npordtests: An R Package of Nonparametric Tests for Equality of Location Against Ordered Alternatives

by Bulent Altunkaynak and Hamza Gamgam

Abstract Ordered alternatives are an important statistical problem in many situation such as increased risk of congenital malformation caused by excessive alcohol consumption during pregnancy life test experiments, drug-screening studies, dose-finding studies, the dose-response studies, age-related response. There are numerous other examples of this nature. In this paper, we present the **npordtests** package to test the equality of locations for ordered alternatives. The package includes the Jonckheere-Terpstra, Beier and Buning's Adaptive, Modified Jonckheere-Terpstra, Terpstra-Magel, Ferdhiana-Terpstra-Magel, KTP, S and Gaur's Gc tests. A simulation study is conducted to determine which test is the most appropriate test for which scenario and to suggest it to the researchers.

Introduction

Ordered alternative tests are employed to evaluate if a quantitative feature is linked to an ordinal trait, as in the association between ammonia levels and the severity of hepatic encephalopathy (Ong et al., 2003), the association of abnormal MRI findings with bone-marrow-related disease (Bredella et al., 2006), and the association between single nucleotide polymorphisms in human genes and quantitative phenotypes (Hoffmeyer et al., 2000; Cheng et al., 2005; Kawaguchi et al., 2012; Uchiyama et al., 2012; Tan et al., 2014; Yorifuji et al., 2018)

There are parametric and nonparametric methods to test ordered alternatives. Nevertheless, the statistical validity of parametric methods depends upon distributional assumptions, such as normality or equality of variances. However, nonparametric tests do not necessitate assumptions about the distribution of the data and are robust to outliers and influential values (Lin et al., 2017b).

Several nonparametric tests were developed to test the equality of locations against ordered alternatives. These tests can be grouped under three headings such as linear combination of two sample statistics, linear rank statistics, and statistics based on k-tuplet.

The tests proposed by Terpstra (1952), Jonckheere (1954), Puri (1965), Govindarajulu and Haller (1971), Tryon and Hettmansperger (1973), Cuzick (1985), Le (1988), Neuhäuser et al. (1998), Gaur (2014), Shan et al. (2014), Gaur (2017) are based on a linear combination of two sample statistics with pairs of samples of $k(k-1)/2$. The problem of testing homogeneity against ordered alternatives was considered for the first time by Terpstra (1952) and Jonckheere (1954). They suggested the nonparametric test (JT) based on a sum of $k(k-1)/2$ Mann-Whitney (MW) statistics for the ordered alternatives.

Linear rank statistics consist of a combination of the rank scores obtained from the combined data and the regression constants. These statistics were originally named as the Left Skewed (LS) and Right Skewed (RS) scores as proposed by Hogg et al. (1975). Gastwirth (1965), Buning and Kossler (1996), and Beier and Buning (1997) proposed Short-Tailed (ST), Long-Tailed (LT), and Wilcoxon (WS) scores, respectively. Beier and Buning (1997) proposed a nonparametric Adaptive Test (AT) for the choice of suitable scores based on the underlying distribution.

The k-tuplet tests are based on the information simultaneously obtained across all samples. These tests are determined by adding $N^* = n_1 \times n_2 \times ... \times n_k$ functions. That is, k-tuplet includes one observation from each group. Terpstra and Magel (2003) proposed a test k-tuplet statistic (TM), which is based on the indicator function. Ferdhiana et al. (2008) proposed a test statistic (FTM), which can be viewed as a generalization of the TM test. The FTM test uses Kendall correlation coefficient based on the following data: $(1, X_{1i_1}), (2, X_{2i_2}), ..., (k, X_{ki_k})$, where X_{ij} $i = 1, 2, ..., k$, $j = 1, 2, ..., n_i$ is the sample data. Here, k is the number of groups and n_i denotes the number of observations in the ith group. Similarly, Terpstra et al. (2011) proposed KTP test, which uses Spearman correlation coefficient instead of Kendall correlation coefficient.

JT is the classical and the most common ordered test. It is included in some packages such as **clinfun** (Venkatraman, 2018), **jtGWAS** (Lin et al., 2017a), **fastJT** (Lin et al., 2017b), **kSamples** (Scholz and Zhu, 2018), **StatCharrms** (Swintek et al., 2018), **PMCMRplus** (Pohlert, 2018). However, the other ordered alternative tests considered in this study are not included in any CRAN package other than **npordtests**.

However, there may be more efficient tests than JT for different data scenarios; nonetheless, a perusal of literature does not yield a comprehensive simulation study in which ordered alternative tests are compared for various scenarios. The nonparametric ordered alternative tests have recently been adapted for such big data structures as gene data and machine learning (Lin et al., 2017b), which clearly indicates the significance such a simulation study has.

Our study contributes significantly to the related literature in two ways: 1) This study includes most of the ordered alternative tests in the literature, introduced as an R package, **npordtests** (Altunkaynak and Gamgam, 2019) including the JT, Modified JT, LS, RS, ST, LT, WS, AT, TM, FTM, KTP, S, and Gaur's Gc tests, and presents open source codes. The **npordtests** package is publicly available on the CRAN. 2) This study presents a comprehensive simulation study that compares ordered alternative tests in terms of power, which helps researchers choose the most appropriate test for a given scenario.

The organization of this paper is presented as follows. After the introduction, firstly, we give the theoretical information about the nonparametric tests for ordered alternatives included in this study. Secondly, we introduce the **npordtests** package and demonstrate the applicability of the package using two benchmark datasets. Thirdly, a simulation study is conducted to determine which test is the most appropriate test for which scenario and to give some advice to the researchers. The results of this simulation study and general comments are given in the final section.

Ordered alternative tests

Let $X_{i1}, X_{i2}, ..., X_{in_i}, i = 1, ..., k$ be random independent samples with size n_i from k populations with continuous cumulative distribution function $F_i(x) = F((x - \theta_i)/\sigma_i)$, where $-\infty < \theta_i < +\infty$ and $\sigma_i > 0$ are location and scale parameters, respectively. The null hypothesis to identify whether the populations have common continuous cumulative distribution function can be expressed as

$$H_0 : F_1(x) = F_2(x) = ... = F_k(x) \qquad \forall x. \tag{1}$$

A number of test statistics have been proposed to test the null hypothesis in (1) under certain assumptions and for different forms of H_1. The ordered alternative states that the distributions are stochastically ordered, i.e.,

$$H_1 : F_1(x) \geq F_2(x) \geq ... \geq F_k(x) \;\; \exists x : F_1(x) > F_k(x)\,. \tag{2}$$

Under H_1, X_i tends to be smaller than X_{i+1}, $i = 1, 2, ..., k-1$, since $F_i(x) \geq F_{i+1}(x)$ implies that $P(X_i \leq X_{i+1}) \geq 1/2$. For the special case of the location model, (2) is equivalent to (Terpstra et al., 2011)

$$H_1 : \theta_1 \leq \theta_2 \leq ... \leq \theta_k \quad (\theta_1 < \theta_k). \tag{3}$$

Similarly, the ordered alternative hypothesis

$$H_1 : F_1(x) \leq F_2(x) \leq ... \leq F_k(x) \;\; \exists x : F_1(x) < F_k(x) \tag{4}$$

states that X_i tends to be larger than X_{i+1}, $i = 1, 2, ..., k-1$, since $F_i(x) \leq F_{i+1}(x)$ implies that $P(X_i \geq X_{i+1}) \geq 1/2$ under H_1 given in (4). For the location model, (4) is equivalent to

$$H_1 : \theta_1 \geq \theta_2 \geq ... \geq \theta_k \quad (\theta_1 > \theta_k). \tag{5}$$

Jonckheere-Terpstra test

This classic nonparametric test is typically used for ordered alternatives and was proposed by Terpstra (1952) and Jonckheere (1954). It is known that the Mann-Whitney statistic defines as

$$U_{ij} = \sum_{l=1}^{n_i} \sum_{m=1}^{n_j} I(X_{il} < X_{jm});$$

where n_i and n_j are the sample sizes for the ith and jth populations, respectively, and $I(\psi) = 1$ if ψ is true and 0 otherwise. The test statistic JT corresponds to the sum of the $k(k-1)/2$ Mann-Whitney statistics, i.e.,

$$JT = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} U_{ij}. \tag{6}$$

The statistic JT is approximately normally distributed under H_0. The mean and variance of this statistic are

$$E(JT) = \frac{N^2 - \sum_{i=1}^{k} n_i^2}{4}$$

and

$$V(JT) = \frac{N^2(2N+3) - \sum_{i=1}^{k} n_i^2(2n_i+3)}{72},$$

where $N = n_1 + n_2 + ... + n_k$.

Beier and Buning's Adaptive test

This test is a two-step method based on the selection of the weight coefficients of the linear rank statistics according to the shape of the distribution (Beier and Buning, 1997). A linear rank statistics has the following form:

$$L_N = \sum_{i=1}^{k} \sum_{j=1}^{n_i} c_N(i) a_N(R_{ij}) \tag{7}$$

where N is the combined sample size; $c_N(.)$ are the regression constants; $a_N(.)$ are the scores; R_{ij} is the rank of X_{ij} in the combined data. For an ordered alternative, the following proposal is made:

$$c_N(i) = i, i = 1, 2, ..., k.$$

Under H_0, the mean and variance of linear rank statistics are

$$E(L_N) = N\bar{c}_N \bar{a}_N,$$

and

$$V(L_N) = \frac{1}{N-1} \sum_{i=1}^{k} n_i (c_N(i) - \bar{c}_N)^2 \sum_{r=1}^{N} (a_N(r) - \bar{a}_N)^2$$

where

$$\bar{c}_N = \frac{1}{N} \sum_{i=1}^{k} n_i c_N(i)$$

and

$$\bar{a}_N = \frac{1}{N} \sum_{r=1}^{N} a_N(r).$$

The distribution of a linear rank statistic converges to a normal distribution with mean $E(L_N)$ and variance $V(L_N)$ (Hogg and Craig, 2013; Beier and Buning, 1997).

There are some suggestions for the score $a_N(.)$ according to the shape of the distribution in the literature as follows

$$a_{LS}(r) = \begin{cases} 0 & \text{if } r \leq (N+1)/2 \\ r - (N+1)/2 & \text{if } r > (N+1)/2 \end{cases}$$

These scores are efficient for detecting shifts in distributions that are skewed to the left (Beier and Buning, 1997).

$$a_{ST}(r) = \begin{cases} r - (N+1)/4 & \text{if } r \leq (N+1)/4 \\ 0 & \text{if } (N+1)/4 < r < 3(N+1)/4 \\ r - 3(N+1)/4 & \text{if } r \geq 3(N+1)/4 \end{cases}$$

These scores are particularly good for detecting shifts in short-tailed distributions and were proposed by Gastwirth (1965).

$$a_{WS}(r) = r, \quad r = 1, 2, ..., N$$

These scores are efficient for detecting shifts in symmetric distributions with medium to heavy tails (Beier and Buning, 1997).

$$a_{LT}(r) = \begin{cases} -((N/4)+1) & \text{if } r < (N/4)+1 \\ r - (N+1)/2 & \text{if } (N/4)+1 \leq r \leq 3(N+1)/4 \\ (N/4)+1 & \text{if } r > 3(N+1)/4 \end{cases}$$

These scores are efficient for detecting shifts in long-tail distributions and were proposed by Buning and Kossler (1996).

$$a_{RS}(r) = \begin{cases} r - (N+1)/2 & \text{if } r \le (N+1)/2 \\ 0 & \text{if } r > (N+1)/2 \end{cases}$$

These scores are efficient for detecting shifts in distributions that are skewed to the right (Hogg et al., 1975).

The adaptive test proposed by Beier and Buning (1997) is denoted by the index of their scores. For example, the distribution-free test based on the scores $a_{ST}(.)$ of Gastwirth (1965), which is particularly good for detecting a shift in short-tailed distributions, is denoted by ST. Now, the adaptive test AT is defined by

$$AT = \begin{cases} LS & \text{if } 0 \le \hat{S}_1 \le 0.6, \hat{S}_2 \ge 1 \\ ST & \text{if } 0.6 < \hat{S}_1 \le 2, 1 \le \hat{S}_2 \le 1.5 \\ WS & \text{if } 0.6 < \hat{S}_1 \le 2, 1.5 < \hat{S}_2 \le 1.5 \\ LT & \text{if } 0.6 < \hat{S}_1 \le 2, \hat{S}_2 \ge 2 \\ RS & \text{if } \hat{S}_1 \ge 2, \hat{S}_2 \ge 1 \end{cases} \tag{8}$$

where x_p is the quantile value of the combined data, and the estimation values of the skewness and tailweight of the distribution are

$$\hat{S}_1 = \frac{x_{0.975} - x_{0.5}}{x_{0.5} - x_{0.025}}$$

and

$$\hat{S}_2 = \frac{x_{0.975} - x_{0.025}}{x_{0.875} - x_{0.125}}.$$

Since the adaptive statistic is a linear rank statistic, the distribution of each of these statistics converges to a normal distribution with mean $E(L_N)$ and variance $V(L_N)$.

Modified Jonckheere-Terpstra test

Tryon and Hettmansperger (1973) proposed the modified JT statistic to test H_0 against the ordered alternatives,

$$MJT = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (j-i)U_{ij}, \tag{9}$$

where U_{ij} is the Mann-Whitney statistic computed for the samples from the ith and jth populations. Neuhäuser et al. (1998) suggested that this test be used in place of the JT tests because it often has larger powers.

This statistic has a normal distribution under H_0, and its mean and variance are

$$E(U_{ij}) = \frac{1}{2} n_i n_j, \quad \forall i \ne j$$

$$V(U_{ij}) = \frac{1}{12} n_i n_j (n_i + n_j + 1), \quad \forall i \ne j$$

$$Cov(U_{ij}, U_{il}) = Cov(U_{ji}, U_{li}) = \frac{1}{12} n_i n_j n_l, \text{ if all } i, j, l \text{ are different}$$

$$Cov(U_{ij}, U_{li}) = Cov(U_{ji}, U_{il}) = -\frac{1}{12} n_i n_j n_l, \text{ if all } i, j, l \text{ are different}$$

$$Cov(U_{ij}, U_{lm}) = 0, \text{ if all } i, j, l, m \text{ are different}$$

Terpstra-Magel test

Terpstra and Magel (2003) proposed a test statistic that does not focus on pairwise information. Instead, they use the information present in the $N^* = n_1 \times n_2 \times ... \times n_k$ k-tuplets, where a k-tuplet includes one observation from each treatment group. More specifically, the Terpstra–Magel (TM) test is based on the following statistic:

$$TM = \sum_{i_1=1}^{n_1} ... \sum_{i_k=1}^{n_k} I(X_{1i_1} \le X_{2i_2} \le ... \le X_{ki_k}) \tag{10}$$

where the indicator function is equal to one when $X_{1i_1} < X_{ki_k}$.

The statistic TM is approximately normally distributed under H_0. The mean and variance of this statistic are

$$E(TM) = \frac{N^*}{k!}$$

and

$$V(TM) = N^* \left(\frac{1}{k!}\right)\left(1 - \frac{1}{k!}\right) + \sum_{i=1}^{k-1} v_i^2$$

where

$$v_i^2 = \sum_{1 \le l_1 < \ldots < l_i \le k} N^* \left[\prod_{s=1}^{k} (n_s - 1)^{I(s \ne l_1) \ldots I(s \ne l_i)}\right] \left[\frac{\binom{2(k-l_i)}{k-l_i}}{2k-i} \prod_{s=1}^{i} \binom{2(l_s - l_{s-1} - 1)}{l_s - l_{s-1} - 1} - \frac{1}{(k!)^2}\right]$$

where $l_0 = 0$.

Ferdhiana-Terpstra-Magel test

Ferdhiana et al. (2008) proposed FTM test statistic can be viewed as a generalization of the TM test.

$$FTM = \sum_{i_1=1}^{n_1} \ldots \sum_{i_k=1}^{n_k} \tau(X_{1i_1}, X_{2i_2}, \ldots, X_{ki_k}) \tag{11}$$

where $\tau(X_{1i_1}, X_{2i_2}, \ldots, X_{ki_k})$ denotes the Kendall correlation coefficient based on $(1, X_{1i_1}), (2, X_{2i_2}), \ldots, (k, X_{ki_k})$.

Under H_0, the statistic FTM is approximately normally distributed with zero mean, and its variance is

$$V(FTM) = \left[\frac{2N^*}{\sqrt{3}k(k-1)}\right]^2 \left[\sum_{r=1}^{k-1}\sum_{s=r+1}^{k} \frac{n_r + n_s + 1}{n_r n_s} + 2\sum_{r=1}^{k-2} \frac{1}{n_r}\left(\binom{k}{2} + \frac{r^2 - (2k-1)r}{2}\right) - 2\sum_{r=1}^{k-2}\sum_{s=r+1}^{k-1} \frac{k-s}{n_s} + 2\sum_{r=1}^{k-2}\sum_{s=r+1}^{k-1}\sum_{t=s+1}^{k} \frac{1}{n_i}\right].$$

KTP test

Terpstra et al. (2011) proposed the k-tuplet Terpstra-Page (KTP) test based on the statistic

$$KTP = \sum_{i_1=1}^{n_1} \ldots \sum_{i_k=1}^{n_k} r_s(X_{1i_1}, X_{2i_2}, \ldots, X_{ki_k}) \tag{12}$$

where $r_s(X_{1i_1}, X_{2i_2}, \ldots, X_{ki_k})$ denotes the Spearman rank correlation coefficient based on $(1, X_{1i_1})$, $(2, X_{2i_2}), \ldots, (k, X_{ki_k})$.

Under H_0, the statistic KTP is approximately normally distributed, and its mean and variance are

$$E(KTP) = 0$$

, and

$$V(KTP) = \frac{144(N^*)^2}{k^2(k^2-1)^2} S,$$

where

$$S = \sum_{i_1=1}^{k-1}\sum_{i_2=i_1+1}^{k} \left[\frac{(i_2 - i_1)^2 (n_{i_1} + n_{i_2} + 1)}{12 n_{i_1} n_{i_2}}\right] + \sum_{i_1=1}^{k-2}\sum_{i_2=i_1+1}^{k-1}\sum_{i_3=i_2+1}^{k} \left[\frac{(i_2 - i_1)(i_3 - i_1)}{6n_{i_1}} + \frac{(i_3 - i_2)(i_1 - i_2)}{6n_{i_2}} + \frac{(i_1 - i_3)(i_2 - i_3)}{6n_{i_3}}\right]$$

In the KTP test, Spearman's rank correlation coefficient r_s is given by the following formula:

$$r_s = 1 - \frac{6\sum\limits_{i=1}^{k} d_i^2}{k(k^2-1)}$$

where d_i represents the difference between the rank given to the value of the variable for each item of the particular data with y_i. This formula is applied in cases when there are no tied observations. The formula to use when there are tied observations is:

$$r_s = \frac{\sum\limits_{i=1}^{k} (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum\limits_{i=1}^{k} (y_i - \bar{y})^2 \sum\limits_{i=1}^{k} (x_i - \bar{x})^2}}$$

where $(y, x) = (1, X_{1i_1}), (2, X_{2i_2}), ..., (k, X_{ki_k})$ and x_i is rank of X_i. Note that if all of x_i values is equal, then $\sum (x_i - \bar{x})^2$ is zero. This result is also similar for Kendall correlation coefficient. Therefore, FTM and KTP tests cannot be applied to this type data. See Lehmann's data used in the demonstration of the **npordtests** package.

S test

Shan et al. (2014) proposed the new rank-based nonparametric test by incorporating the actual differences as follows

$$S = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} D_{ij} \tag{13}$$

where

$$D_{ij} = \sum_{l=1}^{n_i} \sum_{m=1}^{n_j} Z_{ijlm}, \quad Z_{ijlm} = (R_{jm} - R_{il}) I(X_{jm} > X_{il})$$

and $R_{il}(R_{jm})$ is the rank of observation $X_{il}(X_{jm})$ in the combined data.

Under H_0, the statistic S has a normal distribution with the following mean and variance

$$E(S) = \frac{N+1}{6} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} n_i n_j$$

$$\begin{aligned} V(S) = &\left(\frac{N^2+N}{12} - \frac{(N+1)^2}{36} \right) \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} n_i n_j \\ &+ 2 \left[\sum_{i=1}^{k-1} n_i \binom{\sum\limits_{j=i+1}^{k} n_j}{2} + \sum_{i=2}^{k} n_i \binom{\sum\limits_{j=1}^{i-1} n_j}{2} \right] CovA + 2 \left(\sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \sum_{l=j+1}^{k} n_i n_j n_l \right) CovB \end{aligned}$$

where $CovA = \frac{2N^2+N-1}{90}$, and $CovB = \frac{-7N^2-11N-4}{360}$.

Gaur's Gc test

Let $(w_1, w_2, ..., w_{k-1})$ be suitably selected real positive constants. Gaur (2017) proposed the G_c statistic to test H_0 against the ordered alternatives,

$$G_c = \sum_{g=1}^{k-1} w_g V_{g,g+1} \tag{14}$$

where

$$V_{g,h} = \left[\binom{n_g}{c} \binom{n_h}{c} \right]^{-1} \sum_{0} \phi_{gh}(X_{g\alpha_1}, ..., X_{g\alpha_c}; X_{h\beta_1}, ..., X_{h\beta_c})$$

for $g < h$; $h = 1, 2..., k$; $\sum_0$ is the sum over all combinations $(\alpha_1, ..., \alpha_c)$ of c integers selected from $(1, ..., n_g)$ and over all combinations $(\beta_1, ..., \beta_c)$ of c integers selected from $(1, ..., n_h)$;

$$\phi_{gh}(X_{g\alpha_1}, ..., X_{g\alpha_c}; X_{h\beta_1}, ..., X_{h\beta_c}) = \begin{cases} 1 & \text{if } \max(X_{g\alpha_1}, ..., X_{g\alpha_c}) \leq \min(X_{h\beta_1}, ..., X_{h\beta_c}) \\ -1 & \text{if } \max(X_{h\beta_1}, ..., X_{h\beta_c}) \leq \min(X_{g\alpha_1}, ..., X_{g\alpha_c}) \\ 0 & \text{otherwise} \end{cases}.$$

The distribution of Gaur's statistic G_c converges to a normal distribution with zero mean under H_0, and the variance of this statistic are obtained as follows

$$V(G_c) = \mathbf{w}^\top \sum \mathbf{w}$$

where $\mathbf{w}^\top = (w_1, w_2, ..., w_{k-1})$ and $\sum = [\sigma_{gh}]$ is the variance-covariance matrix, such as:

$$\sigma_{gh} = \begin{cases} \left(\frac{(c-1)!c!}{(2c-1)!}\right)^2 \left(\frac{1}{\lambda_g} + \frac{1}{\lambda_{g+1}}\right) \delta_c & \text{for } g = h = 1, 2, ..., k-1 \\ -\left(\frac{(c-1)!c!}{(2c-1)!}\right)^2 \frac{\delta_c}{\lambda_{g+1}} & \text{for } h = g+1; g = 1, 2, ..., k-2 \\ -\left(\frac{(c-1)!c!}{(2c-1)!}\right)^2 \frac{\delta_c}{\lambda_g} & \text{for } h = g-1; g = 2, ..., k-1 \\ 0 & \text{otherwise} \end{cases}$$

where

$$\delta_c = -1 + \frac{4}{4c-1} \sum_{i=c}^{2c-1} \sum_{j=c}^{2c-1} \binom{2c-1}{i} \binom{2c-1}{j} \binom{4c-2}{i+j}^{-1}.$$

It is recommended to use G_c tests for light-tailed and moderate-tailed distributions with $c = 2$, whereas for heavy-tailed and long-tailed distributions with large values of c. The optimum weights w_g's in the G_c test are

$$w_g = \frac{g(k-g)}{2k}, \; g = 1, 2, ..., k-1.$$

Demonstration of the npordtests package

The **npordtests** package includes thirteen tests and six datasets for ordered alternatives. In this section, firstly, we introduce the datasets included in the package. Then, we demonstrate the usage of the package by using two of these datasets. All the examples in this section should run if you type them in exactly as printed, provided that you have the **npordtests** package not only installed but also loaded into your current search path. This is done by entering

```
R> library(npordtests)
```

at the command prompt.

Datasets

Jonckheere's data: jdata

This hypothetic data given by Jonckheere (1954) are used to test the hypothesis that the four samples have come from the same population against the alternative that the populations are such that the values from the samples I, II, III, IV are in an expected order of increasing value.

Lehmann's data: lehmann

This dataset was used by Lehmann (1975) to assess if it is possible for a particular diagnostic test to be successfully interpreted without psychological training. This dataset later became one of the classical datasets used to investigate sequential alternatives (Beier and Buning, 1997). The data included 72 evaluators' (21 staff members, 23 trainees and 28 undergraduate psychology majors) assessment scores for the diagnostic test. If training and experience have any effects, the staff members could be expected to perform the most accurately, the trainees next, and the undergraduates the least.

Chicks' weight data: chicks

These data are given by Desu and Raghavarao (2004) to examine the hypothesis that the chicks' mean weight goes up with the increase in the amount of protein. Eighteen chicks were randomly assigned to three treatments with six chicks in each for balanced data. Treatment 1 had the diet with the lowest level of protein; treatment 2 had the diet with a medium level of protein; and treatment 3 had the highest level of protein. After six weeks of feeding, the values of weight gain were recorded. We wanted to test if the mean weight gain increased with the amount of protein (Chang and Yen, 2011).

Hepatic vein waveform index data: hvwi

These data were collected by Pedersen et al. (2008) through doppler waveforms corresponding to 66 patients scheduled for a percutaneous liver needle biopsy. The waveforms were characterized using a hepatic vein waveform index (HVWI), whereas the biopsy specimens were grouped according to the degree of fibrosis. The hypothesis of interest was that the HVWI values would tend to decrease as the degree of fibrosis increases (Terpstra et al., 2011).

Hypertension data: hypertension

These data presented by Dmitrienko et al. (2006) examine the effect of different drug doses on diastolic blood pressure. The patients with hypertension were randomized into four groups with different dose levels, 0, 10, 20, and 40 mg/day, where the group with 0 mg/day was the placebo group. The number of the patients in each group were 17, 17, 18, and 16, respectively. The complete data can be found at the Dmitrienko et al. (2006) or Shan et al. (2014).

Neuhauser's data: neuhauser

These synthetic data are reported by Neuhäuser et al. (1998). The data consist of 4 groups with 10 observations in each.

In order to compare the distributions of groups for each dataset, the boxplots are given in Figure 1. As can be seen from the figure, there is a ordered alternative pattern in all datasets.

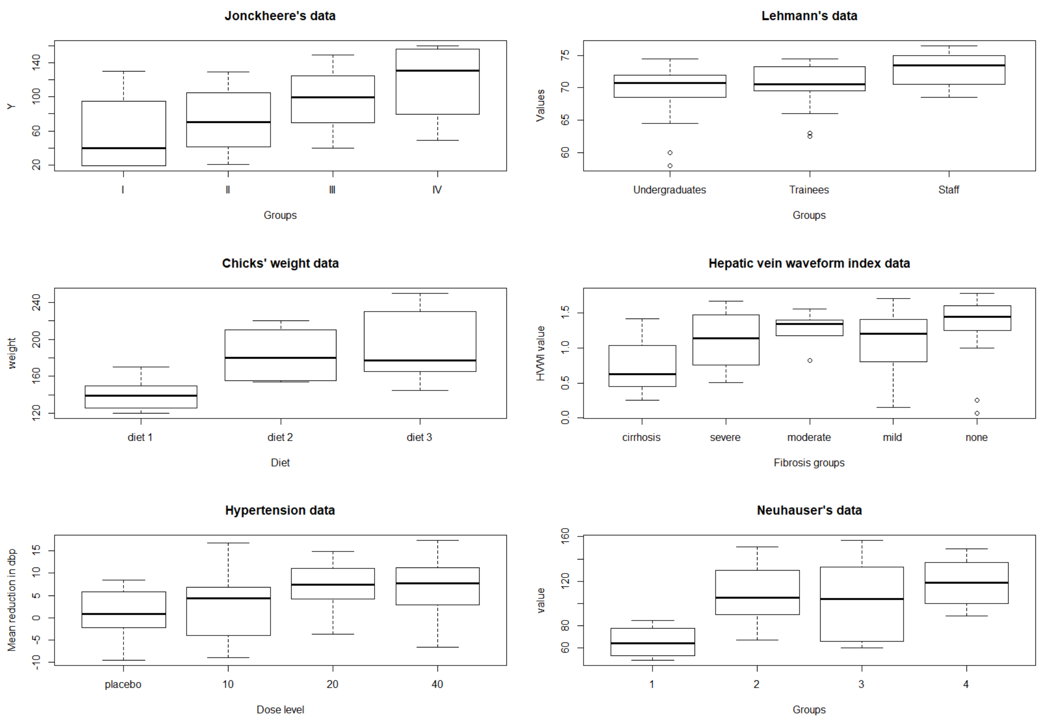

Figure 1: Boxplots for the datasets. Each box plot gives median (the bold line that divides the box into two parts), lower and upper quartiles (start and end points of the box on the vertical axis) and min and max value (the horizontal lines outside the box). The outliers appear as the circles.

Tests

Using the datasets which are named **jdata** and **lehmann**, demonstration of the tests are given below, respectively.

Jonkheere-Terpstra test: JtTest(...)

The `JtTest` function in the **npordtests** package is used to perform the Jonkheere-Terpstra test.

```
R> data(jdata)
R> JtTest(Y~X,jdata,alpha=0.05,na.rm=TRUE,verbose=TRUE)

---------------------------------------------------------
Test : Jonckheere-Terpstra Test
data : Y and X

Statistic = 71
Mean = 48
Variance = 114.6667
Z = 2.147876
Asymp. p-value = 0.0158618

Result     : Null hypothesis is rejected.
---------------------------------------------------------
```

Here, the `JT statistic` is calculated from the Equation (6). Also, the `Mean` and `Variance` are expected value and variance of the JT statistic, respectively. `Z` is calculated from $(JT - E(JT)) / \sqrt{V(JT)}$. `p-value` is the significance value for the JT test. Because this p-value is smaller than $\alpha = 0.05$, the hypothesis of the equality of locations against the ordered alternative is rejected.

`alpha` is the level of significance to assess the statistical difference. Default is set to `alpha = 0.05`. `na.rm` is a logical value indicating whether NA values should be stripped before the computation proceeds. Default is `na.rm = TRUE`. `verbose` is a logical for printing output to R console. Default is set to `verbose = TRUE`. These arguments are available in the functions for ordered alternatives. The users who would like to use the statistics in the output in their programs can use the following codes.

```
R> res<-JtTest(Y~X,jdata,alpha=0.05,na.rm=TRUE,verbose=FALSE)

R> res$statistic
[1] 71

R> res$mean
[1] 48

R> res$variance
[1] 114.6667

R> res$Z
[1] 2.147876

R> res$p.value
[1] 0.0158618
```

Here, the codes for how to obtain the statistics from the Jonckheere-Terpstra test output are given. Since all ordered alternative tests return similar outputs, similar codes are not repeated in the other tests. For all tests, the level of significance is taken as 0.05.

Beier and Buning's Adaptive test: AtTest(...)

The `AtTest` function in the **npordtests** package is used to perform the Adaptive test. The LS, RS, ST, WS and LT tests are also available as functions in the package.

```
R> LsTest(Y~X,jdata)
```

```
-------------------------------------------------------
Test : LS test
data : Y and X

Statistic = 68
Mean = 48
Variance = 141.3333
Z = 1.682316
Asymp. p-value = 0.04625375

Result     : Null hypothesis is rejected.
-------------------------------------------------------
```

Here, the `Statistic` is calculated from the Equation (7) using the score $a_{LS}(r)$. Also, the `Mean` and `Variance` are the expected value and variance of the this statistic, respectively. `Z` is calculated from $(LS - E(LS))/\sqrt{V(LS)}$. `p-value` is the significance value for the LS test. Since this `p-value` is smaller than $\alpha = 0.05$, the null hypothesis against the ordered alternative is rejected.

```
R> RsTest(Y~X,jdata)

-------------------------------------------------------
Test : RS test
data : Y and X

Statistic = -27
Mean = -48
Variance = 141.3333
Z = 1.766432
Asymp. p-value = 0.03866168

Result     : Null hypothesis is rejected.
-------------------------------------------------------
```

In the output, similar to `LsTest`, the `Statistic` is calculated from the Equation (7) using the score $a_{RS}(r)$. `Z` is calculated from $(RS - E(RS))/\sqrt{V(RS)}$. `p-value` is the significance value for the RS test. According to these results, because the `p-value` is smaller than $\alpha = 0.05$, the hypothesis of the equality of locations against the ordered alternative is rejected.

```
R> StTest(Y~X,jdata)

-------------------------------------------------------
Test : ST test
data : Y and X

Statistic = 17.25
Mean = 0
Variance = 46
Z = 2.543374
Asymp. p-value = 0.005489386

Result     : Null hypothesis is rejected.
-------------------------------------------------------
```

In the output, the `Statistic` is calculated from the Equation (7) using the score $a_{ST}(r)$. `Z` is calculated from $(ST - E(ST))/\sqrt{V(ST)}$. `p-value` is the significance value for the ST test. Here, the `Statistic` is calculated value of the test statistic. The `p-value` for the TM test is 0.005489386. Thus, we can conclude that the null hypothesis of the equality of locations is rejected under setting $\alpha = 0.05$.

```
R> WsTest(Y~X,jdata)

-------------------------------------------------------
Test : WS test
data : Y and X

Statistic = 245
```

```
Mean = 204
Variance = 453.3333
Z = 1.92564
Asymp. p-value = 0.02707469

Result     : Null hypothesis is rejected.
---------------------------------------------------------
```

Here, the `WS statistic` is calculated from the Equation (7) using the score $a_{WS}(r)$. `Z` is calculated from $(WS - E(WS))/\sqrt{V(WS)}$. `p-value` is the significance value for the WS test. Because this `p-value` is smaller than $\alpha = 0.05$, the hypothesis of the equality of locations against the ordered alternative is rejected.

```
R> LtTest(Y~X,jdata)

---------------------------------------------------------
Test : LT test
data : Y and X

Statistic = 27.5
Mean = 0
Variance = 322.6667
Z = 1.530931
Asymp. p-value = 0.06289321

Result     : Null hypothesis is not rejected.
---------------------------------------------------------
```

The `LT statistic` is calculated from the Equation (7) using the score $a_{LT}(r)$. `Z` is calculated from $(LT - E(LT))/\sqrt{V(LT)}$. `p-value` is the significance value for the LT test. According to these results, because the `p-value` is not smaller than $\alpha = 0.05$, the hypothesis of the equality of locations (null hypothesis) is not rejected.

```
R> AtTest(Y~X,jdata)

---------------------------------------------------------
Test : Adaptive Test
data : Y and X

Statistic = 17.25
Mean = 0
Variance = 46
Z = 2.543374
Asymp. p-value = 0.005489386

Result     : Null hypothesis is rejected.
---------------------------------------------------------
```

Here, the `Statistic` is calculated from the Equation (8). Note that the AT `Statistic` is equal to the ST `Statistic` for this example. Since this `p-value` is smaller than $\alpha = 0.05$, the null hypothesis against the ordered alternative is rejected.

Modified Jonkheere-Terpstra test: MjtTest(...)

The `MjtTest` function in the **npordtests** package is used to perform the MJT test.

```
R> MjtTest(Y~X,jdata)

---------------------------------------------------------
Test : Modified Jonckheere-Terpstra Test
data : Y and X

Statistic = 121
Mean = 80
Variance = 453.3333
```

```
Z = 1.92564
Asymp. p-value = 0.02707469

Result     : Null hypothesis is rejected.
---------------------------------------------------------
```

Here, the `Statistic` is calculated from the Equation (9). According to these results, because the `p-value` is smaller than $\alpha = 0.05$, the hypothesis of the equality of locations against the ordered alternative is rejected.

Terpstra-Magel test: TmTest(...)

The `TmTest` function in the **npordtests** package is used to perform the TM test.

```
R> TmTest(Y~X,jdata)

---------------------------------------------------------
Test : Terpstra-Magel Test
data : Values and Group

Statistic = 78
Mean = 10.66667
Variance = 151.327
Z = 5.473586
Asymp. p-value = 2.205097e-08

Result : Null hypothesis is rejected.
---------------------------------------------------------
```

In the output, the `Statistic` is calculated from the Equation (10). `Z` is calculated from $(TM - E(TM))/\sqrt{V(TM)}$. `p-value` is the significance value for the TM test. The `p-value` for the TM test is 0.00000002205097. Thus, we can conclude that the null hypothesis of the equality of locations is rejected under setting $\alpha = 0.05$.

Ferdhiana-Terpstra-Magel test: FtmTest(...)

The `FtmTest` function in the **npordtests** package is used to perform the FTM test.

```
R> FtmTest(Y~X,jdata)

---------------------------------------------------------
Test : Ferdhiana, Terpstra and Magel Test
data : Y and X

Statistic = 122.6667
Mean = 0
Variance = 3261.63
Z = 2.147876
Asymp. p-value = 0.0158618

Result     : Null hypothesis is rejected.
---------------------------------------------------------
```

Here, the `Statistic` is calculated from the Equation (11). `Z` is calculated from $FTM/\sqrt{V(FTM)}$. `p-value` is the significance value for the FTM test. Because this `p-value` is smaller than $\alpha = 0.05$, the hypothesis of the equality of locations against the ordered alternative is rejected.

KTP test: KtpTest(...)

The `KtpTest` function in the **npordtests** package is used to perform the KTP test.

```
R> KtpTest(Y~X,jdata)
```

```
--------------------------------------------------------
Test : KTP Test
data : Y and X

Statistic = 131.2
Mean = 0
Variance = 4642.133
Z = 1.92564
Asymp. p-value = 0.02707469

Result     : Null hypothesis is rejected.
--------------------------------------------------------
```

Here, the `Statistic` is calculated from the Equation (12). `Z` is calculated from $KTP/\sqrt{V(KTP)}$. `p-value` is the significance value for the KTP test. Since this `p-value` is smaller than $\alpha = 0.05$, the null hypothesis against the ordered alternative is rejected.

S test: SsTest(...)

The `SsTest` function in the **npordtests** package is used to perform the S test.

```
R> SsTest(Y~X,jdata)

--------------------------------------------------------
Test : Shan's S test
data : Y and X

Statistic = 436
Mean = 272
Variance = 1973.511
Z = 3.69168
Asymp. p-value = 0.0001113888

Result     : Null hypothesis is rejected.
--------------------------------------------------------
```

In the output, the `Statistic` is calculated from the Equation (13). `Z` is calculated from $(S - E(S))/\sqrt{V(S)}$. `p-value` is the significance value for the S test. According to these results, because the `p-value` is smaller than $\alpha = 0.05$, the hypothesis of the equality of locations against the ordered alternative is rejected.

Gaur's Gc test: GcTest(...)

The `GcTest` function in the **npordtests** package is used to perform the Gaur's Gc test.

```
R> GcTest(Y~X,jdata)

--------------------------------------------------------
Test : Gaur's Gc Test
data : Values and Group

Statistic = 0.375
Mean = 0
Variance = 0.06746032
Z = 1.4438
Asymp. p-value = 0.0743976

Result : Null hypothesis is not rejected.
--------------------------------------------------------
```

Here, the `Statistic` is calculated from the Equation (14). `Z` is calculated from $G_c/\sqrt{V(G_c)}$. `p-value` is the significance value for the G_c test. Here, the `Statistic` is calculated value of the test statistic. The `p-value` for the G_c test is 0.0743976. Thus, we can conclude that the null hypothesis of the equality of locations is not rejected under setting $\alpha = 0.05$.

Jonkheere-Terpstra test: JtTest(...)

The `JtTest` function in the **npordtests** package is used to perform the JT test.

```
R> data(lehmann)
R> JtTest(Values~Group,lehmann)

---------------------------------------------------------
Test : Jonckheere-Terpstra Test
data : Values and Group

Statistic = 1159
Mean = 857.5
Variance = 9305.917
Z = 3.125415
Asymp. p-value = 0.0008877709

Result     : Null hypothesis is rejected.
---------------------------------------------------------
```

Here, the `Statistic` is calculated value of the test statistic. `p-value` is the significance value for this test. The `p-value` for the JT test is 0.0008877709. Thus, we can conclude that the null hypothesis of the equality of locations is rejected under setting $\alpha = 0.05$.

Beier and Buning's Adaptive test: AtTest(...)

The `AtTest` function in the **npordtests** package is used to perform the AT test.

```
R> AtTest(Values~Group,lehmann)

---------------------------------------------------------
Test : Adaptive Test
data : Values and Group

Statistic = 851
Mean = 583.1944
Variance = 6570.726
Z = 3.303794
Asymp. p-value = 0.0004769302

Result     : Null hypothesis is rejected.
---------------------------------------------------------
```

Here, the `Statistic` is calculated value of the test statistic. `p-value` is the significance value for this test. The `p-value` for the AT test is 0.0004769302. Because this `p-value` is smaller than $\alpha = 0.05$, the hypothesis of the equality of locations against the ordered alternative is rejected.

Modified Jonkheere-Terpstra test: MjtTest(...)

The `MjtTest` function in the **npordtests** package is used to perform the MJT test.

```
R> MjtTest(Values~Group,lehmann)

---------------------------------------------------------
Test : Modified Jonckheere-Terpstra Test
data : Values and Group

Statistic = 1610
Mean = 1151.5
Variance = 20771.92
Z = 3.181274
Asymp. p-value = 0.0007331448

Result     : Null hypothesis is rejected.
---------------------------------------------------------
```

Here, the `Statistic` is calculated value of the test statistic. `p-value` is the significance value for the MJT test. The `p-value` for the MJT test is 0.0007331448. Since this `p-value` is smaller than $\alpha = 0.05$, the null hypothesis against the ordered alternative is rejected.

Terpstra-Magel test: TmTest(...)

The `TmTest` function in the **npordtests** package is used to perform the TM test.

```
R> TmTest(Values~Group,lehmann)

---------------------------------------------------------
Test : Terpstra-Magel Test
data : Values and Group

Statistic = 5173
Mean = 2254
Variance = 405043.8
Z = 4.586518
Asymp. p-value = 2.253498e-06

Result : Null hypothesis is rejected.
---------------------------------------------------------
```

Here, the `Statistic` is calculated value of the test statistic. `p-value` is the significance value for this test. The `p-value` for the TM test is 0.000002253498. Thus, we can conclude that the null hypothesis of the equality of locations is rejected under setting $\alpha = 0.05$.

Ferdhiana-Terpstra-Magel test: FtmTest(...)

The `FtmTest` function in the **npordtests** package is used to perform the FTM test.

```
R> FtmTest(Values~Group,lehmann)

---------------------------------------------------------
Test : Ferdhiana, Terpstra and Magel Test
data : Values and Group

Statistic = NA
Mean = 0
Variance = 2294071
Z = NA
Asymp. p-value = NA

Error in if (p-value > alpha) { : missing value where TRUE/FALSE needed
In addition: Warning message:
In cor(t(Xmat), Ymat, method = "kendall") : the standard deviation is zero
```

As seen in the output, the error `standard deviation is zero` is encountered. This error occurs because the values of 68.5, 69.0, 70.5, 71.5, 73.0, 74.0, 74.5 are included in all groups.

KTP test: KtpTest(...)

The `KtpTest` function in the **npordtests** package is used to perform the KTP test.

```
R> KtpTest(Values~Group,lehmann)

---------------------------------------------------------
Test : KTP Test
data : Values and Group

Statistic = NA
Mean = 0
Variance = 2897517
```

```
Z = NA
Asymp. p-value = NA

Error in if (p-value > alpha) { : missing value where TRUE/FALSE needed
In addition: Warning message:
In cor(t(Xmat), Ymat, method = "spearman") : the standard deviation is zero
```

In the output, similar to `FtmTest`, the error `standard deviation is zero` is encountered.

S test: SsTest(...)

The `SsTest` function in the **npordtests** package is used to perform the S test.

```
R> SsTest(Values~Group,lehmann)

---------------------------------------------------------
Test : Shan's S test
data : Values and Group

Statistic = 32234
Mean = 20865.83
Variance = 6929623
Z = 4.318527
Asymp. p-value = 7.853701e-06

Result     : Null hypothesis is rejected.
---------------------------------------------------------
```

Here, the `Statistic` is calculated value of the test statistic. `p-value` is the significance value for the S test. The `p-value` for the S test is 0.000007853701. According to these results, because the p-value is smaller than $\alpha = 0.05$, the hypothesis of the equality of locations against the ordered alternative is rejected.

Gaur's Gc test: GcTest(...)

The `GcTest` function in the **npordtests** package is used to perform the Gaur's Gc test.

```
R> GcTest(Values~Group,lehmann)

---------------------------------------------------------
Test : Gaur's Gc Test
data : Values and Group

Statistic = 0.1506891
Mean = 0
Variance = 0.03597884
Z = 0.7944348
Asymp. p-value = 0.2134712

Result : Null hypothesis is not rejected.
---------------------------------------------------------
```

Here, the `Statistic` is calculated value of the test statistic. `p-value` is the significance value for the G_c test. The `p-value` for the G_c test is 0.2134712. Because this `p-value` is not smaller than $\alpha = 0.05$, the hypothesis of the equality of locations against the ordered alternative is not rejected.

Simulation study

In this section, we compared the JT, AT, Modified JT, TM, FTM, KTP, S and Gaur's Gc tests in terms of power and Type I error under some selected scenarios. Since the AT test includes the LS, RS, ST, LT, WS tests, these tests do not need to be compared. The number of iterations and nominal type I error are 10000 and .05, respectively. The five design factors manipulated in this simulation study are:

- number of samples ($k = 3$ and 4),
- average number of observations per group ($n = 5$, 10, 20, 30, and 50),
- sample size patterns (progressive, equal, and one extreme),
- distribution shapes (symmetric, left skewed, and right skewed),
- ordered alternatives shapes (linear, convex, and concave).

The sample size patterns in this simulation study are shown in Table 1. We used log-$F(v_1, v_2)$ distributions to generate the random variable $X_{ij} = \theta_i + \varepsilon_{ij}$, where ε_{ij} is the *iid log-F* distribution, and θ_i is the location parameter; which is symmetric when $v_1 = v_2$, right skewed when $v_1 > v_2$, and left skewed when $v_1 < v_2$ (Terpstra et al., 2011).

Table 1: Simulation study sample size patterns. k is number of samples and n is average number of observations per group. The values in the table are sample sizes. For example, in case of $k = 3$, $n = 5$ and progressive pattern, the sample sizes of groups are 4, 5 and 6, respectively.

	Sample size patterns														
	Progressive					Equal					One extreme				
$k = 3$															
1	4	9	19	29	49	5	10	20	30	50	2	4	8	12	20
2	5	10	20	30	50	5	10	20	30	50	2	4	8	12	20
3	6	11	21	31	51	5	10	20	30	50	11	22	44	66	110
Average n	5	10	20	30	50	5	10	20	30	50	5	10	20	30	50
$k = 4$															
1	2	7	14	21	35	5	10	20	30	50	3	6	12	18	30
2	4	9	18	27	45	5	10	20	30	50	3	6	12	18	30
3	6	11	22	33	55	5	10	20	30	50	3	6	12	18	30
4	8	13	26	39	65	5	10	20	30	50	11	22	44	66	110
Average n	5	10	20	30	50	5	10	20	30	50	5	10	20	30	50

In order to evaluate the performances of the tests, we consider the cases of $(v_1, v_2) = (5, 5)$, $(1, 10)$ and $(10, 1)$ for the symmetric, left skewed and right skewed populations, respectively.

While the location parameters of populations are equal, simulated type I error rates are calculated. Otherwise, in case the location parameters of the populations are not equal, the simulated powers of the tests are computed. In order to assess the robustness of the tests in terms of Type I error rate, we used the robustness criterion recommended by Bradley (1978). This liberal criterion for the robustness is set at $\pm .5\alpha$ around the nominal alpha level. For instance, using the alpha level of .05, a test is considered robust when the simulated Type I error rates fall between .025 and .075.

Results

Figure 2 presents a set of boxplots based on the simulated Type I error rates for all scenarios considered while the nominal alpha level is .05. As shown in Figure 2, although all of the tests ensure the Bradley's liberal criterion, the JT, MJT, and FTM tests are the three best performing approaches that controlled nominal Type I error in all simulation scenarios. On the other hand, the TM test has a wider range than the others for the simulated type I error rates.

The simulated power values of the tests for the simulation scenarios above are given in Table 2-4. The results in these tables can be interpreted as follows:

- As seen in Table 2, when the data is generated from the symmetric distribution (log-$F(5, 5)$), the most powerful test changes according to the shape of ordered alternative. When the shape of ordered alternative is linear, the MJT test are more powerful test than the other tests for all sample size patterns. On the other hand, when the shape of ordered alternative is convex, the S test has the highest power among all tests considered for all sample size patterns. Beside these, the simulated power values of KTP test for ordered alternative with concave shape are higher than those of the other tests when sample size patterns are progressive or one extreme. But, the S test is better than the other tests in terms of power when the sample size pattern is equal. On the other hand, when the average sample size for all distributions was quite large such as 50, the simulated power values for all tests were found to be quite close to 1.

Table 2: Simulated power values $(1-\beta)$ of the test for log-F(5,5) distribution.

		Sample size pattern														
		Progressive					Equal					One Extreme				
		Average n					Average n					Average n				
k	Test	5	10	20	30	50	5	10	20	30	50	5	10	20	30	50
ordered alternatives shape=linear																
3	JT	.422	.706	.932	.979	1.00	.424	.701	.930	.980	1.00	.364	.586	.846	.945	.992
	AT	.372	.666	.922	.977	.998	.373	.661	.917	.967	1.00	.264	.549	.838	.952	.992
	MJT	.442	.709	.934	.993	1.00	.449	.722	.930	.984	1.00	.383	.606	.872	.999	1.00
	TM	.431	.665	.904	.948	.991	.434	.678	.907	.950	.994	.348	.512	.751	.888	.972
	FTM	.431	.696	.933	.978	.999	.415	.700	.929	.974	.997	.316	.542	.816	.979	.988
	KTP	.435	.704	.934	.978	.999	.436	.715	.930	.971	.996	.339	.552	.823	.900	.990
	S	.440	.702	.926	.990	1.00	.448	.699	.926	.967	1.00	.375	.595	.847	.973	.995
	Gc	.359	.674	.915	.924	.990	.388	.656	.914	.958	.990	.233	.439	.757	.841	.902
4	JT	.672	.960	.999	1.00	1.00	.766	.963	.999	1.00	1.00	.769	.962	.999	1.00	1.00
	AT	.616	.939	.999	1.00	1.00	.708	.952	.999	1.00	1.00	.700	.953	.999	1.00	1.00
	MJT	.678	.961	.999	1.00	1.00	.784	.978	.999	1.00	1.00	.782	.967	.999	1.00	1.00
	TM	.597	.894	.993	.999	1.00	.684	.919	.999	1.00	1.00	.586	.815	.999	1.00	1.00
	FTM	.621	.955	.999	1.00	1.00	.778	.972	.999	1.00	1.00	.719	.924	.999	1.00	1.00
	KTP	.592	.958	.999	1.00	1.00	.775	.965	.999	1.00	1.00	.731	.951	.999	1.00	1.00
	S	.667	.951	.999	1.00	1.00	.763	.961	.999	1.00	1.00	.768	.961	.999	1.00	1.00
	Gc	.485	.916	.999	1.00	1.00	.630	.948	.999	1.00	1.00	.455	.869	.992	.999	1.00
ordered alternatives shape=convex																
3	JT	.477	.716	.937	.988	1.00	.405	.682	.920	.977	.999	.486	.756	.955	.985	.999
	AT	.387	.674	.925	.981	1.00	.365	.662	.912	.980	.999	.311	.659	.920	.977	.998
	MJT	.476	.715	.936	.990	1.00	.411	.705	.926	.990	.999	.448	.708	.932	.984	.999
	TM	.388	.577	.827	.972	1.00	.395	.599	.813	.967	.984	.332	.448	.661	.969	.988
	FTM	.391	.672	.922	.980	1.00	.392	.679	.907	.978	.998	.284	.509	.789	.970	.995
	KTP	.421	.686	.921	.979	1.00	.410	.698	.930	.978	.992	.316	.530	.813	.972	.990
	S	.510	.742	.943	.992	1.00	.439	.716	.932	.992	1.00	.519	.765	.958	.988	.999
	Gc	.348	.648	.903	.974	.999	.371	.644	.905	.975	.992	.189	.440	.734	.945	.989
4	JT	.586	.771	.966	.999	1.00	.389	.623	.887	.996	1.00	.655	.894	.994	1.00	1.00
	AT	.468	.702	.938	.997	1.00	.360	.627	.887	.996	1.00	.527	.829	.980	1.00	1.00
	MJT	.525	.732	.950	.999	1.00	.399	.635	.887	.997	1.00	.585	.847	.986	1.00	1.00
	TM	.328	.476	.871	.989	1.00	.329	.473	.812	.990	1.00	.284	.421	.798	.985	1.00
	FTM	.286	.602	.940	.999	1.00	.380	.620	.893	.997	1.00	.333	.569	.914	1.00	1.00
	KTP	.279	.620	.941	.999	1.00	.384	.638	.880	.998	1.00	.320	.594	.838	.990	1.00
	S	.595	.780	.971	1.00	1.00	.401	.642	.899	.999	1.00	.659	.899	.995	1.00	1.00
	Gc	.222	.519	.827	.980	1.00	.309	.597	.833	.977	1.00	.219	.487	.765	.952	1.00
ordered alternatives shape=concave																
3	JT	.341	.633	.914	.984	1.00	.386	.688	.922	.992	1.00	.227	.363	.591	.704	.812
	AT	.305	.612	.897	.978	.998	.350	.642	.905	.987	1.00	.200	.418	.694	.816	.929
	MJT	.384	.661	.923	.993	1.00	.421	.689	.926	.994	1.00	.256	.453	.729	.837	.943
	TM	.381	.605	.839	.919	.952	.375	.595	.850	.919	1.00	.270	.426	.691	.801	.942
	FTM	.399	.678	.930	.998	1.00	.380	.679	.919	.999	1.00	.269	.515	.805	.925	.999
	KTP	.422	.683	.931	.999	1.00	.413	.690	.927	.999	1.00	.302	.538	.819	.931	.999
	S	.371	.663	.920	.990	1.00	.453	.714	.933	1.00	1.00	.245	.392	.639	.738	.846
	Gc	.361	.633	.904	.974	.995	.378	.626	.912	.980	1.00	.234	.433	.755	.863	.983
4	JT	.154	.449	.704	.782	.815	.376	.630	.893	.784	.816	.252	.401	.652	.583	.715
	AT	.158	.468	.755	.835	.862	.342	.604	.887	.842	.862	.250	.459	.757	.752	.872
	MJT	.176	.511	.777	.858	.878	.391	.643	.895	.860	.878	.300	.484	.766	.771	.889
	TM	.290	.484	.720	.810	.830	.332	.468	.821	.811	.842	.275	.405	.744	.774	.905
	FTM	.278	.611	.814	.884	.904	.376	.636	.877	.883	.934	.339	.575	.842	.888	.998
	KTP	.296	.619	.882	.952	.992	.388	.639	.902	.950	1.00	.359	.577	.860	.948	.999
	S	.145	.452	.719	.789	.809	.395	.658	.903	.961	1.00	.247	.417	.666	.614	.724
	Gc	.241	.535	.840	.912	.942	.363	.589	.838	.921	.962	.245	.496	.767	.890	.999

Table 3: Simulated power values $(1-\beta)$ of the test for log-F(1,10) distribution.

		Sample size pattern														
		Progressive					Equal					One Extreme				
		Average n					Average n					Average n				
k	Test	5	10	20	30	50	5	10	20	30	50	5	10	20	30	50
ordered alternatives shape=linear																
3	JT	.179	.309	.495	.718	.941	.174	.296	.492	.688	.884	.145	.230	.383	.537	.694
	AT	.181	.318	.544	.815	1.00	.203	.316	.535	.765	.995	.146	.230	.411	.592	.813
	MJT	.198	.307	.496	.723	.950	.187	.304	.494	.684	.882	.145	.243	.398	.553	.708
	TM	.214	.319	.474	.684	.894	.217	.322	.457	.592	.727	.177	.250	.366	.482	.598
	FTM	.187	.313	.484	.689	.894	.185	.295	.475	.655	.835	.137	.209	.369	.529	.689
	KTP	.193	.303	.499	.734	.969	.197	.314	.500	.686	.872	.134	.220	.371	.522	.673
	S	.188	.303	.481	.695	.908	.199	.297	.483	.669	.855	.146	.225	.367	.509	.651
	Gc	.156	.284	.451	.651	.852	.178	.267	.467	.667	.867	.112	.188	.324	.460	.596
4	JT	.293	.538	.801	1.00	1.00	.333	.549	.816	1.00	1.00	.347	.549	.828	1.00	1.00
	AT	.282	.578	.851	1.00	1.00	.353	.599	.863	1.00	1.00	.312	.596	.875	1.00	1.00
	MJT	.285	.545	.802	1.00	1.00	.346	.560	.816	1.00	1.00	.328	.569	.836	1.00	1.00
	TM	.302	.473	.761	1.00	1.00	.328	.491	.742	.993	1.00	.284	.444	.661	.878	1.00
	FTM	.231	.533	.772	1.00	1.00	.335	.562	.720	.878	1.00	.305	.522	.669	.816	.963
	KTP	.247	.530	.770	1.00	1.00	.359	.553	.815	1.00	1.00	.302	.517	.801	1.00	1.00
	S	.267	.510	.769	1.00	1.00	.338	.534	.801	1.00	1.00	.327	.521	.774	1.00	1.00
	Gc	.173	.446	.710	.994	1.00	.298	.493	.749	1.00	1.00	.204	.424	.670	.916	1.00
ordered alternatives shape=convex																
3	JT	.212	.323	.500	.697	.910	.181	.309	.478	.649	.826	.208	.303	.540	.779	1.00
	AT	.227	.352	.566	.794	1.00	.220	.342	.559	.790	1.00	.215	.316	.543	.782	1.00
	MJT	.209	.317	.499	.687	.893	.190	.306	.489	.674	.877	.180	.292	.493	.704	.927
	TM	.220	.330	.502	.676	.866	.219	.330	.506	.688	.890	.199	.265	.420	.577	.738
	FTM	.191	.295	.485	.685	.899	.176	.298	.495	.702	.927	.144	.220	.392	.570	.766
	KTP	.187	.299	.481	.673	.881	.190	.316	.486	.668	.858	.142	.213	.364	.521	.688
	S	.217	.324	.501	.680	.863	.204	.305	.484	.681	.898	.206	.313	.490	.687	.902
	Gc	.164	.288	.442	.602	.768	.186	.280	.462	.660	.874	.117	.192	.322	.462	.620
4	JT	.264	.349	.555	.769	.989	.186	.283	.450	.631	.824	.275	.440	.701	.978	1.00
	AT	.244	.371	.614	.861	1.00	.209	.324	.521	.738	.957	.244	.442	.705	.980	1.00
	MJT	.223	.331	.521	.713	.911	.184	.283	.447	.629	.827	.237	.389	.641	.897	1.00
	TM	.224	.287	.560	.839	1.00	.203	.282	.468	.656	.860	.191	.258	.567	.894	1.00
	FTM	.140	.257	.546	.841	1.00	.180	.277	.457	.647	.851	.156	.258	.544	.848	1.00
	KTP	.142	.285	.509	.749	.993	.193	.300	.468	.644	.824	.160	.240	.412	.592	.776
	S	.251	.335	.538	.753	.984	.182	.275	.439	.609	.795	.259	.407	.658	.919	1.00
	Gc	.127	.226	.372	.522	.686	.167	.289	.369	.457	.551	.116	.209	.377	.559	.747
ordered alternatives shape=concave																
3	JT	.150	.274	.467	.670	.891	.165	.285	.481	.679	.887	.118	.163	.245	.331	.429
	AT	.140	.270	.485	.716	.955	.179	.294	.511	.736	.977	.135	.149	.284	.437	.596
	MJT	.172	.278	.470	.678	.904	.178	.294	.485	.696	.909	.119	.182	.296	.422	.552
	TM	.181	.256	.375	.508	.655	.194	.256	.386	.526	.678	.154	.190	.293	.404	.525
	FTM	.166	.285	.476	.679	.896	.176	.288	.492	.712	.948	.119	.203	.360	.527	.710
	KTP	.175	.288	.490	.722	.966	.179	.315	.516	.743	.982	.118	.212	.369	.542	.725
	S	.162	.268	.461	.666	.889	.190	.300	.484	.688	.900	.127	.157	.245	.343	.457
	Gc	.163	.274	.436	.610	.796	.141	.242	.406	.572	.752	.130	.167	.294	.433	.582
4	JT	.093	.195	.305	.433	.567	.167	.269	.445	.641	.839	.123	.181	.286	.405	.528
	AT	.082	.205	.332	.479	.636	.160	.271	.459	.649	.843	.105	.190	.325	.462	.617
	MJT	.101	.208	.350	.504	.668	.172	.277	.448	.633	.830	.126	.201	.337	.485	.647
	TM	.166	.197	.288	.389	.508	.175	.181	.351	.529	.709	.153	.177	.330	.495	.668
	FTM	.119	.247	.356	.473	.598	.162	.248	.458	.686	.930	.139	.219	.391	.583	.791
	KTP	.117	.279	.452	.639	.836	.169	.298	.462	.696	.932	.141	.225	.396	.585	.794
	S	.087	.178	.294	.428	.576	.168	.255	.433	.617	.807	.125	.173	.266	.371	.490
	Gc	.114	.203	.333	.473	.619	.124	.260	.381	.514	.659	.107	.216	.311	.426	.547

Table 4: Simulated power values $(1-\beta)$ of the test for log-F(10,1) distribution.

		Sample size pattern														
		Progressive					Equal					One Extreme				
		Average n					Average n					Average n				
k	Test	5	10	20	30	50	5	10	20	30	50	5	10	20	30	50
ordered alternatives shape=linear																
3	JT	.190	.310	.491	.682	.893	.179	.298	.491	.690	.909	.180	.256	.404	.566	.742
	AT	.182	.315	.526	.747	.976	.203	.323	.527	.741	.967	.204	.302	.473	.654	.847
	MJT	.208	.304	.490	.690	.896	.191	.318	.491	.676	.863	.176	.266	.415	.574	.753
	TM	.215	.293	.465	.645	.845	.221	.298	.438	.590	.750	.213	.251	.350	.455	.564
	FTM	.187	.280	.482	.700	.924	.186	.290	.463	.642	.827	.158	.242	.390	.540	.698
	KTP	.184	.304	.484	.682	.888	.185	.305	.493	.689	.897	.178	.249	.393	.553	.721
	S	.199	.305	.477	.663	.863	.195	.304	.487	.688	.907	.208	.291	.433	.593	.773
	Gc	.153	.279	.454	.649	.856	.168	.263	.458	.655	.862	.127	.195	.324	.461	.614
4	JT	.305	.537	.796	1.00	1.00	.340	.554	.819	1.00	1.00	.350	.555	.806	1.00	1.00
	AT	.319	.570	.834	1.00	1.00	.344	.581	.851	1.00	1.00	.379	.594	.848	1.00	1.00
	MJT	.304	.544	.801	1.00	1.00	.342	.550	.819	1.00	1.00	.360	.558	.816	1.00	1.00
	TM	.313	.485	.800	1.00	1.00	.319	.478	.812	1.00	1.00	.297	.428	.788	1.00	1.00
	FTM	.262	.535	.812	1.00	1.00	.337	.557	.822	1.00	1.00	.323	.510	.797	1.00	1.00
	KTP	.269	.505	.785	1.00	1.00	.339	.566	.835	1.00	1.00	.334	.511	.782	1.00	1.00
	S	.295	.518	.804	1.00	1.00	.332	.532	.805	1.00	1.00	.364	.576	.823	1.00	1.00
	Gc	.193	.442	.697	.960	1.00	.266	.533	.760	.999	1.00	.207	.430	.670	.916	1.00
ordered alternatives shape=convex																
3	JT	.206	.310	.505	.714	.933	.167	.286	.474	.670	.880	.232	.337	.533	.741	.963
	AT	.176	.307	.510	.757	1.00	.175	.289	.501	.725	.953	.228	.349	.530	.729	.948
	MJT	.210	.310	.497	.696	.913	.177	.302	.487	.690	.895	.214	.320	.482	.646	.816
	TM	.182	.235	.406	.587	.788	.186	.271	.394	.527	.678	.193	.216	.300	.396	.510
	FTM	.174	.253	.503	.723	.956	.167	.304	.499	.708	.935	.162	.228	.354	.492	.646
	KTP	.178	.297	.481	.673	.873	.174	.303	.484	.683	.890	.175	.245	.379	.533	.695
	S	.219	.320	.501	.692	.897	.189	.311	.484	.661	.852	.261	.381	.573	.767	.973
	Gc	.135	.263	.443	.629	.821	.154	.253	.446	.657	.882	.113	.187	.315	.463	.621
4	JT	.254	.344	.566	.806	1.00	.161	.260	.447	.648	.855	.284	.455	.697	.941	1.00
	AT	.227	.326	.546	.782	1.00	.156	.262	.451	.642	.837	.263	.425	.670	.929	1.00
	MJT	.221	.317	.522	.747	.992	.171	.276	.452	.662	.882	.247	.397	.626	.857	1.00
	TM	.198	.208	.540	.892	1.00	.164	.211	.365	.521	.679	.158	.202	.324	.464	.620
	FTM	.149	.280	.556	.834	1.00	.172	.272	.444	.636	.842	.169	.259	.432	.617	.818
	KTP	.155	.250	.515	.800	1.00	.195	.268	.450	.652	.870	.167	.279	.462	.649	.850
	S	.257	.354	.572	.898	1.00	.162	.257	.437	.629	.841	.304	.479	.728	.995	1.00
	Gc	.109	.213	.331	.461	.601	.145	.242	.383	.526	.675	.113	.226	.332	.452	.580
ordered alternatives shape=concave																
3	JT	.165	.286	.467	.654	.857	.179	.299	.485	.675	.877	.139	.178	.265	.354	.445
	AT	.179	.316	.505	.696	.903	.216	.331	.524	.737	.952	.176	.249	.375	.521	.673
	MJT	.192	.286	.475	.670	.883	.184	.306	.480	.670	.862	.153	.218	.329	.458	.605
	TM	.226	.341	.508	.700	.916	.223	.339	.528	.741	.960	.187	.265	.398	.593	.782
	FTM	.196	.306	.474	.662	.868	.180	.300	.498	.710	.930	.160	.211	.385	.573	.771
	KTP	.194	.313	.486	.663	.842	.184	.310	.471	.644	.825	.183	.250	.393	.548	.707
	S	.177	.290	.477	.674	.877	.204	.306	.486	.682	.894	.157	.208	.304	.412	.532
	Gc	.150	.288	.453	.638	.831	.162	.272	.420	.588	.760	.152	.181	.353	.541	.731
4	JT	.104	.213	.324	.447	.576	.180	.280	.456	.638	.822	.138	.191	.295	.405	.535
	AT	.131	.257	.416	.589	.778	.188	.318	.480	.644	.810	.167	.263	.406	.561	.720
	MJT	.117	.232	.365	.518	.679	.179	.281	.455	.637	.839	.156	.232	.356	.500	.662
	TM	.208	.280	.433	.602	.779	.210	.332	.491	.652	.855	.186	.287	.441	.613	.795
	FTM	.160	.276	.425	.576	.737	.181	.271	.445	.635	.841	.171	.269	.410	.561	.724
	KTP	.160	.271	.426	.585	.746	.192	.301	.466	.641	.826	.172	.283	.432	.601	.778
	S	.103	.207	.322	.455	.600	.184	.273	.440	.609	.798	.140	.200	.302	.424	.554
	Gc	.144	.246	.345	.446	.557	.174	.289	.398	.523	.654	.141	.213	.329	.455	.583

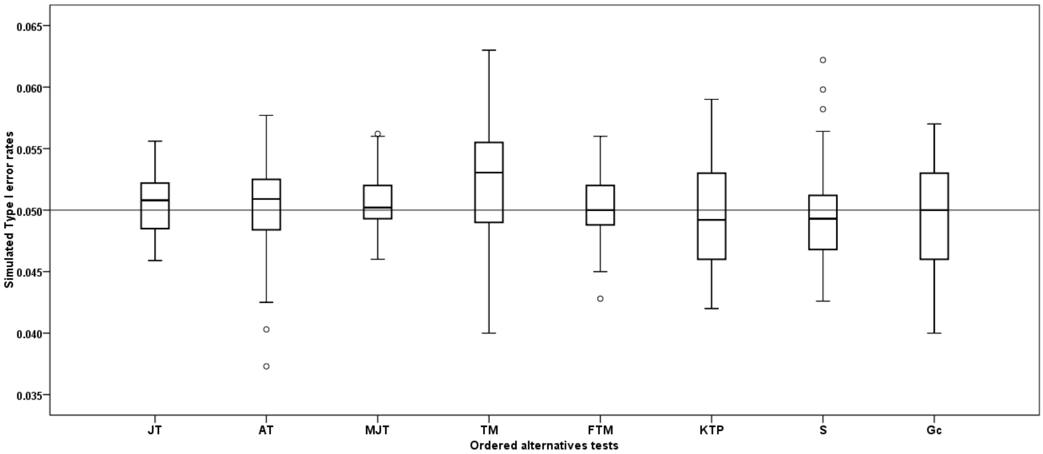

Figure 2: Distributions of simulated Type I error rates across all simulation scenarios when nominal alpha is .05. Each box plot gives median (the bold line that divides the box into two parts), lower and upper quartiles (start and end points of the box on the vertical axis) and min and max value (the horizontal lines outside the box). The outliers appear as the circles.

- For the data generated from the log-$F(1,10)$ distribution which is a skewed to the left, when the shape of ordered alternative is linear, and average sample size is 5 or 10, the TM test for $k = 3$ gives better results, however, the AT test has the highest powers among the whole tests when average sample size is 20, 30, and 50. On the other hand, when $k = 4$ and average sample size is 10, the AT test has the highest powers among the whole tests. For the data generated from this distribution, the AT test, generally, is the most powerful test for ordered alternative with convex shape as seen in Table 3. For the data generated from the log-$F(1,10)$ distribution which is a skewed to the left, the TM test has the highest powers among the whole tests when the shape of ordered alternative is concave and average sample size is 5, but the KTP test for ordered alternative with concave shape is the most powerful among the whole tests when average sample size is 10, 20, 30, and 50. On the other hand, when the average sample size increased in all scenarios considered, the power values of all tests increased as expected.
- When the data is generated from the log-$F(10,1)$ distribution which is a skewed to the right, Table 4 shows that the AT test for ordered alternatives with linear shape, generally, gives better results. On the other hand, when average sample size is 5 and $k = 3$ the TM test for this situation is the most powerful test. As seen in Table 4, when the shape of ordered alternative is a convex, it is observed that the S test generally yields the highest power values. In addition, while the sample size patterns are progressive and equal, and average sample size is 20, 30, and 50, the power values of the AT test for this situation are greater than those of the others. By the examination of the results in Table 6, when ordered alternative has a concave shape, it is seen that the TM test is the most powerful test among the whole tests.

Table 5 gives decision rules indicating which test is more appropriate for which design.

When the ordered alternative has a linear shape and the distribution is symmetric, the MJT test should be preferred. However, when the ordered alternative has a linear shape and the distribution is skewed to left and average sample size is 5 or 10, it can be stated that the TM test has a more significant power advantage than the others. On the other hand, average sample size is 20, 30, or 50, it can be said that the AT test has a more significant power advantage than the others.

On the other hand, when the ordered alternative has a convex shape, the AT test is recommended for the distributions skewed to left. However, if these distributions are symmetric, the S test is proposed. Besides this, if the distributions are skewed to right and the sample size pattern is equal, then the MJT test is recommended. Further, if the distributions are skewed to right and the sample size pattern is progressive or one extreme, then S test is used.

When the ordered alternative has a concave shape and the sample size pattern is equal, then the S test is used for symmetric distribution. In addition, when the ordered alternative has a concave shape and the sample size pattern is progressive or one extreme, then the KTP test is recommended for symmetric distribution. Moreover, if the distributions are skewed to left and the sample size is 5, TM test is recommended, but in the case of 10, 20, 30, 50 for the sample size, the KTP test is recommended. Finally, if the distributions are skewed to right, the TM test is recommended.

Table 5: The rules based on the simulation results for choice the test. For example, when the ordered alternative has a linear shape and the distribution is symmetric, the MJT test should be preferred.

Alternative hypothesis	Distribution shape	Sample size pattern	Average sample size	Test
Linear	symmetric	-	-	MJT
	skewed to left	-	5, 10	TM
	skewed to left	-	20, 30, 50	AT
	skewed to right	-	-	AT
Convex	symmetric	-	-	S
	skewed to left	-	-	AT
	skewed to right	-	-	S
Concave	symmetric	Equal	-	S
	symmetric	Progressive or One Extreme	-	KTP
	skewed to left	-	5	TM
	skewed to left	-	10, 20, 30, 50	KTP
	skewed to right	-		TM

Summary

Tests for ordered alternative are the most frequently used nonparametric methods in a wide range of statistical and medical applications. For example, the evaluation of preclinical studies, clinical dose-finding trials, typical toxicity studies, education studies, agricultural studies and etc. We present the **npordtests** package to test the equality hypothesis of the locations against ordered alternative.

In this paper, we compared the tests included in the **npordtests** package in terms of Type I error rate and power. With the results of the simulation study, when the data is generated from a symmetric distribution, we propose that the use of the MJT test for ordered alternatives with linear shape and the S test for ordered alternatives with convex shape. On the other hand, when ordered alternative has a concave shape, the S test for equal sample size patterns is suggested, but the KTP test is recommended when sample size pattens are progressive and one extreme. For the data generated from a left skewed distribution, when $k = 3$ and shape of ordered alternative is linear, we recommend that the use of the TM test for small sample sizes such as $n = 5$ and 10, and the AT test for sample size 20, 30, and 50. However, when $k = 4$ and sample sizes are 10, 20, 30, and 50, we propose to prefer the AT test. For this kind of data, we propose the use of the AT test when the ordered alternative has a convex shape. On the other hand, if ordered altenative has a concave shape, we propose that the use of the KTP test for sample sizes such as $n = 10, 20, 30$, and 50 and the TM test for small sample size such as $n = 5$. For the data generated from a right skewed distribution, when $k = 4$, we recommend that the use of the AT test for ordered alternative with linear shape. However, when $k = 3$, and the shape of ordered alternative is linear, we propose to choose the AT test for sample sizes $n = 10, 20, 30$, and 50 and the TM test for sample size 5. On the other hand, when ordered alternative has a concave shape, the TM test is the most powerful test in all simulation scenarios. Besides these, for this kind of data, it is understood that it is appropriate to prefer the S test for ordered alternative with convex shape.

To test the equality hypothesis of locations parameters against ordered alternatives, the **npordtests** package covers the prominent nonparametric tests such as Jonckheere-Terpstra test, Beier and Buning's Adaptive test, Modified Jonckheere-Terpstra test, Terpstra-Magel test, Ferdhiana-Terpstra-Magel test, KTP test, S test and Gaur's G_c test. According to the authors knowledge, the tests which are present in the **npordtests** package, except the JT test, are not available in any other R tool. The package will be updated at regular intervals.

Acknowledgments

The authors are genuinely grateful to anonymous reviewer and the Executive Editor (Dianne Cook) for their invaluable contributions to the improvement of our paper.

Bibliography

B. Altunkaynak and H. Gamgam. *npordtests: An R Package for Nonparametric Tests for Equality of Location Against Ordered Alternatives*, 2019. URL https://CRAN.R-project.org/package=npordtests. R package version 1.2. [p148]

F. Beier and H. Buning. An adaptive test against ordered alternatives. *Computational Statistics and Data Analysis*, 25(4):441–452, 1997. URL https://doi.org/10.1016/S0167-9473(97)00014-5. [p147, 149, 150, 153]

J. V. Bradley. Robustness? *British Journal of Mathematical and Statistical Psychology*, 31:144–152, 1978. URL https://doi.org/10.1111/j.2044-8317.1978.tb00581.x. [p163]

M. A. Bredella, L. S. Steinbach, S. Morgan, M. Ward, and J. C. Davis. Mri of the sacroiliac joints in patients with moderate to severe ankylosing spondylitis. *American Journal of Roentgenology*, 187(6): 1420–1426, 2006. URL https://doi.org/10.2214/AJR.05.1423. [p147]

H. Buning and W. Kossler. Robustness and efficiency of some tests for ordered alternatives in the c-sample location problem. *Journal of Statistical Computation and Simulation*, 55(4):337–352, 1996. URL https://doi.org/10.1080/00949659608811774. [p147, 149]

C. H. Chang and C. H. Yen. A nonparametric test for the ordered alternative based on fast discrete fourier transform coefficient. *Journal of Testing and Evaluation*, 39(6):1131–1143, 2011. URL https://doi.org/10.1520/Jte103464. [p154]

Q. Cheng, W. Yang, S. C. Raimondi, C.-H. Pui, M. V. Relling, and W. E. Evans. Karyotypic abnormalities create discordance of germline genotype and cancer cell phenotypes. *Nature Genetics*, 37(8):878–882, 2005. URL https://doi.org/10.1038/ng1612. [p147]

J. Cuzick. A Wilcoxon-Type test for trend. *Statistics in Medicine*, 4(1):87–90, 1985. URL https://doi.org/10.1002/sim.4780040112. [p147]

M. Desu and D. Raghavarao. *Nonparametric Statistical Methods for Complete and Censored Data*. Chapman and Hall/CRC, New York, 1st edition, 2004. [p154]

A. Dmitrienko, C. Chuang-Stein, and R. D'Agostino. *Pharmaceutical Statistics using SAS®: A practical guide (SAS Press)*. SAS Institute, NC, 1st edition, 2006. ISBN 159047886X. [p154]

R. Ferdhiana, J. Terpstra, and R. C. Magel. A nonparametric test for the ordered alternative based on Kendall's Correlation Coefficient. *Communications in Statistics-Simulation and Computation*, 37(6): 1117–1128, 2008. URL https://doi.org/10.1080/03610910801894870. [p147, 151]

J. L. Gastwirth. Asymptotically most powerful rank tests for the two-sample problem with censored data. *Ann. Math. Statist.*, 36(4):1243–1247, 1965. URL https://doi.org/10.1214/aoms/1177699995. [p147, 149, 150]

A. Gaur. A new class of distribution-free tests for testing ordered location parameters based on sub-samples. *Statistics and Probability Letters*, 90(1):53–59, 2014. URL https://doi.org/10.1016/j.spl.2014.03.011. [p147]

A. Gaur. A class of k-sample distribution-free tests for location against ordered alternatives. *Communications in Statistics-Theory and Methods*, 46(5):2343–2353, 2017. URL https://doi.org/10.1080/03610926.2015.1041986. [p147, 152]

Z. Govindarajulu and H. S. Haller. c-sample tests of homogeneity against ordered alternatives. In J. S. Rustagi, editor, *Optimizing Methods in Statistics*, page 479. Academic Press, 1971. ISBN 978-0-12-604550-5. URL https://doi.org/10.1016/B978-0-12-604550-5.50030-4. [p147]

S. Hoffmeyer, O. Burk, O. von Richter, H. P. Arnold, J. Brockmöller, A. Johne, I. Cascorbi, T. Gerloff, I. Roots, M. Eichelbaum, and U. Brinkmann. Functional polymorphisms of the human multidrug resistance gene: Multiple sequence variations and correlation of one allele with p glycoprotein expression and activity in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 97(7):3473–3478, 2000. URL https://doi.org/10.1073/pnas.050585397. [p147]

J. W. Hogg, Robert V. McKean and A. T. Craig. *Introduction to Mathematical Statistics*. Pearson, Boston, 7 edition, 2013. ISBN 9780321795434. [p149]

R. V. Hogg, D. M. Fisher, and R. H. Randles. A two-sample adaptive distribution-free test. *Journal of the American Statistical Association*, 70(351):656–661, 1975. URL https://doi.org/10.2307/2285950. [p147, 150]

A. R. Jonckheere. A distribution-free k-sample test against ordered alternatives. *Biometrika*, 41(1): 133–145, 1954. URL https://doi.org/10.2307/2333011. [p147, 148, 153]

T. Kawaguchi, Y. Sumida, A. Umemura, K. Matsuo, M. Takahashi, T. Takamura, K. Yasui, T. Saibara, E. Hashimoto, M. Kawanaka, S. Watanabe, S. Kawata, Y. Imai, M. Kokubo, T. Shima, H. Park, H. Tanaka, K. Tajima, R. Yamada, F. Matsuda, T. Okanoue, and D. Japan Study Group of Nonalcoholic Fatty Liver. Genetic polymorphisms of the human PNPLA3 gene are strongly associated with severity of non-alcoholic fatty liver disease in Japanese. *PloS one*, 7(6):e38322–e38322, 2012. URL https://doi.org/10.1371/journal.pone.0038322. [p147]

C. Le. A new rank test against ordered-alternatives in k-sample problems. *Biometrical Journal*, 30(1): 87–92, 1988. URL https://doi.org/10.1002/bimj.4710300116. [p147]

E. Lehmann. *Nonparametrics: Statistical Methods based on Ranks*. Holden-Day, San Francisco, 1st edition, 1975. ISBN 9780387352121. [p153]

J. Lin, A. Sibley, I. Shterev, and K. Owzar. *jtGWAS: Efficient Jonckheere-Terpstra Test Statistics*, 2017a. URL https://CRAN.R-project.org/package=jtGWAS. R package version 1.5.1. [p147]

J. Lin, A. Sibley, I. Shterev, and K. Owzar. *fastJT: Efficient Jonckheere-Terpstra Test Statistics for Robust Machine Learning and Genome-Wide Association Studies*, 2017b. URL https://CRAN.R-project.org/package=fastJT. R package version 1.0.4. [p147, 148]

M. Neuhäuser, P.-Y. Liu, and L. A. Hothorn. Nonparametric tests for trend: Jonckheere's Test, a modification and a maximum test. *Biometrical Journal*, 40(8):899–909, 1998. URL https://doi.org/10.1002/(SICI)1521-4036(199812)40:8<899::AID-BIMJ899>3.0.CO;2-9. [p147, 150, 154]

J. P. Ong, A. Aggarwal, D. Krieger, K. A. Easley, M. T. Karafa, F. Van Lente, A. C. Arroliga, and K. D. Mullen. Correlation between ammonia levels and the severity of hepatic encephalopathy. *The American Journal of Medicine*, 114(3):188–193, 2003. URL https://doi.org/10.1016/S0002-9343(02)01477-8. [p147]

J. F. Pedersen, L. G. Madsen, V. A. Larsen, O. Hamberg, T. Horn, B. Federspiel, and P. Bytzer. A doppler waveform index to characterize hepatic vein velocity pattern and evaluate hepatic fibrosis. *Journal of Clinical Ultrasound*, 36(4):208–211, 2008. URL https://doi.org/10.1002/jcu.20446. [p154]

T. Pohlert. *PMCMRplus: Calculate Pairwise Multiple Comparisons of Mean Rank Sums Extended*, 2018. URL https://CRAN.R-project.org/package=PMCMRplus. R package version 1.4.1. [p147]

M. L. Puri. Some distribution-free k-sample rank tests of homogeneity against ordered alternatives. *Communications on Pure and Applied Mathematics*, 18(1):51–63, 1965. URL https://doi.org/10.1002/cpa.3160180108. [p147]

F. Scholz and A. Zhu. *kSamples: K-Sample Rank Tests and Their Combinations*, 2018. URL https://CRAN.R-project.org/package=kSamples. R package version 1.2-8. [p147]

G. G. Shan, D. Young, and L. Kang. A new powerful nonparametric rank test for ordered alternative problem. *Plos One*, 9(11):1–10, 2014. URL https://doi.org/10.1371/journal.pone.0112924. [p147, 152, 154]

J. Swintek, K. Flynn, and J. Haselman. *StatCharrms: Statistical Analysis of Chemistry, Histopathology, and Reproduction Endpoints Including Repeated Measures and Multi-Generation Studies*, 2018. URL https://CRAN.R-project.org/package=StatCharrms. R package version 0.90.91. [p147]

H.-L. Tan, S. M. Zain, R. Mohamed, S. Rampal, K.-F. Chin, R. C. Basu, P.-L. Cheah, S. Mahadeva, and Z. Mohamed. Association of glucokinase regulatory gene polymorphisms with risk and severity of non-alcoholic fatty liver disease: An interaction study with adiponutrin gene. *Journal of Gastroenterology*, 49(6):1056–1064, 2014. URL https://doi.org/10.1007/s00535-013-0850-x. [p147]

J. T. Terpstra. The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae*, 14(3):327–333, 1952. [p147, 148]

J. T. Terpstra and R. C. Magel. A new nonparametric test for the ordered alternative problem. *Journal of Nonparametric Statistics*, 15(3):289–301, 2003. URL https://doi.org/10.1080/1048525031000078349. [p147, 150]

J. T. Terpstra, C. H. Chang, and R. C. Magel. On the use of Spearman's Correlation Coefficient for testing ordered alternatives. *Journal of Statistical Computation and Simulation*, 81(11):1381–1392, 2011. URL https://doi.org/10.1080/00949655.2010.485316. [p147, 148, 151, 154, 163]

P. Tryon and T. Hettmansperger. A class of nonparametric tests for homogeneity against ordered alternatives. *Annals of Statistics*, 1:1061–1070, 1973. [p147, 150]

T. Uchiyama, H. Kanno, K. Ishitani, H. Fujii, H. Ohta, H. Matsui, N. Kamatani, and K. Saito. An snp in CYP39A1 is associated with severe neutropenia induced by docetaxel. *Cancer Chemotherapy and Pharmacology*, 69(6):1617–1624, 2012. URL https://doi.org/10.1007/s00280-012-1872-4. [p147]

E. Venkatraman. *clinfun: Clinical Trial Design and Data Analysis Functions*, 2018. URL https://CRAN.R-project.org/package=clinfun. R package version 1.0.15. [p147]

K. Yorifuji, Y. Uemura, S. Horibata, G. Tsuji, Y. Suzuki, K. Miyagawa, K. Nakayama, K.-i. Hirata, S. Kumagai, and N. Emoto. CHST3 and CHST13 polymorphisms as predictors of bosentan-induced liver toxicity in Japanese patients with pulmonary arterial hypertension. *Pharmacological Research*, 135:259–264, 2018. URL https://doi.org/10.1016/j.phrs.2018.08.011. [p147]

Bulent Altunkaynak
Department of Statistics
Gazi University Faculty of Science, Yenimahalle, Ankara
Turkey
bulenta@gazi.edu.tr

Hamza Gamgam
Department of Statistics
Gazi University Faculty of Science, Yenimahalle, Ankara
Turkey
gamgam@gazi.edu.tr