

## Paper 163-30

## A Macro for Computing Point Estimates and Confidence Intervals for Mahalanobis Distance

Kristine Y. Hogarty

John M. Ferron

Melinda R. Hess

Jeffrey D. Kromrey

University of South Florida, Tampa, FL

### ABSTRACT

This paper discusses a SAS® macro that provides three approaches to statistical inferences about Mahalanobis distance. Mahalanobis distance is useful as a multivariate effect size, being an extension of the standardized mean difference (i.e., Cohen's  $d$ ). This program calculates three point estimates of  $D^2$  (a sample estimate, a jackknife estimate, and an adjusted estimate advanced by Rao, 1973). Further, the program computes a series of confidence bands around each estimated value of  $D^2$ . These confidence intervals are calculated employing two estimation procedures (an interval inversion approach and a common bootstrapping method). The paper presents a demonstration of the SAS/IML® code employed in a macro entitled D2BAND, and provides a table with the resulting output.

### INTRODUCTION

As is true of scholars in many fields, increasing energy has been directed toward encouraging educational researchers to provide comprehensive and accurate analyses of empirical data and thorough, informative reporting of results. Two aspects of statistical analysis that have gained attention over recent years are the use of effect sizes and the application of confidence intervals. The fifth edition of the American Psychological Association's style manual for publication (APA, 2001) cites the failure to report effect sizes, as well as other research issues, as defects in reporting research. However, Thompson (1998) has noted that the 'encouragement' of the APA has not seemed to induce sufficient leverage for researchers to consistently report these informative statistics. As an incentive for researchers to report and interpret effect sizes, some professional journals require authors to supply this information before consideration for publication. Unfortunately, recent research has indicated that while many journals technically require this information, few are enforcing their own mandates (McMillan, Snyder & Lewis, 2002).

The report by Wilkinson and the APA Task Force on Statistical Inference (1999) not only addresses the need for effect size reporting but also stresses the obligation of researchers to provide estimates of confidence intervals for all principal outcomes, including, but not limited to, effect size information. In recent years, research methodologists have investigated the use of confidence intervals (in lieu of traditional point estimates) for the estimation of various parameters of interest (see, for example, Nix & Barnette, 1998; Grissom & Kim, 2001). The use of confidence intervals for effect sizes has also been a topic of empirical study (Hess & Kromrey, 2003; Hess & Kromrey 2002; Cumming & Finch, 2001; Fidler & Thompson, 2001; Steiger & Fouladi, 1992). Such investigations, however, have focused on univariate effect sizes.

In a recent Monte Carlo Study we investigated the accuracy and precision of confidence bands for two-group multivariate effect sizes under a variety of data conditions, including nonnormality of population distributions and covariance matrix heterogeneity. Our findings led to recommendations regarding promising methods for calculating point estimate of multivariate effect size as well as useful approaches to constructing confidence bands. Therefore, the purpose of this paper is to provide a mechanism for other researchers who wish to estimate multivariate estimates of effect size and their corresponding confidence intervals.

### MULTIVARIATE EFFECT SIZES

A commonly used univariate effect size for the two-group case is the standardized mean difference, Cohen's  $d$ . Cohen's  $d$  is defined as the difference in means between groups divided by the pooled standard deviation and is given by:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_p}$$

where  $\bar{X}_1$ ,  $\bar{X}_2$  and  $s_p$  are the sample means of group 1 and 2, and the pooled standard deviation, respectively.

A multivariate extension of the standardized mean difference is the Mahalanobis Distance:

$$D^2 = (\bar{X}_1 - \bar{X}_2)' \hat{\Sigma}^{-1} (\bar{X}_1 - \bar{X}_2)$$

where  $\bar{X}_i$  is the sample mean vector of group  $i$  and  $\hat{\Sigma}$  is the pooled covariance matrix.

As a point estimate, the statistic  $D^2$  is known to evidence small sample bias (Krzanowski & Marriott, 1995). To reduce such bias, an 'adjusted' value of  $D^2$  was suggested by Rao (1973):

$$D_u^2 = \left( \frac{n_1 + n_2 - p - 3}{n_1 + n_2 - 2} \right) D^2 - p \left( \frac{1}{n_1} + \frac{1}{n_2} \right)$$

where  $n_1$  and  $n_2$  are the sample sizes of the two groups and  $p$  is the number of variables in the covariance matrix.

A second approach to the reduction of bias in  $D^2$  is the Jackknife estimator (Rao & Dorvlo, 1985):

$$D_j^2 = (n_1 + n_2) D^2 - (n_1 + n_2 - 1) \bar{D}^2$$

where  $\bar{D}^2$  is the mean value of  $D^2$  obtained from successively removing each observation (i.e., the mean of the leave-one-out values of  $D^2$ ). The leave-one-out values of  $D^2$ , the  $D_{ij}^2$ , can be more explicitly defined by

$$D_{ij}^2 = (\bar{X}_{ij} - \bar{X}_m)' S_{ij}^{-1} (\bar{X}_{ij} - \bar{X}_m),$$

where  $\bar{X}_{ij}$  and  $S_{ij}$  are the sample mean vector and covariance matrix of group  $i$  after observation  $j$  has been removed from group  $i$ , and  $\bar{X}_m$  is the sample mean vector for the other group.

With input from the researcher, the MACRO D2BAND currently computes all three of the aforementioned measures of multivariate effect size.

#### CONFIDENCE INTERVALS FOR MULTIVARIATE EFFECT SIZES

Several methods for constructing confidence bands around the multivariate point estimates of effect size have been suggested. An elegant approach for constructing confidence intervals around  $D^2$  is the interval inversion approach (Reiser, 2001). The computational details of such an approach were lucidly described by Steiger and Fouladi (1992, 1997), although this approach to confidence interval estimation was detailed earlier (see, for example, Venables, 1975; Serlin & Lapsley, 1985). Interval inversion has been shown to provide accurate confidence bands around univariate effect sizes in a variety of conditions (Kromrey & Hess, 2001; Hess & Kromrey, 2002). This approach seeks the population values of  $D^2$  for which the sample value would be expected (for example) 2.5% of the time and 97.5% of the time. Such values provide the endpoints of a 95% confidence interval. In its execution, however, the approach is based on a function of  $D^2$ :

$$\frac{n_1 n_2 (n_1 + n_2 - p - 1)}{(n_1 + n_2)(n_1 + n_2 - 2)p} D^2$$

which is distributed as  $F$  with  $df = p, n_1 + n_2 - p - 1$  and the noncentrality parameter  $\lambda$

where  $\lambda = \frac{n_1 n_2}{n_1 + n_2} \delta^2$  and  $\delta^2$  is the population value of the Mahalanobis Distance.

Thus, the interval inversion approach seeks the values of noncentrality that correspond to the target percentiles of the appropriate  $F$  distribution.

#### BOOTSTRAP METHODS

Bootstrapping is commonly recognized as an efficient method for providing estimates for, among other things, confidence intervals (Efron & Gong, 1983; Efron & Tibshirani, 1986; Stine, 1990). In this procedure, samples of observations are drawn with replacement from the observed data, and the statistic of interest (e.g.,  $D^2$ , Rao's adjusted  $D^2$ ) is computed on the resampled observations. Over the years, a variety of algorithms have been proposed to construct bootstrap intervals. Carpenter and Bithell (2000) investigated eight different bootstrap algorithms using data from a medical study to address this issue. While their study used a different statistic of interest, the methods they

used appear to be generalizable to other statistics, such as multivariate effect sizes. For our initial investigation, four methods were identified as promising for further research. Here we present the bootstrapping method that provided the highest level of accuracy and precision.

This basic method of bootstrapping (called the naïve bootstrap or the percentile method) consists of repeatedly drawing samples of size  $n$  with replacement (e.g., 1000 times or 5000 times) from a single sample of  $n$  observations. Each bootstrap sample provides an estimate of the effect size, and the set of estimates provides an empirical sampling distribution for the effect size. Percentile points in this empirical sampling distribution (e.g., the 2.5<sup>th</sup> percentile and the 97.5<sup>th</sup> percentile) provide the endpoints of the confidence interval. The percentile method is a relatively simple calculation and has the advantage of not requiring an estimate of the standard error. Unfortunately, what this method provides in simplicity and appeal also results in a tendency to be less successful for non-normal distributions.

The MACRO D2BAND currently computes three different band-widths (i.e., 80%, 95%, 99%) using the aforementioned methods of estimation (i.e., the interval inversion approach for  $D^2$ , and the bootstrapping method for the two modified multivariate effect sizes).

The remainder of the paper will describe and present the MACRO D2BAND.

### MACRO D2BAND

The macro D2BAND is called from SAS after the data have been read into a SAS data set. Inputs to the macro include the name of the SAS data set that contains the data to be analyzed (DSN), the name of the independent variable (Group) and the names of the dependent variables (DVs).

```
%macro D2BAND(DSN,Group,DVs);

proc iml;

* +-----+
  Subroutine MAHL computes Mahalanobis Distance and the Jackknife Distance.
  Inputs are data matrices for two groups (G1 and G2).
  Outputs are D2 and D2_J
+-----+;

START MAHL(G1,G2,D2,D2_J);
p=ncol(G1);n1=nrow(G1);n2=nrow(G2);
AG1 = J(n1,1,1)||G1;
AG2 = J(n2,1,1)||G2;
SSCP1 = AG1` * AG1;
SSCP2 = AG2` * AG2;
W1 = SSCP1[2:P+1,2:P+1];
do r = 1 to p;
do c = 1 to p;
if r=c then W1[r,c]=W1[r,c]- (SSCP1[1,c+1]##2/SSCP1[1,1]);
if r^=c then W1[r,c]=W1[r,c]- (1/SSCP1[1,1]#SSCP1[1,c+1]# SSCP1[1,r+1]);
end;
end;

W2 = SSCP2[2:P+1,2:P+1];
do r = 1 to p;
do c = 1 to p;
if r = c then W2[r,c] = W2[r,c] - (SSCP2[1,c+1]##2 / SSCP2[1,1]);
if r^=c then W2[r,c]=W2[r,c]-(1/SSCP2[1,1]#SSCP2[1,c+1]# SSCP2[1,r+1]);
end;
end;

W = W1 + W2;
cent1 = J(1,n1,1) * G1 * (1/n1);
cent2 = J(1,n2,1) * G2 * (1/n2);
Cov_matrix=(1/(n1+n2-2))*W;
D2=(cent1-cent2)*INV(Cov_matrix) *(cent1-cent2)`;

* +-----+
  Remove each observation from G1 and recompute Mahalanobis Distance
+-----+;

D2_1 = 0;
do i = 1 to n1;
one_obs = AG1[i,];
```

```

one_sscp = one_obs` * one_obs;
SSCP1_i = SSCP1 - one_sscp;
W1_i = SSCP1_i[2:P+1,2:P+1];
do r = 1 to p;
  do c = 1 to p;
    if r=c then W1_i[r,c]=W1_i[r,c]-(SSCP1_i[1,c+1]##2/SSCP1_i[1,1]);
    if r^=c then W1_i[r,c]=W1_i[r,c]-(1/SSCP1_i[1,1]#SSCP1_i[1,c+1]# SSCP1_i[1,r+1]);
  end;
end;
W_i = W1_i + W2;

if i=1 then G1_i=G1[2:n1,];
if (i>1 & i<n1) then G1_i=G1[1:i-1,]/ G1[i+1:n1,];
if i=n1 then G1_i=G1[1:n1-1,];
cent1 = J(1,n1-1,1) * G1_i * (1/(n1-1));
cent2 = J(1,n2,1) * G2 * (1/n2);
Cov_matrix = (1/(n1 + n2 - 3)) * W_i;
D2_1=D2_1+((cent1-cent2)*INV(Cov_matrix)*(cent1-cent2`));
end;

* +-----+
  Remove each observation from G2 and recompute Mahalanobis Distance
+-----+;
D2_2 = 0;
do i = 1 to n2;
  one_obs = AG2[i,];
  one_sscp = one_obs` * one_obs;
  SSCP2_i = SSCP2 - one_sscp;
  W2_i = SSCP2_i[2:P+1,2:P+1];
  do r = 1 to p;
    do c = 1 to p;
      if r=c then W2_i[r,c]=W2_i[r,c]-(SSCP2_i[1,c+1]##2/SSCP2_i[1,1]);
      if r^=c then W2_i[r,c]=W2_i[r,c]-(1/SSCP2_i[1,1]#SSCP2_i[1,c+1]# SSCP2_i[1,r+1]);
    end;
  end;
  W_i = W1 + W2_i;
  if i=1 then G2_i=G2[2:n2,];
  if (i>1 & i<n2) then G2_i=G2[1:i-1,]/ G2[i+1:n2,];
  if i = n2 then G2_i = G2[1:n2-1,];
  cent1 = J(1,n1,1) * G1 * (1/n1);
  cent2 = J(1,n2-1,1) * G2_i * (1/(n2-1));
  Cov_matrix = (1/(n1 + n2 - 3)) * W_i;
  D2_2=D2_2+((cent1-cent2)*INV(Cov_matrix)*(cent1-cent2`));
end;
D_bar = (D2_1 + D2_2)/(n1 + n2);
D2_J = (n1+n2)#D2-(n1+n2-1)#D_bar;
FINISH;

* +-----+
  Subroutine BOOTSTRP draws a random sample with replacement from a data matrix.
  Inputs are the original data matrix (RWDATA) and desired size of bootstrap sample (BOOTN).
  Output is the random sample (B_X)
+-----+;

START BOOTSTRP (RWDATA, BOOTN, B_X);
ORIG_X = RWDATA;
BIG_N = NROW(ORIG_X);
DO i = 1 to BOOTN;
  ranrow=round((ranuni(0)*BIG_N)+0.4999);
  if ranrow = 0 then ranrow = 1;
  if i = 1 then do;
    B_X = ORIG_X[ranrow,];
  end;
  if i>1 then do;
    B_X = B_X/ORIG_X[ranrow,];
  end;
end;
free big_n i ranrow;
FINISH;

```

```
* +-----+
Subroutine BOOTP evaluates percentiles from a vector of bootstrapped statistics.
Input is the vector of bootstrapped statistics (md_vec).
Outputs are values of the statistic at defined percentiles.
+-----+;
```

```
START BOOTP(md_vec,lo99,hi99,lo95,hi95, lo80,hi80);
Nboots = nrow(md_vec);
```

```
r_vec = rank(md_vec);
pct025 = round(nboots#.025);
pct975 = round(nboots#.975);
pct005 = round(nboots#.005);
pct995 = round(nboots#.995);
pct100 = round(nboots#.100);
pct900 = round(nboots#.900);
do bb= 1 to nboots;
if r_vec[bb]=pct025 then lo95=MD_vec[bb];
if r_vec[bb]=pct975 then hi95=MD_vec[bb];
if r_vec[bb]=pct005 then lo99=MD_vec[bb];
if r_vec[bb]=pct995 then hi99=MD_vec[bb];
if r_vec[bb]=pct100 then lo80=MD_vec[bb];
if r_vec[bb]=pct900 then hi80=MD_vec[bb];
end;
FINISH;
```

```
* +-----+
Subroutine MAHL_D_PCTL computes percentiles from the sampling distribution of Mahalanobis Distance using the interval
inversion method.
Inputs are the sample value of D2 (smpL_MD), sample sizes in the two groups (n1 and n2), the number of variables in each
group (p), and the desired percentile from the sampling distribution (pctl).
Output is population value of D2 that provides smpL_MD at the pctl percentile.
Subroutine returns the value 999 if the percentile cannot be obtained numerically
+-----+;
```

```
START MAHL_D_PCTL(smpL_MD,n1,n2,p,pctl, mahl_d_t);
if smpL_MD<0 then smpL_MD=0;
d2_tilde=((n1*n2)*(n1+n2-p-1))/((n1+n2)*(n1+n2-2)*p)*smpL_MD;
df1 = p;
df2 = n1+n2-p-1;
```

```
* Step 1: Find value of D2 that is a little too high;
```

```
OK = 0;
oops = 0;
mahl_d_t = 0; * start the loop with D2 = 0;
do until (OK = 1);
nc = mahl_d_t # ((n1#n2)/(n1+n2));
if nc < 5500 then do;
cumprob = PROBF(d2_tilde,df1,df2,nc);
if cumprob<pctl then OK = 1;
if cumprob>pctl then mahl_d_t=mahl_d_t+1;
end;
if nc >= 5500 then do;
mahl_d_t = 999;
OK = 1;
end;
end;
high = mahl_d_t;
```

```
* Step 2: Establish value of D2 that is a little too low (D2 = 0);
low = 0;
```

```
* Step 3: Successively halve the interval between low and high to obtain final value of percentile;
```

```
change = 1;
loop = 0;
small = .00000000001;
do until (change<small);
half = (high + low)/2;
```

```

nc = half # ((n1#n2)/(n1+n2));
if nc < 5500 then do;
  cumprob = PROBF(d2_tilde,df1,df2,nc);
  if cumprob < pctl then high = half;
  if cumprob > pctl then low = half;
  change = abs(high - low);
  loop = loop + 1;
  if loop > 1500 then small = .000000001;
  if loop > 3000 then small = .01;
  mahl_d_t = (high + low)/2;
end;
if nc >= 5500 then do;
  mahl_d_t = 999;
  change = 0;
end;
end;
FINISH;

* +-----+
Main Program
+-----+;

* +-----+
Read data from regular SAS into IML
+-----+;
use &DSN;
read all var{&DVs} where (&Group=1) into G1;
read all var{&DVs} where (&Group=2) into G2;
n1 = nrow(G1);
n2 = nrow(G2);
p = ncol(G1);

* +-----+
Compute D2, Jackknife D2, and Rao adjusted D2
+-----+;
run mahl(G1, G2, MD, MD_J);
D2_u1=((n1 + n2 - p -3)/(n1 + n2 -2))*MD;
D2_u2= p*(1/n1 + 1/n2);
D2_u= D2_u1 - D2_u2;

* +-----+
Compute endpoints of interval inversion confidence bands for D2
+-----+;
run mahl_d_PCTL(MD,n1,n2,p,0.005,hi_l_99);
run mahl_d_PCTL(MD,n1,n2,p,0.995,lo_l_99);
run mahl_d_PCTL(MD,n1,n2,p,0.025,hi_l_95);
run mahl_d_PCTL(MD,n1,n2,p,0.975,lo_l_95);
run mahl_d_PCTL(MD,n1,n2,p,0.100,hi_l_80);
run mahl_d_PCTL(MD,n1,n2,p,0.900,lo_l_80);

* +-----+
Generate 5000 Bootstrap Samples and place bootstrap statistics in matrix boot_mtx
+-----+;
do bb = 1 to 5000;
run bootstrp (G1, n1, b1);
run bootstrp (G2, n2, b2);
run mahl(b1, b2, D2b, MDJ_b);
D2_u1b=((n1 + n2 - p -3)/(n1 + n2 -2))*D2b;
D2_ub= D2_u1b - D2_u2;
if bb=1 then BOOT_MTX = MDJ_b||D2_ub;
if bb>1 then BOOT_MTX=BOOT_MTX// (MDJ_b||D2_ub);
free b1 b2 MD_b MDJ_b D2_ub;
end;

* +-----+
Compute endpoints of bootstrap CI for Jackknife and Rao adjusted D2
+-----+;
test_vec = BOOT_MTX[,1];
run BOOTP(test_vec,l9j,h9j,l5j,h5j,l0j,h0j);
test_vec = BOOT_MTX[,2];

```

```

run BOOTP(test_vec,l9r,h9r,l5r,h5r,l0r,h0r);

* +-----+
  Print output results
+-----+;
file print;
put @1 'Mahalanobis Distance Estimates & Intervals' //
@1 'Mahalanobis Distance' /
@1 '-----' /
@1 'N of Variables' @47 p 3. //
@1 'Sample Sizes' @41 n1 4. @45 ', ' @46 n2 4. //
@1 '-----' /
@1 '    Interval Inversion Confidence Bands' /
@1 '-----' ///
@1 'Sample Value of D2' @42 MD 8.5 //
@1 'Confidence Bands' @25 'Lower Limit'
@40 'Upper Limit' /
@1 '-----' @25 '-----'
@40 '-----' /
@7 '80%' @27 lo_l_80 8.5 @42 hi_l_80 8.5 /
@7 '95%' @27 lo_l_95 8.5 @42 hi_l_95 8.5 /
@7 '99%' @27 lo_l_99 8.5 @42 hi_l_99 8.5 ///
@1 '-----' /
@1 '    Bootstrap Confidence Bands' /
@1 '-----' ///
@1 'Sample Value of Jackknife D2' @42 MD_J 8.5 //
@1 'Confidence Bands' @25 'Lower Limit'
@40 'Upper Limit' /
@1 '-----' @25 '-----'
@40 '-----' /
@7 '80%' @27 l0j 8.5 @42 h0j 8.5 /
@7 '95%' @27 l5j 8.5 @42 h5j 8.5 /
@7 '99%' @27 l9j 8.5 @42 h9j 8.5 //
@1 '-----' ///
@1 'Sample Value of Rao Adjusted D2' @42 D2_U 8.5 //
@1 'Confidence Bands' @25 'Lower Limit'
@40 'Upper Limit' /
@1 '-----' @25 '-----'
@40 '-----' /
@7 '80%' @27 l0r 8.5 @42 h0r 8.5 /
@7 '95%' @27 l5r 8.5 @42 h5r 8.5 /
@7 '99%' @27 l9r 8.5 @42 h9r 8.5 //
@1 '-----';
quit;
%mend D2BAND;

```

### EXAMPLE OF MACRO INVOCATION

An example of the macro's use is provided by a study of gender differences on four psychological traits. A sample of 32 undergraduate male and 32 undergraduate female psychology students completed measures of each trait. The data were analyzed by testing for differences in population mean vectors (using Hotelling's  $T^2$ ) and by estimating the magnitude of the multivariate difference using Mahalanobis Distance. As the following code illustrates, the data were read into a SAS data set (GEN\_DIFFS), and the macro D2BAND was called using a run statement. The arguments to the macro included the name of the SAS data set, the name of the grouping variable (GENDER), and the name of the dependent variables for the analysis (T1 – T4). The results of this analysis are presented in Table 1.

```

data gen_diffs;
input gender 1 t1 3-4 t2 6-7 t3 9-10 t4 12-13;
cards;
1 15 17 24 14
1 17 15 32 26
.
.
2 14 16 22 26
2 17 17 22 28
;
%D2BAND(gen_diffs,gender,t1 t2 t3 t4); run;

```

### OUTPUT FROM MACRO D2BAND

Table 1 provides an example of the output produced by the macro D2BAND.

Table 1

## Mahalanobis Distance Estimates &amp; Intervals

## Mahalanobis Distance

N of Variables	4
----------------	---

Sample Sizes	32, 32
--------------	--------

## Interval Inversion Confidence Bands

Sample Value of $D^2$	6.10009
-----------------------	---------

Confidence Bands	Lower Limit	Upper Limit
80%	3.71199	7.80135
95%	2.88844	9.15019
99%	2.23190	10.47390

## Bootstrap Confidence Bands

Sample Value of Jackknife $D^2$	5.13510
---------------------------------	---------

Confidence Bands	Lower Limit	Upper Limit
80%	3.98709	8.68028
95%	3.27355	10.58943
99%	2.74402	12.65897

Sample Value of Rao Adjusted $D^2$	5.35815
------------------------------------	---------

Confidence Bands	Lower Limit	Upper Limit
80%	4.17527	8.98404
95%	3.41742	10.94231
99%	2.81740	13.14760

**RESULTS**

For these data, the sample value of  $D^2$  is 6.10009, a very large multivariate difference between the males and females. Because  $D^2$  is known to be positively biased in small samples, better point estimates of the multivariate difference are provided by the Jackknife  $D^2$  (5.13510) or Rao's adjusted  $D^2$  (5.35815). The interval estimates of the population value of  $D^2$  reflect the large amount of uncertainty about this parameter, because of the small sample of research participants. For example, the 95% confidence interval constructed using interval inversion ranges from 2.88844 to 9.15019. Somewhat wider confidence intervals are obtained from the bootstrap estimates.

**EMPIRICAL STUDIES OF POINT ESTIMATES AND CONFIDENCE INTERVALS**

As stated earlier, in a recent Monte Carlo study we investigated the accuracy and precision of three point estimates and confidence bands for two-group multivariate effect sizes under a variety of data conditions, including nonnormality of population distributions and covariance matrix heterogeneity. The results of our study were initially analyzed by examining confidence band coverage probabilities and confidence band-widths. With respect to confidence band coverage estimates for Mahalanobis Distance ( $D^2$ ), the interval inversion approach emerged as the superior method, outperforming the percentile approach to the bootstrap. When examining the performance of the two computational methods employing the Jackknife estimate of  $D^2$ , a different picture emerged. In this case, the interval inversion method exhibited a rapid deterioration, while the percentile bootstrap method improved dramatically. Similarly, when we compared the two confidence band estimation methods for the Rao adjusted  $D^2$ , the superior performance of the percentile bootstrap was evident, while the interval approach reported probability coverage levels below 50%.



Confidence band-width for the three multivariate effect size indices was computed by employing the interval inversion approach for  $D^2$  and the percentile bootstrap methods for the other two estimates. An examination of the three estimates of  $D^2$ , revealed that the Jackknife and Rao estimates of  $D^2$  evidenced more variability across the conditions than the interval approach for  $D^2$ .

Although the focus of this study was on interval estimates of effect size, the bias in each of the indices as point estimates of multivariate effect size was also examined. The Jackknife estimate of  $D^2$  appeared to exhibit less statistical bias than the other two methods of computing  $D^2$ . The original Mahalanobis  $D^2$  was observed to evidence considerable bias on average and considerable variability in bias across the conditions examined. The Jackknife and Rao estimates appeared to fare better, evidencing little bias and appearing fairly consistent across conditions. However, the box and whisker plots indicate that the Rao estimate was slightly more biased across more conditions than the Jackknife estimate.

## CONCLUSION

Naturally, this macro can also be modified to other contexts depending on the specifications of the user. For example, the macro currently calculates 80%, 95%, and 99% confidence intervals. Simple modifications of the macro will produce 90% confidence intervals or one-sided confidence intervals. Additionally, the macro currently estimates bootstrap confidence intervals based on 5000 bootstrap samples (the value hard coded in the bootstrapping DO loop). If greater precision in the estimation of the endpoints of these bootstrap intervals is desired, this value can be increased. The precision of the estimation of confidence interval endpoints is proportional to the number of bootstrap samples drawn.

## REFERENCES

- American Psychological Association (2001). *Publication manual of the American Psychological Association* (5<sup>th</sup> ed.). Washington, DC: Author.
- Bradley, J.V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144-151.
- Carpenter, J. & Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, 19, 1141 – 1164.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37 -- 46.
- Cumming G. & Finch (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement* 61(4), 532-74.
- Efron, B. & Gong, G. (1983). *A leisurely look at the bootstrap, the jackknife, and cross-validation*. *The American Statistician*, 37(1), pg 36-49.
- Efron, B. & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1(1), p. 54-77.
- Ferron, J., Yi, Q. & Kromrey, J. D. (1997). NNCORR: A SAS/IML program for generating nonnormal correlated data. *Applied Psychological Measurement*, 21, 64.
- Fidler F. & Thompson B. (2001). Computing correct confidence intervals for ANOVA fixed and random-effects effect sizes. *Educational and Psychological Measurement*, 61(4), 575-604
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43(4), p.521-532.
- Ghosh, B. K. (1979). A comparison of some approximate confidence intervals for the binomial parameter. *Journal of the American Statistical Association*, 74, 894 – 900.
- Grissom R.J. & Kim J.J. (2001). Review of assumptions and problems in the appropriate conceptualization of effect size. *Psychological Methods*, 6(2), p. 135-146.
- Hedges L.V. & Olkin I. (1985). *Statistical Methods for Meta-Analysis*. New York: Academic Press.
- Hess, M.R. & Kromrey, J.D. (2003, April). *Confidence Intervals for Standardized Mean Differences: An Empirical Comparison of Bootstrap Methods Under Non-normality and Heterogeneous Variances*. Paper presented at the American Educational Research Association, Chicago, IL.
- Hess, M.R. & Kromrey, J.D. (2002, April). *Confidence intervals for the standardized mean difference: An empirical comparison of methods for interval estimation of effect sizes*. Paper presented at the American Educational Research Association, New Orleans, LA.

- Huberty, C. J. (1994). *Applied discriminant analysis*. N.Y.: Wiley.
- Huberty, C. J. & Lowman, L. L. (2000). Group overlap as a basis for effect size. *Educational and Psychological Measurement*, 60, 543 – 563.
- Inman, H. F. & Bradley, E. L. (1989). The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Communications in Statistics – Theory and Methods*, 18, 3851 – 3874.
- Kromrey, K. Y. & Hess, M. H. (2001, April). *Interval Estimates of  $R^2$ : An empirical comparison of accuracy and precision under violations of the normality assumption*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Krzanowski, W. J. & Marriott, F. H. C. (1995). *Multivariate analysis: Classification, covariance structures and repeated measurements*. London: Arnold.
- McMillan, J.H., Snyder, A., Lewis, K.L., (2002, April). *Reporting Effect Size: The Road Less Traveled*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Nix, T.W. & Barnette, J. J. (1998). The data analysis dilemma: Ban or abandon. A review of null hypothesis significance testing. *Research in the Schools*, 5(2), p. 3-14.
- Rao, C. R. (1973). *Linear statistical inference and its applications* (2<sup>nd</sup> Edition). N.Y.: Wiley.
- Rao, P. S. R. S. & Dorvlo, A. S. (1985). The Jackknife procedure for the probabilities of misclassification. *Communications in Statistics – Simulation and Computation*, 14, 779 – 790.
- Reiser, B. (2001). Confidence intervals for the Mahalanobis Distance. *Communications in Statistics – Simulation and Computation*, 30, 37 – 45.
- Robey, R.R. & Barcikowski, R.S. (1992). Type I error and the number of iterations in Monte Carlo studies of robustness. *British Journal of Mathematical and Statistical Psychology*, 45, 283-288.
- Serlin, R. C. & Lapsley, D. K. (1985). Rationality in psychological research: The good-enough principle. *American Psychologist*, 40, 73-83.
- Steiger, J. H. & Fouladi, R. T. (1992). R2: A computer program for interval estimation, power calculation, and hypothesis testing for the squared multiple correlation. *Behavior Research, Methods, Instruments, and Computers*, 4, 581-582.
- Steiger, J. H. & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In Harlow, L. L., Mulaik, S. A. & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum, p. 221-257.
- Stevens, J. P. (1980). Power of the multivariate analysis of variance tests. *Psychological Bulletin*, 88, 728 - 737.
- Stine, R. (1990). An introduction to bootstrap methods. *Sociological Methods and Research*, 18 (2&3), p. 243-291.
- Thompson, B. (1998). Statistical significance and effect size reporting: Portrait of a possible future. *Research in the Schools*, 5(2), p. 33-38.
- Venables, W. (1975). Calculation of confidence intervals for noncentrality parameters. *Journal of the Royal Statistical Society, Series B*, 37, 406-412.
- Wilkinson & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.

### CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Kristine Y. Hogarty  
 University of South Florida  
 4202 E. Fowler Ave., EDU162  
 Tampa, FL 33620  
 (813) 974-3198  
 khogarty@luna.cas.usf.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Other brand and product names are trademarks of their respective companies.