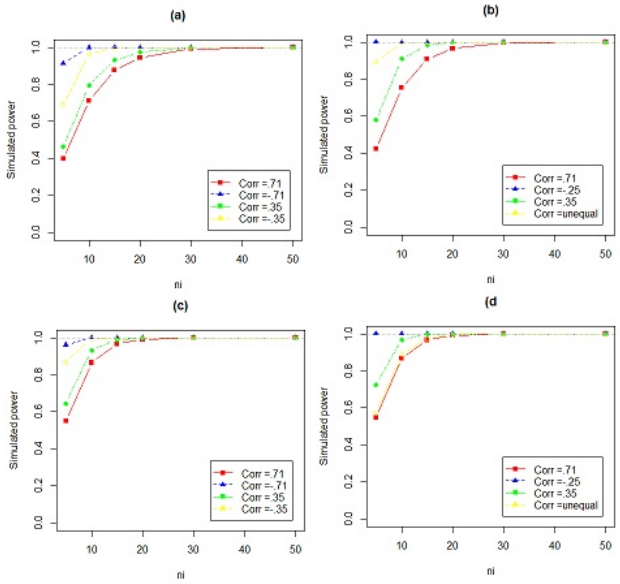

FIGURE 3.8: Simulated power for M_v in $k = 3$ from (a) normal distribution and $p = 2$, (b) normal distribution and $p = 5$, (c) exponential distribution and $p = 2$, and (d) exponential distribution and $p = 5$

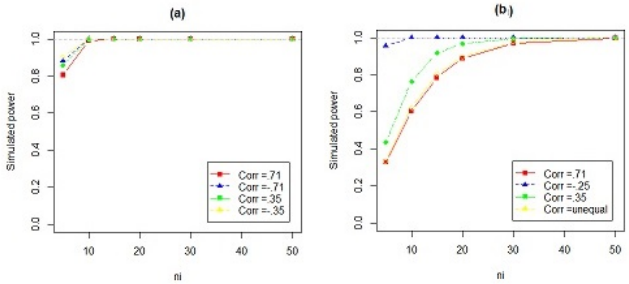

FIGURE 3.9: Simulated power for M_v in $k = 3$ from (a) Lognormal distribution and $p = 2$, and (b) Lognormal distribution and $p = 5$

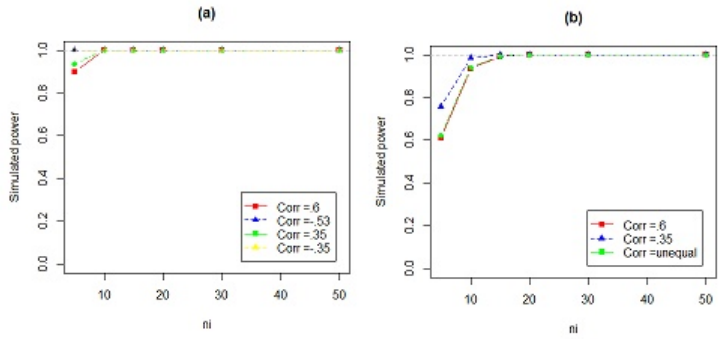

FIGURE 3.10: Simulated power for M_v in $k = 3$ from (a) bivariate ordinal variable, and (b) multivariate ($p = 5$) ordinal variable.

The simulated power curves given in Figures 3.8 - 3.10 reveal that the newly proposed test is efficient in each of the parametric configurations. In particular, the test is powerful when sample size is ten or more, correlation between responses is strong and/or the distribution under consideration is discrete.

3.4.2 Illustration through real data

As a real data illustration, we consider different datasets from clinical trials. The example datasets include electroencephalogram (EEG), psychiatric clinical trial, treatment of acute thrombotic stroke and dose-response trial.

EEG data

Electroencephalography is an electro-physiological method to record the electrical activity of the brain which is generated by the synchronized activity of thousands of neurons (in volts). An instrument called electroencephalograph measures and records voltage fluctuations resulting from ionic current within the neurons of the brain. The result obtained from such recording is called an electroencephalogram (EEG). In order to make the recording conveniently, a neurologist places small flat metal discs called electrodes on the scalp of the patient (see Figure 3.11). Signals

are transmitted to the recording channels of the electroencephalograph through electrodes. These electrodes are mounted in elastic caps for the sake of collecting data from identical scalp positions across all patients [1]. When a regular rhythm is observed on the EEG, it is evident that the neurons synchronize their activity.

The use of electroencephalography varies in different areas of psychiatry. In clinical psychiatry, it is used to detect signs of organic brain damage such as epilepsy, tumors and brain injuries, cerebral circulation and metabolic disorders, and neurodegenerative diseases. However, in biological psychiatry, the EEG is utilized to asses the functional state of certain brain structures and systems in order to study neurophysiological mechanisms of mental disorders, and the effects of psychotropic drugs.

FIGURE 3.11: Pictorial view of performing EEG (Source: www.iLiveOk.com)

Note. From Electroencephalography, by M. Tal, 2020 (https://iliveok.com/health/electroencephalography_105674i15989.html). Copyright 2011 - 2021 by iLive.

Electroencephalogram (EEG) were collected from 160 subjects and stored in the R-package HRM (Happ et al. (2016)). Four EEG variables, namely activity, complexity, mobility and brain rate were measured at ten different brain regions: frontal left, frontal right, parietal left, parietal right, temporal left, temporal right, occipital left, occipital right, central left, central right. Accordingly, forty (4 variables × 10

[1] Bryn Franswoth, What is EEG (Electroencephalography) and How Does it Work?, (accessed on December 24, 2020 online at https://imotions.com/blog/what-is-eeg/)

regions) measurements have been taken from each patient. The patients were classified into four diagnostic groups, i.e., having Alzheimer's disease (AD), mild cognitive impairment (MCI), subjective cognitive complaints with minimal cognitive dysfunction (SCC+) and subjective cognitive complaints without minimal cognitive dysfunction (SCC-). Part of the EEG dataset available in the R-package HRM is presented in the following table.

S.No.	Group	Value	Sex	Subject	Variable	Region	Dimension
1	SCC+	2.8041803	W	1	1	1	1
2	SCC+	2.2850036	W	1	1	2	2
3	SCC+	1.4730937	W	1	1	3	3
4	SCC+	1.5489391	W	1	1	4	4
5	SCC+	2.1603461	W	1	1	5	5
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
6395	AD	1.943609	M	160	4	5	35
6396	AD	1.943292	M	160	4	6	36
6397	AD	1.984193	M	160	4	7	37
6398	AD	1.957243	M	160	4	8	38
6399	AD	1.917165	M	160	4	9	39
6400	AD	1.881346	M	160	4	10	40

TABLE 3.6: EEG dataset taken from the R-package HRM

The objective of the study was to analyze electrical impulses in the brain by electrodes and signals which are transfered to a computer and record the results. The results help to assess whether there are abnormal conditions in the brain such as seizure disorders, head injury, stroke, tumor or dementia.

We calculated the average value of the measurements taken from ten different regions of the brain for each response per subject so that the dimension is reduced to four.

Subject	Group	Activity	Complexity	Mobility	Brain rate
1	AD	2.708176781	0.593478708	-0.639679952	2.08937013
2	AD	3.161949918	1.139704223	-1.119974921	1.914482755
3	AD	2.627534571	0.750321837	-0.877292885	2.042842967
4	AD	3.112878326	1.029464401	-1.100576204	1.97316591
5	AD	3.327205531	0.738105121	-0.790014555	2.000404451
⋮	⋮	⋮	⋮	⋮	⋮
160	SCC-	4.596377089	0.652031578	-0.973908235	2.046397979

TABLE 3.7: Part of the aggregated EEG data set

The groups SCC-, SCC+, MCI, and AD are in natural order, and we anticipate to detect a monotone trend. The alternative pattern across these four groups is described by 1 for AD, 2 for MCI, 3 for SCC+, and 4 for SCC-. Applying the newly proposed test statistic (M_v), that is the formula in equation (3.23), gives $Mv = 2.904$ and $p = 0.00184$. The result confirms a strong decreasing trend in the above groups.

Psychiatric clinical trial data

We apply the test statistic proposed here to the psychiatric clinical trial dataset considered by Munzel and Brunner (2000b) as an illustration for the nonparametric multivariate procedure they proposed. In this dataset, 45 patients with panic disorder were randomized into three treatment groups each with 15 patients. Patients in the physical exercise (PE) group had to perform physical exercise, those in clomipramine (CL) group were given clomipramine which is used to treat symptoms of obsessive-compulsive disorder such as recurrent thoughts or feelings and repetitive actions [2] while the remaining were assigned to a placebo group (PBO). Data were taken from each patient based on two response variables, namely clinical global impression (CGI) and patient's global impression (PGI). After four weeks of

[2] What is clomipramine?, (accessed on February 19, 2021 online at https://www.drugs.com/mtm/clomipramine.html)

treatment, an investigator rated CGI for each patient based on a seven point discrete scale of scores 0 and 6 (0 = not ill, 6 = extremely ill). Each patient assigned a PGU score for herself/ himself based on a seven point ordinal scale as in the case of CGI. The dataset is given in Table 3.8.

Exercise			**Clomipramine**			**Placebo**		
S.No.	CGI	PGI	S.No.	CGI	PGI	S.No.	CGI	PGI
1	4	3	16	1	2	31	5	4
2	1	1	17	1	1	32	5	5
3	2	2	18	2	0	33	5	6
4	2	3	19	2	1	34	5	4
5	2	3	20	2	3	35	2	6
6	1	2	21	2	3	36	4	6
7	3	3	22	3	4	37	1	1
8	2	3	23	1	4	38	4	5
9	5	5	24	1	1	39	2	1
10	2	2	25	2	0	40	4	4
11	5	5	26	2	3	41	5	5
12	2	4	27	1	0	42	4	4
13	2	1	28	1	1	43	5	4
14	2	4	29	1	1	44	5	4
15	6	5	30	2	1	45	3	4

TABLE 3.8: CGI and PGI scores for patients in psychiatric trial after 4 weeks of treatment

The aim of the analysis is to test a potential superiority of one treatment over the others. Munzel and Brunner (2000b) calculated nonparametric relative (marginal) treatment effects for each of the two responses (see Table 3.9).

Score	CL	PE	PBO
CGI	0.34	0.59	0.57
PGI	0.29	0.56	0.65

TABLE 3.9: Relative treatment effects for the three treatment groups psychiatric clinical trial datset.

These treatment effects reveal that the effects of treatment CL are much smaller than the other two, i.e., PE and PBO for both responses while there is only a slight difference between the treatments PE and PBO. To test equality of treatment effects versus the alternative that the effect is different in at least one of the treatments, they considered four statistics, namely the Wald-type statistic (Q_N^W), the ANOVA-type statistic (F_N), and two versions of the Lawley-Hotelling statistic: one is based on chi-squared approximation (LH) and the other is the small sample approximation (LH_c). The results from these test statistics are given in Table 3.10. Each of these statistics indicates that the three treatments are not equally effective. Furthermore, a pairwise comparison of the three treatments by using these statistics was conducted (see Table 3.11).

Statistic	d.f.	p-value
$Q_N^W = 19.10$	4	$= 0.0008$
$F_N = 18.15$	2.83	$= 0.0003$
$LH = 0.46$	4	$= 0.0007$
$LH_c = 0.46$	4	< 0.0050

TABLE 3.10: Results of the statistics Q_N^W, F_N,LH, and LH_c for psychiatric clinical trial dataset.

Statistic	Comparison	p-value
$Q_N^W = 14.77$	CL vs. PBO	$= 0.0006$
$Q_N^W = 2.39$	PE vs. PBO	$= 0.3033$
$Q_N^W = 12.45$	CL vs. PE	$= 0.0020$
$F_N = 13.59$	CL vs. PBO	$= 0.0006$
$F_N = 0.99$	PE vs. PBO	$= 0.4649$
$F_N = 13.46$	CL vs. PE	$= 0.0005$
$LH = 0.53$	CL vs. PBO	$= 0.0006$
$LH = 0.09$	PE vs. PBO	$= 0.3033$
$LH = 0.44$	CL vs. PE	$= 0.0020$
$LH_c = 0.53$	CL vs. PBO	< 0.005
$LH_c = 0.09$	PE vs. PBO	> 0.3
$LH_c = 0.44$	CL vs. PE	< 0.01

TABLE 3.11: Pairwise comparison results of the three treatments based on statistics Q_N^W, F_N,LH, and LH_c for psychiatric clinical trial dataset.

Note in this dataset that there are three samples (groups) in which data were recorded based on two response variables which are measured at an ordinal scale. A reasonable question in study of this type is whether the observed scores (CGI and PGI) remained the same or show some monotone trend across the treatment groups. As a result, rank-based tests which are sensitive to a priori conjectured patterns such as the one proposed in this thesis are desirable. It is anticipated that there is an increasing trend of clinical global impression as well as patient's global impression scores with regard to the three treatment groups, that is, $CL < PE < PBO$. Accordingly, the values 1, 2 and 3 are assigned for the covariate in the CL, PE and PBO groups, respectively. The calculated value of the statistic M_v is found to be 4.009744, and the p-value corresponding to this is $3.04x10^{-5}$. It is thus evident that there is an increasing trend.

Treatment of acute thrombotic stroke data

The test derived in this thesis is also applied to the thrombotic stroke treatment dataset available in Tsai and Koziol (1994). A thrombotic stroke is a type of ischemic stroke when part of the brain gets injured because the artery gets blocked by a blood clot. Acute stroke can be treated by fibrinolytic agents. The relatively fibrin-selective nature of tissue plasminogen activator (t-PA) has made it an attractive thrombolytic agent for intravenous administration in the treatment of stroke. However, researchers suspect that intravenous infusion of recombinant t-PA is associated in a dose dependent way with the central nervous system hemorrhage. To address this issue, the effect of an intravenous infusion of recombinant t-PA on hemorrhagic transformation early after middle cerebral artery territory ischemia was studied in a baboon model. 30 baboons were taken in the study. In each of these baboons, 3 hours occlusion of the middle cerebral artery was induced and 30 minutes re-oxygenation was done. Then a 60-minute infusion of recombinant t-PA (at three doses: Dose 1 - 0.3 mg/kg, n = 6; Dose 2 - 1.5 mg/kg, n = 6; Dose 3 - 10 mg/kg, n = 6) or normal

saline (n = 12) was considered. After 24 hours, the volumes of intracerebral hemorrhage (V) are obtained by computed tomography, and neurological functions (N) were assessed in each animal according to a weighted quantitative neurologic scale. We refer to Tsai and Koziol (1994) for the dataset.

In this example dataset, the 30 baboons were randomly classified into four treatment groups (one normal saline, and three groups based on different doses of fibrin-selective nature of tissue plasminogen activator (t-PA)). For each of these 30 baboons, neurologic functions (N) based on a weighted quantitative neurologic scale and the volumes of intra-cerebral hemorrhage (V) were recorded after middle cerebral artery occlusion (see Table 3.12).

Saline		**t-PA (0.3mg/kg)**		**t-PA (1.5mg/kg)**		**t-PA (10mg/kg)**	
N	V	N	V	N	V	N	V
31	0	45	0	49	2.94	53	5.34
34	0	35	1.74	51	3.54	52	5.7
39	0	44	1.74	46	4.75	53	6.29
40	0	48	3.47	50	5.44	54	6.99
43	0.54	39	5.09	45	5.83	50	10.65
42	0.86	41	6.41	39	7.92	50	11.64
36	2.03						
41	2.73						
42	3.47						
40	3.54						
44	3.77						
38	4.37						

TABLE 3.12: Neurological function (N) and volumes of intracerebral hemorrhage (V) in baboons after middle cerebral artery occlusion.

It is of interest to test the hypothesis that volumes of intra-cerebral hemorrhage increase with increasing dose of t-PA, and hence neurological function should improve (that is, neurologic scores should increase). The calculated value of the test statistic by Tsai and Koziol (1994) = 8.419. The p-value determined from the asymptotic chi-squared distribution with two degrees of freedom is found to be 0.015. This

indicates that the null hypothesis is rejected at 0.05 level. When we employ the newly developed test statistic, we get $M_v = 3.703333$ and p-value = 0.0001063925. This result asserts that the null hypothesis of no treatment differences is rejected. We can, therefore, conclude that both neurological scores as well as volumes of intra-cerebral hemorrhage increase with increasing dose of t-PA.

3.5 Conclusion

In psychological, sociological as well as medical research, two or more treatment groups are compared with respect to several response variables. Such analyses can be performed by using a parametric or nonparametric methods. Parametric methods such as the Hotelling's T^2 require stringent assumptions of normality as well as homogeneity of variance-covariance matrices across groups. In addition, tests of such type which are of a quadratic nature lack power when the alternative hypothesis of interest is directional, that is, an improvement or worsening is demonstrated consistently across treatment groups among the different response variables. In such cases, the most popular procedure is the nonparametric method proposed by O'Brien (1984). The rank-sum test by O'Brien is developed assuming that the underlying distribution as continuous. This multivariate directional test can only be applied if treatment groups are two. As a result, we have proposed and evaluated a new testing procedure that can be applied when treatment groups are two or more in multivariate problems, and when correlations among response variables may be strong, negative, or unequal. The performance of the newly developed test in terms of keeping type-I error rate and power was investigated by generating multivariate data from continuous as well as ordered categorical random variables. In general, it was found that the test reasonable keeps the type-I error rate when data are generated from both continuous as well as discrete distributions when the sample size is ten or more. In addition, the test was found to have good power in each of the simulation situations when sample size of ten or more is considered. Closely looking at the simulation results, the test was more powerful when data under consideration are

ordered categorical. This indeed goes in line with the performance expected theoretically from rank-based inferences. However, the newly proposed test was found to be conservative with low power for extremely small sample sizes ($n < 10$).

We generally recommend the newly proposed test if the primary objective is to detect a directed overall superiority or inferiority of responses for samples of size ten or more and when data under consideration are non-normal in particular ordered categorical and/or the assumption of constant variance is not fulfilled. The proposed test is suggested as an alternative to O'Brien's test when treatment groups are two and when the correlation among responses is not strong negative. The method proposed here is not designed for the situation where different alternative patterns are to be tested for different endpoints at the same time. This would require different techniques and is therefore deferred to the next chapter of this dissertation.

CHAPTER 4

Nonparametric directional testing for multivariate problems based on a closed testing principle

Summary

Comparing several treatments (groups) which involve multiple endpoints is often encountered in biomedical research. When treatment effects are expected to be unidirectional, that is, either increasing or decreasing with respect to the different endpoints, global tests such as the one proposed in this thesis can be used. If on the other hand, treatment effects on the different endpoints are of opposing directions, the power of such equi-directional tests is compromised. Thus, we interchanged the signs of different endpoints by multiplying the values with -1 and made the anticipated direction of treatment effects similar. Following this, we employed the newly proposed test statistic which handles unidirectional alternatives since the direction of treatment effects is made uniform through interchanging the signs. Once

monotonic trend, that is, monotonic increasing for some of the endpoints and monotonic decreasing for others is demonstrated through the global test statistic, an investigator may be further interested in which specific endpoints or sets of endpoints actually these trends are observed. To address this issue, we applied a closed testing principle. The whole procedure is illustrated by reanalyzing a dose-response dataset from a toxicology study carried out by the National Toxicology Program.

4.1 Introduction

Studies focused to test efficacy of a treatment over another or others often involve more than one outcome variable as a single outcome is usually insufficient to describe all the effects. More specifically, clinical trials do not usually select a unique primary response and leave the other(s) as secondary. For example, in testing a heart disease medicine called Repatha, more than one response were considered as primary outcomes [1]. Qizhai et al. (2009) gave an example of clinical trial in which a case control study on the growth and maturation in children with autism or autistic spectrum disorder was conducted. The aim of the analysis was to test whether the levels of a growth-related hormone differ between cases and controls with respect to five response variables, namely insulin-like growth factor-1, insulin-like growth factor 2, insulin-like growth factor binding protein, growth hormone binding protein, and dehydroepiandrosterone. In studies of such type, the multivariate parametric Hotelling's T^2 (Hotelling (1951)) procedure is usually employed to test for the global hypothesis of no difference. However, this standard approach leads to erroneous results if the assumptions of normality and/or constant variance are not satisfied. Furthermore, in a p-dimensional responses for drug-placebo experiment, if it is a priori known that the drug will not be considered effective unless it is significantly superior in at least one of the responses and not worse in any, then a hypothesis of this type cannot be tested by the Hotelling's T^2 as the test is not sensitive for directional alternatives and it lacks power.

[1] https://www.repatha.com/what-is-repatha/ accessed on November 19, 2020

A procedure to test multivariate directional alternatives for which a specified treatment is more effective than another in at least one response was proposed by O'Brien (1984). Alternative tests to handle equi-directional multivariate alternative hypotheses are also proposed by Tsai and Koziol (1994), Dietz (1989), Bloch et al. (2001), Hyun and Song (2009), among others. We also proposed a rank-based test to handle equi-directional alternative hypotheses which is in par with, or sometimes even better than that of the most popular test by O'Brien (see Chapter 3 for more). Each of these multivariate directional tests can only demonstrate a global directional difference while they are restricted to show which specific endpoints or sets of endpoints actually caused this difference. To address this issue, a number of authors applied multiple testing procedures. In a set of n hypotheses $H_1, H_2, \ldots, H_n$ with associated test statistics $T_1, T_2, \ldots, T_n$, a commonly used method is the Bonferroni (B) procedure, which consists of rejecting H_i, for any $i = 1, \ldots, n$, given that the associated test statistic T_i is significant at the level $\alpha^* = \alpha/n$. However, this procedure leads to a very conservative decision, in particular, if the test statistics are highly correlated (cf. Hommel (1989)). A procedure called closed testing which is relatively more powerful than those in common use is proposed by Marcus et al. (1976). In a closed testing procedure, an individual hypothesis is considered significant if all multiple-outcome hypotheses that include this particular hypothesis, beginning with the overall global test, are also significant at level α .

For two-sample comparisons with multiple endpoints, Lehmacher et al. (1991) employed a closed testing procedure in order to identify significant individual or sets of endpoints. Tang and Geller (1999) applied a closed testing procedure in a group sequential setting with the aim of comparing several treatments with a control and detect significant endpoint(s). Also Treondle and Legler (1998) performed a comparison of global tests in conjunction with a closed testing algorithm versus the usual stepwise methods to identify significant individual outcomes in two multivariate samples.

In cases when it is a priori anticipated that treatment effects for some of the endpoints are opposite to that of the others, we cannot directly apply the equi-directional global tests discussed above. Here, we suggest to interchange the sign of those specific endpoints and test for unidirectional case with respect to all endpoints.

The central tenets in the current Chapter are to (1) utilize the equi-directional test statistic proposed in this thesis (see the expression in (3.23)) when treatment effects are a priori anticipated to be positive (measuring good) for some of the responses while they are negative (measuring bad) for the remaining, and (2) apply the principle of closed testing and identify significant individual or sets of responses. We construct sets of hypotheses that are closed under intersection and sequentially test all lower-dimensional marginal hypotheses as long as the global hypothesis is rejected. The procedure is illustrated with a dose-response dataset from a subchronic toxicology study.

4.2 Motivating data

In section 5 of the article on multivariate generalizations of Jonckheere's test for ordered alternatives, Dietz (1989) described a dose-response trial to evaluate the effect of vinylidene fluoride on liver damage. It is often the case that increased levels of serum enzyme are associated with liver damage. The focus of the analysis was to test whether serum enzyme levels are affected by vinylidene fluoride, a chemical suspected of causing liver damage. Three types of enzymes, namely sorbitol dehydrogenase (SDH), serum glutamic oxaloacetic transaminase (SGOT), and serum glutamic pyruvic transaminase (SGPT) were inspected in addition to other response variables. To study whether these enzyme levels increase with increasing dose of vinylidene fluoride, four dosages (0, 1500, 5000, 15000) of this substance in parts/million were examined. In each of these four dosages (treatment groups), ten male Fischer 344 rats were exposed to inhale vinylidene fluoride (see the dataset in Table 4.1). As the aim of the analysis is to test whether increased levels of the three enzymes lead to liver damage, a global test (same direction) on treatment effects with respect to these enzymes was conducted by Dietz (1989).

		Rats within dosage									
Dosage	Enzvme	1	2	3	4	5	6	7	8	9	10
	SDH	18	27	16	21	26	22	17	27	26	27
0	SGOT	101	103	90	98	101	92	123	105	92	88
	SGPT	65	67	52	58	64	60	66	63	68	56
	SDH	25	21	24	19	21	22	20	25	24	27
1500	SGOT	113	99	102	144	109	135	100	95	89	98
	SGPT	65	63	70	73	67	66	58	53	58	65
	SDH	22	21	22	30	25	21	29	22	24	21
5000	SGOT	88	95	104	92	103	96	100	122	102	107
	SGPT	54	56	71	59	61	57	61	59	63	61
	SDH	31	26	28	24	33	23	27	24	28	29
15000	SGOT	104	123	105	98	167	111	130	93	99	99
	SGPT	57	61	54	56	45	49	57	51	51	48

TABLE 4.1: Serum enzyme levels for forty male rats based on different dosages of vinylidene fluoride in parts/million.

When the sum statistic by Dietz is employed to test equi-directional alternative for the three enzyme levels, the calculated value equals 0.25 so that the null hypothesis is retained, that is, there is no evidence for the presence of trend. Application of a univariate test by Jonckheere (1954) to test treatment effect with respect to the three enzyme levels individually yields the following results.

Enzyme	Test Statistic	P-value
SDH	2.45	0.0071
SGOT	1.44	0.0749
SGPT	-3.43	0.0003

The above results reveal that SDH increases with increasing dose while SGPT decreases with increasing dose. Furthermore, there is no evidence of treatment effect with respect to SGOT ($p-$value = 0.0749) at the 5% level of significance. On the

other hand, if the bivariate sum statistic by Dietz is employed for SDH and SGOT, there is evidence of increasing trend ($p-$ value = 0.033).

In the present chapter of this thesis, we reanalyze this dataset by interchanging the sign of SGPT and employing the newly developed test statistic (M_v) together with closed testing principle.

4.3 Related literature

A closed testing procedure was applied to perform multiple group sequential tests and analyze data from clinical trials with multiple endpoints in two samples when the alternative hypothesis of interest is directional as in Tang and Geller (1999). The procedure was also applied to the problem of comparing several treatments with a control based on a single endpoint. $\mathbf{X}_{ij}$, a vector of length k with mean vector $\boldsymbol{\mu}_i$ and unknown variance-covariance matrix Σ was considered from two independent samples in such a way that the global hypothesis of no treatment differences, H_0 was formulated by combining the k single endpoint hypotheses, $H_{0,i} : \mu_{1i} = \mu_{2i}$, $i = 1,\ldots,k$. The global hypothesis was tested against the alternative hypothesis that $H_{1,i} : \mu_{2i} \geq \mu_{1i}$ for all $i = 1,\ldots,k$ with at least one strict inequality.

Two group sequential procedures were proposed: the first one allows stopping the trail as long as a hypothesis $H_{0,K}$ ($K = 1,\ldots,k$ is the set of indices for the k endpoints) is rejected when an interim analysis is performed and then a closed testing procedure is conducted to test all hypotheses implied by it. If the trial continues to the last analysis without rejection of $H_{0,K}$, then none of the hypotheses are rejected. In the second procedure, however, the trial continues to the next stage whenever any hypothesis is not rejected. The procedure is reiterated until all hypotheses are rejected or the last stage is reached.

Lachin et al. (2019) described applications of the closed testing principle using surrogate hypotheses with restricted alternatives. First, the application is demonstrated through analyzing multiple event time outcomes using a Wei-Lachin one-directional multivariate test. Second, longitudinal repeated measures at K time

points are considered in which it is desired to conduct a test of the difference between group means at each of the K points in time. In this case, the tests of intersection hypotheses are conducted using tests of the longitudinal LSMEANS, and a closed testing principle is applied sequentially. The test continues to check for the difference between groups in the means at the j^{th} time provided that all of the intersection hypotheses of LSMEANS containing the j^{th} mean difference are significant at level α. The application was also described to test treatment difference between two groups within multiple subgroups following a test of treatment by subgroup interaction.

A randomized study with two treatment groups having n_1 and n_2 observations from K correlated response variables were considered to test a global null hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ by Lehmacher et al. (1991). They discussed different procedures to test this global null hypothesis. Once this global null hypothesis is rejected, a closed test procedure which involves all lower-dimensional marginal hypotheses was employed step by step. A comparison of different test procedures in conjunction with closed testing was also investigated through a simulation study. They concluded that the application of a closed test procedure for comparing two samples with K endpoints is a reasonable complement to the multivariate method when an investigator is further interested to know which specific endpoints or sets of endpoints actually caused the global difference.

Closed testing in pharmaceutical research has been discussed by Henning and Westfall (2015). A class of closed tests that use additive-combination-based and minimum-based p-value statistics are technically examined. They showed via simulation that the minimum p-value tests for composite hypotheses are generally superior to additive combination tests for multiple comparisons using closure-based testing. The lack of power of additive combination tests is high when the numbers of tests is large. This is not an issue if additive combination tests are used in group sequential trials in which the number of tests is usually small and the main interest is usually in the overall test rather than group-specific tests. It is also concluded that minimum p-value tests can handle dependence structures whereas the additive component tests can only be performed if the correlations are incorporated into the critical values.

Treondle and Legler (1998) compared stepwise testing methods using univariate test statistics to that of a closed testing algorithm with global test statistics. The comparison was made to identify individual significant outcomes from two samples that involve multiple outcome variables. Data were simulated from a multivariate normal distribution with equi-correlated outcomes. Overall, it was found that the power of closed testing methods to reject the global hypothesis is higher than that of the stepwise methods to reject individual hypotheses. When correlation between outcomes is high, the closed testing method with permutation-based approximate likelihood ratio test has higher power given that the number of false null hypotheses is few but the permutational step-up test has higher power when the number of false null hypotheses are many. The reverse happened to be true when the outcomes are uncorrelated. In terms of computer CPU times, closed testing methods are generally preferred to stepwise testing methods though the closed testing methods may involve testing a large number of hypotheses.

A closed multiple testing procedure for hierarchical families of hypotheses to test equality of population means was proposed by Rom and Holland (1995). They considered a finite family of hypotheses of $H_i, i = 1, \ldots, m$. The closure of the family is formed by taking all nonempty intersections $H_K = \bigcap_{i \in K} H_i$ for $K \subseteq A_m$, $A_m = \{1, \ldots, m\}$. For an α level H_K, a closed testing procedure rejects any hypothesis H_K if and only if every H_Q is rejected by its associated α-level test for all $Q \supseteq K$. Their procedure is recursive in such a way that it is applied to successively smaller closures until a decision can be made on H_K. Furthermore, no restveriction is imposed on the order (based on p-values) in which hypotheses are tested given that the coherence requirement is satisfied.

In the context of testing multiple related clinical endpoints for common effect direction, and assuming multivariate normal test statistics with a unit-variance compound symmetric covariance matrix, Bittman et al. (2009) demonstrate that a maximin test for the intersection null $H_i : \theta_i = 0$ for all $i = 1, \ldots, m$ endpoints can be uniformly improved upon if one is interested in making inferences about the individual H_i using closure. Their procedure is based on the consonant property of closed testing principle. That is, a multiple testing procedure is consonant if the rejection of an intersection hypothesis implies the rejection of at least one of its component

hypotheses.

4.4 Method

In this section, we describe the directional test statistic for comparison of multiple endpoints among several independent samples in conjunction with closed testing principle. The newly proposed test statistic to handle multivariate problems when treatment differences of outcomes are in the same direction is revisited in cases when treatment difference in one or more of the outcomes are in different directions.

4.4.1 Notations and hypotheses

Consider a comparison of k samples with respect to two or more response variables potentially measured in different scales. Suppose that $\mathbf{X}_{ij} = \left(X_{ij}^{(1)}, \ldots, X_{ij}^{(p)}\right)^T$ is a p-dimensional outcome for the j^{th} subject in sample i. Let $\mathbf{X}_{ij}$ be independent random vectors with possibly dependent components $X_{ij}^{(d)}$, $d = 1, \ldots, p$, and the marginal distribution function of $X_{ij}^{(d)}$ be $F_i^{(d)}(x) = Pr(X_{ij}^{(d)} < x) + \frac{1}{2}Pr(X_{ij}^{(d)} = x)$ for $i = 1, \ldots, k, j = 1, \ldots, n_i, d = 1, \ldots, p$.

The null hypothesis is formulated as

$$H_0^F : F_1^{(d)}(x) = \ldots = F_k^{(d)}(x) \quad \text{for all x and d = 1, . . . , p}$$

The alternative hypothesis is given by

$$H_1^F : F_1^{(d)}(x) \leq \ldots \leq F_k^{(d)}(x) \quad \text{with at least one strict inequality for at least one d.}$$

If one is interested to test the other direction for each of the response variables, then the above alternative hypothesis can be reformulated by reversing the direction of inequalities as $F_1^{(d)}(x) \geq \ldots \geq F_k^{(d)}(x)$.

We may be interested in the alternative hypotheses of the form: treatment effects are positive (measuring good) with respect to some of the responses while they are negative (measuring bad) for the other(s). Suppose for example that there are two treatment groups (T1 and T2) to be compared with respect to p endpoints in which we would like to test T2 is better than T1 with respect to the entire endpoints except one, say the d^{th} endpoint. In order to use the global test statistic (M_v) for testing T2 is superior than T1 with respect to the entire p endpoints, we interchange the sign of the d^{th} endpoint, i.e., we multiply the values of the d^{th} response by -1. Note that this procedure can only work under the assumption that the effect (direction) of the treatments on each response is suggested a priori, i.e., either measuring good or bad.

4.4.2 Directional test statistics

We consider rank-based test statistics for univariate as well as multivariate datasets together with closed testing principle in the current chapter. In the univariate case, a powerful statistic (see (3.15)) proposed by Bathke (2009) is employed to handle ordered categorical and quantitative variables in a unified form. The newly proposed test statistic in this thesis for equi-directional multivariate problems is also utilized under the assumption that the anticipated treatment effects on different response variables are known a priori. In applying these test statistics, ranking of observations is performed across the k treatment groups and subjects for each endpoint. Let $N = \sum_{i=1}^{k} n_i$ be the total number of subjects in the k samples. Combine the k samples to have the N observations for the d^{th}, $d = 1, \ldots, p$ response variable $X_{11}^{(d)}, \ldots, X_{1n_1}^{(d)}, \ldots, X_{k1}^{(d)}, \ldots, X_{kn_k}^{(d)}$ and obtain the corresponding mid-ranks $R_{11}^{(d)}, \ldots, R_{1n_1}^{(d)}, \ldots, R_{k1}^{(d)}, \ldots, R_{kn_k}^{(d)}$.

The test statistic M_v, which can be applied to test k-sample multivariate problems in which treatment effects are in the same direction is given by (see Chapter 3 Section 3.3 for more):

$$M_v = \frac{\mathbf{1}'\widehat{\mathbf{v}}}{(\mathbf{1}'\widehat{\mathbf{\Sigma}}_{\widehat{v}}\mathbf{1})^{1/2}}, \tag{4.1}$$

where $\mathbf{1}$, $\widehat{\mathbf{v}}$, and $\widehat{\mathbf{\Sigma}}_{\widehat{\mathbf{v}}}$ are as defined in Chapter 3 Section 3.3.

Under the Assumptions 1 - 4 (see Chapter 3 Section 3.3 for more), M_v is asymptotically distributed as standard normal.

4.4.3 Multiple comparisons

A main criticism of global test statistics such as Mv to the multiple endpoints problem is that the rejection of the global hypothesis may not specifically reveal on which endpoint (sets of endpoints) treatment effects are significant. One approach to address this issue is to assess the contribution of each response variable separately by a suitable univariate procedure. This approach, however, does not account for multiplicity. As a result, other procedures which are known as multiple comparisons are designed to take into account of this issue in order to properly control for such multiplicity effect. Multiple comparisons can be performed either in a stepwise or simultaneous manner. In a stepwise procedure, multiple hypotheses are tested sequentially for which critical values are determined separately. In a simultaneous procedure, however, a single critical value is considered for each of the multiple hypotheses.

Multiple decision making procedures

It is evident that every α level test for a single hypothesis controls the type I error rate in a sense that probability of rejecting a true null hypothesis is less or equal α. In multiple hypotheses testing, however, the probability of falsely rejecting one hypothesis assuming that all null hypotheses are simultaneously true exceeds the nominal level α. In order to control such false positive decisions multiple testing procedures can be applied.

When the global hypothesis which includes p response variables is rejected, the individual contribution of each response variable can be tested in a step-down manner. The step-down procedure steps down through the hierarchy of hypotheses implied by the global hypotheses. At any step of testing, if a hypothesis is not rejected,

then all of its implied hypotheses are retained without further tests. Put in other words, a hypothesis is going to be tested if and only if all of its implying hypotheses are rejected.

Suppose that $H_0^{(1)}, \ldots, H_0^{(p)}$ are p hypotheses to be tested and $P_1, \ldots, P_p$ are P-values corresponding to each hypothesis. The Bonferroni procedure is performed by rejecting $H_0 = \{H_0^{(1)}, \ldots, H_0^{(p)}\}$ if any P-value is less than α/p. Specifically, a hypothesis $H_0^{(i)}$ is rejected if $P_i \leq \alpha/p$. If the Bonferroni inequality is applied, one can have a simple α level test, i.e.,

$$Pr\left[\cup_{i=1}^{p}(P_i \leq \alpha/p)\right] \leq \alpha \quad (0 \leq \alpha \leq 1).$$

A disadvantage of the Bonferroni procedure is that it may lead to a conservative decision. The conservative nature of this procedure is more pronounced if the correlation between test statistics is high and/or the number of hypothesis to be tested is increasing.

Holm (1979) proposed a sequentially rejective procedure which has a greater power than the Bonferroni procedure. In Holm's procedure, the P-values are sorted in ascending order $P^{(1)} < P^{(2)} < \ldots < P^{(p)}$ so that $P^{(1)}$ is the smallest and $P^{(p)}$ is the largest P-value. Each P-value is compared to $\alpha/(p-i+1)$, $i = 1, \ldots, p$. That is, the smallest P-value is compared to α/p, the next smallest P-value is compared to $\alpha/(p-1)$, and so on. The denominator is adjusted in each of the steps as long as the null hypothesis under consideration is rejected. The sequential testing stops when a nonsignificant result is obtained, and all untested hypotheses are considered nonsignificant, i.e., a hypothesis $H_0^{(i)}$ is rejected if $(p-i+1)P^{(i)} \leq \alpha$ provided that $(p-j+l)P^{(j)} \leq \alpha$ for all $j < i$. This procedure controls the experiment-wise error rate α. A disadvantage of Holm's sequential procedure is that the test decisions are not necessarily consonant.

Closed testing principle

The closed testing principle is a more general multiple comparison procedure which efficiently controls multiple α-level. Introduced by Marcus et al. (1976), the closed

testing principle leads to stepwise test procedures which allows us to draw individual conclusions about the elementary hypotheses (H_i). Consider any two null hypotheses, H_1 and H_2 where we want to assess whether each of the two treatments, say T1 and T2 are better than a control. The two elementary hypotheses can be formulated as $H_1 : F_1 = F_0$, $H_2 : F_2 = F_0$, where $F_1, F_2,$ and F_0 are the cumulative distribution functions for treatment 1, treatment 2, and the control. The intersection of these hypotheses can be given by $H_{12} = H_1 \cap H_2$. The test strategy here is to first test the intersection hypothesis H_{12} with an appropriate test, and if this is significant, continue testing H_1 and H_2, each at level α. The null hypothesis H_1 is rejected if and only if both H_1 and H_{12} are rejected, each at level α. In turn, H_2 is rejected if both H_2 and H_{12} are rejected. If on the other hand, H_{12} is not rejected, then testing H_1 and H_2 will stop.

In general, suppose there are p hypotheses $H_0^{(1)}, \ldots, H_0^{(p)}$ to be tested and the overall type-I error rate is α. In the closed testing principle, any one of these elementary hypotheses, say $H_0^{(d)}$ is rejected if all possible intersection hypotheses containing $H_0^{(d)}$ are rejected by using a local α level tests. More precisely, the closed testing principle encompasses the following four steps (for more see the book by Bretz et al. (2010), pp. 23 - 29).

1. Consider a set of p elementary hypotheses $H_0 = \left\{H_0^{(1)}, \ldots, H_0^{(p)}\right\}$.
2. Construct a closed set of hypotheses $\overline{H}_0 = \left\{H_0^I = \cap_{i \in I} H_0^{(i)} : I \subset \{1, \ldots, p\}\right\}$.
3. Apply a local level-α test for each $H_0^I \in \overline{H}_0$.
4. Reject $H_0^{(i)}$ if all null hypotheses $H_0^I \in \overline{H}_0$ with $i \in I$ are rejected at the local level-α test.

As indicated above, an elementary null hypothesis H_i will be rejected, if all $H_I \in \bar{\mathcal{H}}$ with $i \in I$ are rejected. In this way, the maximum p-value needs to be less than or equal to α. Therefore, for the null hypothesis H_i, adjusted p-values, say q_i can be computed from

$$q_i = \max_{I : i \in I} p_I, \tag{4.2}$$

where p_I denote the p-value for an intersection hypothesis $H_I, I \subseteq \{1,\ldots,m\}$, and $i = 1,\ldots,m$.

The drawback of a closed testing procedure is that it requires substantial computation as the number of intersection hypotheses to be tested in general is of order 2^m, where m is the number of hypotheses of interest. There are in fact shortcut procedures that can reduce the number of hypotheses to be tested to the order of m as far as the objective is to reach decisions for the elementary hypotheses (for more see Brannath and Bretz (2010)).

Let us clarify the closure principle with m = 4 hypotheses in the family, that is, $\mathcal{H} = \{H_1, H_2, H_3, H_4\}$. The full closure set in this case includes fifteen intersection hypotheses:

$$\bar{\mathcal{H}} = \{H_1, H_2, H_3, H_4, H_{1,2}, H_{1,3}, H_{1,4}, H_{2,3}, H_{2,4}, H_{3,4}, H_{1,2,3}, H_{1,2,4}, H_{1,3,4}, H_{2,3,4}, H_{1,2,3,4}\}.$$

H_1 will be rejected if $H_{1,2,3,4}$, $H_{1,2,3}$, $H_{1,2,4}$, $H_{1,3,4}$, $H_{1,2}$, $H_{1,3}$, $H_{1,4}$ and H_1 are all rejected by their α-level tests (see Figure 4.1). Similarly, H_2 will be rejected if $H_{1,2,3,4}$, $H_{1,2,3}$, $H_{1,2,4}$, $H_{2,3,4}$, $H_{1,2}$, $H_{2,3}$, $H_{2,4}$ and H_2 are all rejected, and H_3 will be rejected if $H_{1,2,3,4}$, $H_{1,2,3}$, $H_{1,3,4}$, $H_{2,3,4}$, $H_{1,3}$, $H_{2,3}$, $H_{3,4}$ and H_3 are all rejected. Finally, H_4 will be rejected if $H_{1,2,3,4}$, $H_{1,2,4}$, $H_{1,3,4}$, $H_{2,3,4}$, $H_{1,4}$, $H_{2,4}$, $H_{3,4}$ and H_4 are all rejected.

Theorem 3 (Marcus et al. (1976)) The closed testing procedure controls the family-wise error rate (FWE) in a strong sense.

Proof.
Consider a collection of true null hypotheses $\{H_i, i \in T\}$ in a set $\mathcal{H} = \{H_1,\ldots,H_m\}$, where rejecting any of these hypotheses is a type I error (α). Let $H_T = \bigcap_{i\in T} H_i$ denotes the intersection of all true elemental hypotheses to be tested. A long as the statistic employed to test H_T keeps the type-1 error rate at a pre-specified level α, Pr(rejecting H_T) $\leq \alpha$. Let H_{T^*} be a true null hypothesis in $\{H_i, i \in T\}$, where $H_{T^*} \neq H_T$. In a closed testing principle described above, a hypothesis H_T is going to be rejected at level of significance α if and only if all hypotheses implying H_T are

also rejected at level α. In particular, the hypothesis H_T should be rejected. Thus,

$$\text{FWE} = \Pr(\text{rejecting } \text{H}_{\text{T}^*}) = \Pr(\text{rejecting } \text{H}_\text{T} \cap \text{rejecting } \text{H}_{\text{T}^*})$$

$$= Pr(\text{rejecting } \text{H}_\text{T})\ \Pr(\text{rejecting } \text{H}_{\text{T}^*} | \text{rejecting } \text{H}_\text{T}) \leq \alpha$$

To demonstrate this, consider a set $m_0 \subseteq m = \{1,2,3,4\}$ of true null hypotheses. Suppose that $H_{1,2,3}$ is true. Then, $m_0 = \{1,2,3\}$ and a Type-1 error occurs if H_1 or H_2 or H_3 is rejected. The following relationships follow when a closure principle is applied.

$$\begin{aligned}\{H_1 \text{ rejected using closure}\} = {} & \{H_{1,2,3,4} \text{ rejected}\} \cap \{H_{1,2,3} \text{ rejected}\} \cap \{H_{1,2,4} \text{ rejected}\} \cap \\ & \{H_{1,3,4} \text{ rejected}\} \cap \{H_{1,2} \text{ rejected}\} \cap \{H_{1,3} \text{ rejected}\} \cap \\ & \{H_{1,4} \text{ rejected}\} \cap \{H_1 \text{ rejected}\} \subseteq \{H_{1,2,3} \text{ rejected}\}\end{aligned}$$

$$\begin{aligned}\{H_2 \text{ rejected using closure}\} = {} & \{H_{1,2,3,4} \text{ rejected}\} \cap \{H_{1,2,3} \text{ rejected}\} \cap \{H_{1,2,4} \text{ rejected}\} \cap \\ & \{H_{2,3,4} \text{ rejected}\} \cap \{H_{1,2} \text{ rejected}\} \cap \{H_{2,3} \text{ rejected}\} \cap \\ & \{H_{2,4} \text{ rejected}\} \cap \{H_2 \text{ rejected}\} \subseteq \{H_{1,2,3} \text{ rejected}\}\end{aligned}$$

$$\begin{aligned}\{H_3 \text{ rejected using closure}\} = {} & \{H_{1,2,3,4} \text{ rejected}\} \cap \{H_{1,2,3} \text{ rejected}\} \cap \{H_{1,3,4} \text{ rejected}\} \cap \\ & \{H_{2,3,4} \text{ rejected}\} \cap \{H_{1,3} \text{ rejected}\} \cap \{H_{2,3} \text{ rejected}\} \cap \\ & \{H_{3,4} \text{ rejected}\} \cap \{H_3 \text{ rejected}\} \subseteq \{H_{1,2,3} \text{ rejected}\}\end{aligned}$$

Note that each of $\{H_1 \text{ rejected using closure}\}$, $\{H_2 \text{ rejected using closure}\}$, and $\{H_3 \text{ rejected using closure}\}$ is a subset of $\{H_{1,2,3} \text{ rejected}\}$, that is, testing each of H_1, H_2, and H_3 is conditional on the rejection of $H_{1,2,3}$. As a result,

$$\begin{aligned}\big\{\{H_1 \text{ rejected using closure}\} \cup \{H_2 \text{ rejected using closure}\} \cup \{H_3 \text{ rejected using closure}\}\big\} \\ \subseteq \{H_{1,2,3} \text{ rejected}\}.\end{aligned}$$

$$\begin{aligned}\implies Pr(\{H_1 \text{ rejected using closure}\} \cup \{H_2 \text{ rejected using closure}\} \cup \\ \{H_3 \text{ rejected using closure}\}) \\ \leq Pr(\{H_{1,2,3} \text{ rejected}\}) = \alpha, \text{ as it is well known}\end{aligned}$$

that the type-1 error or $Pr(\text{rejecting}\{H_{1,2,3}\}|\{H_{1,2,3}\} \text{ is true}) = \alpha$.

Note that the null hypothesis formulated in Section 4.4.1 contains all marginal hypotheses of the p response variables. Subsets of hypotheses in H_0^F which include $q \subset p$ marginal hypotheses are closed under intersection. Thus, we can perform the closed testing procedure based on p hypotheses in the following sequential manner.

- Test H_0 which includes all the p-marginal hypotheses by using M_v. If the test based on the p endpoints is not significant at level α, then the procedure stops. If on the other hand, the test is significant, then the procedure steps down to test all $(p-1)$-dimensional = ${}_pC_{p-1}$ (say n) hypotheses, each at level α.

- Apply the global test statistic (M_v) to test each of the $(p-1)$-dimensional hypotheses and identify those significant hypotheses. In principle, we step down to test (p - 2)-dimensional hypotheses if any of the (p - 1)-dimensional hypotheses is rejected.

- Proceed to test the next all possible subsets of hypotheses $({}_{(p-1)}C_{(p-2)})$ which are implied by those significant $(p-1)$-dimensional hypotheses with the test statistic M_v as long as the number of responses is not less than two.

- Finally, all possible single-dimensional hypotheses implied by significant two-dimensional hypotheses are tested by using the test proposed by Bathke (2009) in order to identify the significant hypotheses. Figure 4.1 shows the hypotheses that must be rejected in order to reject a single outcome hypothesis, say $H_0^{(1)}$: $F_1^{(1)} = \ldots = F_k^{(1)}$.

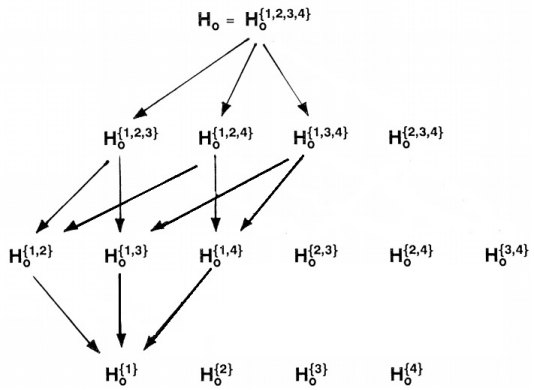

FIGURE 4.1: Sets of hypotheses (connected by arrows) that must be rejected to reject $H_0^{(1)} : F_1^{(1)} = \ldots = F_k^{(1)}$ based on a closed hypothesis testing procedure considering a global hypothesis with four responses {1,2,3,4}.

Note. Adapted from "Procedures for Two-Sample Comparisons with Multiple Endpoints Controlling the Experimentwise Error Rate," by W. Lehmacher, G. Wassmer and P. Reitmeir, 1991, Biometrics, 47(2), pp. 516 (https://www.jstor.org/stable/2532142) CC BY-NC

4.5 Illustration

We consider a dose-response dataset analyzed by Dietz (1989) to illustrate the procedures discussed in the previous sections. The dataset is presented in Table 4.1. The null hypothesis of no treatment difference with respect to the three endpoints versus the alternative that there is a monotone trend with respect to at least one of the endpoints is tested by applying the test statistic M_v. This test failed to show a statistically significant trend (p-value = 0.5619725), and the result goes in line with that of applying the test by Dietz (see Section 4.2 for more). However, the univariate analysis based on Jonckheere (1954) indicates that treatment differences are not in the same direction with respect to each of the three endpoints. Thus, we a priori anticipate that SDH and SGOT levels tend to increase with increasing dose whereas SGPT levels tend to decrease with increasing dose of vinylidene fluoride. Let us

multiply the values of SGPT by -1, and keep SDH and SGOT as they are. We can now employ the unidirectional test statistic M_v since the direction of treatment effects is specified to be the same for the three responses. The statistic M_v yields a p-value = 0.0001117315. This shows a presence of strong trend with respect to at least one of the endpoints. Furthermore, if we apply the closed testing procedure to identify which endpoints (sets of endpoints) are affected by the treatments, then we can have the following results.

Hypotheses to be tested	**P-value**
$H_{\{SDH,SGOT\}}$	0.01190002
$H_{\{SDH,SGPT^*\}}$	$2.385494 * 10^{-05}$
$H_{\{SGOT,SGPT^*\}}$	0.0005085015
$H_{\{SDH\}}$	0.01524419
$H_{\{SGOT\}}$	0.08683297
$H_{\{SGPT^*\}}$	0.001183542

TABLE 4.2: Sequences of intersection hypotheses to be tested uisng closed testing procedure.

Note in the above that we employed the statistic M_v for each of the three hypotheses with bi-dimensional response variables. The hypothesis $H_{\{SDH,SGOT\}}$ refers to testing monotone trend with respect to SDH and SGOT. $H_{\{SDH,SGPT^*\}}$ is tested for monotonic trend on SDH and SGPT* together, where SGPT* is the new variable created after the sign of SGPT is changed. Similarly, we tested the hypothesis $H_{\{SGOT,SGPT^*\}}$ for treatment effects on SGOT and $-1 * SGPT$ together. The test by Bathke (2009) is employed to test for monotonic trends on each of the three responses separately.

Based on the results in the closed testing procedure, we can conclude that SDH and SGOT levels increase with increasing dose of vinylidene fluoride. Although in different directions, SDH and SGPT as well as SGOT and SGPT are jointly affected by the treatments. Specifically, SDH level increases and SGPT level decreases with increasing dose, and SGOT level increases and SGPT level decreases with increasing dose. When the three enzyme levels are taken separately, it is found that SDH level significantly increases with increasing dose, and SGPT level significantly decreases with increasing dose, but SGOT level is not significantly affected by the treatment.

4.6 Conclusion

In this chapter of the thesis, we described an application of the newly proposed test statistic (M_v) together with the closed testing principle. Specific focus was given to test treatment effects with respect to various endpoints that occur in different direction (some positive and others negative), and to identify endpoints (sets of endpoints) affected by two or more treatments. As the test statistic M_v can only handle equi-directional trends with respect to all response variables under consideration, the signs of these responses should be the same. We suggested to interchange the signs of different endpoints and made the anticipated direction of treatment effects uniform on the entire responses. The procedure is demonstrated through re-analyzing a dose-response dataset considered by Dietz (1989). The result of our analysis is found to be in agreement with that of Dietz. Given that treatment effects with respect to different endpoints are specified on a priori basis, that is, either increasing or decreasing and the objective of the test is to identify specific significant endpoints, the application of global test statistic M_v (after the signs of different responses is made uniform) in conjunction with the closed testing algorithm is a good choice.

Although closed testing is a flexible and simple procedure to perform multiple testing, it generally requires to test $\mathcal{O}(2^p)$ hypotheses (p = the number of elementary hypotheses) since there are $2^p - 1$ subset intersection hypotheses. Shortcut methods which reduce the overall complexity can be applied under the assumption that rejection of an intersection hypothesis implies rejection of at least one elementary hypothesis (cf. Hommel et al. (2007)).

Part III

Conclusion and Future Research

CHAPTER 5

Conclusion and Future Research

5.1 Conclusion

This thesis is concerned with fully nonparametric hypothesis testing in k-sample multivariate problems. The popular nonparametric test statistic which is employed to test null hypotheses of no difference among two samples (groups) with respect to two or more response variables versus equi-directional alternatives is developed by O'Brien (1984). This test is proposed with the assumption that the underlying distribution is continuous. Furthermore, the rank sum test statistic by O'Brien lacks power when correlation between response variables is strong positive, weak negative and/or unequal. Motivated by these issues, we proposed a global rank-based test statistic which can handle metric as well as ordered categorical data in a unified manner by generalizing a univariate test statistic proposed by Bathke (2009). Another important improvement which the newly proposed test statistic brings is that the test can be used when the number of multivariate samples exceed two.

The asymptotic distribution of the novel test statistic proposed in this thesis is derived by utilizing asymptotic rank transforms. This is emanated from the fact that asymptotic rank transforms are independent so long as the random variables on

the basis of which they are derived are independent. It is well known that random variables must be independently distributed in order the central limit theorem to be applied. The theoretical derivation of the asymptotic distribution reveals that the test statistic converges in distribution to the standard normal distribution. We employed separate rankings for the p different response variables because (i) practical multivariate data involve a mixed scale of response variables, that is, some are metric and others are ordinal (ii) separate ranking preserves independence when the different response variables are actually independent, and (iii) separate ranking maintains invariance under monotone transformations of different response variables. Note that asymptotic rank transforms are unobservable quantities and hence we used the rank transforms (actual mid-ranks) in the calculation of the test statistic.

The rank-based test developed here is robust to outliers, invariant under any order-preserving transformations of the response variables as well as the weights determining the alternative patters. The test statistic can be employed in more practical data analysis problems, that is, when multivariate samples are unbalanced (number of observations in different samples are not equal). In addition, the test can handle data with ties which comes to happen in most practical situations. Moreover, computation of this test statistic is easy as it does not require any iterative procedure.

The performance of the procedure based on the newly proposed test statistic is evaluated through an extensive simulation study by generating data from continuous as well as ordinal multivariate distributions. In addition, different correlations (weak negative, weak positive, strong negative, strong positive, and unequal) has been considered. The simulation study reveal that the newly proposed test is in a decent control of the nominal type-I error rate with reasonable power when sample size in different groups is more than ten. The test was found to be slightly conservative with low power when sample size in different groups is less than or equal to ten. Applications to several real data sets obtained from clinical trials are analyzed.

Another motivation in this thesis comes from the situation that treatment effects with respect to some of the response variables is a priori anticipated to be positive whereas they remain negative for others. The newly proposed test statistic can only be employed when the alternative hypothesis of interest is unidirectional, that is,

either monotone increasing or monotone decreasing. In some practical cases, however, interest may lie in testing monotonic increasing treatment effect with respect to some of the response variables and monotonic decreasing treatment effect with respect to others. Under the assumption that the pattern to be tested with respect to each response variable is specified on a priori basis, we suggested to interchange the signs of different response variables so as to make the anticipated direction of treatment effects similar. Since the conjectured alternative patterns are made uniform in this way, we can employ an equi-directional test statistic to test monotonic treatment effects. Once the signs of different responses is made uniform, the global null hypothesis of no treatment effect (difference) for all response variables may be rejected by using the newly developed equi-directional test statisti (M_v). Nevertheless, this test cannot tell us in which of the response variable(s) the effect is actually significant. We then turned to closed multiple testing procedures. The closed testing procedure proposed by Marcus et al. (1976) is a general approach which leads to stepwise test procedures in such a way that any H_0 will be rejected at level α if and only if every H_0 containing its subset of responses in the previous steps is rejected. The whole closed testing procedure maintains the experiment-wise or multiple type-1 error rate.

We combined the newly proposed global test statistic with the closed testing principle to test for different directional trends among two or more multivariate samples, and identify on which of the responses such trend is actually significant. The application is demonstrated by reanalyzing a dose-response dataset. The test initially began with the global test involving three response variables and stepped down to test the three intersection hypotheses with two responses. Since these intersection hypotheses were found to be significant, we further performed the test for each of the elemental hypotheses using the univariate statistic proposed by Bathke (2009).

In summary, if it is sensible to a priori specify the alternative patterns to be tested for each response variable (which can be metric or ordered categorical) and the objective of study is to test unidirectional alternatives in two or more sample multivariate problems, the newly developed test (M_v) is a viable alternative. Furthermore, we recommend to use the newly proposed test statistic to test directional alternatives (some positive and others negative) when it is again reasonable to a priori specify

treatment effects on different responses in which the signs of different responses are interchanged and made uniform. If an investigator is further interested to know in which of the responses treatment effects are significant, we recommend to use the newly proposed rank-based test statistic in conjunction with closed testing principle.

5.2 Future research direction

In this section, we briefly introduce some possible extensions of the procedures described in the thesis.

5.2.1 Umbrella alternatives

Testing monotonically ordered alternatives in k-samples involving multiple response variables potentially measured in different scales was the heart of discussion in the present thesis. In some other situations, however, values of response variables may initially increase with the treatment level up to a point and then keeps on decreasing, that is, a monotone alternative is conjectured which is subject to change in direction. Such up-then-down pattern is known as umbrella ordering (alternative). Umbrella alternatives are common in real life problems, for example, to test the effect of age on some physiological parameters that measure the physical efficiency of persons. Two scenarios exist when we consider umbrella alternatives. In the first scenario, the peak of the conjectured umbrella is known while in the second one, which is more practical, it is not necessary to specify the peak of the umbrella prior to data collection.

The alternative hypothesis for umbrella alternatives can be formulated as

$$H_1 : F_1^{(d)}(x) \leq \ldots \leq F_l^{(d)}(x) \geq \ldots \geq F_k^{(p)}(x)$$

for all x and $d = 1, \ldots, p$ with at least one strict inequality for at least one d, and $F(\cdot)$ is the normalized version the distribution function. This alternative is said to have a peak at sample l.

For testing H_0 (see Section 4.4.1) against the above umbrella alternatives, we can separately consider the two situations, that is, monotonically increasing trend until the peak and monotonically decreasing trend after the peak. Define the univariate rank-based test statistic proposed by Bathke (2009) for each of these two situations. For each of the response variables, combine the two statistics corresponding to each case. Following this, it is promising to determine the variance-covariance matrix, and also derive the asymptotic distribution of the coordinate-wise combined test statistic.

In the second scenario, that is, $l \in \{1, \ldots, k\}$, we need to consider a finite collection of possible patterns in $\{1, \ldots, k\}$ and then employ the the peak-known statistic. The result which leads to maximum value is taken as the value of the test statistic (see, for example, Hettmansperger and Norton (1987)).

5.2.2 Missing data

Missing data are a practical problem in multivariate analysis of clinical trials. However, the procedure proposed in this thesis is based on the availability of complete units (subjects). Suppose that measurements are missed in a completely random fashion. Then define indicator functions ($\lambda_{ij}^{(d)}$) corresponding to each response variable as given in Brunner et al. (2001), that is,

$$\lambda_{ij}^{(d)} = \begin{cases} 1 \;\; if \; X_{ij}^{(d)} \text{ is observed} \\ 0 \;\; if \; X_{ij}^{(d)} \text{ is missing,} \end{cases}$$

where $d = 1, \ldots, p$, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$.

We can then redefine the empirical distribution function given in Section 3.3 using indicator and count functions as

$$\widehat{F}_i^{(d)}(x) = \frac{1}{\lambda_{i.}^{(d)}} \sum_{j=1}^{n_i} \lambda_{ij}^{(d)} c(x - X_{ij}^{(d)}),$$

where, $c(\cdot)$ is the normalized count function and $\lambda_{i.}^{(d)} = \sum_{j=1}^{n_i} \lambda_{ij}^{(d)}$ is the total number of available observations for the d^{th} response variable at group i. Note in the above that $X_{ij}^{(d)}$ will be substituted by an arbitrary finite number if it is missing and hence the product $\lambda_{ij}^{(d)} c(x - X_{ij}^{(d)})$ becomes zero.

The asymptotic rank transforms as well as rank transforms for each of the p response variables can be redefined by incorporating the indicator function $\lambda_{ij}^{(d)}$. A global test statistic for k-sample unidirectional multivariate problems adjusted for incomplete data could then be developed based on a coordinate-wise approach. The asymptotic distribution corresponding to the test statistic could be derived in a similar technique used in this thesis. However, the estimation of asymptotic variance-covariance matrix may be complicated.

REFERENCES

Akritas, M. G. (1990). "The Rank Transform Method in Some Two-Factor Designs". In: *Journal of the American Statistical Association,* 85 (409), pp. 73–78.

Akritas, M. G. and S.F. Arnold (1994). "Fully Nonparametric Hypotheses for Factorial Designs I: Multivariate Repeated MeasuresDesigns". In: *Journal of the American Statistical Association,* 89 (425), pp. 336–343.

Akritas, M. G. and E. Brunner (1997). "A Unified Approach to Rank Tests for Mixed Models". In: *Journal of statistical planning and inference,* 61, pp. 249 –277.

Bathke, A. C. (2005). "Testing monotone effects of covariates in nonparametric mixed models". In: *Journal of Nonparametric Statistics,* 17 (4), pp. 423 –439.

Bathke, A. C. (2009). "A unified approach to nonparametric trend tests for dependent and independent samples," in: *Metrika,* 69, pp. 17 –29.

Bathke, A. C. and S. W. Harrar (2008). "Nonparametric methods in Multivariate factorial designs for large number of factor levels". In: *Journal of Statistical Planning and Inference* 138, pp. 588–610.

Bathke, A. C., S. W. Harrar, and L. V. Madden (2008). "How to Compare Small Multivariate Samples Using Nonparametric Tests". In: *Computational Statistics & Data Analysis,* 52 (11), pp. 4951 –4965.

Bittman, R M., J. P. Romano, C. Vallarino, and M. Wolf (2009). "Optimal testing of multiple hypotheses with common effect direction". In: *Biometrika,* 96 (2), pp. 399 –410.

Bloch, D. A., T. L. Lai, and P. Tubert-Bitter (2001). "One-Sided Tests in Clinical Trials with Multiple Endpoints". In: *Annals of Mathematical Statistics,* 57(4), pp. 1039–1047.

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". In: *Annals of Mathematical Statistics,* 25(2), pp. 290 –302.

Brannath, W. and F. Bretz (2010). "Shortcuts for Locally Consonant Closed Test Procedures". In: *Journal of the American Statistical Association,* 105(490), pp. 660 –669.

Bregenzer, T. and W. Lehmacher (1998). "Directional tests for the analysis of clinical trials with multiple endpoints allowing for incomplete data". In: *Biometrical Journal,* 40, pp. 911 –928.

Bretz, F., T. Hothorn, and P. Westfall (2010). *Multiple Comparisons Using R*. Taylor & Francis Group.

Brunner, E., A.C. Bathke, and F. Konietschke (2019). *Rank and pseudo-rank procedures for independent observations in factorial designs using R and SAS*. Springer Nature Switzerland.

Brunner, E., S. Domhof, and F. Langer (2001). *Nonparametric analysis of longitudinal data in factorial experiments*. JOHN WILEY & SONS, INC.

Brunner, E., F. Konietschke, M. Pauly, and M. L. Puri (2016). "Rank-based procedures in factorial designs: hypotheses about nonparametric treatment effects". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Brunner, E., U. Munzel, and M. L. Puri (1999). "Rank-score tests in factorial designs with repeated measures". In: *Journal of Multivariate Analysis,* 70, pp. 286–317.

Callegari, F. and M. G. Akritas (2004). "Rank tests for patterned alternatives in two-way non-parametric analysis of variance". In: *Journal of Statistical Planning and Inference,* 126, pp. 1–23.

Dempster, A. P. (1960). "A significance test for the separation of two highly multivariate small samples". In: *Biometrics,* 16 (1), pp. 41–50.

Dietz, E. J. (1989). "'Multivariate Generalizations of Jonckheere's Test for Ordered Alternatives". In: *Communications in Statistics-Theory and Methods,* 18 (10), pp. 3763–3783.

Dietz, E. J. and T. J. Killeen (1981). "A nonparametric multivariate test for monotone trend with pharmaceutical applications". In: *Journal of the American Statistical Association,* 76 (373), pp. 169 –174.

Follmann, D. (1996). "A simple multivariate test for one-sided alternatives". In: *Journal of the American Statistical Association,* 91 (434), pp. 854 –861.

Gaur, A., K. K. Mahajan, and S. Arora (2012). "New nonparametric tests for testing homogeneity of scale parameters against umbrella alternative". In: *Statistics and Probability Letters,* 82, pp. 1681 –1689.

Gibbons, J. D. and S. Chakraborti (2011). *Nonparametric Statistical Inference*. Fifth. Taylor and Francis Group, LLC.

Happ, M., S.W. Harrar, and A. C. Bathke (2016). *R package HRM*. URL: https://cran.rstudio.com/web/packages/HRM/index.html.

Henning, K. S. S. and P. H. Westfall (2015). "Closed Testing in Pharmaceutical Research: Historical and Recent Developments". In: *Stat Biopharm Res,* 7 (2), 126–147.

Hettmansperger, T. P. and R. M. Norton (1987). "Tests for patterned alternatives in k-sample problems". In: *Journal of the American Statistical Association,* 82 (397), pp. 292–299.

Hollander, M., D. A. Wolfe, and E. Chicken (2014). *Nonparametric Statistical Methods*. Third. John Wiley & Sons, Inc.

Holm, S. (1979). "A Simple Sequentially Rejective Multiple Test Procedure". In: *Scandinavian Journal of Statistics,* 6 (2), pp. 65–70.

Hommel, G. (1989). "A Comparison of Two Modified Bonferroni Procedures". In: *Biometrika,* 76 (3), pp. 624–625.

Hommel, G., F. Bretz, and W. Maurer (2007). "Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies". In: *Statistics in Medicine,* 26, pp. 4063 –4073.

Hotelling, H. (1951). “A Generalized T Test and Measure of Multivariate Dispersion”. In: *Proceedings of the Second Berkeley Symposium on Mathematica Statistics and Probability*.

Hyun, N.R. and H.H. Song (2009). “A nonparametric multivariate test for a monotone trend among k-samples”. In: *The Korean Journal of Applied Statistics,* 22 (5), pp. 1047 –1057.

Jonckheere, A.R. (1954). “A Distribution-Free k-Sample Test Against Ordered Alternatives”. In: *Biometrika,* 41 (1/2), pp. 133–145.

Karrison, T. and P. O’Brien (2004). “A rank-sum-type test for paired data with multiple endpoints”. In: *Journal of Applied Statistics,* 31 (2), pp. 229 –238.

Koziol, J. A., D. A. Maxwell, M. Fukushima, M. E. M. Colmerauer, and Y. H. Pilch (1981). “A distribution-free test for tumor-growth curve analyses with application to an animal tumor immunotherapy”. In: *Biometrics,* 37, pp. 383 –390.

Kruskal, W. H. (1952). “A Nonparametric test for the Several Sample Problem”. In: *The Annals of Mathematical Statistics,* 23 (4), pp. 525–540.

Kruskal, W. H. and W. A. Wallis (1952). “Use of Ranks in One-Criterion Variance Analysis”. In: *Journal of the American Statistical Association,* 47 (260), pp. 583–621.

Lachin, J. M., I. Bebu, M. D. Larsen, and N. Younes (2019). “Closed testing using surrogate hypotheses with restricted alternatives”. In: *PLoS ONE,* 14(7), pp. 1–18.

Lehmacher, W., G. Wassmer, and P. Reitmeir (1991). “Procedures for Two-Sample Comparisons with Multiple Endpoints Controlling the Experimentwise Error Rate”. In: *Biometrics,* 47(2), pp. 511–521.

Liu, C., A.C. Bathke, and S. W. Harrar (2011). “A nonparametric version of Wilks’ lambda —Asymptotic results and small sample approximations”. In: *Statistics and Probability Letters,* 81, pp. 1502–1506.

Mack, G. A. and D. A. Wolfe (1981). “K-sample rank tests for umbrella alternatives”. In: *Journal of the American Statistical Association,* 76 (373), pp. 175–181.

Marcus, R., E. Peritz, and K. R. Gabriel (1976). "On closed testing procedures with special reference to ordered analysis of variance". In: *Biometrika,* 63 (3), pp. 655–660.

Minhajuddin, A.T.M., W.H. Frawley, W. R. Schucany, and W. A. Woodward (2007). "Bootstrap tests for multivariate directional alternatives". In: *Journal of Statistical Planning and Inference,* 137, 2302 – 2315.

Munzel, U. and E. Brunner (2000a). "Nonparametric methods in multivariate factorial designs". In: *Journal of Statistical Planning and Inference,* 88, pp. 117–132.

Munzel, U. and E. Brunner (2000b). "Nonparametric tests in the unbalanced Multivariate one-way design". In: *Biometrical Journal,* 42 (7), pp. 837–854.

Nadar, M. (2013). "Multivariate generalisations of k-sample rank tests for umbrella alternatives". In: *Journal of Nonparametric Statistics,* 25 (1), pp. 91 –107.

O'Brien, P. C. (1984). "Procedures for comparing samples with multiple endpoints". In: *Biometrics,* 40, pp. 1079 –1087.

Pillai, K .C. S. (1955). "Some new test criteria in multivariate analysis". In: *The Annals of Mathematical Statistics,* 26 (1), pp. 117–121.

Puri, M. L. and P. K. Sen (1966). "On a Class of Multivariate Multisample Rank-Order Tests". In: *Sankhya: The Indian Journal of Statistics, Series A (1961-2002),* 28 (4), pp. 353–376.

Qizhai, L., L. Aiyi, Y. Kai, and F.Y. Kai (2009). "A weighted rank-sum procedure for comparing samples with multiple endpoints". In: *Statistics and its inference,* 2, pp. 197 –201.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. URL: https://www.R-project.org.

Reitmeir, P. and G. Wassmer (1996). "One-sided multiple endpoint testing in two-sample comparisons". In: *Communications in Statistics - Simulation and Computation,* 25 (1), pp. 99 –117.

Rom, D. M. and B. Holland (1995). "A new closed multiple testing procedure for hierarchical families of hypotheses". In: *Journal of Statistical Planning and Inference,* 46, pp. 265 –275.

Shan, G., D. Young, and L. Kang (2014). "New powerful nonparametric rank test for ordered alternative problem". In: *PLoS ONE.,* 9 (11).

Silvapulle, M. J. (1995). "A Hotteling's T^2 - type statistic for testing against one-sided hypotheses". In: *Journal of Multivariate Analysis,* 55, pp. 312 –319.

Tang, D. and N. L. Geller (1999). "Closed Testing Procedures for Group Sequential Clinical Trials with Multiple Endpoints". In: *Biometrics,* 55, pp. 1188–1192.

Terpstra, J. T., C.H. Chang, and R. C. Magel (2011). "On the use of Spearman's correlation coefficient for testing ordered alternatives". In: *Journal of Statistical Computation and Simulation,* 81 (11), pp. 1381–1392.

Terpstra, Jeffrey and R. Magel (2010). "A new nonparametric test for the ordered alternative problem". In: *Journal of Nonparametric Statistics,* 15 (3), pp. 289–301.

Terpstra, T.J. (1952). "The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking". In: *Indag Math,* 14(3), 327–333.

Treondle, J. F. and J. M. Legler (1998). "A Comparison of One-Sided Methods to Identify Significant Individual Outcomes in a Multiple Outcome Setting: Stepwise Tests or Global Tests with Closed Testing". In: *Statistics in Medicine,* 17, pp. 1245 –1260.

Tsai, K.T. and J. A. Koziol (1994). "Assessing multiple endpoints with ordered alternatives". In: *Communications in Statistics - Theory and Methods,* 23 (2), pp. 533 –544.

Wilks, S.S. (1946). "Sample Criteria for Testing Equality of Means, Equality of Variances, and Equality ofCovariances in a Normal Multivariate Distribution". In: *The Annals of Mathematical Statistics,* 17 (3), pp. 257–281.

APPENDIX A

R simulation code

```
#########################################################
### Sample R code to simulate Type-I error rates of Mv ##
### basedon data from multivariate (5 responses)       ##
### ordinal variables in 3 samples.                    ##
#########################################################

library(GenOrd)
ni<-c(5,10,15,20,30,50)

sims <- 10000  # Number of simulations to conduct for each ni

g<- 3
p<- 6

# To account equal correlation between responses
R1<-matrix(c(1,0.6,0.6,0.6,0.6,0.6,1,0.6,0.6,0.6,0.6,0.6,
       1,0.6,0.6,0.6,0.6,0.6,1,0.6,0.6,0.6,0.6,0.6,1),5,5)

type_1er.Mv1.4<- rep(NA, length(ni))
```

```
# marginal is the vector of the cumulative probabilities
# defining the marginal distribution of the ith component
# of the multivariate variable.

marginal <- list(0.4, c(0.2,0.9), c(0.3,0.4,0.7),
c(0.1,0.3,0.8,0.9), c(0.1,0.2,0.6,0.7,0.8,0.9))

#### Outer loop to vary the number of subjects

for (j in 1:length(ni)){
N <- ni[j] # Pick the jth value for N

significant.experiments.Mv <- rep(NA, sims)

### Inner loop to conduct experiments "sims" times
### over for each N

for (i in 1:sims){

set.seed(i)
n<-N*g # Total number of observations

#Multivariate (5-variate) ordinal data under H0
random.disc1 <- ordsample(N,marginal,R1)
x12<-rep(1,N) #covariate in the first sample
random.disc2 <- ordsample(N,marginal,R1)
x22<- rep(2,N) #covariate in the second sample
random.disc3 <- ordsample(N,marginal,R1)
x32<- rep(3,N) #covariate in the third sample

Y1<-cbind(random.disc1, x12)
Y1M <-as.data.frame(Y1)
```

```
names(Y1M) <-c("response1","response2","response3",
               "response4","response5", "covariate")
Y2<-cbind(random.disc2, x22)
Y2M <-as.data.frame(Y2)
names(Y2M) <-c("response1","response2","response3",
               "response4","response5", "covariate")
Y3<-cbind(random.disc3, x32)
Y3M <-as.data.frame(Y3)
names(Y3M) <-c("response1","response2","response3",
               "response4","response5", "covariate")


# yM combines the data from THREE samples ##
# for each response & the covariate      ##

yM<-rbind(Y1M,Y2M,Y3M)
rM<-apply(yM,2,rank) # To separately rank the responses

group<-rep(1:g,each=N)

or<-order(group) # To order the groups
group<-group[or]
yM<-as.matrix(yM[or,])
p<-dim(yM)[2] # Number of variables = no. of columns in yM

groupu<-unique(group) #identifies unique groups
g<-length(groupu) #number of groups#
groupind<-sapply(groupu,"==",group) #group indicator#

### calculation of the statistic Mv###

m<- (3*N+1)/2 #expected value of rij or Rij
```

```
Rij1<-rM[,1] # ranks of the first response variable
Rij2<-rM[,2] # ranks of the second response variable
Rij3<-rM[,3] # ranks of the third response variable
Rij4<-rM[,4] # ranks of the fourth response variable
Rij5<-rM[,5] # ranks of the fifth response variable
rij<-rM[,6] # ranks of the covraite

# results of the numerator  for hat(v_i) i = 1, ..., 5
nu1 = sum((Rij1 - m)*(rij-m))
nu2 = sum((Rij2 - m)*(rij-m))
nu3 = sum((Rij3 - m)*(rij-m))
nu4 = sum((Rij4 - m)*(rij-m))
nu5 = sum((Rij5 - m)*(rij-m))

Rij12<-(rM[,1]-m)^2 # square difference of the 1st response
Rij22<-(rM[,2]-m)^2 # square difference of the 2nd response
Rij32<-(rM[,3]-m)^2 # square difference of the 3rd response
Rij42<-(rM[,4]-m)^2 # square difference of the 4th response
Rij52<-(rM[,5]-m)^2 # square difference of the 5th response

rij2<-(rM[,6]-m)^2 # square difference of the covariate

# Rr2 forms a matrix of square differences

Rr2<-cbind(Rij12, Rij22, Rij32,Rij42,Rij52,rij2)

### S finds sums of square rank difference to be
## used in the denominator,i.e., Sum(Rij1-m)^2

S<-t(groupind)%*%Rr2
```

```
S1<-S[,c(1,6)] # Sums of square rank differences for
# 1st response and covariate
S2<-S[,c(2,6)] # Sums of square rank differences for
# 2nd response and covariate
S3<-S[,c(3,6)] # Sums of square rank differences for
# 3rd response and covariate
S4<-S[,c(4,6)] # Sums of square rank differences for
# 4th response and covariate
S5<-S[,c(5,6)] # Sums of square rank differences for
# 5th response and covariate

### Su is to compute product of sums,i.e.,
### sum(Rij1-m)^2*sum(rij-m)^2 for each group ##

Su1<-apply(S1, 1, prod) # For the 1st response & covariate
Su1i<-Su1*(N-1)^-1
Sm1<-sum(Su1i) # To aggregate the sums from the two groups
Su2<-apply(S2, 1, prod) # For the 2nd response & covariate
Su2i<-Su2*(N-1)^-1
Sm2<-sum(Su2i) # To aggregate the sums from the two groups
Su3<-apply(S3, 1, prod) # For the 3rd response & covariate
Su3i<-Su3*(N-1)^-1
Sm3<-sum(Su3i) # To aggregate the sums from the two groups
Su4<-apply(S4, 1, prod) # For the 4th response & covariate
Su4i<-Su4*(N-1)^-1
Sm4<-sum(Su4i) # To aggregate the sums from the two groups
Su5<-apply(S5, 1, prod) # For the 5th response & covariate
Su5i<-Su5*(N-1)^-1
Sm5<-sum(Su5i) # To aggregate the sums from the two groups

# for the denominator
den1<-sqrt(Sm1) # result for the denominator in hat(v_1)
```

```
den2<-sqrt(Sm2) # result for the denominator in hat(v_2)
den3<-sqrt(Sm3) # result for the denominator in hat(v_3)
den4<-sqrt(Sm4) # result for the denominator in hat(v_4)
den5<-sqrt(Sm5) # result for the denominator in hat(v_5)

B1<- nu1/den1 # Calculated value of the test statistic hat(v_1)
B2<- nu2/den2 # Calculated value of the test statistic hat(v_2)
B3<- nu3/den3 # Calculated value of the test statistic hat(v_3)
B4<- nu4/den4 # Calculated value of the test statistic hat(v_4)
B5<- nu5/den5 # Calculated value of the test statistic hat(v_5)

B1.5 = B1+B2+B3+B4+B5 # hat(V_1) +...+ hat(v_5) to be used in
               # numerator of the multivariate test statistic.

nu = B1.5 # value of the numerator for M_v

# To calculate the variance-covariance matrix terms for \hat(V)

Rij1.2<-(rM[,1]-m)*(rM[,2]-m) # Product of corrected
      # rank(response1) and corrected rank(response2)
Rij1.3<-(rM[,1]-m)*(rM[,3]-m) # Product of corrected
      # rank(response1) and corrected rank(response3)
Rij1.4<-(rM[,1]-m)*(rM[,4]-m) # Product of corrected
      # rank(response1) and corrected rank(response4)
Rij1.5<-(rM[,1]-m)*(rM[,5]-m) # Product of corrected
      # rank(response1) and corrected rank(response5)

Rij2.3<-(rM[,2]-m)*(rM[,3]-m) # Product of corrected
     # rank(response2) and corrected rank(response3)
Rij2.4<-(rM[,2]-m)*(rM[,4]-m) # Product of corrected
     # rank(response2) and corrected rank(response4)
Rij2.5<-(rM[,2]-m)*(rM[,5]-m) # Product of corrected
```

```
     # rank(response2) and corrected rank(response5)

Rij3.4<-(rM[,3]-m)*(rM[,4]-m) # Product of corrected
     # rank(response3) and corrected rank(response4)
Rij3.5<-(rM[,3]-m)*(rM[,5]-m) # Product of corrected
     # rank(response3) and corrected rank(response5)

Rij4.5<-(rM[,4]-m)*(rM[,5]-m) # Product of corrected
     # rank(response3) and corrected rank(response4)

Rr3<-cbind(Rij1.2,rij2) # To form a matrix of square corrected
                        # rank(covariate) and the products of
                      # corrected ranks (response1 & response2)
SS<-t(groupind)%*%Rr3

Sp1<-apply(SS, 1, prod) # For the corrected rank responses
                        # and covariate
Sp1i<-Sp1*(N - 1)^-1
Spm1<-sum(Sp1i) # To aggregate the sums from the two groups

nc12<-Spm1 # value of numerator for r_{12}
dc12<-sqrt(Sm1*Sm2) # value of denominator for r_{12}

r12<-nc12/dc12 # covariance between hat(v_1) & hat(v_2)

Rr1.3<-cbind(Rij1.3,rij2) # To form a matrix of square
         # corrected rank(covariate) and the products of
         # corrected ranks (resp1 & resp3)
SS1.3<-t(groupind)%*%Rr1.3

Sp1.3<-apply(SS1.3, 1, prod) # For the corrected rank responses
                             # and covariate
```

```
Sp1.3i<-Sp1.3*(N - 1)^-1
Spm1.3<-sum(Sp1.3i) # To aggregate the sums from the two groups

nc1.3<-Spm1.3 # value of numerator for r_{13}
dc1.3<-sqrt(Sm1*Sm3) # value of denominator for r_{13}

r13<-nc1.3/dc1.3 # covariance between hat(v_1) & hat(v_3)

Rr1.4<-cbind(Rij1.4,rij2) # To form a matrix of square
# corrected rank(covariate) and the products of corrected
# ranks (response1 & response4)
SS1.4<-t(groupind)%*%Rr1.4

Sp1.4<-apply(SS1.4, 1, prod) # For the corrected rank
# responses  and covariate
Sp1.4i<-Sp1.4*(N - 1)^-1
Spm1.4<-sum(Sp1.4i) # To aggregate the sums from the two groups

nc1.4<-Spm1.4 # value of numerator for r_{14}
dc1.4<-sqrt(Sm1*Sm4) # value of denominator for r_{14}

r14<-nc1.4/dc1.4 # covariance between hat(v_1) & hat(v_4)

Rr1.5<-cbind(Rij1.5,rij2) # To form a matrix of square
# corrected rank(covariate) and the products of corrected
# ranks (response1 & response5)
SS1.5<-t(groupind)%*%Rr1.5

Sp1.5<-apply(SS1.5, 1, prod) # For the corrected rank
                             # responses and covariate
Sp1.5i<-Sp1.5*(N - 1)^-1
Spm1.5<-sum(Sp1.5i) # To aggregate the sums from the two groups
```

```
nc1.5<-Spm1.5 # value of numerator for r_{15}
dc1.5<-sqrt(Sm1*Sm5) # value of denominator for r_{15}

r15<-nc1.5/dc1.5 # covariance between hat(v_1) & hat(v_5)

Rr2.3<-cbind(Rij2.3,rij2) # To form a matrix of square
# corrected rank(covariate) and the products of corrected
# ranks (response2 & response3).
SS2.3<-t(groupind)%*%Rr2.3

Sp2.3<-apply(SS2.3, 1, prod) # For the corrected rank responses
# and covariate
Sp2.3i<-Sp2.3*(N - 1)^-1
Spm2.3<-sum(Sp2.3i) # To aggregate the sums from the two groups

nc2.3<-Spm2.3 # value of numerator for r_{23}
dc2.3<-sqrt(Sm2*Sm3) # value of denominator for r_{23}

r23<-nc2.3/dc2.3 # covariance between hat(v_2) & hat(v_3)

Rr2.4<-cbind(Rij2.4,rij2) # To form a matrix of square
# corrected rank(covariate) and the products of corrected
# ranks (response2 & response4)
SS2.4<-t(groupind)%*%Rr2.4

Sp2.4<-apply(SS2.4, 1, prod) # For the corrected rank responses
                             # and covariate
Sp2.4i<-Sp2.4*(N - 1)^-1
Spm2.4<-sum(Sp2.4i) # To aggregate the sums from the two groups

nc2.4<-Spm2.4 # value of numerator for r_{24}
```

```
dc2.4<-sqrt(Sm2*Sm4) # value of denominator for r_{24}

r24<-nc2.4/dc2.4 # covariance between hat(v_2) & hat(v_4)

Rr2.5<-cbind(Rij2.5,rij2) # To form a matrix of square
# corrected rank(covariate) and the products of corrected
# ranks (response2 & response5)
SS2.5<-t(groupind)%*%Rr2.5

Sp2.5<-apply(SS2.5, 1, prod) # For the corrected rank
                             # responses & covariate
Sp2.5i<-Sp2.5*(N - 1)^-1
Spm2.5<-sum(Sp2.5i) # To aggregate the sums from the two groups

nc2.5<-Spm2.5 # value of numerator for r_{25}
dc2.5<-sqrt(Sm2*Sm5) # value of denominator for r_{25}

r25<-nc2.5/dc2.5 # covariance between hat(v_2) & hat(v_5)

Rr3.4<-cbind(Rij3.4,rij2) # To form a matrix of square
          # corrected rank(covariate) and the products
          # of corrected ranks (response3 & response4)
SS3.4<-t(groupind)%*%Rr3.4

Sp3.4<-apply(SS3.4, 1, prod) # For the corrected rank
 # responses & covariate
Sp3.4i<-Sp3.4*(N - 1)^-1
Spm3.4<-sum(Sp3.4i) # To aggregate the sums from the two groups

nc3.4<-Spm3.4 # value of numerator for r_{34}
dc3.4<-sqrt(Sm3*Sm4) # value of denominator for r_{34}
```

```
r34<-nc3.4/dc3.4 # covariance between hat(v_3) & hat(v_4)

Rr3.5<-cbind(Rij3.5,rij2) # To form a matrix of square
          # corrected rank(covariate) and the products
          # of corrected ranks (response3 & response5)
SS3.5<-t(groupind)%*%Rr3.5

Sp3.5<-apply(SS3.5, 1, prod) # For the corrected rank
                             # responses & covariate
Sp3.5i<-Sp3.5*(N - 1)^-1
Spm3.5<-sum(Sp3.5i) # To aggregate the sums from the two groups

nc3.5<-Spm3.5 # value of numerator for r_{35}
dc3.5<-sqrt(Sm3*Sm5) # value of denominator for r_{35}

r35<-nc3.5/dc3.5 # covariance between hat(v_3) & hat(v_5)

Rr4.5<-cbind(Rij4.5,rij2) # To form a matrix of square
          # corrected rank(covariate) and the products
          # of corrected ranks (response4 & response5)
SS4.5<-t(groupind)%*%Rr4.5

Sp4.5<-apply(SS4.5, 1, prod) # For the corrected rank
                             # responses & covariate
Sp4.5i<-Sp4.5*(N - 1)^-1
Spm4.5<-sum(Sp4.5i) # To aggregate the sums from the two groups

nc4.5<-Spm4.5 # value of numerator for r_{45}
dc4.5<-sqrt(Sm4*Sm5) # value of denominator for r_{45}

r45<-nc4.5/dc4.5 # covariance between hat(v_4) & hat(v_5)
```

```
vc1<-c(1,r12,r13,r14,r15)
vc2<-c(r12,1,r23,r24,r25)
vc3<-c(r13,r23,1,r34,r35)
vc4<-c(r14,r24,r34,1,r45)
vc5<-c(r15,r25,r35,r45,1)

vcv<-rbind(vc1,vc2,vc3,vc4,vc5)

l<-c(1,1,1,1,1) # vector of ones with dimension = p-1
de<-sqrt(t(l)%*%vcv%*%l) # value of denominator for M_v

Mv = nu/de # calculated value of M_v

significant.experiments.Mv[i]<-pnorm(Mv, lower.tail=FALSE)
# p-values for Mv
}
type_1er.Mv1.4[j] <- mean(significant.experiments.Mv<0.05)
# store average success rate (type-I error) for each N in Mv}
```