

SAS ® Code for Estimating the Confidence Interval for a Standardized Effect Size, d

S. Jonathan Mends-cole, Wayne State University, Detroit, MI

ABSTRACT

The inferential confidence interval provides an alternative to the significance test. The range of confidence limits indicates the veracity of the null hypothesis (Hoenig & Heisey, 2001). The width of the interval indicates the precision of estimating the population parameter (Hoenig & Heisey, 2001). Effect sizes aid decisions regarding the effectiveness of an experimental treatment (Kirk, 1996). The interval inversion principle brings together the explanatory capacity of the confidence Interval and the effect size (Steiger and Fouladi, 1997).

The observed test statistic is used to find confidence limits about the noncentrality parameter. The SAS (2001) function TNONCT provides the limits. The equation for the noncentrality parameter is modified to solve for the effect size limits.

INTRODUCTION

This document intends to show how the interval inversion principle connects the confidence Interval and effect size (Steiger and Fouladi, 1997). Secondly, it intends to show how the SAS (2001) function TNONCT can be used to solve for the confidence limits about a standardized effect size.

The first section outlines the interpretation of findings using effect sizes and confidence intervals. The second section outlines the theory and equations used to derive the limits of a standardized effect size. In the last section, data is simulated using SAS (2001) code to arrive at the confidence limits of a standardized effect size.

INTERVAL ESTIMATION

INTERPRETATION OF FINDINGS USING INFERENTIAL CONFIDENCE INTERVALS

Inferential confidence intervals are a viable alternative to significance tests. Inferential confidence intervals aid reaching data-driven conclusions about hypotheses or research questions. For example:

[1]...if a confidence interval does not cover a hypothesized parameter value [PV, i.e., $PV \text{Upper Limit (UL) or } PV \text{Lower Limit (LL)}$], then the value is refuted by the observed data...

[2] Conversely all values covered by the observed confidence interval could not be rejected [i.e., $UL \leq PV \leq LL$]; we refer to these as the set of nonrefuted values...

[3] If the nonrefuted states are clustered tightly about a specific null value [i.e., the interval width is small], one has confidence that nature is near the null value...

[4] If the nonrefuted states range widely from the null [i.e., the interval width is large], one must obviously be careful about interpreting the nonrejection as an indication of a near "null state". (Hoenig & Heisey, 2001, p. 22)

In addition, standardized effect sizes have been used when one is trying to reach data-driven conclusions about hypotheses or research questions.

INTERPRETATION OF FINDINGS USING EFFECT SIZES

For example, Kirk (1996) has suggested that:

Decisions regarding scientific hypotheses and practical usefulness are less straightforward when a measuring scale involves unfamiliar units. In such cases, it is necessary to (a) compute an effect magnitude and a confidence interval for that effect magnitude and (b) develop guidelines for deciding whether the effect magnitude is useful...With respect to determining practical significance of results, Cohen's definitions of small medium, and large effects represent a good beginning. (p. 756)

Similarly, Wilkinson and the APA Task Force on Statistical Inference (1999) recommended: "We must stress again that reporting and interpreting effect sizes in the context of previously reported effects is essential to good research. It enables readers to evaluate the stability of results across samples, designs, and analyses." (p. 599).

INTERVAL ESTIMATION ABOUT EFFECT SIZES

Confidence intervals and effect sizes were appropriately linked through the confidence interval inversion principle. Steiger and Fouladi (1997) defined the confidence interval inversion principle:

Let $f(\theta)$ be a monotonic, strictly increasing continuous function of θ . Let l_1 and l_2 be endpoints of a $1-\alpha$ confidence interval on quantity θ . Then $f(l_1)$ and $f(l_2)$ are endpoints of a $1-\alpha$ confidence interval on $f(\theta)$. (p. 234)

That is, if the values of $f(\theta)$ and θ are reliably increasing and not decreasing, (or reliably decreasing and not increasing) then confidence limits for θ can be used to derive confidence limits for $f(\theta)$. The importance of the confidence interval inversion principle is that it allows one to calculate exact confidence intervals for effect sizes, using the noncentrality parameter.

THE NONCENTRALITY PARAMETER & EFFECT SIZE

In order to compute the interval for an effect size, the observed test statistic (i.e., in this situation the independent samples t-test) is used to derive the interval for the noncentrality parameter.

$$t_{obs} = \frac{\overline{X}_t - \overline{X}_c}{S_{err}}$$

Where $\overline{X}$ represents the mean for the treatment (subscript t) and control (subscript c) groups;

S_{err} is the standard error of the estimate:

$$S_{err} = \sqrt{S_p \left(\frac{1}{n_t} + \frac{1}{n_c} \right)}$$

S_p is the pooled variance estimate:

$$S_p = \frac{(n_t - 1) S_t^2 + (n_c - 1) S_c^2}{(n_t - 1) + (n_c - 1)}$$

The expected value of the noncentral t-distribution for which t_{obs} intersects the $1-\alpha/2$ region is the lower limit for the observed noncentrality parameter. The expected value of the noncentral t-distribution for which t_{obs} intersects the $\alpha/2$ region is the upper limit for the observed noncentrality parameter (See figure 1.).

The SAS (2001) function TNONCT (see Table 1) enables one to compute the upper and lower limits of the noncentrality parameter. For example, if $t_{obs}=1.96$, $df=30$, $\alpha/2=.025$, the upper limit for the noncentrality parameter is

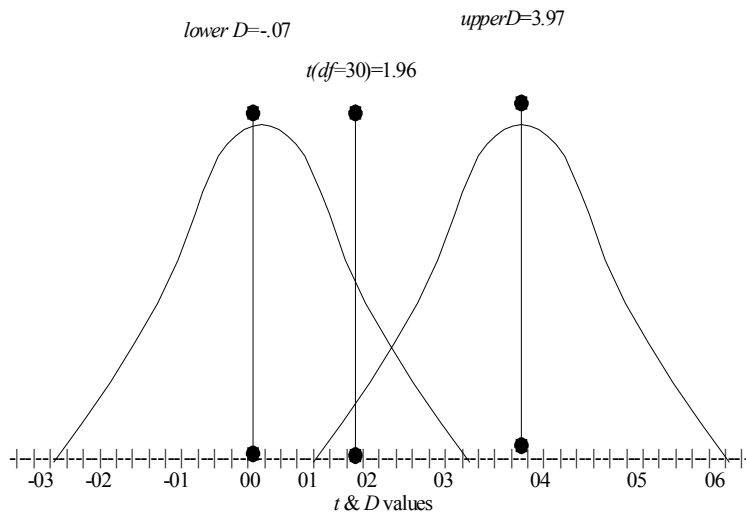


Figure 1. Hypothetical Depiction of the Noncentral t Distribution

$D=3.96$. The lower limit is $D= -.08$.

Table 1. Program to compute Upper & Lower Limits of the Noncentrality Parameter.

Code	Annotation
DATA NCP;	Creates a data file called NCP,
NCPLO=TNONCT(1.96,30,.975);	Computes the lower limit of the noncentrality parameter,
NCPHI=TNONCT(1.96,30,.025);	Computes the upper limit of the noncentrality parameter
OUTPUT;	Writes results to the file NCP,
RUN;	Executes preceding commands,

ILLUSTRATION WITH SIMULATED DATA

SIMULATING DATA

Data has been simulated to illustrate the 95% interval for a standardized effect size. (1) A random sample, of size n (i.e., $n=10, 30, 50, \& 70$, respectively), was generated from a uniform distribution (i.e., mean $M_u=0$, standard deviation $SD_u=1$). (2) The mean M_y and the standard deviation SD_y were set using the equation (see equation 2.1 of Fan, Felsovalyi, Sivo, & Keenan, 2002):

$$\hat{i}=M_i+SD_i(y)$$

where $\hat{i}$ is a new random variable of specified mean and standard deviation;

M_i is the mean to be assigned to the new random variable;

SD_i is the standard deviation assigned to the new random variable; and

y is a random variable from a distribution having mean 0, and standard deviation 1.

Two conditions were simulated, e.g., treatment and control. Means will be set at 3.00 and 1.00, respectively. Both standard deviations were set at 1.17. For example,

$$\hat{i}=3+1.17(y); \text{ and,}$$

$$\hat{i}=1+1.17(y).$$

See table 2.1

Table 2.1 Annotated Code for Computing the Interval for Standardized Effect Sizes for Four Sample Size Conditions

Code	Annotation
DATA MYDAT1;	
DO I=1 TO 5;	-Generates 10 observations,
X1=0;	-Seed values are I*27 for group 0
Y1=1+RANUNI(I*54)*1.17;	and I*54 for group 1 ¹ ,
OUTPUT MYDAT1;	-Writes results to the file MYDAT1,
X1=1;	
Y1=3+RANUNI(I*27)*1.17;	
OUTPUT MYDAT1;	
END;	
RUN; QUIT;	

The observed test statistic (i.e., in this situation the independent samples t-test), t_{obs} , is used to derive the interval for the noncentrality parameter. See table 2.2.

¹ I used the index as the seed value for generating the sample. This is not recommended. One should use seed values that ensure the observations are independent (see Fan, Felsovalyi, Sivo, & Keenan, 2002)

Table 2.2 Annotated Code for Computing the Interval for Standardized Effect Sizes for Four Sample Size Conditions

Code	Annotation
PROC MEANS DATA=MYDAT1; CLASS X1; VAR Y1; OUTPUT OUT=MYOUT1 T=TOBS N=NSIZE; RUN; QUIT;	-Computes the <i>t</i> -statistics using data from MYDAT1. -Writes the observed <i>t</i> and total sample size, <i>n</i> , to file MYOUT,

The equation for the noncentrality (see equation 4.6 of Smithson, 2003)¹ parameter is:

$$D = d \left(\frac{n_t n_c}{n_t + n_c} \right)^{1/2}$$

where *D* is the noncentrality parameter;

d is the measure of effect size (see equation 4.5 of Smithson 2003);

$$d = \frac{\bar{X}_t - \bar{X}_c}{S_p}$$

n_i is the group sample size.

The SAS (2001) function TNONCT returns the value of the noncentrality parameter. The relevance of the equation for the noncentrality parameter is the effect size. Rearranging the equation for the noncentrality parameter, the effect size is:

$$d = \frac{D}{\left(\frac{n_t n_c}{n_t + n_c} \right)^{1/2}}$$

The equation allows one to solve for the standardized effect size given the noncentrality parameter and sample sizes. See table 2.3.

Table 2.3 Annotated Code for Computing the Interval for Standardized Effect Sizes for Four Sample Size Conditions

Code	Annotation
DATA MYDATOUT; SET MYOUT1; IF _FREQ_=10; T=TOBS; N=NSIZE; DLO=TNONCT(T, N-2, .975)/ (((N/2)*(N/2))/((N/2)+(N/2)))**.5) ; DUP=TNONCT(T, N-2, .025)/ (((N/2)*(N/2))/((N/2)+(N/2)))**.5) ; WDTH=DUP-DLO; OUTPUT MYDATOUT; KEEP T N DLO DUP WDTH; RUN; QUIT;	-Sets MYOUT1 as the data set that succeeding statements should refer, -Writes lower (DLO) and upper (DUP) bounds of the standardized interval, the interval width (WDTH), <i>t_{obs}</i> , & <i>N</i> to file MYDATOUT,

¹Roman characters are substituted for Smithson's (2003) Greek characters.

Similar code was written to derive the standardized interval width for total sample sizes of 30, 50, and 70.

RESULTS

Results show that as sample size increases from 10 to 70, the lower limit increases from 2.63 to 3.79 (see table 3). The upper limit decreases from 8.57 to 5.63.

Table 3. Interval Limits for Standardized Effect Sizes

<i>tobs</i>	<i>N</i>	<i>Interval Limits</i>	
		<i>Lower</i>	<i>Upper</i>
8.89	10.00	2.63	8.57
13.49	30.00	3.45	6.38
17.36	50.00	3.78	6.03
19.72	70.00	3.79	5.63

This results in decreasing the interval width. The width is 5.94 when $N=10$. It reaches 1.84, for $N=70$. See figure 2.

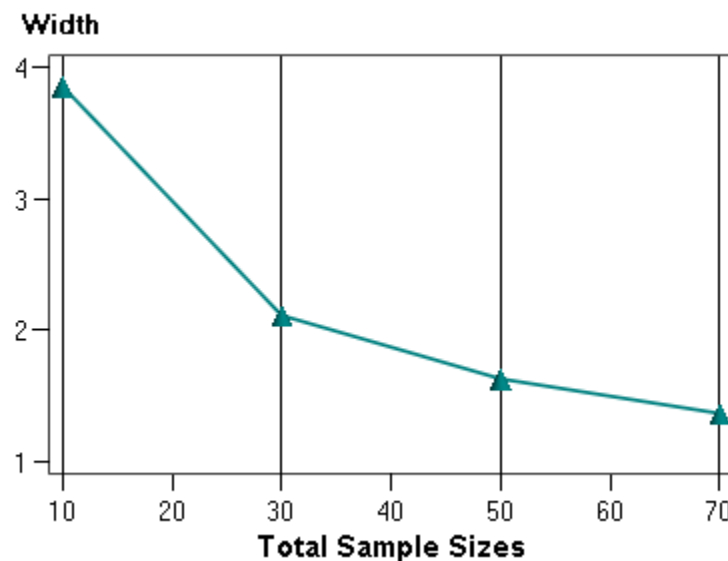


Figure 2. Relation between Interval Width & Size

CONCLUSION

The Inferential confidence interval provides an alternative to the significance test. The range of the limits indicates the veracity of the null hypothesis (Hoenig & Heisey, 2001). The width of the interval indicates the precision of estimating the population parameter (Hoenig & Heisey, 2001). Effect sizes aid decisions regarding the effectiveness of a treatment (Kirk, 1996). The interval inversion principle brings together the explanatory capability of the confidence interval and effect size (Steiger and Fouladi, 1997). The observed test statistic is used to find confidence limits about the noncentrality parameter. The SAS (2001) function TNONCT provides the limits. The equation for the noncentrality parameter is then modified to solve for the effect size limits.

REFERENCES

- Fan, X., Felsovalyi, A., Sivo, S. A. & Keenan, S. C. (2002). *SAS for monte carlo studies: A guide for quantitative researchers*. Cary, NC: SAS
- Hoenig, J. M. & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *American Statistician*, 55, 19-23.
- Kirk, Roger E. 1995. *Experimental design: Procedures for the behavioral sciences*. Pacific Grove, CA: Brooks/Cole.
- SAS (2001). The SAS System for Windows (Version 8.02). [Computer Software]. Cary, NC: SAS Institute Inc.
- Smithson, M. (2003). *Confidence intervals*. Thousand Oaks, CA: Sage Publications, Inc.
- Steiger, J. H., & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221–257). Hillsdale, NJ: Erlbaum.
- Wilkinson, L. & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*. 594-604.