



A Note on Determining Which of J Parameters Has the Largest or Smallest Value

Rand Wilcox

Dept of Psychology, University of Southern California

Orcid: 0000-0002-2524-2976

ABSTRACT

Let $\theta_1, \dots, \theta_J$ denote J parameters that are associated with J random variables. Let $\theta_{(1)} \leq \dots \leq \theta_{(J)}$ denote the parameters written in ascending order. This note suggests two simple approaches to determining whether it is reasonable to make a decision about which random variable corresponds to $\theta_{(J)}$. Both are related to Tukey's three-decision rule, they are easily adapted to a wide range of situations, and they have certain advantages over extant techniques. The focus is on trimmed means, but the method is easily adapted to a wide range of situations. One version of the proposed approach is based on a variation of the percentile bootstrap that has not been previously studied.

Keywords: multiple comparisons, familywise error, robust methods, ranking and selection

INTRODUCTION

Consider J independent groups having measures of location $\theta_1, \dots, \theta_J$ and let $\theta_{(1)} \leq \dots \leq \theta_{(J)}$ denote these measures written in ascending order. As is evident, in many situations there is interest determining which group has the largest measure of location, $\theta_{(J)}$. Let $\hat{\theta}_j$ ($j = 1, \dots, J$) denote an estimate of θ_j based on a random sample of size n_j and denote the estimates written ascending order by $\hat{\theta}_{(1)} \leq \dots \leq \hat{\theta}_{(J)}$. This paper deals with the issue of determining whether it is reasonable to conclude that the group corresponding to $\hat{\theta}_{(J)}$ is in fact the group having the largest population measure of location $\theta_{(J)}$.

Bechhofer [1] was the first to address this issue assuming that observations are randomly sampled from normal distributions having a common known variance σ^2 and that the goal is to identify the group with the largest population mean. His approach was to determine the common sample size n so that the probability of a correct decision (PCD), meaning the probability that the group with the largest sample mean is the group with the largest population mean, is at least $1 - \beta$. As is evident, the value of n depends in part on the unknown population means $\mu_1, \dots, \mu_J$. Bechhofer addressed this issue in the context of what is known as an indifference zone. That is, for some specified constant δ , the common sample size was determined when $\mu_1 = \dots = \mu_{J-1} = \mu$, say, and $\mu_J = \mu + \delta$. The solution is conservative in the sense that if $\mu_{(J-1)} \leq \mu_{(J)} + \delta$, the PCD will be greater than or equal to $1 - \beta$. Subsequently, two-stage procedures were derived for dealing with situations where the variances are unknown, still assuming normality [2, 3, 6, 17, 18]. That is, pilot data are used to estimate the variances,

which in turn can be used to determine how many more observations are needed so that the PCD is at least $1 - \beta$. These methods are a subset of statistical techniques generally known as ranking and selection methods [3, 7, 8, 16].

The first goal in this paper is to suggest a method that deals with two issues that have not been addressed. First, there is the concern that the population mean and variance are not robust [9, 14, 20]. Roughly, this means that an arbitrarily small change in a distribution can have an arbitrarily large impact on the values of the population mean and variance. In particular, even a small departure from a normal distribution can result in relatively poor power when testing hypotheses [21]. Moreover, the population mean can poorly reflect the typical response. These results have obvious negative implications about ranking and selection techniques based on means and variances. So the first goal is to consider methods for identifying the group having the largest measure of location $\theta_{(J)}$, where $\theta_{(J)}$ is some robust measure of location. Typically, the estimator $\hat{\theta}_j$ has a reasonably high breakdown point, where the breakdown point refers to the minimum proportion of values that must be altered to make it arbitrarily large or small. The sample mean and variance, for example, have a breakdown point of only $1/n$. That is, even a single outlier has the potential of giving a distorted view of the typical response and it can destroy power. The sample median has the highest possible breakdown point, 0.5.

The second goal is to suggest a single-stage method that does not require the specification of an indifference zone, δ , and simultaneously allows heteroscedasticity. This is done in the context of testing relevant hypotheses in a manner similar to comparing groups to a control, but the details differ in a crucial way as will be evident. The fundamental goal is to determine the extent it is reasonable to make a decision about which group has the largest measure of location.

There is a substantial literature on methods for comparing robust measures of location [21]. Based on extant results, a solution for the problem at hand might seem trivial. Imagine that group I corresponds to $\hat{\theta}_{(J)}$, the group with the largest estimate. Then one might proceed by testing.

$$H_0: \theta_j = \theta_I \quad (1)$$

for each $j \neq I$. If all $J - 1$ of these hypotheses are rejected, decide that group I has the largest measure of location, otherwise no decision is made. However, there is the issue of controlling the familywise error (FWE) rate, meaning the probability of one or more Type I errors.

A simple approach to controlling the FWE rate is to use the Bonferroni method, meaning that each test is performed at the $\alpha/(J - 1)$ level. It is well known that if the Type I error is controlled for each of the $J - 1$ tests, then the Bonferroni method guarantees that FWE is less than or equal to α . Various methods aimed at improving on the Bonferroni have been derived [12, 13, 19], which are based on p-values. For brevity, details about these methods are not provided because they can be highly unsatisfactory for the problem at hand. Simulations indicated that with $J = 4$, there are situations where FWE exceeds 0.070 when using the Bonferroni method and when testing at the 0.05 level, but estimates did not exceed 0.075. With $J = 5$, FWE can exceed 0.08.

The next section describes the proposed approach. Initially attention is focused on trimmed means. Then a more general approach is described that can deal with other robust measures of location. This is followed by simulation results and illustrations demonstrating the practical importance of the proposed method.

THE PROPOSED APPROACH

It is convenient to first focus on trimmed means. Let $X_1, \dots, X_n$ be a random sample. Let $g = [\gamma n]$, where $[\gamma n]$ is the integer portion of γn , $0 \leq \gamma < 0.5$. The γ trimmed mean is

$$\bar{X}_t = \frac{1}{n - 2g} (X_{(g+1)} + \dots + X_{(n-g)}),$$

where $X_{(1)} \leq \dots \leq X_{(n)}$ are the $X_1, \dots, X_n$ written in ascending order. The γ trimmed mean has a breakdown point of γ . The focus here is on 20% trimming, $\gamma = 0.2$, which has good efficiency under normality [21]. Moreover its standard error can be substantially less than the standard error of the sample mean when dealing with heavy-tailed distributions, which are common. The final section of this paper comments on concerns that arise when the amount of trimming is close to zero.

Yuen [22] derived a method for testing the hypothesis that two independent groups have equal population trimmed means. The method is based in part on the Winsorized sample variances. Winsorizing $X_1, \dots, X_n$ means that the g smallest are set equal to $X_{(g+1)}$ and the g largest are set equal to $X_{(n-g)}$. The γ Winsorized sample variance, s_w^2 , is the usual sample variance based on the Winsorized data. Let $\bar{X}_{tj}$ and s_{wj}^2 denote the sample trimmed mean and Winsorized variance, respectively, for the j th group ($j = 1, 2$). Yuen's test statistic is:

$$T_y = \frac{\bar{X}_{t1} - \bar{X}_{t2}}{\sqrt{d_1 + d_2}},$$

where

$$d_j = \frac{(n_j - 1)s_{wj}^2}{h_j(h_j - 1)}$$

and n_j is the sample size for the j th group and $h_j = n_j - 2g_j$, the number of observations left after trimming. The null distribution of T_y is approximated with a Student's t distribution with estimated degrees of freedom:

$$\hat{v}_y = \frac{(d_1 + d_2)^2}{\frac{d_1^2}{h_1 - 1} + \frac{d_2^2}{h_2 - 1}}.$$

The properties of Yuen's method have been studied extensively [21].

Now consider the goal of testing each of the $J - 1$ hypotheses given by (1) using Yuen's method with 20% trimming. If all $J - 1$ of these hypotheses are rejected, decide that group I has the largest measure of location. Otherwise no decision is made. The goal then is to control the FWE rate when testing these $J - 1$ hypotheses. But before describing how this is done, it helps to first consider the goal of controlling the probability of a Type I error for each individual test.

Let $\bar{X}_{t(1)} \leq \dots \leq \bar{X}_{t(J)}$ and let $\mu_{\pi(j)}$ denote the trimmed mean associated with $\bar{X}_{t(j)}$ ($j = 1, \dots, J$). Let p_j ($j = 1, \dots, J - 1$) denote the p-value when testing

$$H_0: \mu_{\pi(j)} = \mu_{\pi(J)}. \quad (3)$$

When $j = 1$, for example, (3) corresponds to comparing the group with the lowest estimate to the group with the highest estimate. Given j , suppose the null hypothesis is rejected if $p_j \leq d_j$. A simulation is readily performed for determining d_j so that the probability of a Type I error is α for each j . Consider for example $J = 4$, $\alpha = 0.05$, all four distributions are normal with a common variance of one and $n_1 = n_2 = n_3 = n_4 = 20$. Based on a simulation with 10,000 replications, the estimate of (d_3, d_2, d_1) was (0.11675, 0.04905, 0.02049). So when comparing the group with the largest estimate to the group with the next largest estimate, the Type I error probability will be 0.05 if the null hypothesis is rejected when the p-value is less than or equal to 0.11675. For $n_1 = n_2 = n_3 = n_4 = 100$ the estimate of (d_3, d_2, d_1) was (0.0922, 0.0295, 0.0053). The result $d_3 > 0.05$ is related to the fact that $\hat{\theta}_{(3)}$ and $\hat{\theta}_{(4)}$ are dependent.

An outline of the proposed method for controlling the FWE rate is as follows. Momentarily assume normality and homoscedasticity, and use a simulation to estimate the joint distribution of the $J - 1$ p-values when the J groups have a common measure of location. That is, $\theta_1 = \dots = \theta_J$, where for the moment the focus is still on 20% trimmed means. Once this joint distribution is estimated, determine constants $c_1, \dots, c_{J-1}$ such that

$$P(p_1 \leq c_1, \dots, p_{J-1} \leq c_{J-1}) = \alpha. \quad (4)$$

That is, the j th hypothesis is rejected if $p_j \leq c_j$ in which case FWE will be approximately α . Extant results suggest that for non-normal distributions, the FWE rate will be reasonably close to the nominal level when using a 20% trimmed mean, an issue that is addressed via simulations in the next section.

To elaborate, given J and $(n_1, \dots, n_J)$, randomly generate n_j observations from the j th group and determine the $J - 1$ p-values based on Yuen's method when testing (3). Repeat this L times yielding a matrix of p-values, say $\hat{\mathbf{P}}$, having L rows and $J - 1$ columns. As is evident, the j th column of $\hat{\mathbf{P}}$ provides an estimate of d_j , the critical p-value so that each of the individual tests has a Type I error of α . That is, d_j is just the α sample quantile based on the j th column of $\hat{\mathbf{P}}$. The idea is to choose the constant a such that $c_j = a d_j$ satisfies (4). This will be called method Y henceforth. Note that $a = 1/(J - 1)$ corresponds to the Bonferroni method. So the strategy is to take advantage of the estimate of the joint distribution of the p-values with goal of finding a better choice for a .

Dealing with Other Robust Measures of Location: Method PB

Here, three additional location estimators are considered: a one-step M-estimator, the median, and the estimate of the population median derived by Harrell and Davis [10]. Generally, when dealing with robust estimators that deal effectively with outliers, inferences based on a percentile bootstrap have been found to provide relatively good control over the Type I error probability, so this approach is used here. This section reviews the one-step M-estimator and the Harrell-Davis estimator followed by a review of the percentile bootstrap method.

The one-step M-estimator based on Huber's Ψ , is computed as follows. Let MAD (median absolute deviation) denote the median of $|X_1 - M|, \dots, |X_n - M|$, where M is the usual sample median. Let $MADN = MAD/0.6745$. Under normality, $MADN$ estimates the standard deviation. Let i_1 be the number of observations X_i for which $(X_i - M)/MADN < -1.28$, and let i_2 be the number of observations such that $(X_i - M)/MADN > 1.28$. The one-step M-estimate of location is

$$\tilde{\mu} = \frac{1.28(MADN)(i_2 - i_1) + \sum_{i=i_1+1}^{n-i_2} X_{(i)}}{n - i_1 - i_2}.$$

It has the highest possible breakdown point, 0.5, and has good efficiency under normality. To describe the Harrell and Davis estimate of the median, let U be a random variable having a beta distribution with parameters $(n+1)q$ and $(n+1)(1-q)$. Let

$$W_i = P\left(\frac{i-1}{n} \leq U \leq \frac{i}{n}\right).$$

Then the Harrell–Davis estimate of the q th quantile is

$$\hat{\theta}_q = \sum_{i=1}^n W_i X_{(i)}.$$

Although the Harrell–Davis estimator estimates the median when $q = 0.5$, a criticism is that it has a breakdown point of only $1/n$ because all of the order statistics get a positive weight. However, a possible appeal of the Harrell–Davis is that it can be more efficient than the usual sample median when dealing with distributions that are relatively light tailed.

For convenience, the percentile bootstrap method is described when dealing with the one-step M-estimator. It is readily extended to situations where the sample median and the Harrell–Davis estimator are used.

To apply the percentile bootstrap method, first take a bootstrap sample from the j th group by randomly sampling with replacement n_j values from $X_{1j}, \dots, X_{n_j j}$. Compute the M-estimate of location based on this bootstrap sample yielding $\tilde{\mu}_j^*$. When comparing groups j and k , let $D^* = \tilde{\mu}_j^* - \tilde{\mu}_k^*$. Repeat this process B times yielding D_b^* ($b = 1, \dots, B$). Let

$$A = \frac{1}{B} (\sum I(D_b^* < 0) + 0.5I(D_b^* = 0.5)),$$

where the indicator function $I(D_b^* < 0) = 1$ if $D_b^* < 0$ otherwise $I(D_b^* < 0) = 0$. The indicator function $I(D_b^* = 0.5)$ is defined in an analogous function. Then $p^* = 2\min(A, 1 - A)$ is a (generalized) p-value (Liu & Singh, 1997).

In principle, method Y for testing (1) based on trimmed means could be extended to the situation where the percentile bootstrap method is used. But a practical concern is the execution time needed to compute the matrix $\hat{\mathbf{P}}$. The approach used here is to reject the j th hypothesis if $p^* \leq c_j$, where c_j is the critical p-value used by method Y with 20% trimming. The next section reports simulation results on how well this approach performs.

SIMULATIONS RESULTS

Simulations were used to assess the actual FWE rate when dealing with non-normal distributions and there is heteroscedasticity. The focus here is on $J = 4$ groups, but some results for $J = 8$ are reported as well. For each group, data were generated from one of four distributions: normal, symmetric and heavy-tailed, skewed and light-tailed, and skewed and heavy-tailed. More precisely, data were generated from g-and-h distributions [11], which arise as follows. Let Z be a random variable having a standard normal distribution. Then

$$\frac{\exp(gZ) - 1}{g} \exp(hZ^2/2),$$

if $g > 0$ and

$$Z \exp(hZ^2/2),$$

if $g = 0$ has a g-and-h distribution, where g and h are parameters that determine the first four moments. The four distributions used here are the standard normal ($g = h = 0$), a symmetric heavy-tailed distribution ($h = 0.2, g = 0$), an asymmetric distribution with relatively light tails ($h = 0, g = 0.2$), and an asymmetric distribution with heavy tails ($g = h = 0.2$). Table 1 summarizes the skewness (κ_1) and kurtosis (κ_2) of these distributions.

Table 1: Some properties of the g-and-h distribution.

g	h	κ_1	κ_2	
0.0	0.00	3.00		0.0
0.2	0.00	21.46		0.0
0.0	0.61	3.68		0.2
0.2	2.81	155.98		0.2

The sample sizes were taken to be $(n_1, n_2, n_3, n_4) = (20, 20, 20, 20)$ and $(15, 20, 30, 40)$. Heteroscedasticity was modeled by centering the generated data so that the population measure of location is zero and then multiplying the data for the j th group by σ_j . Here $(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ was taken to be $(1, 1, 1, 1)$ and $(1, 2, 3, 4)$. For unequal sample sizes, $(4, 3, 2, 1)$ was used as well. For method Y, 5000 replications were used to estimate the actual probability of at least one Type I error when testing at the 0.05 level. When using the percentile bootstrap method, 2000 replications were used with number of bootstrap samples taken to be $B = 2000$.

Table 2 shows the simulation results. The columns headed by PB are the results using the percentile bootstrap method, MED indicates that the usual sample median was used, M-EST is the one-step M-estimator and HD is the Harrell and Davis (1982) estimate of the median. The column headed by Y is method Y with a 20% trimmed mean. Note that for unequal sample sizes, no results are reported based on the one-step M-estimator. The reason is that with a sample size of fifteen, bootstrap samples can have MAD=0 resulting in division by zero when computing the M-estimate of location.

Although the seriousness of a Type I error can depend on the situation, Bradley [4] suggests that as a general guide, when testing at the 0.05 level, the actual level should be between 0.025 and 0.075. As can be seen, this criterion is met for all of the situations considered. All of the estimates are less than 0.070 and the smallest estimate is 0.036. If the goal is to have FWE less

than 0.060, method PB with the median was best among the situations considered. But for light-tailed distributions, the standard error of the median can be relatively large compared to the other location estimators used here.

Table 2: Estimates of α when testing at the 0.05 level using method PB as well as method Y, $J = 4$. The columns headed by MED, M-EST and HD are results for the median, one-step M-estimator and the Harrell–Davis estimator, respectively, based on method PB. The column headed by Y is method Y based on Yuen’s method for a 20% trimmed mean.

n	g	h	σ	MED	M-EST	HD	Y	
(20, 20, 20, 20)	0.0	0.0	(1, 1, 1, 1)	0.036	0.050	0.045	0.050	(20,
20, 20, 20)	0.0	0.2	(1, 1, 1, 1)	0.036	0.049	0.049	0.044	(20,
20, 20, 20)	0.2	0.0	(1, 1, 1, 1)	0.036	0.067	0.042	0.046	(20,
20, 20, 20)	0.2	0.2	(1, 1, 1, 1)	0.036	0.064	0.042	0.039	(15,
20, 30, 40)	0.0	0.0	(1, 1, 1, 1)	0.041	...	0.057	0.049	(15,
20, 30, 40)	0.0	0.2	(1, 1, 1, 1)	0.041	...	0.050	0.042	(15,
20, 30, 40)	0.2	0.0	(1, 1, 1, 1)	0.041	...	0.047	0.045	(15,
20, 30, 40)	0.2	0.2	(1, 1, 1, 1)	0.041	...	0.047	0.040	(20,
20, 20, 20)	0.0	0.0	(1, 2, 3, 4)	0.046	0.068	0.052	0.067	(20,
20, 20, 20)	0.0	0.2	(1, 2, 3, 4)	0.048	0.066	0.051	0.060	(20,
20, 20, 20)	0.2	0.0	(1, 2, 3, 4)	0.045	0.065	0.051	0.059	(20,
20, 20, 20)	0.2	0.2	(1, 2, 3, 4)	0.046	0.062	0.049	0.052	(15,
20, 30, 40)	0.0	0.0	(1, 2, 3, 4)	0.037	...	0.055	0.049	(15,
20, 30, 40)	0.0	0.2	(1, 2, 3, 4)	0.038	...	0.050	0.043	(15,
20, 30, 40)	0.2	0.0	(1, 2, 3, 4)	0.034	...	0.055	0.044	(15,
20, 30, 40)	0.2	0.2	(1, 2, 3, 4)	0.035	...	0.051	0.040	(15,
20, 30, 40)	0.0	0.0	(4, 3, 2, 1)	0.059	...	0.059	0.059	(15,
20, 30, 40)	0.0	0.2	(4, 3, 2, 1)	0.059	...	0.061	0.053	(15,
20, 30, 40)	0.2	0.0	(4, 3, 2, 1)	0.052	...	0.067	0.055	(15,
20, 30, 40)	0.2	0.2	(4, 3, 2, 1)	0.049	...	0.055	0.048	

Additional simulations were run with $p = 8$ where all four groups have a common sample size of 20. Under homoscedasticity, the results were similar to those in Table 2. Note that for the heteroscedastic case, method Y has the largest estimate, 0.067, which occurs when $g = h = 0$ and $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1, 2, 3, 4)$. For $p = 8$, if $(\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7, \sigma_8) = (1, 2, 3, 4, 1, 2, 3, 4)$, the estimate was 0.072 using method Y. Using method PB with a 20% trimmed mean does not improve matters, the estimate is now 0.073. Using PB with the usual median, the estimate is 0.059.

Some Illustrations

The proposed methods are illustrated using data from the Well Elderly II study [5], which was generally aimed at improving the physical and mental well-being of older adults. A portion of this study measured depressive symptoms (CESD) after intervention. Here, the focus is on CESD measures for five groups based on education: less than high school, high school graduate. Some college or technical school, four years of college completed and post-graduate study. The sample sizes are 92, 68, 113, 33 and 20, respectively. The first group had the largest 20% trimmed mean, 15.07.

Here are the results using method Y returned by an R function that performs it when testing at the 0.05 level:

Est.Best	Grp	Est	Dif	ci.low	ci.up	p.value	p.crit
15.07	2	11.17	3.90	1.29	6.52	<0.001	0.064209
3	8.73	6.34	3.20	9.49	<0.001	0.027394	15.07
4		7.71	7.36	2.96	11.76	<0.001	0.011883
15.07	5	6.25	8.82	2.83	14.80	<0.001	0.003277

The second column indicates the group being compared to the one with largest estimate. The second line, for example, indicates that group 2 had the second largest estimate, which is 11.17. This is followed by the difference between the trimmed means, a confidence interval for the difference, and the p-value. The confidence intervals are designed to have simultaneous probability coverage $1 - \alpha$ by using Yuen's method at the c_j level when testing the j th hypothesis. The last column indicates the c_j values, which indicate that all four hypotheses are rejected. The corresponding c_j values when testing at the 0.01 level are 0.017500, 0.005095, 0.002155, and 0.000421. That is, method Y rejects at the 0.01 as well and indicates that the first group has the largest population trimmed mean. Method PB gives similar results using the sample median, the one-step M-estimator and the Harrell–Davis estimator. In this particular case, if Hochberg's method is used to adjust the p-values via the R function `p.adjust`, with the goal of controlling FWE, again these four hypotheses would be rejected at the 0.01 level.

Method Y can have more power than Hochberg's method based on the p-values stemming from Yuen's test. As an illustration, consider the case where $J = 4$, the distributions are normal, $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1, 1, 1, 1)$, $(n_1, n_2, n_3, n_4) = (20, 20, 20, 20)$, and the corresponding population trimmed means are 1, 0, 0 and 0. The probability of getting three significant results, that is making a decision about which group has the largest trimmed mean, using Hochberg's method is 0.610. Using method Y it is 0.853. If instead the trimmed means are 1, 0.7, 0, 0 and 0, now the probability of three significant results using Hochberg's method 0.126 versus 0.775 using method Y. Using the methods derived by Hommel and Rom yield very similar results.

CONCLUDING REMARKS

In summary, all indications are that all four methods control the Type I error probability reasonably well based on Bradley's criterion. If the goal is to avoid a FWE rate greater than 0.06, method PB based on the usual sample median is best. If the goal is avoid a FWE rate less than 0.04, method Y or method PB coupled with the Harrell–Davis estimator are good choices. In principle, method Y can be used with less trimming or no trimming at all. However, as the amount of trimming gets close to zero, it becomes increasingly difficult to control the Type I error probability. Wilcox [21] summarizes known concerns. In terms of maximizing power, the best method depends on the unknown characteristics of the distributions being compared.

An open issue is how methods Y and PB might be extended to related issues. For example, when comparing measures of location associated with dependent groups, how might methods Y and PB be generalized to deal with this situation? If they are used by simply ignoring the dependence among the groups, the actual level can be substantially less than the nominal level. As is evident, there are other related issues.

Finally, the R function `anc.best.PV` applies method Y and is stored in the file `Rallfun-v40` located at <https://dornsife.usc.edu/cf/labs/wilcox/wilcox-faculty-display.cfm>. By default, it uses 5000 replications when computing $\hat{P}$. This can be altered via the argument `iter`. The amount of trimming is controlled by the argument `tr` and defaults to 0.2. The R function `anc.bestpb.PV` performs method PB with the number of bootstrap samples controlled by the argument `nboot`, which defaults to 2000. The argument `loc.fun` indicates which measure of location is used. The output includes a p-value related to making a decision about whether the group with the largest measure of location can be indentified. This is done by computing $c_1, \dots, c_{J-1}$ for FWE rates 0.001(.001)0.1(0.01)0.99 and determining the smallest rate for which all $J - 1$ hypotheses are rejected. If the $c_1, \dots, c_{J-1}$ values are known, the functions `anc.best` and `anc.bestpb` can be used instead, which results in lower execution time.

References

- [1] Bechhofer, R.E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Annals of Mathematical Statistics*, 25, 16–39.
- [2] Bechhofer, R.E., Dunnett, C. W. & Sobel, M. (1954) A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance *Biometrika*, 41, 170–176.
- [3] Bechhofer, R.E., Kiefer, J. & Sobel, M. (1968). *Sequential Identification and Ranking Procedures*. Chicago, Ill: University of Chicago Press.
- [4] Bradley, J. V. (1978) Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144–152. doi.org/10.1111/j.2044-8317.1978.tb00581.x
- [5] Clark, F., Jackson, J., Carlson, M., Chou, C.-P., Cherry, B. J., Jordan-Marsh, M., Knight, B. G., Mandel, D. Blanchard, J., Granger, D. A., Wilcox, R. R., Lai, M. Y., White, B., Hay, J., Lam, C., Marterella, A., & Azen, S. P. (2012). Effectiveness of a lifestyle intervention in promoting the well-being of independently living older people: results of the Well Elderly 2 Randomise Controlled Trial. *Journal of Epidemiology and Community Health*, 66, 782–790. doi:10.1136/jech.2009.099754
- [6] Dudewicz, E.J. & Dalal, S. R. (1975). Allocation of observations in ranking and selection with unequal variances. *Sankhya: Indian Journal of Statistics*, 37B, 27–78.
- [7] Gibbons, J., Olkin, I. & Sobel, M. (1987). *Selecting and Ordering Populations: A New Statistical* New York: Wiley.
- [8] Gupta, S. S. & Panchapakesan, S. (1987). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations* SIAM.
- [9] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (1986). *Robust Statistics*. New York: Wiley.
- [10] Harrell, F. E. & Davis, C. E. (1982). A new distribution-free quantile estimator. *Biometrika*, 69, 635–640.
- [11] Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distribution. In D. Hoaglin, F. Mosteller & J. Tukey (Eds.) *Exploring Data Tables Trends and Shapes*. New York: Wiley, pp. 461–515.
- [12] Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75, 800–802.
- [13] Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75, 383–386.
- [14] Huber, P. J. & Ronchetti, E. (2009). *Robust Statistics*, 2nd Ed. New York: Wiley.
- [15] Liu, R. G. & Singh, K. (1997). Notions of limiting P values based on data depth and bootstrap. *Journal of the American Statistical Association*, 92, 266–277. DOI: 10.2307/2291471
- [16] Mukhopadhyay, N. & Solanky. T. (1994). *Multistage Selection and Ranking Procedures: Second Order Asymptotics*. New York: Marcel Dekker.

-
- [17] Paulson, E. (1964). A sequential procedure for selecting the population with the largest mean from k normal populations. *Annals of Mathematical Statistics*, 35, p. 174–180
- [18] Rinott, Y. (1978). On two-stage selection procedures and related probability-inequalities. *Communications in Statistics -Theory and Methods*, A7(8): 799–811.
- [19] Rom, D. M. (1990). A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika*, 77, 663–666.
- [20] Staudte, R. G. & Sheather, S. J. (1990). *Robust Estimation and Testing*. New York: Wiley.
- [21] Wilcox, R. R. (2022). *Introduction to Robust Estimation and Hypothesis Testing*. 5th Ed. San Diego CA: Academic Press.
- [22] Yuen, K. K. (1974). The two sample trimmed t for unequal population variances. *Biometrika*, 61, 165–170.