LEAST COST FORMULATIONS, LTD.
824 Timberlake Drive, Virginia Beach, VA 23464-3239
Tel: (757) 467-0954   Fax: (757) 467-2947
E-mail: office@lcftld.com   URL: http://lcfltd.com/

TECHNICAL REPORT

NUMBER: TR298   DATE: 09 June 17

TITLE: Confidence intervals on variance components.

AUTHOR: R. A. LaBudde

ABSTRACT: Formulas are given for confidence intervals of variance components which are linear combinations of independent normal variances. Simulation using R indicates very accurate coverage for 95% confidence intervals. The Graybill-Wang formulas are recommended for general use as most accurate in coverage, so long as numerical problems do not arise when negative coefficients are present. Otherwise the Ting et al. interval must be used.

KEYWORDS: 1) CI   2) VARIANCE   3) COMPONENT

REL.DOC.:

REVISED:

INTRODUCTION

Variance components are the latent variables estimated in a random effects model, such as

$$Y_{ijk} = \mu + A_i + B_j + E_{ijk} \qquad (1.1)$$

Here, the $A_i \sim N(0, \sigma_A^2)$, $B_j \sim N(0, \sigma_B^2)$ and $E_{ijk} \sim N(0, \sigma^2)$, and σ_A^2, σ_B^2 and σ^2 are the variance components to be estimated based on observations on Y_{ijk}.

All the formulas that follow assume independent normal distributions.

CONFIDENCE INTERVAL FOR A SINGLE VARIANCE

Suppose we wish a 95% confidence interval on a single variance σ^2 estimated by $v = s^2$, where s is the standard deviation estimate. Then the interval is

$$LCL(\sigma^2) = v / (\chi_{0.975;\, f}^2 / f) \qquad (2.1a)$$

$$UCL(\sigma^2) = v / (\chi_{0.025;\, f}^2 / f) \qquad (2.1b)$$

where f is the degrees of freedom associated with v and $\chi_{0.975;\, f}^2$ and $\chi_{0.025;\, f}^2$ are the 97.5% and 2.5% quantiles of the χ^2 distribution for f degrees of freedom. (For definiteness of notation, note that $\chi_{0.025;\, 10}^2 = 3.25$ and $\chi_{0.975;\, 10}^2 = 20.48$. Frequently these values are switched in tables and in software.)

If a 95% confidence on σ is needed, take the square roots of eqs.(2), i.e.:

$$LCL(\sigma) = s / \sqrt{(\chi_{0.975;\, f}^2 / f)} \qquad (2.2a)$$

$$UCL(\sigma) = s / \sqrt{(\chi_{0.025;\, f}^2 / f)} \qquad (2.2b)$$

CONFIDENCE INTERVAL FOR A LINEAR COMBINATION OF VARIANCES

Suppose a variance component v is to be estimated from a linear combination of independent estimated variances $v_1, v_2, ..., v_p$:

$$v = \Sigma a_i v_i \qquad (3.1)$$

The Graybill-Wang approximate 95% confidence interval is [1]

$$LCL(\sigma^2) = v - \sqrt{ \{ \Sigma (g_i a_i v_i)^2 \} } \qquad (3.2a)$$

$$UCL(\sigma^2) = v + \sqrt{ \{ \Sigma (h_i a_i v_i)^2 \} } \qquad (3.2b)$$

where

$$g_i = 1 - 1 / (\chi_{0.975;\, fi}^2 / f_i) \qquad (3.3a)$$

$$h_i = 1 / (\chi_{0.025;\, fi}^2 / f_i) - 1 \qquad (3.3b)$$

and f_i is the degrees of freedom associated with v_i.

As before, if a 95% confidence interval on $s = \sqrt{v}$ is needed, use the square roots of LCL and UCL in eqs.(3.2).

Burdick and Graybill indicate that eqs.(3.2) should not be used when any of the a_i are negative.

Note that eqs.(3.2) include eqs.(2.1) as a subcase.

CONFIDENCE FOR A DIFFERENCE OF VARIANCES

Suppose a variance component v is to be estimated from a difference of independent estimated variances v_1 and v_2:

$$v = a_1 v_1 - a_2 v_2 \quad (4.1)$$

The Ting et al. approximate 95% confidence interval is [1]

$$LCL(\sigma^2) = v - \sqrt{ \{ \Sigma L \} } \quad (4.2a)$$

$$UCL(\sigma^2) = v + \sqrt{ \{ \Sigma U \} } \quad (4.2b)$$

where

$$L = (g_1 a_1 v_1)^2 + (h_2 a_2 v_2)^2 + g_{12} a_1 a_2 v_1 v_2 \quad (4.3a)$$

$$g_1 = 1 - 1 / (\chi_{0.975;\, f1}^2 / f_1) \quad (4.3b)$$

$$h_2 = 1 / (\chi_{0.025;\, f2}^2 / f_2) - 1 \quad (4.3c)$$

$$g_{12} = \frac{\{ (F_{0.975;f1,f2} - 1)2 - g_1^2 F_{0.975;f1,f22}^2 - h_2^2 \}}{F_{0.975;f1,f2}} \quad (4.3d)$$

$$U = (h_1 a_1 v_1)^2 + (g_2 a_2 v_2)^2 + h_{12} a_1 a_2 v_1 v_2 \quad (4.3e)$$

$$g_2 = 1 - 1 / (\chi_{0.975;\, f2}^2 / f_2) \quad (4.3f)$$

$$h_1 = 1 / (\chi_{0.025;\, f1}^2 / f_1) - 1 \quad (4.3g)$$

$$g_{12} = \frac{\{ (F_{0.025;f1,f2} - 1)2 - h_1^2 F_{0.025;f1,f22}^2 - g_2^2 \}}{F_{0.025;f1,f2}} \quad (4.3h)$$

and f_i is the degrees of freedom associated with v_i, and $F_{0.025;f1,f2}$ and $F_{0.975;f1,f2}$ are quantiles of the F distribution.

If σ^2 is known to be positive, then replace any negative LCL or UCL in eqs.(4.2) by 0.

As before, if a 95% confidence interval on $s = \sqrt{v}$ is needed, use the square roots of LCL and UCL in eqs.(3.2).

SIMULATION

To assess the coverage accuracies of the above intervals, two independent normal variances v_1 and v_2 with degrees of freedom $f_1 = 10$ and $f_2 = 30$ were sampled 10,000 times from the relevant χ_f^2 distributions with $\sigma_1^2 = 4$ and $\sigma_2^2 = 2$.

The mean values found from the sampling were 4.023 and 1.999 for v_1 and v_2, respectively, within expected precision.

The confidence intervals from eqs.(2.1) for σ_1^2 averaged [1.96, 12.39], close to the [1.95, 12.32] expected. The interval coverage was 94.8%, very close to the 95% expected.

The confidence intervals from eqs.(2.1) for σ_2^2 averaged [1.28, 3.57], identical to the [1.28, 3.57] expected. The interval coverage was 95.0%.

A confidence interval for v = v1 + v2 was estimated using eqs.(3.2). The mean interval was [3.81, 14.58] with a coverage of 94.8%.

A confidence interval for v = 3 v1 + v2 was estimated using eqs.(3.2). The mean interval was [7.83, 39.23] with a coverage of 94.8%.

A confidence interval for v = v1 - v2 was estimated using eqs.(3.2). The mean interval was [-0.19, 10.58] with a coverage of 95.5%. A similar interval using the supposedly better eqs.(4.2) gave a mean interval of [-1.12, 10.37] with a coverage of 91.5%.

A confidence interval for v = 3 v1 - v2 was estimated using eqs.(3.2). The mean interval was [3.84, 35.23] with a coverage of 95.1%. A similar interval using the supposedly better eqs.(4.2) gave a mean interval of [4.61, 35.12] with a coverage of 91.5%.

It appears that eqs.(3.2) are quite accurate in coverage, and are more accurate than eqs.(4.2), despite the negative coefficient and the recommendation in [1].

REFERENCES

1. Burdick, R.K. and Graybill, F.A. (1992). *Confidence intervals on variance components*. Marcel Dekker, NY. ISBN 0-8247-8644-0.

R SCRIPT FOR SIMULATION

```
#06.16.09 23.10 test-vcCI.r
#copyright 2009 by Robert A LaBudde, all rights reserved
#Test of vcCI.r confidence intervals on variances
#created: 06.16.09 by r.a. labudde
#changes:

source('vcCI.r')

nReal<- 10000
df1<- 10
df2<- 30

V1<- 4
V2<- 2

stats<- matrix(rep(0,nReal*18),ncol=18)
for (iReal in 1:nReal) {  simulate
  v1<- V1*rchisq(1,df1)/df1
  v2<- V2*rchisq(1,df2)/df2
  CI1<- vcCI1(v1,df1) #CI on v1
  CI2<- vcCI1(v2,df2) #CI on v2
  CI3<- vcCI2p(1,v1,df1,1,v2,df2)   #CI on v1 + v2
  CI4<- vcCI2p(3,v1,df1,1,v2,df2)   #CI on 3*v1 + v2
  CI5<- vcCI2n(1,v1,df1,1,v2,df2,bPos=FALSE)   #CI on v1 - v2
  CI6<- vcCI2n(3,v1,df1,1,v2,df2,bPos=FALSE)   #CI on 3*v1 - v2
  CI7<- vcCI2p(1,v1,df1,-1,v2,df2)  #CI on v1 - v2 using vcCI2p
  CI8<- vcCI2p(3,v1,df1,-1,v2,df2)  #CI on 3*v1 - v2 using vcCI2p
  stats[iReal,]<- c(v1,v2,CI1,CI2,CI3,CI4,CI5,CI6,CI7,CI8)
}

colMeans(stats)
sum(V1>=stats[,3] & V1<=stats[,4]) #coverage for V1 using vcCI1
sum(V2>=stats[,5] & V2<=stats[,6]) #coverage for V2 using vcCI1
sum(V1+V2>=stats[,7] & V1+V2 <=stats[,8]) #coverage for V1+V2 using vcCI2p
sum(3*V1+V2>=stats[,9] & 3*V1+V2 <=stats[,10]) #coverage for 3*V1+V2 using vcCI2p
sum(V1-V2>=stats[,11] & V1-V2 <=stats[,12]) #coverage for V1-V2 using vcCI2n
sum(3*V1-V2>=stats[,13] & 3*V1-V2 <=stats[,14]) #coverage for 3*V1-V2 using vcCI2n
sum(V1-V2>=stats[,15] & V1-V2 <=stats[,16]) #coverage for V1-V2 using vcCI2p
sum(3*V1-V2>=stats[,17] & 3*V1-V2 <=stats[,18]) #coverage for 3*V1-V2 using vcCI2p
```