

**FORTRAN 90 AND SAS-IML PROGRAMS FOR  
COMPUTATION OF CRITICAL VALUES  
FOR  
MULTIPLE TESTING AND SIMULTANEOUS CONFIDENCE INTERVALS**

25JUN2001

Paul N. Somerville

Department of Statistics, University of Central Florida, Orlando, FL 32816-2370

Frank Bretz

LG Bioinformatik, Hannover University, Hannover, Germany

Key words: multiple comparisons, multiple testing, step-up and step-down testing, critical values, general linear model, Fortran 90, SAS

### 0. Introduction

Let  $\mathbf{x}$  be an estimate of the  $k$ -vector mean  $\boldsymbol{\mu}$ , where  $\mathbf{x}$  is  $MVN(\boldsymbol{\mu}, \Sigma\sigma^2)$ . We assume  $\Sigma$  is known and if we do not know  $\sigma^2$ , we have an independent estimate  $s^2$  of  $\sigma^2$  with  $v$  degrees of freedom such that  $v s^2 / \sigma^2$  is a chi-square variate. We have a set  $\mathbf{B}$  of contrasts  $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p$ , and are interested either in simultaneous hypotheses testing or obtaining simultaneous confidence intervals.

Without loss of generality, we can assume all the hypotheses  $H_i$  are one-sided and of the form  $H_i: \mathbf{c}_i'\boldsymbol{\mu} \leq \delta_i$ , and all the confidence intervals for  $\mathbf{c}_i'\boldsymbol{\mu}$  are also one-sided and of the form  $(-\infty, \gamma_i)$ . This is because the two-sided hypothesis  $H_i: \mathbf{c}_i'\boldsymbol{\mu} \neq \delta_i$  can be regarded as a set of two one-sided hypotheses  $H_{ia}: \mathbf{c}_i'\boldsymbol{\mu} < \delta_i$ , and  $H_{ib}: -\mathbf{c}_i'\boldsymbol{\mu} < -\delta_i$ . In similar fashion, a two-sided confidence interval  $(\gamma_{ia}, \gamma_{ib})$  for  $\mathbf{c}_i'\boldsymbol{\mu}$  can be regarded as a set of two simultaneous one-sided confidence intervals  $(-\infty, \gamma_{ib})$  for  $\mathbf{c}_i'\boldsymbol{\mu}$ , and  $(-\infty, -\gamma_{ia})$  for  $-\mathbf{c}_i'\boldsymbol{\mu}$ .

To simultaneously test  $H_1, H_2, \dots, H_p$ , we calculate  $t_i = (\delta_i - \mathbf{c}_i'\mathbf{x}) / v_i^{1/2}$  where  $v_i$  is the variance (estimated if necessary) of  $\mathbf{c}_i'\mathbf{x}$ . We will accept  $H_i$  whenever  $t_i \geq$  the "critical value" which we define to be  $d$  (or  $q/2^{1/2}$ ). The value  $d$  (or  $q$ ) is chosen such that the probability of accepting one or more of the  $H_i$  is less than  $\alpha$  (type I error) when in fact  $\mathbf{c}_i'\boldsymbol{\mu} = 0$  for  $i = 1, \dots, p$ . Use of  $d$  is consistent with most of the literature of multiple comparisons. Use of  $q$  is consistent with Tukey's multiple comparison procedure.

Simultaneous  $(1 - \alpha)$  confidence intervals for  $\mathbf{c}_i'\boldsymbol{\mu}$  are given by

$$(-\infty \leq \mathbf{c}_i'\boldsymbol{\mu} \leq \mathbf{c}_i'\mathbf{x} + v_i^{1/2} d) \quad i = 1, \dots, p$$

where  $v_i$  is the variance of  $\mathbf{c}_i'\mathbf{x}$  (estimated if necessary). Combining  $(1 - \alpha)$  confidence intervals for  $\mathbf{c}_i'\boldsymbol{\mu}$  and  $-\mathbf{c}_i'\boldsymbol{\mu}$ , would result in the corresponding two-sided confidence interval

$$(\mathbf{c}_i'\mathbf{x} - v_i^{1/2} d \leq \mathbf{c}_i'\boldsymbol{\mu} \leq \mathbf{c}_i'\mathbf{x} + v_i^{1/2} d).$$

Our problem is obtaining the value of  $d$  (or  $q$ ), which is a function of  $v$ ,  $k$ ,  $\alpha$ , and  $\mathbf{B}$ . With no loss in generality, we may assume  $\boldsymbol{\mu} = \mathbf{0}$ . Then in the  $k$  dimensional space  $x_1, x_2, \dots, x_k$ , we need to choose  $d$  such that the probability content of the region (call it  $A_d$ ) defined by  $\mathbf{c}_i'\mathbf{x} / v_i^{1/2} \leq d$ , for  $\mathbf{c} \in \mathbf{B}$ , is  $1 - \alpha$ .

Two Fortran 90 programs and two SAS-IML programs have been written to implement the procedures. The Fortran 90 program QBATCH4.FOR requires that two files exist prior to making a run. They are QCARB.IN, which contains the input for the program and QCARB.OUT, which is output for the program. QINTER4 is an interactive version of QBATCH4.

The Fortran 90 programs QBATCH4.FOR and QINTER4.FOR, and the SAS-IML programs QBATCH4.SAS and QINTER4.SAS calculate the value of  $q$  necessary for hypotheses testing or simultaneous confidence intervals and multiple comparisons, given  $\mathbf{B}$ ,  $\alpha$ ,  $k$  and  $v$ . If  $\sigma^2$  is known,  $v$  is set to  $-1$ . For 15 different multiple comparison procedures, e.g. those of Tukey, Dunnett, Hsu's MCB, etc., the programs generate the set  $\mathbf{B}$ . The programs also accept user generated sets  $\mathbf{B}$ , and can also be used for "step-down" testing, Dunnett and Tamhane (1991). QBATCH4.FOR and QBATCH4.SAS are batch programs while QINTER4.FOR and QINTER4.SAS are interactive programs.

Two recent extensions of QBATCH4.FOR and QINTER4.FOR are DBATCH2.FOR and DINTER2.FOR. They extend the capabilities to calculating critical values for "step- up" testing. The hypotheses (or confidence intervals) can be one-sided, two-sided or a mixture. They also directly calculate  $d$  rather than  $q$  as the critical constant.

## 1. Methodology

The following is an overview of the methodology involved in calculating the value for  $d$  (or  $q$ ). In what follows we confine ourselves to discussing the calculation of  $d$ .

First, a coordinate transformation (Cholesky) is made so that in the new coordinate system, we have  $k$  uncorrelated normal (or spherically symmetric  $t$ ) random variables with unit variances. The transformed region is now bounded by  $p$  hyperplanes, each of which is at a distance  $d$  from the origin.

Consider a randomly selected point in the  $k$  dimensional space. The square of the distance to the point is the sum of the squares of the individual coordinates. Thus if  $\sigma^2$  is known, it is distributed as  $\chi^2$  with  $k$  degrees of freedom. If  $\sigma^2$  is not known, the distance squared, divided by  $k$ , has the  $F$  distribution with  $k$  degrees of freedom for the numerator and the degrees of freedom for the denominator the same as those for the estimate of  $\sigma^2$ . An unbiased estimate of the probability content of  $A_d$  is the probability of obtaining a distance less than or equal to the distance to the boundary.

The problem is to find the value of  $d$  such that the probability content of the region  $A_d$  is  $1 - \alpha$ . Divide the region  $A_d$  into two subregions, the region inside the hypersphere of radius  $d$  and centered at the origin, and the portion outside the hypersphere (say  $A$ ). The probability content of the hypersphere is  $\Pr[F \leq d^2/k]$  if  $\sigma^2$  is known, and  $\Pr[\chi^2 \leq d^2]$  if  $\sigma^2$  is not known.

It is convenient to use reciprocal distances instead of distances. Let  $H(y)$  be the cumulative distribution function for  $y$ , the reciprocal distance from the origin to the boundary in a randomly selected direction and let  $f(y)$  be the probability density function for  $y$ . Then the probability content of the region  $A$  is given by

$$\begin{aligned} P &= \int_A f(y) dy. \\ &= \int H(y)f(y) dy \end{aligned}$$

where the integration is from 0 to  $1/d$ . Quadrature is used to estimate the value of the integral. The empirical cumulative step function estimate of  $H(y)$  is obtained by Monte Carlo with the location of the steps chosen so as to optimize the quadrature (Gauss-Legendre). "Brent's method" (Press et al., 1986) is used to iteratively solve the integral equation  $P(d) = 1 - \alpha$  for  $d$ . (The SAS programs use "Ridders' method"). An important feature of the methodology is that the empirical cumulative step function estimate of  $H(y)$  need only be estimated one time. This is because the structure of  $A_d$  is independent of the value of  $d$ , and the Gauss-Legendre quadrature scales the range of integration to  $(-1, 1)$ .

A more detailed description of the methodology and a discussion of its efficiency can be found in Somerville (1997,1999).

Sections 2 - 6 refer specifically to the Fortran 90 programs. Although the Fortran 90 and the SAS programs are extremely similar, there are some differences. Section 7 details the differences and gives examples for the SAS-IML programs.

## 2. Usage of the FORTRAN 90 programs

DBATCH2.FOR and DINTER2.FOR calculate the proper  $d$  values for the fifteen different multiple comparison procedures. For these procedures, the programs generate  $\mathbf{B}$ , the appropriate set of contrasts. The programs will also calculate the proper values of  $d$  for an arbitrary set  $\mathbf{B}$ , which of course must be part of the input.

The  $d$  values can be calculated both for the case when the estimates  $x_1, x_2, \dots, x_k$  are uncorrelated with equal variances, and when the estimates are correlated or have different variances. In the latter case either the variance covariance or the correlation matrix needs to be input.

For the Dunnett procedures, if the estimates are not uncorrelated with equal variances, it is necessary to input the population number of the control.

For Hsu's multiple comparison with the best (MCB), when the estimates are not uncorrelated with equal variances, the program calculates the  $q$  values associated with each population.

The user may choose to use "mocar" random directions to obtain "irep" preliminary estimates of  $d$ . The "irep" estimates are then combined and used to obtain standard error of the estimate. (If the user sets "irep" equal to 1, the program uses an empirical formula to obtain the standard error of the estimate.) An alternate procedure is to specify a desired standard error of the estimate ("givense"). The program then uses an empirical formula to find the number of random directions necessary to achieve the specified accuracy. To use the alternate procedure, in addition to specifying a value for "givense", "irep" must be a negative integer.

The variance  $\sigma^2$ , may be known or estimated. The degrees of freedom ("ndenom") of the estimate  $s^2$ , must be included in the input. If  $\sigma^2$  is known, "ndenom" is set to -1.

Values of 1000 and 10 for mocar and irep respectively (a total of 10,000 random directions) should result in accuracies adequate for most applications.

### 3. Description of the input file QCARB.IN

The file QCARB.IN must contain the following:

- i) two lines common to all procedures
- \* ii)  $k$  lines for the lower triangular portion of the variance covariance or correlation matrix.
- \*\* iii) a line containing one or more constants
- \*\*\* iv)  $mm$  lines each containing a contrast from the set **B**

i) The first two lines are as follows:

```
swcov swtype conf givense
k seed ndenom mocar irep mm
```

swcov This is a switch which tells the programs whether the estimates are uncorrelated with equal variances (swcov = 1) or whether the variance covariance or correlation matrix will be included in the input (swcov = 0)

swtype This is a switch which tells which multiple comparison procedure  $q$  values are required

0	Arbitrary set of contrasts (to be included in the input)
1	Tukey (1953)
2	Bofinger (1985)
3	Hsu's MCB (1981)
4	Dunnett (one-sided) (1955)
5	Dunnett (two-sided) (1955)
6	Hayter (1990)

---

\* Use only if the estimates are correlated or have unequal variances.

\*\* Use only for Dunnett's procedures when estimates are correlated or have unequal variances, and also for swtypes 9 - 15.

\*\*\* Use only when swtype = 0, 20, 21 or 22.

- 7 Successive ordered treatments (1-sided), Liu et al (1999)
  - 8 Successive ordered treatments (2-sided), Liu et al (1999)
  - 9 Umbrella contrasts, Mack and Wolfe (1981)
  - 10 Generalized umbrella contrasts Bretz and Hothorn - (submitted)
  - 11 Williams type contrasts, Bretz (1999)
  - 12 Marcus type contrasts, Bretz (1999)
  - 13 Hirotsu contrasts, Hirotsu (1979,1997)
  - 14 McDermott et al (1993)
  - 15 Isotonic contrasts Bretz (1999)
  - 20 Somerville et al (2001)
  - 21 Step-up constants (one-sided)
  - 22 Step-up constants (two-sided)
- (swtypes 20, 21 and 22 are not yet available for the SAS versions)

conf The value of  $1 - \alpha$ , or overall confidence.

givense The user requested se of estimate (used only for alternate procedure described in section 2, in which case irep is set to any negative integer). Otherwise input any integer (say 999).

k Dimension of  $\mathbf{x}$  and  $\boldsymbol{\mu}$ .

seed Seed for the random number generator.

ndenom degrees of freedom associated with  $s^2$ . If  $\sigma^2$  is known, use ndenom = -1.

mozar Number of random directions for each preliminary estimate Not used for the "alternate" procedure, in which case input any integer, e.g. 999.

irep Number of preliminary estimates. For the "alternate" procedure which specifies the required standard error of the estimate by "givense", replace irep with any negative integer, e.g. -999.

mm Number of contrasts. Used only for swtype = 0, 20, 21 or 22. Otherwise input any integer, e.g. 999.

ii) If the estimates  $x_1, x_2, \dots, x_k$  are correlated or their variances are unequal, then the next input is the lower triangular portion of either the variance covariance or correlation matrix.  $k$  lines are used, one line for each row of the matrix. This is true for each of the procedures.

iii) If swtype = 4 or 5, this line contains the number of the control population.  
 If swtype = 9, line contains the number of the peak population.  
 If swtype = 10, line contains the interval containing the number of the "peak" population.  
 If swtype is between 11 and 15, line contains  $n_1, n_2, \dots, n_k$ , the  $k$  sample sizes.  
 If swtype = 20, line contains ifix, qlower.

iv) If swtype = 0, 20, 21, or 22, the  $mm$  contrasts of the set  $\mathbf{B}$  are next input, with one line being used for each contrast. A line consists of the coefficients of  $x_1, x_2, \dots, x_k$ , in that order. For 20, 21 or 22, a "d" value is needed at the end of each line.

Several calculations for  $d$  may be made in the same run, with the instructions for each run being in sequence in QCARB.IN. A line with 4 or more negative ones, for example "-1 -1 -1 -1" is used to denote the end of the run.

A special note is in order regarding swtype 20. It is used for calculation of the special constants needed when combining one-sided and two-sided confidence interval procedures for successive ordered treatment effects, Somerville et al (2000). For swtype 20, the region is bounded by the hyperplanes

$$\begin{aligned} c_i'x/v^{1/2} < d & \quad i = 1, \dots, p\text{-ifix} \\ c_i'x/v^{1/2} < d_{\text{lower}} & \quad i = \text{ifix}, \dots, p. \end{aligned}$$

In the application the value of dlower is the critical value calculated using swtype 7.

For swtype 20, 21 and 22 the “structure” of the region A is a function of d. For this reason the computation is more complicated and running times are considerably greater.

#### 4. Examples

i) Calculate the value of d for the cases in Table 1 using DBATCH2.FOR. Use 1000 random directions for each of 10 preliminary estimates. Assume the estimates  $x_1, x_2, \dots, x_k$  are uncorrelated with equal variance.

<u>Procedure</u>	<u>k</u>	<u>df for <math>s^2</math></u>	<u>confidence</u>	<u>seed</u>
Tukey	4	30	.95	457
Hsu's MCB	5	$\sigma^2$ known	.90	1893
Dunnett (1 sided)	4	19	.85	777
Generalized umbrella [3,4]	5	20	.95	757

Input Values for Example i)  
Table 1

The following lines in QCARB.IN are sufficient. (We use 999 for a value which is a required input, but which is not utilized by the program.)

```

1 1 .95 999
4 457 30 1000 10 999

1 3 .90 999
5 1893 -1 1000 10 999

1 4 .85 999
4 777 19 1000 10 999

1 10 .95 999
5 757 20 1000 10 999
3 4
-1 -1 -1 -1 -1 -1

```

The output in QCARB.OUT is as follows:

```

d-value is 2.72039. Seed is 457
4 populations .950 confidence.
date and time 20010613 212433.920
df for variance estimate is 30
10000 random directions used. swtype= 1
Standard error of estimate is 0.001474
date and time 20010613 212433.970
-----
date and time 20010613 212433.970
d-value is 1.84530. Seed is 1893
5 populations .900 confidence.
variance is assumed known
10000 random directions used. swtype= 3
Standard error of estimate is 0.001619
date and time 20010613 212434.030
-----

```

```

date and time 20010613 212434.030
d-value is 1.56852. Seed is 777
4 populations .850 confidence.
df for variance estimate is 19
10000 random directions used. swtype= 4
Standard error of estimate is 0.002912
date and time 20010613 212434.080
-----
date and time 20010613 212434.080
d-value is 2.67280. Seed is 757
5 populations .950 confidence.
df for variance estimate is 20
10000 random directions used. swtype= 10
Standard error of estimate is 0.002799
date and time 20010613 212434.140
-----

```

ii) If instead of specifying mocar and irep we specify that we wish the estimate of d to have a standard error of estimate of .001, the input in QCARB.IN for the first two calculations is as follows:

```

1 1 .95 .001
4 457 30 999 -10 999

1 3 .90 .001
5 1893 -1 999 -10 999
-1 -1 -1 -1 -1 -1

```

The resulting output in QCARB.OUT is as follows:

```

date and time 20010613 212434.140
d-value is 2.72007. Seed is 457
4 populations .950 confidence.
df for variance estimate is 30
24697 random directions used. swtype= 1
Standard error of estimate is 0.000707
date and time 20010613 212434.250
-----
date and time 20010613 212434.250
d-value is 1.84057. Seed is 1893
5 populations .900 confidence.
variance is assumed known
131212 random directions used. swtype= 3
Standard error of estimate is 0.000707
date and time 20010613 212434.960
-----

```

iii) We now assume the following variance covariance matrix for the first and third example in (i)

```

76.807 36.146 -20.366 7.2384
36.148 27.373 15.176 21.198
-20.366 15.176 64.583 40.605
7.2384 21.198 40.605 31.239

```

For the third example (Dunnett's one-sided), we use population 2 as the control population. The input values for mocar and irep are 1000 and 10 respectively. The input in QCARB.FOR is:

```

0 1 .95 999
4 457 30 1000 10 999
76.807
36.148 27.373
-20.366 15.176 64.583

```

```

7.2384  21.198  40.605  31.239

0  4  .85  999
4  777  19  1000  10  999
76.807
36.148  27.373
-20.366  15.176  64.583
7.2384  21.198  40.605  31.239
2
-1  -1  -1  -1  -1  -1

```

The output is as follows:

```

date and time 20010613  212434.960
d-value is 2.04407. Seed is  457
4 populations .950 confidence.
df for variance estimate is  30
10000 random directions used. swtype= 1
Standard error of estimate is 0.005727
date and time 20010613  212435.020
-----
date and time 20010613  212435.020
d-value is 1.51375. Seed is  777
4 populations .850 confidence.
df for variance estimate is  19
10000 random directions used. swtype= 4
Standard error of estimate is 0.006094
date and time 20010613  212435.070
-----

```

iv) Suppose we wish to use Hsu's MCB for 4 populations, seed of 8763, df for variance estimate of 25, confidence is .90, and the means have the variance covariance matrix of the previous example. Using 10 preliminary estimates each with 1000 random directions, we have the following input in QCARB.IN

```

0  3  .90  999
4  8763  25  1000  10  999
76.807  36.146  -20.366  7.2384
36.146  27.373  15.176  21.198
-20.366  15.176  64.583  40.605
7.2384  21.198  40.605  31.239
-1  -1  -1  -1  -1  -1

```

The output in QCARB.OUT is

```

date and time 20010613  212435.070
d-value is 1.30881. Seed is  8763
4 populations .900 confidence.
df for variance estimate is  25
10000 random directions used. swtype= 3
Standard error of estimate is 0.007340

d-value corresponding to population  2 is 1.697670
Standard error of estimate is 0.006679
10000 random directions used.

d-value corresponding to population  3 is 1.322799
Standard error of estimate is 0.009211
10000 random directions used.

d-value corresponding to population  4 is 1.708277
Standard error of estimate is 0.005714

```

10000 random directions used.  
 date and time 20010613 212435.240  
 -----

v) Suppose we wish two-sided 99% confidence intervals for  $\mu_1 - \mu_2$ , and one-sided confidence intervals for each of  $\mu_1 - \mu_4$ ,  $\mu_2 - \mu_4$  and  $\mu_3 - \mu_4$ . We use the variance covariance example of the previous example and with 1000 random directions repeated 10 times.  $\sigma^2$  is estimated with 25 degrees of freedom. We may use the following input in QCARB.IN:

```

0 0 .99 999
4 911 25 10000 10 5
76.807
36.148 27.373
-20.366 15.176 64.583
7.2384 21.198 40.605 31.239
1 -1 0 0
-1 1 0 0
1 0 0 -1
0 1 0 -1
0 0 1 -1
-1 -1 -1 -1 -1 -1

```

The output in QCARB.OUT is

```

date and time 20010613 212435.240
d-value is 2.79294. Seed is 911
4 populations .990 confidence.
df for variance estimate is 25
100000 random directions used. swtype=0
Standard error of estimate is 0.003489
date and time 20010613 212435.510
-----

```

### 5. Calculation of constants for step-down testing

Let  $t_i$  be the test statistic for  $H_i$  (if  $H_i$  is two-sided use  $|t_i|$ ). Relabel the hypotheses so that  $t_1$  denotes the least significant test statistic, and  $t_p$  denotes the most significant. The  $t$ 's are compared with a set of critical constants  $d_1 \leq d_2 \leq \dots \leq d_p$ . The procedure stops the first time  $t_m < d_m$  is observed, with the hypotheses for which  $i > m$  accepted, and the rest rejected. Dunnett and Tamhane (1991) showed that for step-down testing, the critical constants  $d_m$  are those for which the equations

$$\Pr[t_1 \leq d_m, \dots, t_m \leq d_m] = 1 - \alpha, \quad \text{for } m = 1, \dots, p.$$

Denote by  $\mathbf{B}_m$ , the set of contrasts corresponding to  $H_1, H_2, \dots, H_m$ . The value  $d_m$  can be calculated given the values for  $k, v, \alpha$ , and the set  $\mathbf{B}_m$ .

Let the hypotheses be  $H_1: \mu_1 - \mu_2 \neq 0, H_2: \mu_1 - \mu_4 < 0, H_3: \mu_2 - \mu_4 < 0, H_4: \mu_3 - \mu_4 < 0$ , where  $k = 4$ , and  $|t_1| < t_2 < t_3 < t_4$ . Then  $\mathbf{B}_4$  is the set of contrasts whose coefficients are:

```

1 -1 0 0
-1 1 0 0
1 0 0 -1
0 1 0 -1
0 0 1 -1

```

The sets  $\mathbf{B}_3, \mathbf{B}_2, \mathbf{B}_1$  are the first four, the first three and the first two contrasts of  $\mathbf{B}_4$ , respectively. If the estimates are uncorrelated with equal variances, the variance  $\sigma^2$  is known,  $\alpha$  is .10 and we use 10 independent estimates, each with 100000 random directions, the seed is 373, the input in QCARB.IN for the calculation of  $d_4, d_3$  and  $d_2$  ( $d_1$  is  $z_{.05} = 1.64485$ ) may be given as follows:

```

1 0 .90 999
4 373 -1 100000 10 5
1 -1 0 0
-1 1 0 0
1 0 0 -1

```



```

0 1 0 -1
0 0 1 -1

1 0 .90 999
4 373 -1 100000 10 4
1 -1 0 0
-1 1 0 0
1 0 0 -1
0 1 0 -1

1 0 .90 999
4 373 -1 100000 10 3
1 -1 0 0
-1 1 0 0
1 0 0 -1
-1 -1 -1 -1 -1 -1

```

The values for  $d_2$ ,  $d_3$  and  $d_4$  are 1.80009, 1.89064 and 1.97395, with standard errors of .00061, .00062 and .00038, respectively.

### 6. Calculation of constants for step-up testing

As for step-down testing, let  $t_i$  be the test statistic for  $H_i$  (if  $H_i$  is two-sided use  $|t_i|$ ). Relabel the hypotheses so that  $t_1$  denotes the least significant test statistic, and  $t_p$  denotes the most significant. The  $t$ 's are compared with a set of critical constants  $d_1 \leq d_2 \leq \dots \leq d_p$ . The procedure stops the first time  $t_m \geq d_m$  is observed, with the hypotheses for which  $i \geq m$  accepted, and the rest rejected. Dunnett and Tamhane (1991) showed that for step-up testing, the critical constants  $d_m$  are those for which the equations

$$\Pr[t_1 < d_1, \dots, t_m \leq d_m] = 1 - \alpha, \quad \text{for } m = 1, \dots, p.$$

Denote by  $\mathbf{B}_m$ , the set of contrasts corresponding to  $H_1, H_2, \dots, H_m$ . The value  $d_m$  can be calculated given the values for  $k, v, \alpha, d_1, \dots, d_{m-1}$  and the set  $\mathbf{B}_m$

Again, suppose the hypotheses are  $H_1: \mu_1 - \mu_2 \neq 0, H_2: \mu_1 - \mu_4 < 0, H_3: \mu_2 - \mu_4 < 0, H_4: \mu_3 - \mu_4 < 0$ , where  $k = 4$ , and  $|t_1| < t_2 < t_3 < t_4$ . Then  $\mathbf{B}_1$  is:

```

1 -1 0 0
-1 1 0 0

```

Assuming the estimates are uncorrelated with equal variances, the variance  $\sigma^2$  is known,  $\alpha$  is .10 and we use 10 independent estimates, each with 100000 random directions, the seed is 373, then appending 1.5 (an arbitrary guess for the value of  $d_1$ ) to each line of  $\mathbf{B}_1$  we may use the following as input in QCARB.IN (we use swtype = 22, since  $H_1$  is a two sided hypotheses):

```

1 22 .90 999
4 757 -1 100000 10 2
1 -1 0 0 1.5
-1 1 0 0 1.5
-1 -1 -1 -1 -1 -1

```

or, equivalently, since the contrasts involve only the first 2 means:

```

1 22 .90 999
2 757 -1 100000 10 2
1 -1 1.5
-1 1 1.5
-1 -1 -1 -1 -1 -1

```

The calculation is unnecessary, however, since we know  $d_1$  is  $z_{.05} = 1.64485$ .

To calculate  $d_2$  ( $H_2$  is a one-sided hypotheses so we use swtype = 21. Also since  $d_2 > d_1$  our guess, say 1.7 must be greater than  $d_1$ ) we may use the following as input to QCARB.IN:

```

1 21 .90 999
4 757 -1 100000 10 3
1 -1 0 0 1.64485

```

```

-1 1 0 0 1.64485
1 0 0 -1 1.7
-1 -1 -1 -1 -1 -1

```

We obtain  $d_2 = 1.80314$  with s.e. .00053.

To calculate  $d_3$ , we may use the following as input:

```

1 21 .90 999
4 757 -1 100000 10 4
1 -1 0 0 1.64485
-1 1 0 0 1.64485
1 0 0 -1 1.80314
0 1 0 -1 1.9
-1 -1 -1 -1 -1 -1

```

We obtain  $d_3 = 1.89257$  with s.e. .00040.

Finally to calculate  $d_4$ , we may use as input:

```

1 21 .90 999
4 757 -1 100000 10 5
1 -1 0 0 1.64485
-1 1 0 0 1.64485
1 0 0 -1 1.80314
0 1 0 -1 1.89257
0 0 1 -1 2.
-1 -1 -1 -1 -1 -1

```

We obtain  $d_4 = 1.97611$  with s.s .00046.

We may compare  $d_1, d_2, d_3, d_4$  for step-up testing 1.64485, 1.80314, 1.89257 and 1.97611 with the corresponding constants for step-down testing, namely 1.64485, 1.80009, 1.89064 and 1.97395.

## 7. SAS/IML Programs

This section provides a short description of the SAS/IML programs and provides examples of their use. The programs QBATCH4.SAS and QINTER4.SAS, in general follow closely the coding for QBATCH4.FOR and QINTER4.FOR. The major differences are in some SAS specific modifications. Differences in the q-values between the Fortran 90 and the SAS-IML programs are due to the use of a different routine for the generation of random normal variates. Accuracies for the Fortran 90 and SAS-IML programs are equivalent.

**System requirements:** The SAS programs have been written and tested using Release 6.12. However, users of older releases of Version 6, should be able to execute the programs. The following modules are required for processing of the SAS programs: SAS/BASE, SAS/STAT, SAS/IML.

**Description:** QBATCH4.SAS is a batch program written using SAS/IML. No additional input or output files are required. The user need only insert the correct input statements at the beginning of the program before submission.

The SAS programs use the same variable names as the Fortran programs. Section 3 gives a complete description.

### Use of the Program:

Instead of requiring a separate input file, as in QBATCH4.FOR, the user need only insert the “situation parameters” immediately following the “proc iml” statement. An easy way to do this is to insert statements corresponding to those in QCARB.IN (see 3). For example, if QCARB.IN were:

```

1 5 .9 999
5 1893 -1 1000 10 999
110.25
130.2 255.77

```

```

139.65    318.44    526.74
149.1     340.71    630.64    898.66
1

```

then the following statements could be inserted after "proc iml" as "situation parameters".

```

swcov   = 1      ;
swtype  = 5      ;
conf    = .9     ;
givense = 999   ;
k       = 5      ;
seed    = 1893  ;
ndenom  = -1    ;
mocar   = 1000  ;
irep    = 10    ;
mm      = 999   ;
vc=. ;
icontrol= 1    ;
contr   = .     ; /* contr statement is optional! */

```

The variables givense and mm are not needed for this example and instead of assigning them arbitrary values, could be omitted or treated as missing values, i.e.

```

givense = . ;
mm      = . ;

```

It is permissible to input the complete vc matrix!

It is important to note that that for all variables included in the input for QBATCH4.FOR, i.e. in QCARB.IN, and used, a statement for that variable must be included with the inserted statements.

Submission of QBATCH4.SAS for above example the results in the following output:

```

CONTR
-1      1      0      0      0
-1      0      1      0      0
-1      0      0      1      0
-1      0      0      0      1
 1     -1      0      0      0
 1      0     -1      0      0
 1      0      0     -1      0
 1      0      0      0     -1
 1      0      0      0      0

```

The q-value is 3.0475009 . The seed is 1893

5 populations, 0.9 confidence.

variance is assumed known

10000 random directions used. swtype= 5

Standard error of estimate is 0.0041347

The output resembles the Fortran 90 output. A difference is the addition of the contrast matrix.

Differences in the q-value and the standard error of estimate occur because a different random number generator is used. The SAS code uses some facilities provided by SAS, and not available in Fortran 90, namely:

- PROBNORM for generating standard normal variates,
- ROOT for the Cholesky decomposition,
- PROBF for computing univariate F-probabilities, and
- RIDDERS for obtaining the root of a real continuous function.

QINTER4.SAS is the interactive program. The interactive data entry is done by use of SAS/Macro-%window commands. Vectors or matrices are input one element at a time. Note that for the variance-covariance matrix only the upper (or lower) triangular portion is required. (If the lower triangular portion

is used, the order is row by row). Press 'ENTER' after typing each element. No error control has been included. Pressing 'CTRL' and 'PAUSE' simultaneously interrupts the process and the user may start the program again. The order of events of the queries is given below:

```

/* Order of events, where info_i(var) describe the %window of number i
   asking for the input variable var */

info_1(swtype)
info_2(conf)
info_3(k)
info_4(ndenom)
info_5(irep)
if (irep < 0) then info_6(givense)
else info_7(mocar)
  if (mocar < 0) then
    info_8(irep)
    info_9(mocar)
if ((swcov = 0) and ((swtype = 4) or (swtype = 5))) then
info_12(icontr)
if (swtype = 0) then info_13(mm)
info_15(seed)
info_10 swcov
if (swcov = 0) then info_11(vc)
info_14(contr)

```

**Examples:** Four additional examples are provided. On the left are the statements (situation parameters) which must be inserted after "proc iml". On the right the output is given (omitting the contrast matrix).

**Example 1:**

uncorrelated Bofinger, df = 49, 1 -  $\alpha$  = 0.9, 4 groups, 10 estimates with 1000 replications each

```

swcov = 1;          The q-value is 2.9341036 . The seed is 1893
swtype = 2;
conf = .9;          4 populations,      0.9 confidence.
givense= 999;
k = 4;              df for variance estimate is 49
seed = 1893;
ndenom = 49;        10000 random directions used. swtype= 2
mocar = 1000;
irep = 10;          Standard error of estimate is 0.0016585
mm = 999;

```

**Example 2:**

correlated Hsu's MCB, df = -1, 1 -  $\alpha$  = 0.99, 4 groups, 10 estimates with 1000 replications

```

swcov = 0;          The q-value is 3.6754943 . The seed is 1893
swtype = 3;
conf = .99;         4 populations,      0.99 confidence.
                    variance is assumed known
givense= 999;       10000 random directions used. swtype= 3
k = 4;              Standard error of estimate is 0.0060596
seed = 1893;
ndenom = -1;        q-value corresponding to population 2 is 3.7708395
mocar = 1000;       10000 random directions used. swtype= 3
irep = 10;          Standard error of estimate is 0.0082631
mm = 999;
                    q-value corresponding to population 3 is 3.7544567
vc={110.25 130.2 139.65 149.1,
    130.2 255.77 318.44 340.71,
    139.65 318.44 526.74 630.64,
    149.1 340.71 630.64 898.66};
                    10000 random directions used. swtype= 3
                    Standard error of estimate is 0.0046354
                    q-value corresponding to population 4 is 3.6214269
                    10000 random directions used. swtype= 3
                    Standard error of estimate is 0.0065361

```

**Example 3:**

uncorr. one-sided Dunnett, df = 5, 1 -  $\alpha$  = 0.95, 4 groups, givense = 0.005, icontr = 2

```

swcov = 1;          The q-value is 3.7881064 . The seed is 141071
swtype = 4;
conf = .95;         4 populations,      0.95 confidence.
givense = 0.005;
k = 4;              df for variance estimate is 5

```

```

seed      = 141071;
ndenom   = 5;
mocar    = 1000;
irep     = -10;
mm       = 99;
icontrol = 2;
61098 random directions used. swtype= 4
Standard error of estimate is 0.005

```

#### Example 4:

uncorr. two-sided Dunnett,  $df = 500$ ,  $1 - \alpha = 0.95$ , 5 groups, 100 estimates with 10000 repl.

```

swcov = 1;
swtype = 5;
conf = .95;
givense= 999;
k = 5;
seed = 1893;
ndenom = 500;
mocar = 10000;
irep = 100;
mm = 999;
The q-value is 3.4645534 . The seed is 1893
5 populations, 0.95 confidence.
df for variance estimate is 500
1000000 random directions used. swtype= 5
Standard error of estimate is 0.0006127

```

#### Example 5:

correlated one-sided comparisons between groups 1 & 2, 2 & 3 and 3 & 4 ( $mm = 3$ ),  $df = -1$ ,  $1 - \alpha = 0.95$ , 4 groups, 10 estimates with 1000 repl.

```

swcov = 0;
swtype = 0;
conf = .8;
givense= 999;
k = 4;
seed = 1893;
ndenom = -1;
mocar = 1000;
irep = 10;
mm = 3;
The q-value is 1.9476881 . The seed is 1893
4 populations, 0.8 confidence.
variance is assumed known
10000 random directions used. swtype= 0
Standard error of estimate is 0.0034326

```

```

vc={110.25 130.2 139.65 149.1,
    130.2 255.77 318.44 340.71,
    139.65 318.44 526.74 630.64,
    149.1 340.71 630.64 898.66};
contr={-1 1 0 0,
        0 -1 1 0,
        0 0 -1 1};

```

## 8. Summary and conclusions

Fortran 90 and SAS-IML programs have been written which calculate the constants necessary for several multiple comparison and multiple testing procedures under fairly general conditions. The programs can also be used to obtain the constants for arbitrary sets of contrasts, and also for step-down testing. The Fortran programs can be used to obtain step-up constants. There are batch and interactive versions of both the Fortran 90 and SAS programs. A compiled version of the Fortran 90 programs which should run on any PC with Windows 95 or later can be found at

<http://pegasus.cc.ucf.edu/~somervil/home.html>

It should be observed that the programs can be used to obtain the constants necessary for multiple comparisons in the general linear model, with no approximations being necessary.

## 9. References

Bofinger, E. (1985) Multiple Comparisons and Type III errors. *J Amer, Statist. Assoc.* **80**, 433-437.

Bretz, F. and Hothorn, L.A. (2000) Testing dose-response relationships with a priori unknown umbrella point using multiple contrast tests. (Submitted for publication)

Bretz, F. (1999) Powerful modifications of Williams' test on trend. Ph. D. Dissertation, University of Hannover. (available at <http://www.bioinf.uni-hannover.de/~bretz/thesis/index.html>)

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

Dunnett, C.W. and Tamhane, A.C. (1991) Stepdown multiple tests for comparing treatments with a control in unbalanced one-way layouts. *Statistics in Medicine* **10**, 939-947.

Hayter, A. J. (1990) A one-sided studentized range test for testing against a simple ordered alternative. *J. Amer. Statist. Assoc.* **85**, 778-785.

Hirotsu, C. (1979) The cumulative chi-squares method and a studentised maximal contrast method for testing an ordered alternative in a one-way analysis of variance model. *Reports of Statistical Applications Research* **26**, 12-21.

Hirotsu, C. (1997) Isotonic inference with particular interest to clinical trials. *Industrial Statistics*, Kitsos, C. P. and Edler, L. (eds.), Physica-Verlag, Jena.

Hsu, J. C. (1981) Simultaneous Confidence intervals for all distances from the 'best'. *Annals of Statistics* **9**, 1026-1034.

Hsu, Jason C. and Nelson, Barry (1997) Multiple comparisons in the general linear model. *Journal of Computational and Graphical Statistics*, Vol **7**, 23-41.

Liu, Wei, Miwa, Tetuhisa and Hayter, A. J. (2000) Simultaneous confidence interval estimation for successive comparisons of ordered treatment effects. *Journal of Statistical Planning and Inference* **88** (1), 75-86

Mack, G.A and Wolfe, D.A. (1981) K-sample rank tests for umbrella alternatives. *J. Amer. Statist. Assoc.* **76**, 175-181.

McDermott, M. P. and Mudholker, G. S. (1993) A simple approach to testing homogeneity of order-constrained means. *J. Amer. Statist. Assoc.* **88**, 1371 - 1379.

Press, William H. et al (1986, 1992) Numerical methods in Fortran, the art of scientific computing, *Cambridge University Press*.

Somerville, Paul N. (1997). Multiple testing and simultaneous confidence intervals: calculation of constants. *Computational Statistics and Data Analysis* **25**, 217-223.

Somerville, Paul N. (1998) Numerical computation of multivariate normal and multivariate-t probabilities over convex regions. *Journal of Computational and Graphical Statistics*, Vol. **7**, No. **4**, 529-544.

Somerville, Paul N. (1999) Critical values for multiple testing and comparisons - one step and stepdown procedures. *Journal of Statistical Planning and Inference* **82**, 129-138.

Somerville, Paul N., Liu, Wei, Miwa, Tetsuhisa and Hayter, A. J. (2001) Combining one-sided and two-sided confidence interval procedures for successive comparisons of ordered treatment effects. *Biometrical Journal* (to appear).

Tukey, J. W. (1953) The problem of multiple comparisons, Mimeographed Monograph (1953).