

11-1-2013

# Constructing Confidence Intervals for Effect Sizes in ANOVA Designs

Li-Ting Chen

*Indiana University, Bloomington, IN, [litngc@unr.edu](mailto:litngc@unr.edu)*

Chao-Ying Joanne Peng

*Indiana University, Bloomington, IN, [peng@indiana.edu](mailto:peng@indiana.edu)*

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

## Recommended Citation

Chen, Li-Ting and Peng, Chao-Ying Joanne (2013) "Constructing Confidence Intervals for Effect Sizes in ANOVA Designs," *Journal of Modern Applied Statistical Methods*: Vol. 12 : Iss. 2 , Article 5.

DOI: [10.22237/jmasm/1383278640](https://doi.org/10.22237/jmasm/1383278640)

---

# Constructing Confidence Intervals for Effect Sizes in ANOVA Designs

## **Cover Page Footnote**

This research was supported in part by the Maris M. Proffitt and Mary Higgins Proffitt Endowment Grant of Indiana University, awarded to the second author while the first author worked on the project as a research assistant.

# Constructing Confidence Intervals for Effect Sizes in ANOVA Designs

**Li-Ting Chen**

Indiana University  
Bloomington, IN

**Chao-Ying Joanne Peng**

Indiana University  
Bloomington, IN

---

A confidence interval for effect sizes provides a range of plausible population effect sizes (ES) that are consistent with data. This article defines an ES as a standardized linear contrast of means. The noncentral method, Bonett's method, and the bias-corrected and accelerated bootstrap method are illustrated for constructing the confidence interval for such an effect size. Results obtained from the three methods are discussed and interpretations of results are offered.

*Keywords:* Confidence interval, linear contrast, effect size, bootstrap, noncentral

---

## Introduction

The importance of reporting effect sizes (ESs) and confidence intervals (CIs) has been strongly emphasized in the debate over null hypothesis significance testing as a methodology in social science research (Cohen, 1994; McCartney & Rosenthal, 2000; Nix & Barnette, 1998; Schmidt, 1996, although see Sawilowsky & Yoon, 2002, in this journal for a contrary view). Cumming (2012) characterized the shift from reliance on null hypothesis significance testing to the use of ESs, CIs, and meta-analyses as new statistics. Thompson (2002) stated, "An improved quantitative science would emphasize the use of confidence intervals (CIs), and especially CIs for effect sizes" (p.25), and constructing CIs for ESs facilitates meta-analytic thinking and interpretation. Thompson explained that reporting CIs allows future researchers to incorporate prior knowledge into the estimation of the same population ES. Furthermore, CI is directly related to the precision of ES estimates obtained from different studies. (See Knapp & Sawilowsky, 2001a, 2001b for a contrary view.)

Professional organizations such as the American Psychological Association (APA) and the American Educational Research Association (AERA) have both

---

*Dr. Chen is a recent PhD graduate. Email her at: [litchen@indiana.edu](mailto:litchen@indiana.edu). Dr. Peng is a professor of inquiry methodology and adjunct professor of statistics in the Department of Counseling and Educational Psychology. Email her at: [peng@indiana.edu](mailto:peng@indiana.edu).*

stressed the importance of reporting CIs for ESs, particularly since 1999. According to the APA Task Force Report, “Interval estimates should be given for any effect sizes involving principal outcomes” (Wilkinson and the Task Force on Statistical Inference, 1999, p. 599). The fifth and sixth editions of the APA *Publication Manual* stress that “The inclusion of confidence intervals (for estimates of parameters, for functions of parameters such as differences in means, and for effect sizes) can be an extremely effective way of reporting results” (APA, 6th edition, 2010, p.34). In addition, the sixth edition of the APA *Publication Manual* emphasizes, “Whenever possible, provide confidence interval for each effect size reported to indicate the precision of estimation of the effect size” (APA, 6th edition, 2010, p.34). Likewise, the AERA’s Standards for Reporting on Empirical Social Science Research suggest that, “For each of the statistical results that is crucial to the logic of the design and analysis, there should be included: ... An indication of the uncertainty of that index of effect size ...” (AERA, 2006, p. 37). According to the sixth edition of the APA *Publication Manual*, it is crucial to report confidence intervals because “confidence intervals combine information on location and precision and can often be directly used to infer significance levels, they are, in general, the best reporting strategy” (p. 34). For ways to report CIs, the same APA manual states, “As a rule, it is best to use a single confidence level, specified on an a priori basis (e.g., a 95% or 99% confidence interval), throughout the manuscript. Wherever possible, base discussion and interpretation of results on point and interval estimates” (p. 34).

Despite these efforts, the reporting rate of CIs for ESs in empirical studies is still low (Odgaard & Fowler, 2010; Peng, Chen, Chiang, & Chiang, 2013). This phenomenon may be due to a lack of understanding of the statistical properties of CIs for ESs, or a lack of suitable algorithms for the construction of CIs implemented in commercial statistic software (e.g., SPSS, SAS). Thus, this article aims to present three methods and algorithms for constructing the CI for a standardized linear contrast of means in a one-way fixed-effects univariate ANOVA design. This article defines a standardized linear contrast of means as a measure of ES for fixed-effects ANOVA designs. And the three methods are: the noncentral method, Bonett’s method, and the bias-corrected and accelerated bootstrap method.

To facilitate the understanding of standardized linear contrasts of means and to illustrate the three methods, a sleep deprivation example from Kirk (1995) is used. This example serves as a template for discussing the construction of CI for a standardized linear contrast of means using the three methods.

## CONSTRUCTING CONFIDENCE INTERVALS IN ANOVA DESIGNS

### A sleep deprivation example

This example examines the effects of sleep deprivation on hand-steadiness. According to Kirk (1995): Assume an interest in the effects of sleep deprivation, treatment  $A$ , on hand-steadiness. The four levels of sleep deprivation of interest are 12, 18, 24, and 30 hours, which are denoted by  $a_1$ ,  $a_2$ ,  $a_3$ , and  $a_4$ , respectively. An experiment is conducted in which 32 subjects are randomly assigned to the four levels of sleep deprivation, with the restriction that eight subjects are assigned to each level. The dependent variable is the number of times that a stylus makes contact with the side of a 1/2-inch hole during a 2-minute interval (p. 166). The independent variable is hours of sleep deprivation and the dependent variable is the number of times that a stylus held by a participant makes contact with the side of a 1/2-inch hole. The higher the number, the worse the performance, presumably affected by the deprivation of sleep. Data gathered from this study are shown in Table 1.

**Table 1.** The number of times that a stylus held by a participant makes contact with a 1/2-inch hole during a 2-minute interval from the sleep deprivation sample.

Hours of Sleep Deprivation Treatment Level	12 hours $a_1$	18 hours $a_2$	24 hours $a_3$	30 hours $a_4$
	4	4	5	3
	6	5	6	5
	3	4	5	6
	3	3	4	5
	1	2	3	6
	3	3	4	7
	2	4	3	8
	2	3	4	10
<b>Group Sizes (<math>n_j</math>)</b>	8	8	8	8
<b>Group Means (<math>\bar{Y}_j</math>)</b>	3	3.5	4.25	6.25
<b>Standard deviation (<math>\hat{\sigma}_j</math>)</b>	1.51	0.93	1.04	2.12

Consider the hypothesis that a human's fine motor skill decreases dramatically after being deprived of sleep for 24 hours or longer. Thus, interest lies in the contrast between the average performance of participants after 24 and

30 hours of sleep deprivation versus the average performance of participants after 12 and 18 hours. The linear contrast of means ( $\psi$ ) is written as

$$\psi = 0.5 \times (\mu_{24 \text{ hours}} + \mu_{30 \text{ hours}}) - 0.5 \times (\mu_{12 \text{ hours}} + \mu_{18 \text{ hours}}) = \sum_{j=1}^k c_j \mu_j, \quad (1)$$

where  $\mu_j$  is the population mean for the  $j$ th group,  $k$  is the number of independent groups (= 4 for the sleep deprivation example), and  $c_j$  is the coefficient or weight assigned to the  $j$ th group (= 0.5, 0.5, -0.5, and -0.5 for 24 hours, 30 hours, 12 hours, and 18 hours of sleep deprivation, respectively). Equation 1 and all subsequent equations are written specifically to suit the sleep deprivation example first, followed by a general formulation (in blue).

The value obtained from Equation 1 based on sampled data is an estimate of the corresponding population ES in original units. If a researcher wishes to standardize this ES, he/she needs to divide  $\psi$  with a standardizer. Such a standardizer is usually the population standard deviation, assumed to be equal and expressed as  $\sigma$ . For the specific  $\psi$  defined in Equation 1, its standardized form ( $\delta$ ) is written as follows:

$$\delta = \frac{\psi}{\sigma} = \frac{0.5 \times (\mu_{24 \text{ hours}} + \mu_{30 \text{ hours}}) - 0.5 \times (\mu_{12 \text{ hours}} + \mu_{18 \text{ hours}})}{\sigma} = \frac{\sum_{j=1}^k c_j \mu_j}{\sigma}. \quad (2)$$

Reporting a standardized linear contrast of means is more informative than reporting a linear contrast of means in original units, when (1) the original unit of the dependent variable is not familiar to readers, or (2) a researcher intends to compare ESs obtained from studies that employ different dependent variables.

The following three sections introduce three methods for constructing CIs for standardized linear contrasts of means as ESs. The three methods are the noncentral method, Bonett's method, and the BCa (or the bias-corrected and accelerated bootstrap) method. After obtaining CIs results are compared and proper interpretations of CIs in this context are discussed.

## Methods

### Noncentral Method

Within the null hypothesis significance testing framework, a linear contrast  $\psi$  defined in Equation 1 is tested with a  $t$ -statistic defined as:

## CONSTRUCTING CONFIDENCE INTERVALS IN ANOVA DESIGNS

$$t = \frac{0.5 \times (\bar{Y}_{24 \text{ hours}} + \bar{Y}_{30 \text{ hours}}) - 0.5 \times (\bar{Y}_{12 \text{ hours}} + \bar{Y}_{18 \text{ hours}})}{\hat{\sigma} \times \sqrt{\frac{(0.5)^2 + (0.5)^2 + (-0.5)^2 + (-0.5)^2}{8}}} = \frac{\sum_{j=1}^k c_j \bar{Y}_j}{\hat{\sigma} \times \sqrt{\sum_{j=1}^k \frac{c_j^2}{n_j}}}, \quad (3)$$

where  $\bar{Y}_j$  is the sample mean for the  $j$ th group (= 4.25, 6.25, 3.00, and 3.50 for 24, 30, 12, and 18 hours of sleep deprivation, respectively),  $\hat{\sigma}$  is the pooled standard deviation that is used to estimate the equal population standard deviation ( $= \sqrt{(1.51^2 + 0.93^2 + 1.04^2 + 2.12^2) / 4} = 1.48$ ), and  $n_j$  is the sample size for the  $j$ th group ( $= 8$  for each of the four groups in sleep deprivation example).

Under the null hypothesis of a 0 linear contrast of means, the  $t$  statistic is distributed as a symmetric central  $t$  distribution with a mean of 0. When the null hypothesis is false (meaning the population linear contrast of means does not equal 0), the  $t$  statistic follows a noncentral  $t$  distribution that is centered approximately at the noncentrality parameter  $\lambda$ , when the degree of freedom is large (see Cumming & Finch, 2001). The noncentral  $t$  distribution has two parameters: the degrees of freedom (or  $df$  = the number of participants – the number of independent groups) and  $\lambda$ . When  $\lambda$  is zero, the noncentral  $t$  distribution is the central  $t$  distribution, or simply the  $t$  distribution.

One way to construct the CI for  $\delta$  defined in Equation 2, is to use the noncentral  $t$  distribution. The noncentrality parameter  $\lambda$  of the noncentral  $t$  distribution is related to  $\delta$  as follows,

$$\delta = \lambda \times \sqrt{\frac{0.5^2}{8} + \frac{0.5^2}{8} + \frac{(-0.5)^2}{8} + \frac{(-0.5)^2}{8}} = \lambda \times \sqrt{\sum_{j=1}^k \frac{c_j^2}{n_j}}. \quad (4)$$

And  $\lambda$  is defined as follows,

$$\lambda = \frac{0.5 \times (\mu_{24 \text{ hours}} + \mu_{30 \text{ hours}}) - 0.5 \times (\mu_{12 \text{ hours}} + \mu_{18 \text{ hours}})}{\sigma \times \sqrt{\frac{0.5^2}{8} + \frac{0.5^2}{8} + \frac{(-0.5)^2}{8} + \frac{(-0.5)^2}{8}}} = \frac{\delta}{\sqrt{\frac{0.5^2}{8} + \frac{0.5^2}{8} + \frac{(-0.5)^2}{8} + \frac{(-0.5)^2}{8}}} = \frac{\sum_{j=1}^k c_j \mu_j}{\sigma \times \sqrt{\sum_{j=1}^k \frac{c_j^2}{n_j}}}. \quad (5)$$

Steiger and Fouladi (1997) illustrated how to derive  $\lambda$  from the observed  $t$  statistic obtained from a sample. From  $\lambda$ , using Equation 4,  $\delta$  can be derived. To construct a 95% confidence interval for  $\delta$ , first compute the lower and the upper limits of  $\lambda$  from the observed  $t$  statistic. The lower limit for  $\lambda$  is the noncentrality parameter of the noncentral  $t$  distribution in which the observed  $t$  statistic is at the 97.5<sup>th</sup> percentile. The upper limit for  $\lambda$  is the noncentrality parameter of the noncentral  $t$  distribution in which the observed  $t$  statistic is at the 2.5<sup>th</sup> percentile. From the two limits of  $\lambda$ , the limits for  $\delta$  can be derived.

The use of noncentral distributions in constructing the CI for ESs involves sequence of iterations. In recent years, the computational difficulty for the noncentral  $t$  distribution has been overcome by algorithmic improvement. For example, the lower and upper limits of  $\lambda$  can be obtained in SAS® with the following syntax:

```
lamda_lower=TNONCT (t_observed, df, .975);
```

and

```
lamda_upper=TNONCT (t_observed, df, .025);
```

The  $df$  for the current example is  $32 - 4 = 28$ . Once the lower limit and the upper limit of  $\lambda$  are obtained from SAS®, the lower limit and the upper limit of  $\delta$  can be computed from the following according to Equation 4:

$$\delta_{\text{lower}} = \lambda_{\text{lower}} \times \sqrt{\frac{0.5^2}{8} + \frac{0.5^2}{8} + \frac{(-0.5)^2}{8} + \frac{(-0.5)^2}{8}} = \lambda_{\text{lower}} \times \sqrt{\sum_{j=1}^k \frac{c_j^2}{n_j}}, \text{ and} \quad (6)$$

$$\delta_{\text{upper}} = \lambda_{\text{upper}} \times \sqrt{\frac{0.5^2}{8} + \frac{0.5^2}{8} + \frac{(-0.5)^2}{8} + \frac{(-0.5)^2}{8}} = \lambda_{\text{upper}} \times \sqrt{\sum_{j=1}^k \frac{c_j^2}{n_j}}. \quad (7)$$

Applying the noncentral method for constructing the CI for a standardized linear contrast of means is discussed in the literature (Cumming & Finch, 2001; Kline, 2004; Steiger, 2004). Liu (2010) illustrated the geometric meaning of the noncentrality parameter for a linear contrast in a Euclidian space. Kelley and Rausch (2006) and Lai and Kelley (2012) considered the sample size required to achieve the desired accuracy in CI estimations. Readers should note that there are



## CONSTRUCTING CONFIDENCE INTERVALS IN ANOVA DESIGNS

two statistical assumptions associated with the noncentral confidence intervals for  $\delta$ . These two assumptions are (1) normality for each population distribution and (2) equal variances for all population distributions.

The SAS® macro “cinoncentral” (See [Appendix A](#)) yields the noncentral CI for a standardized linear contrast of means ( $\delta$ ). To execute this SAS® macro, readers first create a SAS data set in the DATA step of SAS®, or import the data into SAS®. This step is followed by the specification of a level of confidence, such as .95, and a coefficient for each group.

### Bonett’s Method

Bonett (2008) proposed a more general definition of the standardized linear contrast of means in order to deal with unequal variances across populations:

$$\delta_{\text{Bonett}} = \frac{0.5 \times (\mu_{24 \text{ hours}} + \mu_{30 \text{ hours}}) - 0.5 \times (\mu_{12 \text{ hours}} + \mu_{18 \text{ hours}})}{\sqrt{\frac{\sigma_{24 \text{ hours}}^2 + \sigma_{30 \text{ hours}}^2 + \sigma_{12 \text{ hours}}^2 + \sigma_{18 \text{ hours}}^2}{4}}} = \frac{\sum_{j=1}^k c_j \mu_j}{\sqrt{\frac{\sum_{j=1}^k \sigma_j^2}{k}}} = \frac{\psi}{\sqrt{\frac{\sum_{j=1}^k \sigma_j^2}{k}}}. \quad (8)$$

It is worth noting that, when population variances are equal (i.e.,  $\sigma_{24 \text{ hours}}^2 = \sigma_{30 \text{ hours}}^2 = \sigma_{12 \text{ hours}}^2 = \sigma_{18 \text{ hours}}^2 = \sigma^2$ ),

$$\delta_{\text{Bonett}} = \frac{0.5 \times (\mu_{24 \text{ hours}} + \mu_{30 \text{ hours}}) - 0.5 \times (\mu_{12 \text{ hours}} + \mu_{18 \text{ hours}})}{\sqrt{\frac{4\sigma^2}{4}}} = \frac{0.5 \times (\mu_{24 \text{ hours}} + \mu_{30 \text{ hours}}) - 0.5 \times (\mu_{12 \text{ hours}} + \mu_{18 \text{ hours}})}{\sigma} = \frac{\sum_{j=1}^k c_j \mu_j}{\sigma} = \delta. \quad (9)$$

In other words,  $\delta$  is a special case of  $\delta_{\text{Bonett}}$  when population variances are all equal. Based on a large sample approximation, Bonett derived the CI for  $\delta_{\text{Bonett}}$  as follows:

$$\hat{\sigma}_{\text{Bonett}} \pm z_{\text{critical}} \left[ \text{var}(\hat{\delta}_{\text{Bonett}}) \right]^{1/2}, \quad (10)$$

where  $z_{\text{critical}}$  is the critical value from the standard normal distribution,  $\hat{\delta}_{\text{Bonett}}$  is the sample estimate for the corresponding population  $\delta_{\text{Bonett}}$ , and  $\text{var}(\hat{\delta}_{\text{Bonett}})$  is

the sample variance of  $\hat{\delta}_{\text{Bonnett}}$ . The sample estimate for Bonett's standardized linear contrast of means, i.e.,  $\hat{\delta}_{\text{Bonnett}}$ , is obtained from the following equation:

$$\hat{\delta}_{\text{Bonnett}} = \frac{0.5 \times (\bar{Y}_{24 \text{ hours}} + \bar{Y}_{30 \text{ hours}}) - 0.5 \times (\bar{Y}_{12 \text{ hours}} + \bar{Y}_{18 \text{ hours}})}{\sqrt{\frac{\hat{\sigma}_{24 \text{ hours}}^2 + \hat{\sigma}_{30 \text{ hours}}^2 + \hat{\sigma}_{12 \text{ hours}}^2 + \hat{\sigma}_{18 \text{ hours}}^2}{4}}} = \frac{\sum_{j=1}^k c_j \bar{Y}_j}{\sqrt{\frac{\sum_{j=1}^k \hat{\sigma}_j^2}{k}}} = \frac{\sum_{j=1}^k c_j \bar{Y}_j}{\hat{\sigma}_{\text{Bonnett}}}. \quad (11)$$

It is worth noting that when sample sizes are all equal,  $\hat{\sigma}_{\text{Bonnett}} = \hat{\sigma}$ . And  $\text{var}(\hat{\delta}_{\text{Bonnett}})$  is obtained from the following equation:

$$\begin{aligned} \text{var}(\hat{\delta}_{\text{Bonnett}}) &= \left( \frac{\hat{\delta}_{\text{Bonnett}}^2}{4^2 \hat{\sigma}_{\text{Bonnett}}^4} \right) \times \left[ \frac{\hat{\sigma}_{24 \text{ hours}}^4}{2 \times (8-1)} + \frac{\hat{\sigma}_{30 \text{ hours}}^4}{2 \times (8-1)} + \frac{\hat{\sigma}_{12 \text{ hours}}^4}{2 \times (8-1)} + \frac{\hat{\sigma}_{18 \text{ hours}}^4}{2 \times (8-1)} \right] \\ &+ \left[ \frac{(0.5)^2 \times \hat{\sigma}_{24 \text{ hours}}^2}{(8-1)} + \frac{(0.5)^2 \times \hat{\sigma}_{30 \text{ hours}}^2}{(8-1)} + \frac{(-0.5)^2 \times \hat{\sigma}_{12 \text{ hours}}^2}{(8-1)} + \frac{(-0.5)^2 \times \hat{\sigma}_{18 \text{ hours}}^2}{(8-1)} \right] \frac{\hat{\sigma}_{\text{Bonnett}}^2}{\hat{\sigma}_{\text{Bonnett}}^4} \quad (12) \\ &= \left( \frac{\hat{\delta}_{\text{Bonnett}}^2}{k^2 \hat{\sigma}_{\text{Bonnett}}^4} \right) \sum_{j=1}^k \frac{\hat{\sigma}_j^4}{2 \times df_j} + \frac{\sum_{j=1}^k \frac{c_j^2 \hat{\sigma}_j^2}{df_j}}{\hat{\sigma}_{\text{Bonnett}}^2}, \end{aligned}$$

where  $df_j$  = the number of participants in the  $j$ th group minus 1.

Bonett's method assumes normality, but not equal variances for the population distributions. When population variances are equal,  $\delta$  becomes a special case of  $\delta_{\text{Bonnett}}$ . The SAS® macro "cibonett" (See Appendix B) yields Bonett's CI for a standardized linear contrast of means ( $\delta_{\text{Bonnett}}$ ). To execute this SAS® macro, readers first create a SAS data set in the DATA step of SAS®, or import the data into SAS®. This step is followed by the specification of a level of confidence, such as .95, and a coefficient for each group.

### The BCa Bootstrap Method

The bootstrap method is a resampling technique that constructs an empirical distribution of estimates from data already collected. Thus, the bootstrap method does not require assumptions of either normality or equal variances. Nor does it rely on a theoretical sampling distribution, such as  $t$  or normal, to derive the lower or the upper limits of a confidence interval.

Several methods of constructing CIs based on bootstrapping have been developed. These include the symmetric percentile bootstrap method, the bias-corrected and accelerated (BCa) bootstrap method, and the approximate bootstrap confidence (ABC) interval method. The BCa bootstrap method introduced here corrects the bias in the symmetric percentile bootstrap method. To provide the general idea of bootstrapping technique, the symmetric percentile bootstrap method is presented first, followed by the BCa bootstrap method.

The symmetric percentile bootstrap method constructs the CI by finding the  $\alpha/2 \times B$ th and  $[1 - (\alpha/2)] \times B$ th ranked values of the empirical distribution of the sample estimates. Here,  $\alpha$  is the Type I error rate, such as .05;  $B$  is the number of bootstrap samples, such as 1,000. A bootstrap sample is a random sample of size  $n$ , drawn with replacement from the observed  $n$  scores. After a large number of bootstrap samples (e.g., 1,000) are formed, an empirical bootstrap distribution of the estimated effect sizes is constructed. From the empirical bootstrap distribution, the lower and upper confidence limits are derived. If  $\alpha = 0.05$  and  $B = 1,000$ , the lower limit of a 95% bootstrap confidence interval is the  $0.05/2 \times 1,000^{\text{th}}$  ranked value of the empirical bootstrap distribution and the upper limit is the  $[1 - (0.05/2)] \times 1,000^{\text{th}}$  ranked value.

Readers can apply the bootstrap technique to construct the CI for either  $\delta$  (Equation 2) or  $\delta_{\text{Bonett}}$  (Equation 8). A step-by-step instruction for obtaining the CI for  $\delta_{\text{Bonett}}$  using the symmetric percentile bootstrap method is presented. Readers can use this instruction to construct the CI for  $\delta$  as well. Assuming that a 95% CI is to be constructed for  $\delta_{\text{Bonett}}$ , based on the sample estimator  $\hat{\delta}_{\text{Bonett}}$ . Three steps are completed  $B$  times (e.g., 1,000) in order to yield  $B$  bootstrap estimates,  $\hat{\delta}_{\text{Bonett}}^*$ . The first step is to randomly sample eight scores with replacement from each of the four groups, where eight is the number of scores in each group. The second step is to compute the mean and the standard deviation of these eight scores from each group. The third step is to compute  $\hat{\delta}_{\text{Bonett}}^*$  by plugging the means and the standard deviations obtained from the second step into Equation 11. After obtaining the  $B$  (e.g., 1,000) bootstrap  $\hat{\delta}_{\text{Bonett}}^*$ , one lists them in an ascending order.

The  $.025 \times B$ th (=25th) and the  $.975 \times B$ th (=975th) ranked  $\hat{\delta}_{\text{Bonett}}^*$  are, respectively, the 95% lower and upper confidence limits. When estimates are tied, an average rank is assigned to the tied estimates.

The BCa bootstrap method is an improvement over the symmetric percentile bootstrap method. Specifically, it constructs the CI for the standardized linear contrast of means ( $\delta$  or  $\delta_{\text{Bonett}}$ ) using  $\text{CI}_{\text{lower}} \times B$ th and  $\text{CI}_{\text{upper}} \times B$ th ranked values of bootstrapped estimates. The values of  $\text{CI}_{\text{lower}}$  and  $\text{CI}_{\text{upper}}$  depend on acceleration and bias-correction numbers,  $\hat{a}$  and  $\hat{z}_0$ , respectively. According to Efron and Tibshirani (1993),  $\hat{a}$  refers to the rate of change in the standard error of the estimated parameter (i.e.,  $\hat{\delta}_{\text{Bonett}}$ ) with respect to the true population value (i.e.,  $\delta_{\text{Bonett}}$ ). The bias-correction number,  $\hat{z}_0$ , is interpreted as the median bias of the sample bootstrapped estimates. When exactly 50% of bootstrapped estimates are less than or equal to the observed estimate,  $\hat{z}_0 = 0$ . Using the notations described above for the symmetric percentile bootstrap method,

$$\hat{z}_0 = \Phi^{-1} \left[ \frac{\#(\hat{\delta}_{\text{Bonett}}^* < \hat{\delta}_{\text{Bonett}})}{B(=1,000)} \right], \quad (13)$$

where  $\Phi^{-1}$  is the inverse of the standard normal cumulative function and  $\#(\hat{\delta}_{\text{Bonett}}^* < \hat{\delta}_{\text{Bonett}})$  is the frequency of those bootstrap estimates (i.e.,  $\hat{\delta}_{\text{Bonett}}^*$ ) that are less than the observed estimate (i.e.,  $\hat{\delta}_{\text{Bonett}}$ ). The acceleration  $\hat{a}$  is obtained using the jackknife method that takes the form:

$$\hat{a} = \frac{\sum_{i=1}^{32} \left( \bar{\hat{\delta}}_{\text{Bonett}} - \hat{\delta}_{\text{Bonett}-i} \right)^3}{6 \left\{ \sum_{i=1}^{32} \left( \bar{\hat{\delta}}_{\text{Bonett}} - \hat{\delta}_{\text{Bonett}-i} \right)^2 \right\}^{\frac{3}{2}}}, \quad (14)$$

where  $\hat{\delta}_{\text{Bonett}-i}$  is the value of  $\hat{\delta}_{\text{Bonett}}$  with the  $i$ th score removed from the entire data ( $i = 1, \dots, 32$  in the sleep deprivation example), and  $\bar{\hat{\delta}}_{\text{Bonett}}$  is the average of all possible  $\hat{\delta}_{\text{Bonett}-i}$ . The  $\text{CI}_{\text{lower}}$  and  $\text{CI}_{\text{upper}}$  for a 95% confidence interval are given by

## CONSTRUCTING CONFIDENCE INTERVALS IN ANOVA DESIGNS

$$CI_{\text{lower}} = \Phi \left[ \hat{z}_0 + \frac{\hat{z}_0 + z^{\left(\frac{\alpha}{2}=.025\right)}}{1 - \hat{a} \left( \hat{z}_0 + z^{\left(\frac{\alpha}{2}=.025\right)} \right)} \right], \text{ and} \quad (15)$$

$$CI_{\text{upper}} = \Phi \left[ \hat{z}_0 + \frac{\hat{z}_0 + z^{\left(1-\frac{\alpha}{2}=.975\right)}}{1 - \hat{a} \left( \hat{z}_0 + z^{\left(1-\frac{\alpha}{2}=.975\right)} \right)} \right]. \quad (16)$$

Here,  $z^{\left(\frac{\alpha}{2}\right)}$  is the  $(100 \times \frac{\alpha}{2})$ th percentile of a standard normal distribution.

For a 95% confidence interval,  $z^{\left(\frac{\alpha}{2}=.025\right)} = -1.96$  and  $z^{\left(1-\frac{\alpha}{2}=.975\right)} = 1.96$ . The BCa confidence intervals yield the same results as the symmetric percentile bootstrap confidence intervals, when  $\hat{z}_0$  and  $\hat{a}$  both equal 0. In other words,  $CI_{\text{lower}} = \Phi(-1.96) = .025$  and  $CI_{\text{upper}} = \Phi(1.96) = .975$ , when  $\hat{z}_0$  and  $\hat{a}$  both equal 0. The BCa method is superior to the symmetric percentile bootstrap method because it leads to better approximations to the lower and upper limits. However, Efron and Tibshirani (1993) stated, “their [the BCa] coverage accuracy can still be erratic for small sample sizes” (p.178). Chen’s dissertation (2013) uncovered that the coverage probability produced by the BCa method was satisfactory when each group size was 30. Kelley’s (2005) simulation found that BCa method’s coverage probability was poor when each group size was eight.

The process for constructing the BCa CI for  $\delta_{\text{Bonett}}$  may appear complex to some readers. However, a SAS® macro “cibca” (See [Appendix C](#)) based on the SAS program written by Barker (2005) is provided here to assist researchers in constructing BCa CIs for  $\delta_{\text{Bonett}}$ , of which  $\delta$  is a special case. To execute this SAS® macro, readers first create a SAS data set in the DATA step of SAS®, or import the data into SAS®. This step is followed by the specification of a number of bootstrap estimates (e.g., 1,000), a coefficient for each group, and a level of confidence, such as .95.

## Results

Given the standardized linear contrast of means ( $\delta$  in Equation 2) from the sleep deprivation example, the 95% noncentral CI was computed to be [0.56758, 2.12318]. Likewise, given the standardized linear contrast of means ( $\delta_{\text{Bonett}}$  in Equation 8), the 95% Bonett's CI was computed to be [0.50045, 2.20958] and the BCa bootstrap CI to be [0.51967, 2.13941]. Thus, the CIs constructed by the three methods are slightly different from each other. The noncentral CI is the narrowest ( $= 2.12318 - 0.56758 = 1.55560$ ) or most precise, followed by the BCa bootstrap CI ( $= 2.13941 - 0.51967 = 1.61973$ ), and the Bonett's CI ( $= 2.20958 - 0.50045 = 1.70913$ ). These results are consistent with findings obtained by Chen (2013) in a thorough investigation of these three methods under a variety of conditions.

In actuality, it is not necessary to compute more than one CI for a standardized ES. It is however necessary for researchers to be informed of the optimal method for a particular research context. For the purpose of demonstration, the correct interpretation of the noncentral CI for the contrast of interest (i.e.,  $\delta$  in Equation 2) is described.

### How to interpret confidence intervals for a standardized linear contrast of means?

The 95% noncentral CI ranges from 0.56758 (or 0.57) to 2.12318 (or 2.12). Derived from the data presented in Table 1, all values contained in this interval cannot be rejected with a Type I error rate of 5%, if they are placed in a null hypothesis. Furthermore, all values in this interval are greater than 0. Thus, a null hypothesis of 0 standardized mean difference should be rejected at an  $\alpha$  level of .05. Based on the data and the noncentral CI, readers can conclude that the difference in the number of times that a stylus touched the sides of a  $\frac{1}{2}$  hole between people deprived of sleep for 24 hours or longer and people deprived of sleep less than 24 hours can be as large as twice of the standard deviation of the data, or as small as a half of the standard deviation.

## Discussion

A measure of an ES gives a point estimate of a treatment effect, whereas a CI of such an ES provides the precision of the estimation. Although both the APA and the AERA have encouraged researchers to report CIs for ESs, Odgaard and Fowler's (2010) study found that the reporting rate of CIs for ESs was only 40% in the *Journal of Consulting and Clinical Psychology*—the first APA journal that

## CONSTRUCTING CONFIDENCE INTERVALS IN ANOVA DESIGNS

required the reporting of CIs for ESs for primary results. A similar finding is reported by Peng et al. (2013) across a variety of psychology and education journals. The reporting of CI can be encouraged through accessible and reliable computing algorithms.

This article (1) illustrates the need to report CIs for ESs, (2) addresses the importance of reporting the CIs for ESs, (3) introduces, demonstrates, and compares three methods (the noncentral method, Bonett's method, and the BCa bootstrap method) for constructing the CI for a standardized linear contrast of means (a measure of the ES), and (4) provides SAS programming codes for these methods. The readers should note that the SAS programming codes provided in Appendices A – C are applicable for unequal sample sizes as well. It is hoped that this paper facilitates researchers' understanding of these three methods and enables them to report the CIs for ESs, defined as standardized linear contrasts of means in fixed-effects ANOVA designs.

### Acknowledgements

This research was supported in part by the Maris M. Proffitt and Mary Higgins Proffitt Endowment Grant of Indiana University, awarded to the second author while the first author worked on the project as a research assistant.

### References

- American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, 35(6), 33-40. doi: 10.3102/0013189X035006033
- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: American Psychological Association.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.
- Barker, N. (2005). A practical introduction to the bootstrap using the SAS system. *Proceedings of SAS conference: Phuse*. Retrieved from <http://www.lexjansen.com/phuse/2005/pk/pk02.pdf>
- Bonett, D.G. (2008). Confidence intervals for standardized linear contrasts of means. *Psychological Methods*, 13, 99-109. doi: 10.1037/1082-989X.13.2.99

Chen, L.-T. (2013). *Effect size measures and their interval estimations: The multi-independent group case*. (Doctoral dissertation). Indiana University, Bloomington.

Cohen, J. (1994). The earth is round ( $p < .05$ ). *American Psychologist*, *49*, 997-1003. doi: 10.1037/0003-066X.49.12.997

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge/Taylor & Francis Group.

Cumming, G., & Finch, S. (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*, 532-574. doi: 10.1177/0013164401614002

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Kelley, K., & Rausch, J. R. (2006). Sample size planning for the standardized mean difference: Accuracy in parameter estimation via narrow confidence intervals. *Psychological Methods*, *11*, 363-385. doi: 10.1037/1082-989X.11.4.363

Kelley, K. (2005). The effects of nonnormal distributions on confidence intervals around the standardized mean difference: Bootstrap and parametric confidence intervals. *Educational and Psychological Measurement*, *65*, 51-69. doi: 10.1177/0013164404264850

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Kline, R. B. (2004). *Beyond significance testing*. Washing, DC: American Psychological Association.

Knapp, T. R., & Sawilowsky, S. S. (2001a). Strong arguments: Rejoinder to Thompson. *Journal of Experimental Education*, *70*, 94-95.

Knapp, T. R., & Sawilowsky, S. S. (2001b). Constructive criticisms of methodological and editorial practices. *Journal of Experimental Education*, *70*, 65-79.

Lai, K., & Kelley, K. (2012). Accuracy in parameter estimation for ANCOVA and ANOVA contrasts: Sample size planning via narrow confidence intervals. *British Journal of Mathematical & Statistical Psychology*, *65*, 350-370. doi: 10.1111/j.2044-8317.2011.02029.x



## CONSTRUCTING CONFIDENCE INTERVALS IN ANOVA DESIGNS

Liu, X. S. (2010). A note on noncentrality parameters for contrast tests in a one-way analysis of variance. *The Journal of Experimental Education*, 78, 53-59. doi: 10.1080/00220970903224669

McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, 71, 173-180. doi: 10.1111/1467-8624.00131

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, 3-14.

Odgaard, E. C., & Fowler, R. L. (2010). Confidence intervals for effect sizes: Compliance and clinical significance in the *Journal of Consulting and Clinical Psychology*. *Journal of Consulting and Clinical Psychology*, 78, 287-297. doi: 10.1037/a0019294

Peng, C.-Y. J., Chen, L.-T., Chiang, H.-M., & Chiang, Y.-C. (2013). The impact of APA and AERA guidelines on effect size reporting. *Educational Psychology Review*, 25, 157-209. doi: 10.1007/s10648-013-9218-2

Sawilowsky, S. S., & Yoon, J. S. (2002). The trouble with trivials. *Journal of Modern Applied Statistical Methods*, 1(1), 143-144.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1, 115-129. doi: 10.1037/1082-989X.1.2.115

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, 9, 164-182. doi: 10.1037/1082-989X.9.2.164

Steiger, J. H., & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. Harlow, S. Mulaik & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221-257). Hillsdale, NJ: Erlbaum.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, 31(3), 25-32. doi: 10.3102/0013189X031003025

Wilkinson, L., and the APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals - Guidelines and explanations. *American Psychologist*, 54, 594-604. doi: 10.1037//0003-066X.54.8.594

## Appendix A

### SAS® Macro “cinoncentral”

```

*****;
*   format for data set: first variable = group, second variable = scores   ;
*   data = data set name                                                    ;
*   nominal = nominal confidence level                                       ;
*   contrast = a coefficient for each group, e.g., {.5 .5 -.5 -.5}         ;
*****;
%MACRO cinoncentral(data,nominal,contrast);

PROC IML;
USE &data;
READ ALL INTO datain;
nn=NROW(datain);
groups = UNIQUE(datain[,1]);          *get group information of the data;
ngroups = NCOL(groups);              *get number of groups of the data;
CALL SYMPUTX("n_groups",ngroups);    *set the macro variable n_groups to be the
                                     number of groups;
%DO i = 1 %TO &n_groups;              *loop* for groups;
group&i = datain[LOC(datain[,1]=&i),2]; /*obtain all the scores for each group*/
mu&i=mean(group&i);
sum&i = sum(group&i);
v_sum&i =(sum&i)**2;
n&i = nrow(group&i);
v_sum_n&i=v_sum&i/ n&i;
%END;
mu=mu1;
v_sum=v_sum1;
n=n1;
v_sum_n=v_sum1/n1;
%DO i=2 %TO &n_groups;
mu=mu//mu&i;
v_sum = v_sum//v_sum&i;
n = n//n&i;
v_sum_n=v_sum_n//v_sum_n&i;
contrast = t(&contrast);
%END;
df=n-1;
numerator=(contrast)`*mu;
mse1=(datain[,2])`*(datain[,2]); *squared values of all scores;
mse2=sum(v_sum_n);
mse=(mse1-mse2)/(nn-ngroups);
contrast_square=(contrast)##2;
n_1=1/n;
nu=(contrast_square)`*(n_1);
t=numerator/(SQRT(mse*nu));
lamda_lower = TNONCT(t,nn-ngroups,1-(1-&nominal)/2); /*compute the lower
                                                         noncentrality*/
lamda_upper = TNONCT(t,nn-ngroups,(1-&nominal)/2); /*compute the upper
                                                         noncentrality*/
coe = sqrt(nu);

```

## CONSTRUCTING CONFIDENCE INTERVALS IN ANOVA DESIGNS

```
NcT_lower=lamda_lower*coe;          /*Lower limit of noncentral ES*/
NcT_upper=lamda_upper*coe;          /*Upper limit of noncentral ES*/
Width= NcT_upper- NcT_lower;
TITLE "The confidence interval based on the noncentral method";
TITLE2 "Coefficient=&contrast Confidence level=&nominal";
PRINT NcT_lower NcT_upper Width;
QUIT;
%MEND;
data a;
input group y @@;
cards;
1 3 2 5 3 4 4 4
1 5 2 6 3 5 4 6
1 6 2 5 3 4 4 3
1 5 2 4 3 3 4 3
1 6 2 3 3 2 4 1
1 7 2 4 3 3 4 3
1 8 2 3 3 4 4 2
1 10 2 4 3 3 4 2
;
run;
%cinoncentral(a,.95,{.5 .5 -.5 -.5});
```

## Appendix B

### SAS® Macro “cibonett”

```
*****;
*   format for data set: first variable = group, second variable = scores   ;
*   data = data set name                                                    ;
*   nominal = nominal confidence level                                       ;
*   contrast = a coefficient for each group, e.g., {.5 .5 -.5 -.5}         ;
*****;

%macro cibonett(data,nominal,contrast);
TITLE "Bonett &nominal confidence interval";
proc IML;
use &data;
read all into datain;
groups = UNIQUE(datain[,1]);          *get group information of the data;
ngroups = NCOL(groups);              *get number of groups of the data;
CALL SYMPUTX("n_groups",ngroups);    *set the macro variable n_groups to be the
                                     number of groups;
%DO i = 1 %TO &n_groups;              *loop for groups;
group&i = datain[LOC(datain[,1]=&i),2]; /*obtain all the scores for each group*/
mu&i = mean(group&i);
var&i = var(group&i);
n&i = nrow(group&i);
%END;
mu=mu1;
var=var1;
n=n1;
```

```

%DO i=2 %TO &n_groups;
mu = mu//mu&i;
var = var//var&i;
n = n//n&i;
contrast = t(&contrast);
%END;
df=n-1;
delta_bonett=sum(mu#contrast)/sqrt(mean(var));
k = ngroups;
v1=(delta_bonett**2/(k**2*(mean(var))**2));
v2=sum((var##2)/(2*df));
v3=sum(((contrast##2)#var)/df)/mean(var);
var_delta_bonett=v1*v2+v3;
bonett_upper = delta_bonett + PROBIT(1-(1-&nominal)/2)*SQRT(var_delta_bonett);
bonett_lower = delta_bonett - PROBIT(1-(1-&nominal)/2)*SQRT(var_delta_bonett);
width=bonett_upper-bonett_lower;
PRINT delta_bonett bonett_lower bonett_upper width;
quit;
%mend;
data a;
input group y @@;
cards;
1 3 2 5 3 4 4 4
1 5 2 6 3 5 4 6
1 6 2 5 3 4 4 3
1 5 2 4 3 3 4 3
1 6 2 3 3 2 4 1
1 7 2 4 3 3 4 3
1 8 2 3 3 4 4 2
1 10 2 4 3 3 4 2
;
run;

%cibonett(a,.95,{.5 .5 -.5 -.5});

```

## Appendix C

### SAS® Macro “cibca”

```

*****;
*   format for data set: first variable = group, second variable = scores   ;
*   data = data set name                                                    ;
*   b = the number of bootstrap sample                                       ;
*   con = a coefficient for each group, e.g., {.5 .5 -.5 -.5}                ;
*   nominal = nominal confidence level                                       ;
*****;
%MACRO cibca(data=,b=,con=,nominal=);
/*****/
/*This section of IML do the bootstrap resampling with B replications and   */
/*save the samples into zboots&i data sets (&i = 1 to number of groups)     */
/*It also calculate the delta_bonett for the original data set              */
/*****/

```

## CONSTRUCTING CONFIDENCE INTERVALS IN ANOVA DESIGNS

```

PROC IML;
USE &data;
READ ALL INTO datain;
groups = UNIQUE(datain[,1]);          *get group information of the data;
ngroups = NCOL(groups);              *get number of groups of the data;
CALL SYMPUTX("n_groups",ngroups);    *set the macro variable n_groups to be
                                     the number of groups;
%DO i = 1 %TO &n_groups;              *loop for groups;
group&i = datain[LOC(datain[,1]=&i),2]; *obtain all the scores for each group;
mu&i = mean(group&i);
var&i = var(group&i);
group_n=NROW(group&i);
CALL SYMPUTX("n",group_n);
z&i = j(1,2,.);                       *obtain the b times bootstrap sample for each group;
%DO m = 1 %TO &b;                       *loop for the bootstrap samples;
  y = group&i; /*This part of code (bootstrap) is adapted from */
  z = j(&n, 1, 0) ; /*http://www.biostat.umn.edu/~john-c/5421/notes.016b*/
  ite = J(&n,1,&m); /*Identify the nth bootstrap sample */
  do j = 1 to &n ;
    yrandindex = 1 + int(&n * ranuni(-1)) ;
    z[j] = y[yrandindex] ;
  end ;
  z = ite||z;
  z&i=z&i//z ;
%END;
CREATE zboots&i FROM z&i; /*Put matrices to data sets the first column
                           is the index for boot sample*/;
APPEND FROM z&i;          *the second column is the boot sample data;
                           *Save the matrices from IML to SAS data set;
%END;
mu=mu1; /*do the delta_bonett for original data; */for jackknife*/
var=var1;
%DO i=2 %TO &n_groups;
mu = mu||mu&i;
var = var||var&i;
%END;
o_delta_bonett=j(&b,1,(mu#&con)[,+)/sqrt((var[,+]/&n_groups));
CREATE origbonett FROM o_delta_bonett ;
APPEND FROM o_delta_bonett; /* save the delta_bonett for original data
                             to origbonett data*/
QUIT;

/*****
/* Below loop calculate the means and variances by each
boot index for each groups */
*****/
%DO i = 1 %TO &n_groups;
proc means data = zboots&i noprint; /*Compute the group means and vars for each
bootstrp sample*/
class COL1;
var COL2;
output out=meanvar&i mean=mu var=sigmasq;
run;
data meanvar&i;set meanvar&i;          *delete unused information;

```

## CHEN & PENG

```

if _type_=1;
drop _type_ _freq_;
run;
%END;
/*****
/*Below IML calculate the delta_bonett for the bootstrap data set*/
*****/
PROC IML;
%DO i = 1 %TO &n_groups;
USE meanvar&i;
READ ALL INTO mv&i;
%END;
mu = mv1[,2];          *assign the means of first group to mu;
var = mv1[,3];        *assign the vars of first group to var;
%DO i = 2 %TO &n_groups; /*this loop add means and vars of other groups
                        to mu and var*/
mu = mu || mv&i[,2];
var = var || mv&i[,3];
%END;
con=repeat(&con,&b,1);  *make the contrast to a matrix for calculation;
delta_bonett=mv1[,1] || (mu#&con)[,+] / sqrt((var[,+]) / &n_groups);
                        /*calculate delta_bonett for each bootstrap sample
                        and then add the bootstrap index to the first
                        column for later use*/;

CREATE delta_bonett FROM delta_bonett ;
APPEND FROM delta_bonett;
QUIT;
/*COMPUTE BIAS*/
data bonett           /* data set containing bootstrap values */
bias (keep=bias);    /* data set containing bias correction value */
merge delta_bonett(rename=(COL1=sample COL2=delta_bonett))
origbonett(rename=(COL1=origbonett)) end=eof;
if delta_bonett lt origbonett then lessthan=1;    /*flag if bootstrap sample
                                                gives lower */

else lessthan=0;    /*value than original sample */
retain nless 0;    /*retain variable nless with starting value 0,
                    the second value of nless will be 0 add to the
                    first value of lessthan*/

if sample gt 0 then nless=nless+lessthan; /* count samples with flag lessthan */
if sample ne 0 then output bonett; /* output only bootstrap sample statistics */
if eof then do; /* for the last value calculate: */
propless=nless/sample; /* 1. proportion of values below original estimate */
bias=probit(propless); /* 2. inverse normal of that proportion */
output bias;          /* 3. output only that record to new data set */
end;
run;
/*JACKKNIFING ACCELERATION*/
data origjack;      /* create a new data set which contains observation */
set &data end=eof; /* numbers 1 to &nobs (no. obs in data set) */
obsnum=_n_;
if eof then call symput('nobs', put(obsnum, 2.)); /*assign the characterstring
                                                _n_ to macro variable nobs"*/

run;
%macro jackdata;    /* use macro for %do processing utility */

```

## CONSTRUCTING CONFIDENCE INTERVALS IN ANOVA DESIGNS

```

data jackdata;
set
%do i=1 %to &nobs;          /* do loop to create all samples */
origjack (in = in&i
where=(obsnum ne &i))      /* remove a different value each time */
%end;;
%do i=1 %to &nobs;
if in&i then repeat=&i;    /* add repeat number for each sample */
%end;
run;
%mend;
%jackdata;
proc means data=jackdata noprint nway; /*Group means for each JACKKNIFE sample*/
class repeat group;
var Y;
output out=jackmeanvar mean=mu var=sigmasq;
run;
data jackmeanvar;set jackmeanvar;          /*delete unused information*/
if _type_=3;
drop _type_ _freq_;
run;
proc transpose data=jackmeanvar out=jackmeanvar let; /*data restructure*/
by repeat;
id group;
run;
data jackmean ; set jackmeanvar;
if _NAME_="mu";
drop repeat _NAME_;
run;
data jackvar ; set jackmeanvar; /*Group variance for each JACKKNIF sample*/
if _NAME_="sigmasq";
drop repeat _NAME_;
run;
/*****
/*Below IML calculate the delta_bonett for JACKKNIFING ACCELERATION*/
*****/
PROC IML;
USE jackmean;
READ ALL INTO jackmean;
USE jackvar;
READ ALL INTO jackvar;
n_jack=NROW(jackmean);
CALL SYMPUTX("n_j",n_jack);
con=repeat(&con,&n_j,1);
delta_bonett=(jackmean#&con)[,+]/sqrt((jackvar[,+])/&n_groups);
CREATE jackbonett FROM delta_bonett ;
APPEND FROM delta_bonett;
QUIT;
DATA jackbonett (rename=(COL1 = delta_bonett));
SET jackbonett;
PROC SQL NOPRINT;
select mean(delta_bonett)      /* put mean of jackknifed values
                               into macro variable */
into :mean_delta_bonett

```

```

from jackbonett;
quit;
data meanbonett;
set jackbonett;
cubed=(&mean_delta_bonett - delta_bonett)**3;      /* create cubed value of
difference */
squared=(&mean_delta_bonett - delta_bonett)**2;     /* create squared value of
difference */

run;
proc means data=meanbonett noprint;
output out=sumbonett
sum(cubed)=sumcube      /* find sum of cubed values */
sum(squared)=sumsuar;   /* find sum of squared values */
run;
data accel;
set sumbonett;
accel=sumcube / (6 * (sumsuar**1.5));      /* plug values into equation for */
keep accel;                               /* the acceleration statistic */
run;
data ciends;
merge accel
bias;
part1=(bias + probit((1-&nominal)/2)) / (1 - (accel*(bias + probit((1-
&nominal)/2))));
part2=(bias + probit(1-(1-&nominal)/2)) / (1 - (accel*(bias + probit(1-(1-
&nominal)/2))));
alpha1=probnorm(bias + part1);
alpha2=probnorm(bias + part2);
n1=alpha1*&b;
n2=alpha2*&b;
if n1 < 1 then n1 = 1;
call symput('n1', put(floor(n1), 5.)); /* Create macro variables with values */
call symput('n2', put(ceil(n2), 5.)); /* of N1 and N2 for later use */
run;
proc sort
data=bonett;
by delta_bonett;
run;
data ci_bca;
set bonett end=eof;
retain conf_lo conf_hi width;
if _n_=&n1 then conf_lo=delta_bonett; /* select values for upper and lower */
if _n_=&n2 then conf_hi=delta_bonett; /* limits using N1 and N2 values */
if eof then output;
width=conf_hi-conf_lo;
run;
proc print data=ci_bca;
title "The Confidence interval based on BCa bootstrap method";
title2 "B=&b Coefficeint=&con Confidence level=&nominal";
var conf_lo conf_hi width;
run;
%MEND;
Data BCa;
INPUT group y @@;

```



## CONSTRUCTING CONFIDENCE INTERVALS IN ANOVA DESIGNS

```
DATALINES;  
1 3 2 5 3 4 4 4  
1 5 2 6 3 5 4 6  
1 6 2 5 3 4 4 3  
1 5 2 4 3 3 4 3  
1 6 2 3 3 2 4 1  
1 7 2 4 3 3 4 3  
1 8 2 3 3 4 4 2  
1 10 2 4 3 3 4 2  
;  
%cibca(data=Bca,b=1000,con={.5 .5 -.5 -.5},nominal=.95); /*con is the  
coefficient  
for contrast*/
```