APPROXIMATE QUANTILES FOR THE MULTIVARIATE STUDENTIZED RANGE IN THE CASE OF THREE UNEQUAL GROUPS[1]

Otto Schwalb and James R. Thompson[2]
Department of Statistics
Rice University
6100 Main Street MS-138
Houston, TX 77005

ABSTRACT

In this paper, we develop equations which provide approximate quantiles for the multivariate studentized range in the case of three unequal groups. These equations are ready to use and are available via a website. In the equal groups case, our approximations are more accurate than existing approximations by a factor of ten. In the unequal groups case, there are no direct competitors to our approximations.

INTRODUCTION

The theory of multiple comparisons has a rich history in statistics. There is usually no debate about which multiple comparisons procedure is best for a given situation. In the multivariate case, for example, it is known that the T^2_{max} procedure (the simplest case of which is originally due to Roy (1953) and generalizations of which are attributed to Krishnaiah (1969)) will always outperform the equivariant procedures (among which are the procedures based upon Roy's root, the Pillai-Bartlett trace, the Hotelling-Lawley trace, and Wilks' Lambda) in cases where there are a fixed number of contrasts specified in advance and either confidence intervals or confidence regions (for particular parametric families) are desired. This claim is strong indeed and is perhaps not widely known, but it has been established for quite some time now. It requires more effort to demonstrate than just citing a few references here in the introduction (and some of the results required are fairly obscure), and so will be discussed below. That the T^2_{max} procedure possesses such outstanding properties is completely analogous to the relationship between Scheffe's procedure and the studentized range procedure in ANOVA. The reason that the T^2_{max} procedure has not received much attention seems to be due to the problems associated with obtaining its distribution in closed form. Siotani (1992) is one of the major investigators of the T^2_{max} procedure. His breakthrough in providing an equation for approximate quantiles for the general case of equal groups (after the closed form solution has remained an unsolved problem for nearly 40 years) is remarkable and should hopefully serve to bring renewed attention to a procedure which is so useful and easy to understand.

In this paper, we develop equations which provide approximate quantiles for the multivariate studentized range in the case of one-way MANOVA with three unequal groups. The equations apply to the following situations:

- quantiles: 0.8, 0.825, 0.85, ..., 0.925, 0.95, 0.975, 0.99 (e.g. for a significance level of $\alpha = 0.1$, we desire the $1 - \alpha = 0.9$ quantile)
- dimensions: 1, 2, 3, 4, 5
- group sizes: arbitrary group sizes between 4 and 100 (subject to the restriction that the largest group be no more than 4 times the size of the smallest group).

A JavaScript available at the web site:

[1]Approved for public release; distribution is unlimited.

[2]Otto Schwalb is a graduate student in the Statistics Department, Rice University, 6100 Main Street, MS-138, Houston, TX, 77005 (E-mail:schwalb@stat.rice.edu). James R. Thompson is Professor, Department of Statistics, Rice University, 6100 Main Street, MS-138, Houston, TX, 77005 (E-mail:thomp@stat.rice.edu).

http://www.stat.rice.edu/~schwalb

allows the interested reader easy access to these quantiles. The importance of these approximate quantiles is that, until now, no quantiles have been available for the procedure in the case of unequal groups. Our approximations also apply to the case of equal groups, and, for the equal groups cases we consider, the accuracy of our approximations improves upon the accuracy reported by others by a factor of ten.

Section 2 treats the theory associated with the T^2_{max} procedure in the general case, as well as the special cases of two and three groups in 1-way MANOVA. Section 2 also demonstrates the claims made above concerning the performance of the T^2_{max} procedure with respect to the equivariant procedures. In section 3, the practical utility of the procedure is demonstrated with an example using the Fisher-Anderson Iris data (1939). The example illustrates how the T^2_{max} procedure is a strongly indicated choice in many realistic situations. Section 4 discusses the development of our equations. Most of the details are omitted from this section for the sake of brevity, but may be obtained from the first author. The details are straightforward, however.

THEORY

OVERVIEW

The derivation of the T^2_{max} procedure to follow is rather general, that is to say, very little of the theory to follow is specific to the 1-way case or the case of all pairwise comparisons. Parts of what follow can be found scattered throughout the literature, but the results are certainly not easy to find. For example, the results we synthesize on general confidence regions are very obscure. They are certainly beyond the reach of anyone who simply needs to use the procedures in an applied situation, but might have very little time to search the literature. Hopefully, the presentation here will make these results more accessible to a wider audience.

The MANOVA setup starts with a matrix Y which is $n \times p$ and whose rows are each distributed independently as p-variate normal vectors with common unknown covariance matrix Σ. It is further assumed that $E(Y) = X\beta$ where X is a design matrix of order $n \times l$ and β is the $l \times p$ matrix of unknown parameters. We assume that X has full column rank. We first define Roy's root and the related multivariate test statistics. We wish to test the hypothesis $H_0 : C\beta = 0$ where C is $r \times l$ and has full row rank. Define the hypothesis and error matrices (respectively) as

$$H = \hat{\beta}^T C^T W^{-1} C\hat{\beta} \quad \text{and} \quad E = Y^T \left(I - X(X^T X)^{-1} X^T \right) Y$$

where $\hat{\beta} = (X^T X)^{-1} X^T Y$ and $W = C(X^T X)^{-1} C^T$. Let the eigenvalues of HE^{-1} be denoted by $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$. Then Roy's root is λ_1, the Hotelling-Lawley trace is $\sum_{i=1}^p \lambda_i$, the Pillai-Bartlett trace is $\sum_{i=1}^p \lambda_i/(1+\lambda_i)$, and Wilks' lambda is $\prod_{i=1}^p 1/(1+\lambda_i)$. These statistics can be defined in similar ways using the eigenvalues of $H(E+H)^{-1}$, which tends to make the notation in the literature sometimes confusing. In fact, Roy's root seems to be defined in different ways depending on the author. We follow Wijsman (1979) and Kres (1983) in defining Roy's root as λ_1. For convenience, we will subsequently refer to these four statistics as the equivariant statistics.

For the T^2_{max} procedure, however, the hypothesis is stated in a different form as $H_0 : L\beta = 0$, where L is not required to have full row rank. The matrix $L\beta$ is an explicit representation of each and every contrast of interest for the problem at hand. Since anything of the form $a^T C\beta$, $\forall\, a \in \Re^r$, is protected by procedures based on the equivariant statistics, the family of linear combinations considered by those procedures is vast in comparison with the family of linear combinations considered by the T^2_{max} procedure. It should be apparent that the null hypothesis considered by the T^2_{max} procedure and the null hypothesis for the classical formulation are not equivalent. This might seem confusing, however, the problem of multiple comparisons deals with obtaining the best (usually in the sense of the smallest) set of confidence regions for a given parametric family. Within the classical framework then, we start with the hypothesis $H_0 : C\beta = 0$. Once a parametric family is specified (e.g. $a^T C\beta$, $\forall\, a \in \Re^r$), the discussion of competing multiple comparison procedures follows next. Such a discussion cannot be carried out prior to specifying the parametric family of interest (see Wijsman (1979) for instance). With these subtleties in mind, the T^2_{max} procedure tests the null hypothesis that $H_0 : L\beta = 0$ vs $H_1 : L\beta \neq 0$ where L is a $t \times l$ matrix of contrasts. For example, in the case of 1-way MANOVA with 3 groups, $l = 3$ (the number of groups) and $t = 3$ (the number of pairwise comparisons). Applying a variation on the union-intersection approach (originally due to Roy (1953)), one sees that:

$$\{H_0 : L\beta = 0\} \Leftrightarrow \bigcap_{i=1}^{t} \{H_{0i} : L_{[i,\ ]}\beta = 0\}$$

where $L_{[i,\ ]}$ represents the ith row of the contrast matrix L (e.g. in the case of 1-way MANOVA with 3 groups under the "cell means" formulation of the design matrix X, $L_{[i,\ ]} = (1,\ -1,\ 0)$ compares group 1 and group 2). Define $\delta_i^T := L_{[i,\ ]}\beta$. As usual, the least squares estimators $\hat{\beta}$ are a linear combination of Y. By the Gauss-Markov Theorem, the minimum variance unbiased estimators for the δ_i's are $\hat{\delta}_i^T = L_{[i,\ ]}\hat{\beta}$ which are also linear combinations of Y. It is easily seen that

$$\hat{\delta}_i \sim N_p(\delta_i, w_{ii}\Sigma)$$

where w_{ii} is the ith diagonal element of $L(X^TX)^{-1}L^T$. Let $\delta_i^* = \delta_i/\sqrt{w_{ii}}$ and $\hat{\delta}_i^* = \hat{\delta}_i/\sqrt{w_{ii}}$. Then

$$\hat{\delta}_i^* \sim N_p(\delta_i^*, \Sigma)$$

so that

$$T_i^2 = (\hat{\delta}_i^* - \delta_i^*)^T \left(\frac{E}{n-l}\right)^{-1} (\hat{\delta}_i^* - \delta_i^*)$$

is distributed as Hotelling's T^2 where E is the error matrix (defined above). The T^2_{max} statistic is then defined as

$$T^2_{max} = \max_{1 \leq i \leq t} T_i^2$$

The procedure rejects H_0 if any of the t Hotelling T^2's exceed $T^2_{max}(\alpha)$, where $T^2_{max}(\alpha)$ denotes the $1-\alpha$ quantile of the distribution of T^2_{max}, i.e. the value such that $P(T^2_{max} \leq T^2_{max}(\alpha)) = 1-\alpha$. As will be discussed immediately, each T^2 which exceeds $T^2_{max}(\alpha)$ is flagged as a significant contrast. It should be noted here that we will speak of the comparison of T_i^2 with $T^2_{max}(\alpha)$, but our equations actually give values for $T_{max}(\alpha)$ due to the conventions which are already established for the univariate studentized range, $q(\alpha)$, and the equal-groups multivariate studentized range, $R_{MAX}(\alpha)$. Thus, in practice, one would need to take square roots of the T_i^2's before comparing them to $T_{max}(\alpha)$.

The T^2_{max} procedure has the intuitive appeal that it rejects iff $T_i^2 > T^2_{max}(\alpha)$ for some i, $i = 1, \ldots, t$, which occurs iff at least one of the p-dimensional confidence ellipsoids does not contain the 0-vector. That this follows is a simple consequence of the fact that $T^2_{max}(\alpha)$ is chosen to satisfy

$$P_{H_0}[\ T_i^2 \leq T^2_{max}(\alpha), \quad i = 1, \ldots, t\ ] = 1 - \alpha$$

which is the same as writing

$$P_{H_0}[\ (\hat{\delta}_i^* - \delta_i^*)^T \left(\frac{E}{n-l}\right)^{-1} (\hat{\delta}_i^* - \delta_i^*) \leq T^2_{max}(\alpha), \quad i = 1, \ldots, t\ ] = 1 - \alpha \quad .$$

Beyond simply rejecting, however, each T_i^2 which exceeds $T^2_{max}(\alpha)$ will be flagged as significant. For example, if $t = 3$ and we are given data and find that T_1^2 and T_2^2 exceed $T^2_{max}(\alpha)$ but T_3^2 does not, then we would flag δ_i^* and δ_2^* as the culprits. The practical interpretation of this is subject to the limitations of any confidence statement as follows. If repeated samples are taken and the 3 ellipsoids are constructed each time, then $(1-\alpha)\times 100\%$ of the time *all* 3 of these ellipsoids will contain the true δ_i^*'s, $i=1,2,3$. Equivalently, $(1-\alpha)\times 100\%$ of the time, all 3 of the T_i^2's will reflect the location of the true δ_i^*'s, $i=1,2,3$, as indicated by their relationship to the $T^2_{max}(\alpha)$ quantile. For example, if δ_1^* and δ_2^* were truly non zero, but δ_3^* was zero, then, under repeated sampling, $(1-\alpha)\times 100\%$ of the time we would find that $T_1^2 > T^2_{max}(\alpha)$, $T_2^2 > T^2_{max}(\alpha)$, *and* $T_3^2 < T^2_{max}(\alpha)$.

In addition to the procedure described above, one can also construct confidence intervals for all pairwise contrasts. Specifically, it can be shown - see Krishnaiah (1969, p 126) - that confidence intervals for the T^2_{max} procedure are given by

$$c^T\delta_i \in c^T\hat{\delta}_i \pm \sqrt{T^2_{max}(\alpha)\, c^T E c\, w_{ii}/(n-l)} \quad , \quad \forall\, c \in \Re^p$$

Confidence intervals for the pairwise differences on each variable can be obtained by substituting $c_j^T = (0,\cdots,0,1,0,\cdots,0)$, where the 1 is in the jth position, for c in the previous equation. This leads to confidence intervals of the form

$$\delta_{ij} \in \hat{\delta}_{ij} \pm \sqrt{T^2_{max}(\alpha)\, e_{jj}\, w_{ii}/(n-l)} \quad , \quad \begin{array}{l} i=1,\ldots,t \\ j=1,\ldots,p \end{array} \tag{1}$$

where δ_{ij} is the jth component of δ_i and e_{jj} is the jth diagonal element of E. Notice in constructing these intervals that whatever is to be gained over a 1-d battery of univariate procedures (the expression "1-d battery" will be described below in the example using the Iris data) depends completely upon the distribution of T^2_{max}. This follows since e_{jj} is just the univariate SSE on the jth dependent variable. The covariance structure within the data is largely reflected in the off-diagonal elements of E, and the off-diagonal elements completely dominate the tilting of the p-dimensional ellipsoids. But the off-diagonal elements are not used in (1), hence important information is being ignored if one looks at confidence intervals (along the dimensions of the variables as given) alone. This particular issue lies at the heart of the example we give involving the Iris data (below). We will revisit it once we get there.

CONFIDENCE INTERVALS

For the case of a fixed number of contrasts specified in advance, Krishnaiah (1969, p. 131) demonstrates that the T^2_{max} confidence intervals are smaller than those for Roy's root (analogous to the ellipsoid discussion below, the T^2_{max} confidence intervals provide protection for the parametric family of $a_i^T C\beta b$, for the pre-specified $\Re^r$ vectors $a_1,\ldots,a_t$ and $\forall\, b\in\Re^p$). In turn, Wijsman (1979) demonstrates that the confidence intervals based upon Roy's root are the smallest among the class of all equivariant procedures which provide protection for the parametric family $a^T C\beta b$, $\forall\, a\in\Re^r$ and $\forall\, b\in\Re^p$. This class of equivariant procedures includes those based on Wilks' Lambda, the Hotelling-Lawley trace, and the Pillai-Bartlett trace. Hence the confidence intervals based upon T^2_{max} are always shorter than those based upon these equivariant procedures when the number of contrasts considered is fixed and specified in advance.

CONFIDENCE REGIONS

We start this section by stating (with a reference) why the T^2_{max} ellipsoids are smaller than the ellipsoids based upon Roy's root. We also illustrate the relationship between Roy's root and T^2_{max} which dramatically emphasizes the conservative nature of Roy's procedure. After that, we cite the results which show that the Roy's root ellipsoids are the smallest (for all practical purposes) among the equivariant class of MANOVA procedures which provide protection for the parametric family $a^T C\beta$, $\forall\, a\in\Re^r$. The conclusion then is that the T^2_{max} ellipsoids will always be smaller than ellipsoids based on equivariant MANOVA procedures in all practical cases (provided we make additional mild restrictions to the equivariant class under consideration; one such sufficient condition is that the confidence regions in p space be convex).

The fact that the ellipsoids (for a fixed number of contrasts specified in advance) for the T^2_{max} procedure are always smaller than the ellipsoids based on Roy's root is an immediate consequence of the fact that $T^2_{max}(\alpha)/(n-l) \leq \lambda_\alpha$ (see Krishnaiah (1969, p. 131)). In the presentation below, we illustrate the conservative nature of the equivariant procedures via a reexpression of the Roy's root ellipsoidal simultaneous confidence regions. For clarity here, we are using λ_α to denote the $1-\alpha$ quantile of Roy's root (i.e. $P(\lambda_1 \leq \lambda_\alpha) = 1-\alpha$ where we defined λ_1 at the beginning of the THEORY section); our λ_α coincides with the λ_α of Wijsman (1979) and Kres (1983).

The ellipsoids for Roy's root are given by (see Wijsman (1979), equation (2.4))

$$(a^T C\hat{\beta} - a^T C\beta)E^{-1}(a^T C\hat{\beta} - a^T C\beta)^T \leq \lambda_\alpha a^T C(X^T X)^{-1} C^T a \quad , \quad \forall\, a \in \Re^r \tag{2}$$

where the C matrix is the full rank version of the contrast matrix for H_0 as described previously, and r is the rank of H_0 (for the 1-way case $r = l-1$). Now let $\lambda^*_\alpha = \lambda_\alpha(n-l)$. Rewrite (2) as

$$\frac{1}{a^T C(X^T X)^{-1} C^T a}\,(a^T C\hat{\beta} - a^T C\beta)\left(\frac{E}{n-l}\right)^{-1}(a^T C\hat{\beta} - a^T C\beta)^T \leq \lambda^*_\alpha \quad , \quad \forall\, a \in \Re^r$$

which is equivalent to

$$\sup_{a\in\Re^r} T^2_a \leq \lambda^*_\alpha \quad ,$$

where

$$T^2_a := \frac{1}{a^T C(X^T X)^{-1} C^T a}\,(a^T C\hat{\beta} - a^T C\beta)\left(\frac{E}{n-l}\right)^{-1}(a^T C\hat{\beta} - a^T C\beta)^T \quad .$$

We have shown in a different section that the T^2_{max} ellipsoids are given by

$$(\hat{\delta}^*_i - \delta^*_i)^T\left(\frac{E}{n-l}\right)^{-1}(\hat{\delta}^*_i - \delta^*_i) \leq T^2_{max}(\alpha) \quad , \quad i = 1, \ldots, t$$

which after some rearranging can be written

$$\frac{1}{a_i^T C(X^T X)^{-1} C^T a_i}\,(a_i^T C\hat{\beta} - a_i^T C\beta)\left(\frac{E}{n-l}\right)^{-1}(a_i^T C\hat{\beta} - a_i^T C\beta)^T \leq T^2_{max}(\alpha) \quad , \quad i = 1, \ldots, t \tag{3}$$

where a_i^T is a row vector in $\Re^r$ chosen so that $a_i^T C = L_{[i,\ ]}$ (it is always possible to choose such an a_i^T, since given L, one can choose C to be a matrix such that all the rows of L are in the span of C and C has full row rank). Rewrite (3) as

$$\max_{a\in\{a_1,a_2,\ldots,a_t\}} T^2_a \leq T^2_{max}(\alpha)$$

where

$$T^2_{a_i} := \frac{1}{a_i^T C(X^T X)^{-1} C^T a_i}\,(a_i^T C\hat{\beta} - a_i^T C\beta)\left(\frac{E}{n-l}\right)^{-1}(a_i^T C\hat{\beta} - a_i^T C\beta)^T$$

And, expressed in this fashion, it becomes obvious that

$$T^2_{max} = \max_{a\in\{a_1,a_2,\ldots,a_t\}} T^2_a \leq \sup_{a\in\Re^r} T^2_a \quad . \tag{4}$$

The relationship as expressed in (4) clearly shows that the T^2_{max} procedure will always outperform Roy's procedure for the situation of a fixed number of contrasts specified in advance. It also emphasizes dramatically the conservative nature of the equivariant procedures, particularly when the number of contrasts t is very small.

Finally, Wijsman (1979) demonstrates that the simultaneous ellipsoidal confidence regions based upon Roy's root are the smallest among all equivariant procedures (which provide protection for the family $a^T C\beta, \forall\, a \in \Re^r$) when $p < l - r$, where l is the space spanned by the columns of X and r is the number of constraints in the null hypothesis. For the case of 1-way MANOVA with 3 groups, for example, $l = 3$ and $r = 2$, since the hypothesis testing in this case essentially asks whether the 3 p-variate location vectors for the groups can be adequately described by a single p-variate location vector plus noise (see the canonical form of the MANOVA model in Lehmann (1986)). Wijsman (1979) demonstrates further that the Roy's root ellipsoids are also the smallest in the equivariant class of procedures in the case where $p \geq l - r$, provided some very mild conditions are imposed upon the shapes of the confidence regions desired. Requiring the confidence regions to be convex is one example of such a mild condition. Another example is requiring the confidence regions to be connected and contain their "center" (i.e. contain the point estimate for the mean vector). From an applied point of view, it would be difficult to characterize these additional conditions as restrictions at all, and it would be fair to say that Wijsman's results bring a rich history of discussion on the subject to a close.

THE CASE OF TWO GROUPS

When there are two groups, there is only one contrast to consider. Let n_i be the sample size of group i and $n = n_1 + n_2$. Here $T^2_{max} = n_1 n_2/(n_1 + n_2)(\bar{x}_1 - \bar{x}_2)^T (n-2)E^{-1}(\bar{x}_1 - \bar{x}_2)$, which one may easily verify is the 2-sample Hotelling's T^2. Hence, $T^2_{max}(\alpha, n, p) = p(n-2)/(n-p-1)F_{p,n-p-1}(\alpha)$ and the quantiles are available in closed form. If $T^2_{max} > T^2_{max}(\alpha)$, then the procedure rejects.

THE CASE OF THREE GROUPS

In the case of three groups, there are 3 contrasts, and the distribution is no longer available in closed form although there has been a great deal of effort in that direction (see Siotani (1959) and (1992)). Using the "cell means" formulation of the design one is led to $(X^T X)^{-1} = diag(1/n_1, 1/n_2, 1/n_3)$ and the contrast matrix

$$L = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$$

which leads to

$$L(X^T X)^{-1} L^T = \begin{pmatrix} 1/n_1 + 1/n_2 & 1/n_1 & -1/n_2 \\ 1/n_1 & 1/n_1 + 1/n_3 & 1/n_3 \\ -1/n_2 & 1/n_3 & 1/n_2 + 1/n_3 \end{pmatrix}$$

So in this simple case, $w_{11} = 1/n_1 + 1/n_2$, $w_{22} = 1/n_1 + 1/n_3$, and $w_{33} = 1/n_2 + 1/n_3$. Also, $\delta_1 = \mu_1 - \mu_2$, $\delta_2 = \mu_1 - \mu_3$, and $\delta_3 = \mu_2 - \mu_3$. It is also clear (this is similar to the general case already discussed) that $\hat{\delta}_1 = (\bar{x}_1 - \bar{x}_2) \sim N_p(\delta_1, w_{11}\Sigma)$. Rewrite this as $\hat{\delta}^*_1 \sim N_p(\delta^*_1, \Sigma)$ (where $\hat{\delta}^*_1 := \hat{\delta}_1/\sqrt{w_{11}}$). This yields

$$T^2_1 = (\hat{\delta}^*_1 - \delta^*_1)^T \left(\frac{E}{n-l}\right)^{-1} (\hat{\delta}^*_1 - \delta^*_1) = \frac{1}{w_{11}} (\bar{x}_1 - \bar{x}_2)^T \left(\frac{E}{n-l}\right)^{-1} (\bar{x}_1 - \bar{x}_2)$$

Similarly for T^2_2 and T^2_3. Then T^2_{max} is just the largest of these 3 statistics. The procedure rejects, then, if any of the three T^2_i's exceed $T^2_{max}(\alpha, n_1, n_2, n_3, p)$. Additionally (as explained above) each T^2_i which exceeds the $T^2_{max}(\alpha)$ quantile is flagged as significant. As a point of reference it is helpful to note that if the group sizes are equal, then

$$T_{max}(\alpha, n_1, n_1, n_1, 1) = \frac{1}{\sqrt{2}}\, q(\alpha, 3, n-3)$$

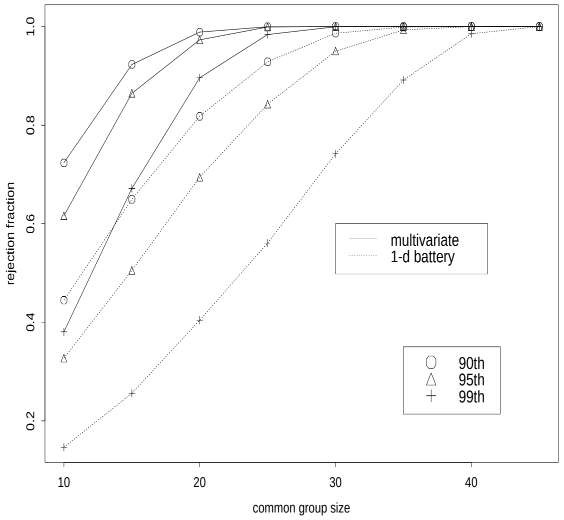

Figure 1: Plotted are the fractions of the time (out of $10,000$ resamples) when the given procedure found the (shifted) versicolor-virginica difference.

where $q(\alpha, 3, n-3)$ is the $1-\alpha$ quantile of the univariate studentized range with 3 groups and $n-3$ error degrees of freedom ($n := n_1 + n_1 + n_1$). And also for the case of equal group sizes, it is helpful to note that

$$T_{max}(\alpha, n_1, n_1, n_1, p) = \frac{1}{\sqrt{2}}\, R_{MAX}(\alpha, m, \nu, p)$$

where R_{MAX} is the *multivariate studentized range* as defined in Siotani (1992) with parameters m and ν. Under Siotani's definition $\nu = \sum_i n_i - 3$ and $m = 3$.

PRACTICAL UTILITY OF THE T^2_{max} PROCEDURE

It is a characteristic of the MANOVA problem that no single procedure is going to be the best for all situations. With this in mind, it seems appropriate to demonstrate the practical utility of the T^2_{max} procedure by comparing its performance with a reasonable competitor. The fact that the T^2_{max} procedure will always outperform Roy's root and, in turn, the other equivariant test statistics (for the all pairwise comparisons situation) clearly rules out procedures based on those statistics as competition. It is clear that a battery of univariate procedures protected against type I inflation via the Bonferroni method is a reasonable competitor. The studentized range procedure is known to provide the shortest confidence intervals (of the available procedures) for all pairwise comparisons for equal groups in the univariate case, and it is a remarkable fact that the procedure is conservative when the groups are not equal. Thus a battery (across dimensions) of studentized range procedures will be used. It should be noted in passing that there are situations where the studentized range 1-d battery will find differences when the T^2_{max} procedure does not, and vice versa. It should also be pointed out here that for all the cases covered by our approximations for T^2_{max}, confidence intervals based on the studentized range 1-d battery will always be shorter than confidence intervals based on T^2_{max}. Nonetheless, by comparing each T^2_i with $T^2_{max}(\alpha)$ there are many situations where the T^2_{max} procedure will find group differences when the Bonferroni-studentized range procedure will not, which this example will show.

We consider the Fisher-Anderson Iris data (1939). This data set consists of 3 groups with 4 dimensions. There is no problem in declaring the setosa group to be different from the versicolor and virginica groups. It may be surprising that the 1-d studentized range battery also finds the versicolor-virginica difference

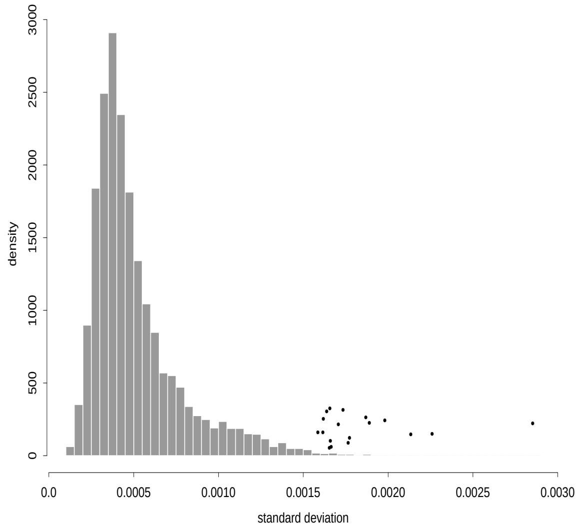

Figure 2: Histogram of all 4500 standard deviations for the validation runs. Overlaid on the plot are the largest 18 standard deviations.

quite easily, even at very high levels of significance. However, marginal plots of petal length and petal width clearly show these two groups to be different. Shifting the versicolor group closer to the virginica group provides a more interesting example for the procedures at hand. We shift the versicolor group by adding $(1, 0, 1.5, 0.5)$ to each observation where the coordinates are (sepal l., sepal w., petal l., petal w.). It might be even more surprising that the 1-d battery still finds the versicolor-virginica difference to be highly significant.

Now suppose we took random samples without replacement of size $n_0 < 50$ (say $n_0 = 45$, for now) from each group (we stick with equal samples since this is where the univariate studentized range procedure is the most powerful) and each time applied (1) the 1-d battery and (2) the T^2_{max} procedure. We can record whether each procedure finds the versicolor-virginica difference for each sample. We do this some large number of times. We can then repeat the whole setup for $n_0 = 40, 35, 30, \ldots$, etc. to get some idea of how the procedures perform as we decrease the sample size.

We carry this out $10,000$ times at each value of n_0 and summarize the results in Figure 1. Even with only a subsample of 25 points from each group, the T^2_{max} procedure suffers practically no loss in performance, declaring the versicolor-virginica difference about 98% of the time using the $\alpha = 0.01$ critical value. On the other hand, the 1-d battery declares the versicolor-virginica difference only about 56% of the time using its (conservative) $\alpha = 0.01$ critical value for subsamples of size 25 from each group. A rotating 3-d plot of the shifted data in software such as Xgobi illustrates dramatically how it is the covariance structure which separates the groups. The battery of 1-d tests is essentially blind to this type of multivariate structure. In all fairness, the T^2_{max} procedure would be blind in the case where the difference in the groups is small along a single dimension and many redundant dimensions (which contribute no help through their covariance structure - independence as one example) with no differences are included. Nevertheless, this example using the Iris data demonstrates the practical utility of the T^2_{max} procedure.

Returning to an issue mentioned earlier, what lies at the heart of this example is the fact that

$$\frac{1}{\sqrt{2}} q(\alpha/4, 3, n-3) < T_{max}(\alpha, n_1, n_1, n_1, 4)$$

so that the intervals in (1) are always smaller for the 1-d battery of studentized range procedures than they are for the T^2_{max} procedure. Hence, the only way to outperform the 1-d battery is by taking advantage of the multivariate structure as we have done here (i.e. in comparing the T_i^2's to $T^2_{max}(\alpha)$ we are using the multivariate information which is contained in the off-diagonals of E). As mentioned above, for all the cases our equations cover,

$$\frac{1}{\sqrt{2}} q(\alpha/p, 3, n-3) < T_{max}(\alpha, n_1, n_1, n_1, p)$$

so that the 1-d studentized range battery always leads to shorter confidence intervals than T^2_{max}. This leads us to conjecture that in the 1-way case, the 1-d studentized range battery will always lead to shorter confidence intervals than T^2_{max}. This conjecture is clearly of great relevance from a practical standpoint.

DEVELOPING THE EQUATIONS

The equations we developed are based upon Monte Carlo simulation of the quantiles of T^2_{max}. It would have been possible of course to provide the reader with FORTRAN code which performs simulations of T^2_{max} in a very general setting (e.g. for an arbitrary design X and an arbitrary contrast matrix L). However, it was decided that such a code would not be so widely used due to the fact that it requires the user to wait for a simulated quantile (the waiting time could be as much as an hour or so). To overcome this liability, it was decided to provide approximate quantiles in some format which does not require the user to wait. One immediately thinks of tables (of approximate quantiles) as a potential such format. One prohibitive difficulty of providing tables in a very general setting is that the completely general T^2_{max} formulation is almost as general as the (multivariate) linear model. To make any headway at all in a table-type direction, one must consider particular cases of the general formulation of T^2_{max}.

The simplest case of T^2_{max} is the case of one-way MANOVA. For the unequal groups case, however, one-way MANOVA is still too general, and one must restrict the generality even further to make any progress. Within one-way MANOVA, the natural place to start is 3 unequal groups, since quantiles for 2 unequal groups are available in closed form (this was described above). For 3 groups of unequal sizes, there are really too many combinations of the group sizes to be able to construct a table (of approximate quantiles) within any reasonable number of pages. Based on this difficulty, it was decided that we should offer the user a set of equations rather than a lengthy table. Not only can the equations provide a more concise summary of the approximate quantiles, but they can also interpolate (approximate) quantiles at which no simulations have been performed. Further, the user is not required to wait at all. Thus, equations for approximate quantiles of T^2_{max} seemed the appropriate format; this is the format we provide for the case of 3 unequal groups.

The equations were developed as follows: (a) we chose a (large) grid, say $\mathcal{G}$, of (n_1, n_2, n_3) triples at which to perform simulations of T^2_{max} and (b) after simulating quantiles at each point of $\mathcal{G}$, we fit (interpolating) equations to these simulated quantiles. We will refer to the simulations used in constructing these equations as the *model-building* simulations. The fitting of these equations is described in the first subsection (ABRIDGED SUMMARY OF THE FITTING PROCESS) below.

To obtain some idea of the accuracy of our equations, we performed additional simulations. These additional simulations will be referred to as the *validation* runs. In the validation runs, we performed a much larger number of simulations over a much smaller subset of points of $\mathcal{G}$. This is described in the second subsection (ASSESSING THE ACCURACY OF THE APPROXIMATIONS) below.

ABRIDGED SUMMARY OF THE FITTING PROCESS

It was necessary to restrict the approximations to the region (a) $4 \le n_i \le 100$, $i = 1, 2, 3$, and (b) $\max_{i \in \{1,2,3\}}(n_i) \le 4 \min_{i \in \{1,2,3\}}(n_i)$ in order to reduce the number of (n_1, n_2, n_3) triples considered to something manageable. Even with this reduction, 9357 (n_1, n_2, n_3) triples (about 5,000 distinct points) were used in the model-building simulations.

For a given dimension (recall that dimension corresponds to p), at least 100,000 simulations per point (or triple) were used, more simulations being required for the higher dimensions . Using an adaptive number of simulations was attempted but proved to be too slow and unstable. Thus the number of simulations per point had to be fixed for each dimension.

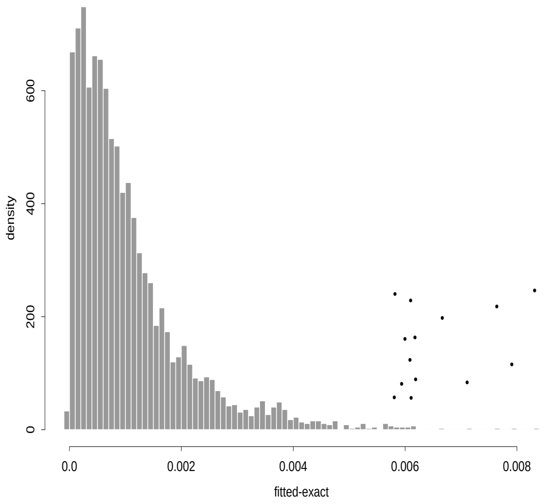

Figure 3: Histogram of the (absolute values) of all 4500 of the fitted values - the "exact" values. Overlaid on the plot are the 14 largest absolute errors.

It became clear from the inspection of many scatter plots that the function might be well approximated by a linear combination of terms like $1/n_1$, $1/n_2$, etc. Least squares was used to fit equations for each quantile and each value of p. Thus 45 equations were fit (5 dimensions: $p = 1, \ldots, 5$, and 9 quantiles: $0.8, 0.825, 0.85, \ldots, 0.925, 0.95, 0.975, 0.99$). The equations took the form

$$\sqrt{T^2_{max}(n_1, n_2, n_3)} = a_0 + \sum_{i=1}^{3}\sum_{s=1}^{k} a_{is}\left(\frac{1}{n_i}\right)^s + \sum_{1\le i<j\le 3}\sum_{s=1}^{k} a_{ijs}\left(\frac{1}{n_i n_j}\right)^s + \sum_{s=1}^{k} a_s\left(\frac{1}{n_1 n_2 n_3}\right)^s \quad (5)$$

where k denotes the order of the polynomial used (we eventually arrived at $k = 4$). This form was determined via a standard regression approach. A second order model was actually enough to remove all lack-of-fit; however, adding the third and fourth order terms provided enormous improvements in accuracy (accuracy is discussed below). It is well known that a Laurent series expansion can sometimes lead to a more parsimonious expression with fewer coefficients. While a second order Laurent series expansion provided similar regression results, the accuracy of those equations proved to be much worse than the fourth order model (5), not to mention the fact that the Laurent series expansion required just as many coefficients.

ASSESSING THE ACCURACY OF THE APPROXIMATIONS

To evaluate the accuracy of the fitted equations, 100 of the original 9357 points were chosen at random (subject to some restrictions requiring that more points be chosen from the region of the smaller sample sizes, due to the fact that the region of the smaller sample sizes is harder to fit). At each of these 100 points, at least $1,000,000$ simulations were performed (again, more simulations were performed for higher dimensions). Each point was replicated 10 times in order to provide an estimate of variability at each of the 100 points. As an example, one of the 100 random points chosen was $(n_1, n_2, n_3) = (8, 12, 29)$. For $p = 1$, $1,000,000$ simulations of T^2_{max} were performed with $(n_1, n_2, n_3) = (8, 12, 29)$. The 9 sample quantiles $0.8, 0.825, \ldots, 0.95, 0.975, 0.99$ were recorded. This gave estimates of the true quantiles, call these estimates $T^2_{max}(\alpha, 8, 12, 29, 1)_{(1)}$, $\alpha = 0.2, 0.175, \ldots, 0.025, 0.01$. We repeat this 9 more times to get $T^2_{max}(\alpha, 8, 12, 29, 1)_{(2)}, \ldots, T^2_{max}(\alpha, 8, 12, 29, 1)_{(10)}$, $\alpha = 0.2, 0.175, \ldots, 0.025, 0.01$. Then (use $\alpha = 0.2$ for

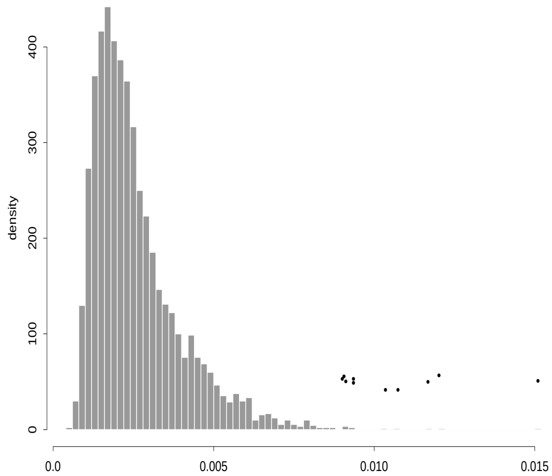

Figure 4: Histogram of 3*(standard deviation of "exact") + |fitted-"exact"|. Overlaid on the plot are the 10 largest such values.

example) we use the average of $T^2_{max}(0.2, 8, 12, 29, 1)_{(1)}, \ldots, T^2_{max}(0.2, 8, 12, 29, 1)_{(10)}$ as our estimate of the true quantile (we will refer to this average as the "exact" quantile). We use the standard deviation of $T^2_{max}(0.2, 8, 12, 29, 1)_{(1)}, \ldots, T^2_{max}(0.2, 8, 12, 29, 1)_{(10)}$ divided by $\sqrt{10}$ as an estimate of the standard deviation of the "exact" quantiles.

The increase in the number of simulations per point with the increase in dimension was adequate to keep these standard deviations constant across dimensions. And, again for brevity, a histogram of the standard deviations of the "exact" quantiles for all 4500 validation points (i.e. 100 points $\times$ 9 quantiles $\times$ 5 dimensions) is given in Figure 2. The largest standard deviation is 0.002853, but it is clear from the histogram that performance is generally much better than that. These "exact" estimates will be referred to as the *validation runs*.

To estimate the accuracy of the fitted equations, we calculate fitted-"exact" for each of the 100 validation points. A histogram of the absolute values of these results is given in Figure 3. The largest absolute error (i.e. |fitted - exact|) is 0.0083. The absolute errors rarely exceed 0.006, which can also be observed from the figure. In fact, the absolute errors never exceed 0.006 for $p = 1$ and $p = 2$. To be very conservative, if we believed that the true quantile were as far as ± 3 standard deviations from the "exact" quantile, then the furthest our fitted values could be from the truth would be *3*(standard deviation of "exact") + | fitted-"exact" |*. A histogram of these values is given in figure 4. The largest such value is 0.0151, but the performance is clearly better than that in general.

We may use the same analysis as in the preceding paragraph on the results Siotani (1992) reports. Due to his definitions, we must rescale what he reports by $\sqrt{2}$ (the rescaling is clearly indicated in the table below). Nevertheless, a direct comparison between his results and our results is appropriate. The worst case he does report is for $p = 5$ and $\alpha = 0.01$; we reproduce this case here (the first four columns here are taken from Siotani's Table 1 and Table 3).

ν	fitted	"exact"	s.d.("exact")	fitted-"exact"	$3 * \frac{s.d.}{\sqrt{2}} + \left\lvert \frac{f-e}{\sqrt{2}} \right\rvert$
10	14.359	14.31	0.297	0.049	0.6647
20	8.340	8.339	0.077	0.001	0.1641
60	6.551	6.548	0.051	0.003	0.1103

The 0.6647 value in the last column of this table compares directly to our "worst case" value of 0.0151.

Giving Siotani more than the benefit of the doubt, 0.1641 is his second worst value which is still more than 10 times larger than our worst value of 0.0151.

COMPUTING PLATFORM AND OTHER SOFTWARE

The far majority of the computing for these simulations was done on 2 computers, named *student* and *gauss*. *Student* is a Silicon Graphics Indigo 2 High Impact Workstation, and *Gauss* is a Pentium Pro 150 running FreeBSD 2.2. The running times were roughly 3 weeks for the validation runs and roughly 2.5 weeks for the model building runs. The computer code for the simulations consisted of roughly 300 lines of FORTRAN. The code calls a solve routine from LAPACK, as well as a few basic routines from the BLAS. The 45 separate regressions were carried out using ordinary proc reg in SAS, while diagnostic plots for the regressions (as well all of the plots in this paper) were done in Splus. The code used for the *random number generators* was written by Barry Brown, James Lovato, and Kathy Russell. It is all contained in the "randlib.f" library. The latest version (version 1.3) of the library is now available at *http://odin.mdacc.tmc.edu/anonftp/source.html*.

ACKNOWLEDGEMENTS

The first half of this work was supported by NIH Grant 5 R01 GM53545-02. The second half was supported by U.S. Army Research Office DAAH04-95-1 0665. The authors would like to thank Marek Kimmel, Dennis Cox, David Andrews, David Scott and Kathy Ensor for their help and support.

REFERENCES

—–Anderson, E. (1939). "The Irises of the Gaspé Peninsula," *Bulletin of the American Iris Society*, **59**, 2-5.
—–Kres, Heinz (1983). *Statistical Tables for Multivariate Analysis*, Springer-Verlag.
—–Krishnaiah, P.R. (1969). "Simultaneous Test Procedures under General MANOVA Models." *Multivariate Analysis II*, 121-143.
—–Lehmann, E.L. (1986). *Testing Statistical Hypotheses*, Wadsworth & Brooks/Cole.
—–Roy, S.N. and Bose, R.C. (1953). "Simultaneous Confidence Interval Estimation." *Annals of Mathematical Statistics*, **24**, 513-536.
—–Siotani, M. (1959)."The extreme value of the generalized distances of the individual points in the multivariate normal sample." *Annals of the Institute of Statistical Mathematics*, **76**, 725-728.
—–Siotani, M. and Seo, Takashi (1992). "The multivariate studentized range and its upper percentiles." *Journal of the Japan Statistical Society*, **22**, 123-137.
—–Wijsman, Robert A. (1979). "Constructing all simultaneous confidence sets in a given class, with applications to MANOVA", *Annals of Statistics*, **7**, 1003-1018.