Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 48 (5) (2019), 1547 – 1559
DOI : 10.15672/hujms.481380

Research Article

A random decision for testing of the homogeneity of normal means against the tree order alternative

Reza Momeni*, Javad Etminan and Mohammad Khanjari Sadegh

Department of Statistics, University of Birjand, Birjand, Iran

Abstract

A typical problem of interest is to compare the $k+1$ normal means under the tree order restriction $\theta_0 \leq \theta_i$ for $i = 1, \ldots, k$. In this paper, we propose new multiple comparisons procedures for testing of the tree order constraint. New test procedures along with the corresponding simultaneous confidence intervals are motivated by some new estimation methods which are constructed based on a random decision and the Bayesian approach. Also, these procedures are developed for two-sided tree order alternatives. We compare the performance of the proposed methods with some existing test procedures, such as likelihood ratio test and some multiple comparisons tests for the tree order constraint. In some cases, the gains in power due to the proposed procedures are substantial. The results for two sided alternative are similar to the one-sided hypotheses and new procedures perform well for almost every configuration. We illustrate the efficiency of the proposed methods by analyzing of the two bioassay numerical examples.

Mathematics Subject Classification (2010). 62F30, 62H15.

Keywords. Dunnett-type test, likelihood ratio test, restricted maximum likelihood estimator, tree order constraint

1. Introduction

In many problems researchers are interested in testing equality of $k+1$ normal means constrained with the tree order restriction $\theta_0 \leq \theta_i$ for $i = 1, \ldots, k$. This restriction arises naturally in experimental situations in which one wishes to compare several treatments with a control when it is believed a priori that all of the treatments are as effective as the control. Inclusion of this prior information about the mean parameters leads to more powerful tests than the usual unrestricted tests which do not take the tree ordering of the means into account. Some authors [1, 16] demonstrated that these order restricted tests can be substantially more powerful than their usual unconstrained tests. The most well known approach is the likelihood ratio test (LRT) method. Bartholomew [2] has developed the LRT for testing equality of ordered normal means. But, LRT is not widely used in practical applications. One crucial drawback lies in the difficulty of evaluating the null distribution. Because of such practical problems in implementing the LRT, researchers

*Corresponding Author.
Email addresses: rezamomeni@birjand.ac.ir (R. Momeni), jetminan@birjand.ac.ir (J. Etminan), mkhanjari@birjand.ac.ir (M. K. Sadegh)
Received: 11.11.2018; Accepted: 12.03.2019

have tried to develop alternative testing procedures. An influential approach is Dunnett's type-test. It is commonly used for testing the null hypothesis $H_0 : \theta_0 = \theta_1 = \ldots = \theta_k$ against $H_1 : \theta_0 \leq \theta_i$ for all $i \geq 1$ and for at least one i the inequality is strict. It has been widely used in dose response studies. When the populations have a common known variance, Dunnett's test provide the simultaneous confidence intervals for contrast of $\theta_i - \theta_0$. Marcus and Talpaz [11] have proposed a class of multiple contrast tests, in which each test is based on the maximum of $k2^{k-1}$ contrast statistics. They recommended to use the restricted maximum likelihood estimator (RMLE) instead of the unrestricted estimator in Dunnett's test statistic. Since under the tree order restriction, RMLE of the control mean may perform poorly, so their procedure has a lower power for some patterns of parameters. Hence, Peddada et al. [15] have proposed a test statistic based on the estimation procedure of Hwang and Peddada [8]. Based on the modified RMLE in analysis of covariance models, Betcher and Peddada [4] developed a general test procedure and showed that their modified test has higher power than the other test procedures.

In this paper, we introduce three test statistics via the randomized, smoothed and the weighted estimators which are constructed based on a random decision and a Bayesian approach. By using a random device, we allocate a probability to the control group mean and therefore the mean squared error (MSE) of the proposed estimator decreases. By using of this estimator the first test statistic is constructed and for second test statistic we apply a smoothed combination using of the expectation of actions in decision theory. In the third test statistic, we apply a weighted restricted estimator in which the weights are the posterior probabilities of the population orders and therefore the third test statistic is constructed based on this weighted estimator.

The rest of this paper is as follows. In Section 2, we present several classical and well known procedures for testing the equality of normal means against the tree order constraint. The LRT in Bartholomew [2] and then the multiple comparisons tests in [4, 11, 15] are also discussed. In Section 3, we propose three alternative test procedures due to the proposed test statistic based on the nature of pair-wise comparisons of Dunnett's test. By using of a random decision, two test procedures are described in this section,. The third test procedure provides a weighted test statistic based on the Bayesian approach. Also by using of the improved procedures, the two-sided test statistic is constructed. In Section 4, a large simulation study is carried out to compare the error rate and power of the proposed methods with those of the other existing methods. Since similar results were obtained for some patterns, we just report small representative sample of these results. For various patterns of population parameters, it is concluded that the new test procedures based on the random decision method, are more powerful than the other tests. Unlike the LRT in [2] and test of Marcus and Talpaz [11], the power values of the proposed tests appear to stabilize and slightly increase when k increases. As expected, the confidence intervals constructed by these new statistics have smaller size than the alternative intervals. In Section 5, two bioassay data sets discussed in [9, 15] are analyzed in details. A brief discussion and concluding remarks are provided in Section 6.

2. LRT and three tests via to the multiple comparisons

In this section, we discuss LRT and several well known multiple comparisons tests which are introduced by authors in the normal model. Suppose random samples $X_{i1}, X_{i2}, \ldots, X_{in_i}$, $i = 0, 1, \ldots, k$, are taken from independent normal populations with means θ_i and common known variance σ^2, where the mean parameters constrained with the tree order restriction $\theta_0 \leq \theta_i$. The sample mean $\bar{X}_i = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i}$ denotes the unrestricted maximum likelihood estimator (UMLE) of θ_i and the index $i = 0$ refers to the control group. Under the tree order restriction, the LRT for testing homogenous normal means (H_0), versus H_1, reject

the null hypothesis for large values of

$$\bar{\chi}^2 = \sum_{i=0}^{k} \frac{n_i}{\sigma^2}(\hat{\theta}_i^{RMLE} - \hat{\theta})^2, \tag{2.1}$$

where $\hat{\theta} = \frac{\sum_{i=0}^{k} n_i \bar{X}_i}{\sum_{i=0}^{k} n_i}$ is the overall mean estimator of the common population mean under H_0 and $\hat{\theta}_i^{RMLE}$ is the RMLE of θ_i under H_1 which is the isotonic regression of $\bar{\boldsymbol{X}} = (\bar{X}_0, \bar{X}_1, \ldots, \bar{X}_k)$ with weights $w_i = \frac{n_i}{\sigma^2}, i = 0, 1, \ldots, k$. Analytically, RMLE of $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_k)$ can be derived as follow [10]:

$$\hat{\theta}_0^{RMLE} = \min_{S \subseteq K} \frac{\sum_{j \in S} w_j \bar{X}_j}{\sum_{j \in S} w_j}, \tag{2.2}$$

where the minimization is taken over all subsets S of $K = \{0, 1, \ldots, k\}$ containing element 0. The RMLE of treatment means are then:

$$\hat{\theta}_i^{RMLE} = \max\{\hat{\theta}_0^{RMLE}, \bar{X}_i\} \quad for\ i = 1, \ldots, k. \tag{2.3}$$

The null distribution of the LRT statistic is a mixture of chi-squared distributions in which the mixing coefficients are the level probabilities obtained by a recursive approach for $k \leq 4$ in [1]. Unfortunately, its difficulty in computation of the distribution restricts the practical application, especially for the general unequal sample sizes (unbalanced case). Also, Cohen [5] observed that the inference based on the likelihood methodology in the tree order restriction is unsatisfactory.

Another approach for testing of the tree order restriction is due to Dunnett [7] who gives the simultaneous multiple comparisons between each treatment mean θ_i and the control mean θ_0, whereas Dunnett's procedure is based on the UMLEs of θ_i. Under the tree order restriction, Marcus and Talpaz [11] used RMLE of the population means θ_i instead of the UMLE, which is a reasonable assumption for constrained parameter space. Their test statistic (MT test) for equal sample sizes and common known variance σ^2 is as follow:

$$MT = \max_{1 \leq i \leq k} \left\{ \frac{\sqrt{n}(\hat{\theta}_i^{RMLE} - \hat{\theta}_0^{RMLE})}{\sigma} \right\}. \tag{2.4}$$

The MT test statistic is similar to that of the multiple comparison test in [7], and can be regarded as the maximum overall one-sided pair-wise comparisons $\hat{\theta}_i^{RMLE}$ versus $\hat{\theta}_0^{RMLE}$. The MT test can handle for the case of unequal sample sizes that is general unbalanced case design.

Under the tree order constraint, Lee [10] noticed that the RMLE of control group θ_0 fails to dominate corresponding UMLE in sense of MSE. Hence, the MT test performs poorly in terms of the power. Peddada et al. [15] extended Dunnett's procedure for modified restricted estimators instead of the RMLEs. The basic idea is to replace the numerator of the Dunnett's test statistic by a suitable restricted estimator given by Hwang and Peddada [8]. But their statistic depends on the chosen simple order between the treatment means $\theta_1, \ldots, \theta_k$, and it is not defined uniquely. For this reason, Betcher and Peddada [4] generalized this procedure by setting of the modified RMLE in the Dunnett's test statistic instead of their estimators. So, their test statistic is an analogue to MT-test, where $\hat{\theta}_i^{RMLE} - \hat{\theta}_0^{RMLE}$ is replaced by the modified RMLEs, which are in the unequal variances case:

$$\hat{\theta}_0^{BP} = \frac{\sum_{j=0}^{k} (\sigma_j^2)^{-1} \hat{\theta}_0^{RMLE}(\bar{X}_0, \bar{X}_j)}{\sum_{j=0}^{k} (\sigma_j^2)^{-1}}; \quad \hat{\theta}_i^{BP} = \max\{\hat{\theta}_0^{BP}, \hat{\theta}_i^{RMLE}\}, \quad for\ i \geq 1. \tag{2.5}$$

We adopt the same technique, except that we shall use improved estimators by suitable approaches rather than the RMLEs, as described in the following section.

3. Proposed test procedures

In this section under the tree order restriction on the mean parameters $\theta_0 \leq \theta_i$ for all $i \geq 1$, we propose a new test statistic for testing equality of normal means $H_0 : \theta_0 = \theta_1 = \ldots, \theta_k$, when the common variance σ^2 across populations is known. Based on the considerations of Section 2 and by using of the proposed estimation methods which are described in the next two subsections, we derive new test statistics. In the final subsection a brief overview of two-sided test procedure is given.

3.1. Randomized and smoothed test statistics

The basic idea of the proposed test procedure is to invoke a probabilistic method by using of a random device. Because of the restricted estimator $\hat{\theta}_0^{RMLE}$ is decreasing in k, by increasing of the number of treatments the chance of the occurrence of inappropriate event $\{\bar{X}_i \leq \bar{X}_0\}$ for some $i = 1, \ldots, k$ may increased. In this case, the pressure of the tree order constraint is on the control group estimator to satisfy the tree order restriction. By consecutive occurrence of these events, $\hat{\theta}_0^{RMLE}$ tends to infinity and therefore the corresponding test statistic has no good performance. Motivated by the randomized decision used in $\hat{\theta}_0^{RMLE}$, we give a chance to the sample mean of the control group via to the allocated probability such that the obtained estimator does not have the drawback of the RMLE and performs well in terms of the MSE and coverage probability ([13]). From this stochastic mechanism, the pressure of the tree order on the control group mean is decreased and hence the tree ordering is satisfied. Thus, on basis of the decision theory we propose the randomized estimator as follows:

$$\hat{\theta}_0^{RE} = \begin{cases} \bar{X}_0 & \textit{with propability}; \quad p \\ \hat{\theta}_0^{RMLE} & \textit{with probability}; \quad 1-p, \end{cases} \tag{3.1}$$

where p is derived based on the preliminary hypothesis testing in [13] which is given by:

$$p = \frac{\sum_{i=0}^{k} n_i I_{\{\bar{X}_0 \leq \bar{X}_i\}}}{\sum_{i=0}^{k} n_i}, \tag{3.2}$$

and the treatment estimators are:

$$\hat{\theta}_i^{RE} = \max\{\hat{\theta}_0^{RE}, \bar{X}_i\}, \quad for \quad i = 1, \ldots, k. \tag{3.3}$$

Hence, similar to Dunnett's method and by using of the randomized estimator, the proposed test statistic for testing H_0 in favor of H_1 is given by:

$$R = \max_{1 \leq i \leq k} \Big\{ \frac{\hat{\theta}_i^{RE} - \hat{\theta}_0^{RE}}{\sigma \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}} \Big\}. \tag{3.4}$$

We can modify the randomized estimator $\hat{\theta}_0^{RE}$ to a smoothed estimator which is the expectation of randomized estimator given sampling means. Hence, we use the expectation of the randomized estimator based on the Theorem 3.1 in [3], whereas the obtained smoothed estimator has a smaller MSE than that of the randomized estimator. The smoothed estimators for θ_i's are as follows:

$$\hat{\theta}_0^{SE} = \Big(\frac{\sum_{i=0}^{k} n_i I_{\{\bar{X}_0 \leq \bar{X}_i\}}}{\sum_{i=0}^{k} n_i} \Big) \bar{X}_0 + \Big(\frac{\sum_{i=0}^{k} n_i I_{\{\bar{X}_0 > \bar{X}_i\}}}{\sum_{i=0}^{k} n_i} \Big) \hat{\theta}_0^{RMLE}, \tag{3.5}$$

$$\hat{\theta}_i^{SE} = \max\{\hat{\theta}_0^{SE}, \bar{X}_i\}, \quad for \quad i = 1, \ldots, k. \tag{3.6}$$

Unlike the RMLE of θ_0, $\hat{\theta}_0^{SE}$ does not fail for large k. By replacing $(\hat{\theta}_0^{SE}, \hat{\theta}_i^{SE})$ with $(\hat{\theta}_0^{RE}, \hat{\theta}_i^{RE})$ in the Equation (3.4), we derive the following test statistic,

$$S = \max_{1\le i\le k}\Big\{\frac{\hat{\theta}_i^{SE} - \hat{\theta}_0^{SE}}{\sigma\sqrt{\frac{1}{n_i}+\frac{1}{n_0}}}\Big\}. \tag{3.7}$$

In a simulation study in [13], it is found that the quantity p in (3.2) is a suitable choice to reduce the MSE of the randomized and hence smoothed estimator, significantly. So, we expect the test statistic (3.7) performs well in terms of the power and error rate.
In general unknown variance case, the null distribution of S-test statistic can be approximated by the following theorem.

Theorem 3.1. *Suppose* $\bar{Z}_i \sim N(0,1)$, $i = 1,\ldots,k$ *and* $U \sim \chi^2_{N-k-1}$ *where* $N = \sum_{i=0}^{k} n_i$. *Let* $\bar{Z}_i$ *and* U *are independently distributed and* $\bar{\boldsymbol{Y}} = (\bar{Y}_0,\ldots,\bar{Y}_k)$ *is the smoothed estimator which is constructed by using* $\bar{\boldsymbol{Z}} = (\bar{Z}_0,\ldots,\bar{Z}_k)$, *and satisfies in the tree order restriction. Then under* H_0 *we have,*

$$Z = \max_{1\le i\le k}\Big\{\frac{\hat{\theta}_i^{SE} - \hat{\theta}_0^{SE}}{\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_0}}}\Big\} \overset{d}{=} \max_{1\le i\le k}\left\{\frac{\sqrt{df}(\bar{Y}_i - \bar{Y}_0)}{U\sqrt{\frac{1}{n_i}+\frac{1}{n_0}}}\right\}. \tag{3.8}$$

For a proof of the theorem, one can use a similar argument in [4]. The upper percentile of Z-test statistic denoted by z_α and must be simulated. Also, the null distribution of R-test statistic can be obtained similarly. Thus, we reject H_0 at a level of significance α, if

$$\max_{1\le i\le k}\Big\{\frac{\hat{\theta}_i^{SE} - \hat{\theta}_0^{SE}}{\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_0}}}\Big\} \ge z_\alpha, \tag{3.9}$$

where the critical point z_α is defined by

$$P(Z \ge z_\alpha|\theta_0 = \theta_1 = \ldots = \theta_k) = P\Big(\max_{1\le i\le k}\Big\{\frac{(\hat{\theta}_i^{SE} - \hat{\theta}_0^{SE})}{\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_0}}}\Big\} \ge z_\alpha|\theta_0 = \theta_1 = \ldots = \theta_k\Big) = \alpha. \tag{3.10}$$

As done in [4], the critical values of Z-test can be applied for constructing $(1-\alpha)100\%$ simultaneous confidence bounds for contrast between each treatment with the control, i.e., $\theta_i - \theta_0, i = 1,\ldots,k$. These simultaneous confidence bounds are

$$\cap_{i\in\{1,2,\ldots,k\}}\Big\{(\theta_i - \theta_0) \in (\hat{\theta}_i^{SE} - \hat{\theta}_0^{SE}) \pm z_\alpha\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_0}}\Big\}, \tag{3.11}$$

and the size of the above confidence hypercube is given by,

$$Size = \prod_{i=1}^{k}\Big\{2(z_\alpha\hat{\sigma})\sqrt{\frac{1}{n_i}+\frac{1}{n_0}}\Big\} \tag{3.12}$$

3.2. A test statistic based on the weighting method

Lee [10] demonstrated that by increasing of the weight of the control group mean w_0, the MSE reduction can be achieved. We propose an intuitively reasonable test procedure utilizing the weighted restricted estimator ([12]). The weights are constructed by using of the Bayesian approach. Under the tree order restriction with maintaining known inequalities between θ_0 and $\theta_i, i \ge 1$, all possible $k!$ simple orderings of the parameters $\theta_1,\ldots,\theta_k$ are considered. Then by use of a non-informative prior distribution i.e., $\pi(\theta_0,\ldots,\theta_k) = 1$, the joint posterior distribution is given by,

$$\pi(\theta_0,\theta_1,\ldots,\theta_k|\bar{\boldsymbol{x}}) = N(\bar{x}_0,\sigma^2_{\bar{x}_0}).\prod_{i=1}^{k} TN_{\theta_0^-}(\bar{x}_i,\sigma^2_{\bar{x}_i}). \tag{3.13}$$

where $TN_{\theta_0^-}(\bar{x}_i, \sigma^2_{\bar{x}_i})$ is the truncated normal distribution with left truncation point θ_0^-. Therefore, these posterior probabilities are proportional to $N(\bar{x}_i, \sigma^2_{\bar{x}_i})$ subject to the tree order restriction. By considering all $k!$ simple orders between $\theta_1, \ldots, \theta_k$, the posterior probability for each simple order is estimated. The posterior probabilities of the order of populations are now used to make weightings about $k!$ simple orders and make evidence about order of parameters $\theta_1, \ldots, \theta_k$. Therefore, by using of these posterior probabilities as the weights, the weighted proposed estimators are as follows:

$$\hat{\theta}_0^{WA} = \sum_{j=1}^{k!} \pi(\theta_1^{(j)} \leq \ldots \leq \theta_k^{(j)} | \bar{\boldsymbol{x}}) \hat{\theta}_0^{(j)}. \tag{3.14}$$

$$\hat{\theta}_i^{WA} = \max\{\hat{\theta}_0^{WA}, \bar{X}_i\}, \quad for \quad i = 1, \ldots, k, \tag{3.15}$$

where $\hat{\theta}_0^{(j)}$ is the RMLE of the smallest parameter (i.e., control group) in jth simple order and $\pi(\theta_1^{(j)} \leq \ldots \leq \theta_k^{(j)} | \bar{\boldsymbol{x}})$ is the posterior probability of the jth simple order, where j is corresponding to a permutation of the treatment groups $\{1, 2, \ldots, k\}$. So, as in (3.4) and according to the weighted estimator we now derive third new test statistic that is given by:

$$WA = \max_{1 \leq i \leq k} \left\{ \frac{\hat{\theta}_i^{WA} - \hat{\theta}_0^{WA}}{\sigma \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}} \right\}. \tag{3.16}$$

We reject H_0 if $WA \geq z_\alpha^*$, hence for WA-test statistic we have a similar expression as in (3.10) to obtain the upper percentile z_α^*.

In many applications for practical reasons the number of treatments k is usually small, frequently $k \in \{2, 3, 4, 5\}$. In these cases, the WA test statistic can be applied.
Due to the (3.11) in a similar way, by using of the WA statistic $\hat{\theta}_i^{WA} - \hat{\theta}_0^{WA}$ in place of $\hat{\theta}_i^{SE} - \hat{\theta}_0^{SE}$ we propose the $(1-\alpha)100\%$ simultaneous confidence bounds and that is given by:

$$\cap_{i \in \{1,2,\ldots,k\}} \left\{ (\theta_i - \theta_0) \in (\hat{\theta}_i^{WA} - \hat{\theta}_0^{WA}) \pm z_\alpha^* \sigma \sqrt{\frac{1}{n_i} + \frac{1}{n_0}} \right\}, \tag{3.17}$$

where z_α^* is the upper α quantile of the WA-test statistic distribution.

3.3. Two-sided test procedures

In some applications, it is impossible to decide a priori on the direction of the inequalities between the control and treatment groups. As dosage increases, it is unlikely that the direction of the response mean changes arbitrarily, at different doses relative to the control. Currently, the National Toxicology Program (NTP) [14] addresses this problem by performing a test of two-sided alternative $\theta_0 \neq \theta_i$ for each dose group i, that is ANOVA test method. But, such an alternative does not consider a priori knowledge information available in dose-response studies. So, the inclusion of prior information as the two-sided alternative in the following leads to more powerful tests than those that do not take the ordering of the means into account. To construct a two-sided test of size α, it is not appropriate to combine the one-sided tests each of size $\frac{\alpha}{2}$. The union of the critical regions of each hypothesis $\theta_0 \leq \theta_i$ and $\theta_0 \geq \theta_i$ for $1 \leq i \leq k$ leads to a size which is strictly less than α for all simple members of the composite null hypothesis. The two-sided tree order alternative is as follow:

$$H_1 : \{\theta_0 \leq \theta_i, i = 1, \ldots, k\} \cup \{\theta_0 \geq \theta_i, i = 1, \ldots, k\} - \{\theta_0 = \theta_1 = \ldots = \theta_k\}. \tag{3.18}$$

Due to the two-sided reasonable idea and by using of the proposed smoothed estimators in the two types of the tree orderings i.e., down-turn tree order $\theta_0 \leq \theta_i$ and up-ward tree

order $\theta_0 \geq \theta_i$ for $1 \leq i \leq k$, we define the two-sided critical region via to the random decision as follow:

$$S_{two-sided} = \max_{1\leq i\leq k} \left\{ \frac{\max(\hat{\theta}_i^{SE} - \hat{\theta}_0^{SE}, \tilde{\theta}_i^{SE} - \tilde{\theta}_0^{SE})}{\sigma\sqrt{\frac{1}{n_i} + \frac{1}{n_0}}} \right\} \geq z_\alpha, \tag{3.19}$$

where $\tilde{\theta}_i^{SE}, i = 0, 1, \ldots, k$ denote the proposed smoothed estimators subject to the up-ward tree order restriction, $\theta_0 \geq \theta_i$ for $1 \leq i \leq k$, i.e.,

$$\tilde{\theta}_0^{SE} = \Big(\frac{\sum_{i=0}^k n_i I_{\{\bar{X}_0 \geq \bar{X}_i\}}}{\sum_{i=0}^k n_i}\Big)\bar{X}_0 + \Big(\frac{\sum_{i=0}^k n_i I_{\{\bar{X}_0 < \bar{X}_i\}}}{\sum_{i=0}^k n_i}\Big)\tilde{\theta}_0^{RMLE}, \tag{3.20}$$

$$\tilde{\theta}_i^{SE} = \min\{\tilde{\theta}_0^{SE}, \bar{X}_i\}, \quad for \quad i = 1, \ldots, k, \tag{3.21}$$

and RMLE in the $\tilde{\theta}_0^{SE}$ is derived based on the up-ward tree order restriction $\theta_0 \geq \theta_i$ for $1 \leq i \leq k$, which is given by:

$$\tilde{\theta}_0^{RMLE} = \max_{S\subseteq k} \frac{\sum_{j\in S} n_j \bar{X}_j}{\sum_{j\in S} n_j}. \tag{3.22}$$

Similar to the one-sided case, we note that the simultaneous confidence bounds for the two-sided multiple comparisons can be obtained. As before, the critical values are computed approximately. Similarly, we can test the aforementioned two-sided test by using of the WA-test statistic instead of the S-test statistic as in (3.19). All the remaining tests that were mentioned in the previous section, were compared for one-sided alternative. Peddada et al. [15] by using of the Hwang and Peddada's estimator in [8] constructed the two-sided test statistic.

4. Comparison of the power and error rate

In this section the power values and error rates of the three proposed tests (R, S and WA) are compared with those of the LRT, Marcus and Talpaz (MT) in [11], Peddada et al. (P) in [15] and Betcher and Peddada (BP) in [4]. Various configurations of mean parameters are considered in Tables 1 and 3 for both one and two-sided tests, respectively. The number of replications for each estimate is 10,000 iterations. We estimate the type I error rate by using 10000 simulation runs. According to the first rows in Tables 1 and 3, it is seen in each case that the type I error is close to the nominal level of 0.05. For $k = 5$ and 10 normal populations with sample sizes $n_i = 5$, mean parameters $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_k)$ and $\sigma^2 = 1$, the power values are estimated at the 5% level of significance for both one and two-sided alternatives in Tables 1 and 3, respectively.

Table 1, gives the power comparison of the new proposed tests based on the randomized, smoothed and weighted estimators with the other test methods that are listed in the literatures for $k = 5$ and 10 treatment groups. Numerical results indicate that the gains in the power of the proposed S-test can be substantial. For $k = 5$, the powers of the LRT and MT are larger at the center than the edges of parameter space. But, the S-test performs uniformly well, in both cases of the parameter space. For the least favorable case $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_k) = (0, 0, \ldots, 0)$, there is a little difference in the power values for all test procedures. For the extreme points such as $\theta = (0, 0, \ldots, 0, 1)$ the LRT and MT procedures have higher power than the other tests, in which the proposed tests are better than the BP test. When the mean parameter $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_k)$ is near to the center of the tree order cone, both LRT and proposed tests are more powerful, and would be preferred. However, for the interior configuration such as $\theta = (0, 1, \ldots, 1)$ the proposed S-test is more powerful than the other competitors. From Table 1, it is obvious that the proposed tests compete well in terms of the power with the other tests. The S-test appears to have the most power in the center points of the parameter space e.g., $(0, 0.5, \ldots, 0.5, 1, \ldots, 1)$. For the extreme

points such as $(0, 0, \ldots, 1)$ the MT procedure and then the proposed S-test have higher power than the other tests, which is consistent with the previously published results in [4]. Only, in the near of the center parameter space e.g., $(0, 0, 0, 1, 1, 1)$, the power of the LRT is large. It is seen that the power of the proposed tests increases with k, but the opposite is true for the LRT and MT tests.

Table 1. Power comparisons of LRT, MT, BP, R, S and WA. For k=5 and 10 treatments.

$\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_5)$	LRT	MT	BP	R	S	WA
$(0, 0, 0, 0, 0, 0)$	0.049	0.050	0.051	0.051	0.048	0.049
$(0, 0, 0, 0, 0, 1)$	0.311	0.324	0.287	0.320	0.323	0.319
$(0, 0.2, 0.4, 0.6, 0.8, 1)$	0.315	0.312	0.355	0.345	0.355	0.350
$(0, 0.5, 0.5, 0.5, 1, 1)$	0.421	0.417	0.510	0.420	0.508	0.490
$(0, 0, 0, 1, 1, 1)$	0.471	0.459	0.445	0.459	0.463	0.460
$(0, 1, 1, 1, 1, 1, 1)$	0.412	0.507	0.541	0.523	0.542	0.531
$\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_{10})$	LRT	MT	BP	R	S	WA
$(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$	0.050	0.048	0.052	0.050	0.051	0.051
$(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1)$	0.528	0.555	0.497	0.531	0.550	0.541
$(0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2)$	0.999	0.998	0.997	0.997	0.999	0.990
$(0, 0.5, 0.5, 0.5, 0.5, 0.5, 1, 1, 1, 1, 1)$	0.729	0.733	0.749	0.736	0.758	0.745
$(0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1)$	0.970	0.923	0.819	0.814	0.848	0.890
$(0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$	0.639	0.786	0.842	0.825	0.849	0.836

In Table 2, the estimated size of simultaneous confidence intervals for $\theta_i - \theta_0, i = 1, \ldots, k$ as in (3.12) are obtained. The critical constants are chosen so that the coverage rate of the simultaneous confidence interval is 0.95. It is seen that the size of the proposed simultaneous confidence intervals is smaller than that of MT procedure and compete well with the BP test procedure. These results are true for a variety of patterns in this table. Hence, the test based on the proposed simultaneous confidence intervals has larger power than the corresponding test based on the MT test procedure and competes well with the BP test method. In overall, as k increases the maximum powers of the S-test and WA-test are always slightly higher than those of the LRT and tend to be slightly higher than those of the MT tests.

Table 2. Size of simultaneous confidence intervals for $\theta_i - \theta_0, i = 1, \ldots, k$. (in log scale).

σ	k=5					k=10				
	MT	BP	R	S	WA	MT	BP	R	S	WA
0.50	7.73	7.67	7.70	7.65	7.68	14.49	14.34	14.40	14.33	14.38
0.60	8.14	8.01	8.09	8.02	8.09	15.35	14.99	15.20	15.00	15.12
0.75	8.75	8.49	8.56	8.45	8.50	16.58	15.92	16.04	15.90	15.99
1	9.63	9.27	9.35	9.25	9.29	18.35	17.38	17.40	17.36	17.38
1.50	11.01	10.62	10.70	10.62	10.65	20.96	19.85	19.91	19.84	19.88

The results of our simulation experiment for two-sided alternative are summarized in Table 3. Similar to the LRT, the BP and MT tests are not available in the literatures for two-sided alternative hypotheses, therefore in our simulation study we investigate the performance of the proposed procedures only with the Peddada's procedure in [15] which is constructed by the Hwang and Peddada's estimator in [8]. The overall conclusions are very similar for both one-sided as well as two-sided alternatives. It seems that all two-sided test procedures (T, R, S and WA) attain the nominal type I error rate of 0.05,

approximately. However, in some cases the gain in power due to the proposed procedures R, S and WA is substantial for two-sided alternative hypotheses. As seen in Table 3, in some situations the T-test in [15] may have smaller power than the proposed procedures that are introduced in this paper. In contrast, the two new test procedures, R and S, perform well for almost every configuration in two-sided hypothesis. Also, among the two-sided testing procedures, the proposed S-test appears to have the largest power in all configurations.

Table 3. Power comparisons of T, R, S and WA tests for two-sided alternative. For $n_i = 5$, $\sigma = 1$.

$\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_5)$	T	R	S	WA
$(0,0,0,0,0,0)$	0.046	0.051	0.050	0.049
$(0,0,0,0,0,2)$	0.791	0.885	0.909	0.890
$(0,0,0,0,2,0)$	0.785	0.876	0.896	0.880
$(0,0,0,2,0,0)$	0.768	0.884	0.900	0.886
$(0,0,2,0,0,0)$	0.751	0.885	0.903	0.891
$(0,2,0,0,0,0)$	0.736	0.881	0.899	0.889
$(0,0,0,0,1,2)$	0.801	0.868	0.892	0.872
$(0,1,2,0,0,0)$	0.729	0.867	0.884	0.875
$(0,0,0,0,2,2)$	0.912	0.948	0.965	0.956
$(0,0,0,2,2,2)$	0.943	0.965	0.975	0.969
$(0,0,2,2,2,2)$	0.945	0.970	0.975	0.970
$(2,2,2,2,2,2)$	0.905	0.974	0.977	0.973

5. Application to analysis of bioassay data

In this section, by using of two bioassay data sets which are obtained from the NTP [9, 14], the proposed methods are illustrated. In the first subsection, the one-sided test procedure and in the second the two-sided test procedure are explained.

5.1. Application to the Estrogen-like compounds data

Kanno et al. [9] conducted a large Uterotrophic bioassay to evaluate the effects of different compounds i.e., (1) Bisphenol A, (2) DBP, (3) DDT, (4) Genestein, (5) Ethinly Estradiol (EE high dose), (6) Ethinly Estradiol (EE low dose) and (7) Methoxychloron rat uterine weight of the Estrogen data. Their study consisted of n_i animals across per groups. Biological, Estrogen levels are effective on the rat uterine weight than the control level (placebo). Hence when comparing seven compounds (treatments) with the control group, we consider a tree order restriction on the mean uterine weight of log-transformed weights $\theta_i, i = 0, 1, \ldots, 7$, as an alternative hypothesis. Based on the two-way ANOVA, Betcher and Peddada [4] found that the body weight is a significant covariate ($p - value < 0.05$). Hence, when the comparing of the above seven compounds with the control group is interested, the uterine weight adjusted for the body weight of the animal. The sample mean uterine weights (i.e., UMLEs) and standard errors (in parentheses) along with the mean body weight necropsy and the corresponding standard errors (in parentheses) are summarized in Table 4.

Since, the uterine weight has a skewed distribution, we then transform the logarithm of the data. Therefore, $\theta_i, i = 0, 1, \ldots, 7$ represent the mean uterine weight of log transformed weights. In this experiment, under the tree order restriction $\theta_0 \leq \theta_i$ for $1 \leq i \leq 7$, the two proposed estimators, smoothed (SE) and weighted average (WA), are given in Table 4. For $k = 7$ and using of the simulated distribution of S-test statistic and WA statistic, we obtain the critical values equal to 2.41 and 2.42, respectively. These critical values are

Table 4. Comparison of the mean uterine weight relative to the control group based on the confidence intervals of S and WA statistics.

Compound	Sample size	Uterine weight	UMLE of Uterine weight	Proposed estimation (SE)	Proposed estimation (WA)	Confidence intervals for $\theta_i - \theta_0$ (S-statistic)
Control	4	19(4.21)	2.90(0.12)	2.90	2.90	− − −−
Bisphenol A	4	27.75(5.35)	3.31(0.12)	3.31	3.315	$0.41 \pm 2.41(0.17)$
DBP	6	17.33(3.79)	2.80(0.10)	2.90	2.908	$0 \pm 2.41(0.17)$
DDT	4	70.25(8.50)	4.35(0.12)	4.35	4.348	$1.45 \pm 2.41(0.16)$
Genestein	5	58.20(3.72)	3.99(0.11)	3.99	3.998	$1.09 \pm 2.41(0.16)$
EE high dose	5	77.40(10.37)	4.33(0.10)	4.33	4.320	$1.43 \pm 2.41(0.16)$
EE low dose	6	47.67(6.13)	3.79(0.10)	3.79	3.805	$0.89 \pm 2.41(0.16)$
Methoxychlor	6	53.33(4.17)	3.98(0.10)	3.98	3.983	$1.08 \pm 2.41(0.16)$

used for constructing 95% simultaneous confidence intervals for contrasts $\theta_i - \theta_0, 1 \leq i \leq 7$. For summary, since the results are same, these confidence intervals are given in Table 4 only for S-statistic method. Note that the 95% confidence intervals for the mean difference of uterine weight of the DBP group relative to the control group contain, 0. Thus, there exists no significance difference in the mean uterine weights between the DBP treated animals and the control group. Also, none of the remaining confidence intervals contain 0. Therefore, the mean uterine volumes for all other Estrogen compounds differ from that of the control group. These results are Consistent with previously published results in [4] and except for treatment Bisphenol A, our results about other compounds agree with the results in [9]. They did not conclude any significance difference between the test compound Bisphenol A, with the control group.

5.2. Application to the red blood cells data

To evaluate the performance of the two-sided proposed procedures (R and S-test) with the Peddada's test procedure (i.e., T-test) in [15], we apply the blood count data from core clinical pathology evaluation in [14]. The complete blood count contains a variety of variables. Since the Haematology data consisted of multiple variables, the data for Mean Corpuscular Volume (MCV) were arbitrarily selected for illustration. The MCV data for male rats in the Prechronic study of Anthraquinone (NTP [14]) for this survey are utilized. The corresponding group means (with standard errors in parentheses) presented in Table 5.

Table 5. Mean Corpuscular Volume (MCV) and confidence intervals of $\theta_i - \theta_0$ at different levels.

Dose groups	Mean and (SE) of MCV	Confidence intervals Using T	Confidence intervals Using R	Confidence intervals Using S
0	61.1(0.2)	− − −−	− − −−	− − −−
1875 ppm	60.2(0.2)	$(-0.344, 2.120)$	$(-0.344, 2.11)$	$(-0.342, 2.09)$
3750 ppm	61.3(0.2)	$(-0.324, 1.448)$	$(-0.315, 1.445)$	$(-0.304, 1.435)$
7500 ppm	61.4(0.2)	$(-0.228, 1.548)$	$(-0.220, 1.535)$	$(-0.218, 1.522)$
15000 ppm	61.7(0.2)	$(0.072, 1.848)$	$(0.079, 1.845)$	$(0.080, 1.826)$
30000 ppm	62.4(0.3)	$(0.772, 2.548)$	$(0.781, 2.511)$	$(0.785, 2.521)$

Since often there exists a little information about direction of the toxicity compounds, we tested a two-sided hypothesis. In theory, MCV would be expected to demonstrate Erythrocyte volumes which are normally distributed. Thus, two proposed test methods (S and WA) are compared with the T-test in [15] that obtained for two-sided alternative in the normal distributions. The unknown population means for the dose groups (treatments) are in the $0, 1875, 3750, 7500, 15000$ and 30000 parts per million (ppm) which are denoted by $\theta_0, \theta_1, \ldots, \theta_5$, respectively. To illustrate the proposed methodology, we compute 95% simultaneous confidence intervals for $\theta_i - \theta_0, i = 1, \ldots, 5$. Therefore, in Table 5

the estimated 95% simultaneous confidence intervals for contrasts $\theta_i - \theta_0$ for $1 \leq i \leq 5$ are given. These procedures identified a significance increase ($\alpha = 0.05$) in MCV in the 30000 ppm level.

Furthermore, Dunnett's test identified a significant decrease ($\alpha = 0.05$) in MCV in 1875 ppm level. Also, both of proposed test procedures, R and S, detected a decrease in MCV in 1875 ppm and increases in 15000 and 30000 ppm levels ($\alpha = 0.05$). In fact, the proposed methods (R and S) for these data identified an additional (lower) dose group that may have been affected by anthraquinone treatment but was undetected by currently methods in NTP studies (i.e., Dunnett's test). Therefore, in this survey, the proposed tests detected an additional affected dose level, i.e. 15000 ppm level. From a practical point of view, both proposed R and S tests have shorter confidence intervals than that of the T-test in [15]. Although, the T-test of [15] in the literature, were designed to test the two-sided alternative, but this test is not unique. Because of the T- test statistic is based on the arbitrary choice of the simple order. If this order departures from the corresponding actual order, then the T-test is very liberal (i.e., exceeds of 0.05), and so perform poorly. The randomized (R) and smooth (S) proposed methods have not these drawbacks and detect the significance increase or decrease by the random device from decision theory. So, similar to the results of simulation study for two-sided alternative in Table 3, the proposed confidence intervals have larger power than the T-test in [15]. Thus, the proposed methods in this paper will also be useful in the analysis of NTP data.

6. Concluding remarks

In this paper, we developed new test procedures as well as Dunnett's test for comparing the means of several dose groups with the mean of a control group, known as the tree order restriction. For this purpose, the possibility of using the maximum of a finite number of contrast statistic is considered for testing equality of normal means against the tree order restriction among of mean parameters. We have successfully extended the technique of building the maximum over several contrasts by comparison of each treatment with to the control group. Also, we have considered both one-sided and two-sided tree alternatives. For ordered hypotheses, because of the difficulties involved in applying LRT, several researchers including [11, 15] and more recently [4] considered the testing of the equality between normal means against the tree order constraint. The new tests are particularly attractive when the LRT cannot be readily conducted due to difficulties in the determination of the null distribution of the test statistic. These new test methods are based on the estimators which are constructed via to a random decision and the Bayesian method.

If it is believed that the treatment effects are reasonably homogeneous, this situation of the restriction could be employed since it is quite powerful at the center of the tree order cone. At the other extreme points, the proposed test procedures have good power if one of the treatments is greater than the others and the other treatments are fairly homogeneous that is $\theta_0 < \theta_1 = \theta_2 = \ldots = \theta_k$.

Simulation study revealed that the proposed tests have better ability to detect the significant differences and have more power in comparison with the existing procedures which are listed in the literatures. In some situations, the test based on the S statistic would seem to be the preferred test. Although the proposed tests are not uniformly more powerful than the competitor, they have higher power at some points in the alternative.

In general, the patterns of a dose-response can not arbitrarily change with dose. In dose-response studies, an investigator may not be sure about the direction of the inequalities between the control group and dose groups. In some experiments there exists a significant dose-related increase in response mean, conversely in other trials there exists a significant dose-response reduction in the mean of response variable. So, it is impossible to decide a priori on the direction of the inequalities between the control group and dose groups.

Since dose-related increases (or decreases) may be occur in response mean but cannot be predicted, and little information often exists concerning the direction of compounds in NTP studies, so we test a two-sided tree hypothesis based on the random decision in which performed better than the methods that do not take into account the tree order constraint. An important conclusion from proposed two-sided tests was the acceptable power when the experimenter is unsure about the direction of the inequality in the tree order restriction.

On the basis of two numerical examples that were studied in this paper, we find that the new procedures compete well with Betcher and Peddada's method in [4] for the one-sided hypothesis, and they perform well with Peddada's procedure in [15] when our interest is two-sided hypothesis.

To summarize the results concerning the proposed tests for testing equality of tree order normal means, note that the new procedures are easy to use and have a higher power than the other procedures. As expected, the proposed methods perform accurately to achieve the nominal level type I error rate. However, according to the underlying patterns of parameters, the best test varies among the aforementioned tests, but in most cases the new proposed tests perform better than both the MT and BP tests. If the alternative is two-sided tree order restriction, then the proposed two-sided tests perform substantially better than the Peddada's test [15] in almost every situation.

Acknowledgment. The authors thanks the referee for carefully reading this manuscript and giving the useful comments that improved the presentation of the manuscript.

References

[1] R. E. Barlow, D. J. Bartholomew, J. M. Bremner and H. D. Brunk, *Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression*, Wiley, New York, NY, 1972.

[2] D. J. Bartholomew, *A test of homogeneity of means under restricted alternatives*, J. R. Stat. Soc. Ser. B. Stat. Methodol., **23**, 239-281, 1961.

[3] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. Second edition. New York: Springer, 1985.

[4] J. Betcher and S. D. Peddada, *Statistical inference under order restrictions in analysis of covariance using a modified restricted maximum likelihood estimator*, Sankhya A, Volume **71-B**, Part 1, 79-96, 2009.

[5] A. Cohen, J. H. B. Kemperman and H. B. Sackrowitz, *Properties of likelihood inference for order restricted models*, J. Multivariate Anal., **72**, 50-77, 2000.

[6] S. Dunbar, M. Conaway and S. D. Peddada, *On improved estimation of parameters subject to order restrictions*, Statist. Applic., **3**, 121-128, 2001.

[7] C. W. Dunnett, *A multiple comparison procedure for comparing several treatments with a control*, J. Amer. Statist. Assoc., **50**, 1096-1121, 1955.

[8] J. T. G. Hwang and S. D. Peddada, *Confidence interval estimation subject to order restrictions*, Ann. Statist., **22**, 67-93, 1994.

[9] J. Kanno, L. Onyon, S. D. Peddada, J. Ashby, E. Jacob and W. Owens, *The OECD program to validate the rat uterotrophic bioassay: Phase two-coded single dose studies*, Env. Health Persp., **111**, 1550-1558, 2002b.

[10] C. I. C. Lee, *The quadratic loss of order restricted estimators for several treatment means and a control mean*, Ann. Statist., **16**, 751-758, 1988.

[11] R. Marcus and H. Talpaz, *Further results on testing homogeneity of normal means against simple tree alternatives*, Comm. Statist. Theory Methods, **21**, 2135-2149, 1992.

[12] R. Momeni, J. Etminan and M. Khanjari Sadegh, *Estimation of normal means in the tree order by the weighting methods*, Comm. Statist. Simulation Comput., DOI: 10.1080/03610918.2018.1554115. 2018.

[13] R. Momeni, J. Etminan and M. Khanjari Sadegh, *Estimation of parameters in the tree order restriction by a randomized decision*, Revised manuscript in J. Stat. Comput. Simul., 2019.

[14] National Toxicology Program. *Toxicology and carcinogenesis studies of anthraquinone in F344/N rats and B6C3F1 mice*. Technical Report 494. National Toxicology Program, Research Triangle Park, 1999a.

[15] S. D. Peddada, J. Haseman, X. Tan and G. Travlos, *Tests for simple tree order restriction with application to dose-response studies*, J. Roy. Statist. Soc., Ser. C, **55**, 493-506, 2006.

[16] T. Robertson, F. T. Wright and R. L. Dykstra, *Order Restricted Statistical Inference*. New York: John Wiley, 1988.

[17] M. J. Silvapulle and P. K. Sen, *Constrained Statistical Inference: Inequality, Order and Shape Restrictions*, Wiley, New York, NY, 2005.