Confidence Intervals Within Hypnosis Research

Marty Sapp, Ed.D.

Confidence intervals within hypnosis would improve research, especially confidence intervals for effect sizes. Moreover, within hypnosis research, confidence intervals are needed for reliability and validity measures. This paper describes definitions and issues related to confidence intervals. Confidence intervals can be used with one sample and two sample cases, and they can facilitate meta-analytic thinking. Finally, this paper provides definitions of confidence intervals that decrease misunderstanding and SPSS computer codes for calculating them. **(Sleep and Hypnosis 2004;6(4):169-176)**

Key words: *Confidence intervals, hypnosis research, effect sizes, reliability, validity*

INTRODUCTION

This paper is divided into eight sections. Section One provides definitions of confidence intervals. Section Two discusses the concept of the standard error of the mean. Section Three describes confidence intervals for the one-sample case, and Section Four describes confidence intervals for the two-sample case. Section Five describes confidence intervals around reliability. Confidence intervals around validity indices is described in Section Six, and Section Seven discusses confidence intervals around effect sizes. Finally, Section Eight provides a discussion of the implications of confidence intervals within hypnosis research.

From the Department of Educational Psychology, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, USA

Address reprint requests to: Dr. Marty Sapp, Department of Educational Psychology, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, USA
Phone: (414) 229-6347
Fax: (414) 229-4939
E-mail: Sapp@uwm.edu

Accepted November 13, 2004

DEFINITIONS OF CONFIDENCE INTERVALS

Thompson (1) described confidence intervals as an interval among an infinite number of intervals, or least large numbers of intervals for a given parameter, such as population mean, population reliability, population difference score, in which one minus the alpha level would capture the population parameters. In other words, for a 95% interval for a population mean, 95 percent of the intervals would capture the population parameter or mean and 5% of the intervals would not. Within the context of confidence intervals, parameters describe population characteristics. Stated somewhat differently, confidence intervals allow researchers to put lower limits and upper limits around population parameters. The 95% and the 99% intervals are the most commonly used intervals. With a 99% confidence interval, a researcher is assuming that 99% of the confidence intervals contain the population parameters, and 1% would not. In essence, a 99% confidence interval is wider than a 95% interval, and with wider intervals the more certain one can be.

THE STANDARD ERROR OF THE MEAN

The upper and lower limits of confidence intervals for one-sample and two-sample cases are found by the standard error of the mean, and this refers to the plus and minus (±) part of a confidence interval. (Technically, this is referred to as plus or minus, but often individuals fail to understand that addition and subtraction must occur, so colloquially this sign is referred to as plus and minus.). The central limit theorem and the t-distribution provide the rational for confidence intervals for the one-sample and two-sample cases. Sapp (2) defined the central limit theorem as random samples of a fixed size drawn randomly from a population and as the number of cases get large, the sample means approach normality, regardless of the shape of the population distribution. With randomly selected means of a fixed size that are randomly selected from a population, the mean of the sample means will approach the population mean and the standard error of the mean equals the population standard deviation divided by the square root of the number of cases. Technically, the standard deviation of the sample means is the standard error, but within practical terms, one does not have access to a sampling distribution of means.

CONFIDENCE INTERVALS FOR THE ONE-SAMPLE CASE

Suppose 10 participates had the following scores on the Stanford Hypnotic Susceptibility Scale, Form C: (SHSS:C) :

Participants' Scores	Frequency
1 1	1
10	2
9	1
8	3
7	3

Sapp (3) provided the steps for conducting such an analysis using the Texas Instrument TI-30xa calculator. The mean for these data is 8.5, the standard deviation is 1.43, and the standard error of the mean is the standard deviation divided by the square of the number of cases which is 10; therefore, the standard error of the mean equals 1.43/3.16=.47. The general formula for a confidence interval is mean plus and minus the critical value of t times the standard error of the mean, or $\bar{X} \pm (t)(S)$. In order to find the t value we need the degrees of freedom, which is the number of cases minus one, or 9. The critical t values for the .05 level is 2.26 and 3.25 for the .01 level. S is the standard error of the mean. One-hundred minus .05 provides the 95% confidence level and one-hundred minus .01 provides the 99% confidence interval. For the 95% confidence interval, the lower limit is 7.5 and the upper limit is 9.5, and for the 99% confidence interval, the lower limit is 7 and the upper limit is 10. The reader should notice that the 99% confidence is larger and that neither interval contains zero which means that the null hypothesis was rejected and statistical significance was assumed at the .05 and .01 levels.

The SPSS control lines for running these analyses are the following:

```
 T-TEST
   /TESTVAL=0
   /MISSING=ANALYSIS
    /VARIABLES=SHSS
    /CRITERIA=CIN (.95).*

T-TEST
   /TESTVAL=0
   /MISSING=ANALYSIS
    /VARIABLES=SHSS
   /CRITERIA=CIN (.99).*
```

Table 2 Note the subcommand /criteria=CIN (.95) provides the confidence interval for the 95% level and the subcommand /criteria=CIN (.99) provides the confidence interval for the 99% level.

CONFIDENCE INTERVALS FOR THE TWO-SAMPLE CASE

The two-sample case proceeds the same way as the one-sample case. First, the means for each group is found, the difference between means, and next the standard error of the means. With the following example, two groups of participants were tested on the SHSS:C

Group 1	Group 2
0	10
1	12
2	10
3	9

The mean difference is 10.25-1.5 or 8.75. The standard error is just the variance of each group divided by the group size and then take the square root of the sum of the variances divided by the appropriate group sizes. For example, the variance for group 1 is 1.67 and it is 1.58 for group 2. Thus, the standard error is .89. Now, the 95% confidence interval is the following:

Mean difference plus and minus (critical value of t)(standard error) 8.75 plus and minus 2.447(.89)=6.57 for the lower limit and 10.93 for the upper limit. The critical value of t is found by the degrees of freedom which is the number of cases minus 2 or 6. The 99% confidence interval is the following:

8.75 plus and minus 3.707(.89)=3.29923 (standard error) and 5.45 for the lower limit and 12.05 for the upper limit. As previously stated, the reader can see that the 99% confidence interval provides a larger interval than the 95% one. The SPSS control lines for this analysis are the following:

```
T-TEST
    GROUPS=var00001(1 2)
    /MISSING=ANALYSIS
    /VARIABLES=var00002
    /CRITERIA=CIN(.95).
```

To obtain the 99% confidence interval the only change in the control lines is the CIN (.99).

CONFIDENCE INTERVALS AROUND RELIABILITY

Thompson (1) has made a number of recommendations for social sciences researchers. One, he recommends putting confidence intervals around reliabilities like coefficient alpha. Sapp (3) defined reliability as the function of test items and it measures the consistency of items. In addition, reliability is always squared area or a squared correlation. Confidence intervals around reliabilities require non-centralized distributions-which allow one to perform power analysis, or the probability of rejecting a false null hypothesis (no treatment effect). Unlike centralized distributions like the z and t distributions, which have a mean of zero, non-centralized distributions have a mean of some hypothesized value. Fan and Thompson (4) provided the following SPSS codes for confidence intervals for reliability around coefficient alpha. Suppose 10 participants had completed 3 items from SHSS: C, what is the reliability and confidence interval around reliability and does the valued obtained differ from a hypothesized value of .70? And the SPSS codes are the following:

```
Title "Reliability and Confidence."
Data list free/v1 v2 v3.
Begin data.
1 2 1
7 4 6
3 1 1
3 2 5
7 4 4
7 4 6
5 3 4
3 2 2
2 1 1
0 1 2
End data.
List.
Reliability variables=v1 to v3/
  Scale(TOTAL)=v1 to V3/
  Statistics=corr cov/summary=means var total/
  Icc=model(random) type(consistency) cin=95
  testval=.70/
  Model=alpha.
```

The reader should will notice two things from this output. First, coefficient alpha-the point estimate is .8987, and the 95% confidence interval is .7033, .9726. Where

.7033 is the lower limit value of the confidence interval, and .9726 is the upper limit or value. Second, testing the hypothesized value of coefficient alpha of .7000 against the calculated value of .8987 did differ significantly, because F=2.9617 and the probability level is .0239.

CONFIDENCE INTERVALS AROUND VALIDITY INDICES

One simple definition of validity is the correlation among a set of items that have been shown to be valid with a set of items are being tested to determine their validity; hence, validity can be defined as a Pearson product-moment correlation (2,3,5). Unfortunately, the sampling distribution of the population Pearson product-moment correlation is skewed when the value is other than zero, and when the value is zero it is symmetrical and almost normal; moreover, the degree of skewness of the sampling distribution of the population correlation is also a function of sample size.

For example, when the population correlation coefficient is not zero and the sample size is small, there is greater skewness within the sampling distribution. Since the sampling distribution of the population correlation coefficient is not normally distribution, the statistican R. A. Fisher developed a logarithmic transformation of r with tables that can be found in Sapp (5, p. 124). Suppose a researcher found that the items of two hypnotizability scales had a correlation of .30, how would one construct a 95% confidence interval. First, using the table from Sapp (5, p 124), we find the Fisher's z transformation of .30 which is .31. Next, we calculate the standard error which is one divided by the square root of the number of cases minus 3. If we assume that the number of cases is 25, the standard error is . 21. The 95% confidence interval is the following:

.31 plus and minus 1.96(.21), with the lower confidence limit is -.10 and the upper limit .72, but we have to transform the Fisher's z back to the regular r which become -.10 for the lower limit and .62 for the upper limit. The 99% confidence interval can be found by substituting 2.58 for 1.96 and following the same steps. The reader should notice that the confidence interval, -.10, .62 contains zero therefore the population correlation coefficient does not differ significantly from zero.

CONFIDENCE INTERVALS AROUND EFFECT SIZES

Sapp (6) reported that are over 40 different effect size measures. Effect sizes are important because they allow for meta-analytic thinking, or they provide a quantitative way of summarizing or synthesizing the literature within an area of hypnosis. Cohen (7) was one of the first researchers to describe effect sizes and a related area called power analysis. Power is the probability of rejecting a false null hypothesis (population means are equal). Even though there are a variety of effect sizes, two are commonly used within meta-analysis, the correlation, which was discussed within the section on confidence intervals around validity, and the d effect size. The Pearson product-moment correlation or Pearson r is familiar to readers, and Rosenthal (1984) defined value of r of .1 as a small effect size, and a value of .3 as a medium effect size, and a value of .5 as a large effect size. Rosenthal is clear that these are rough rules of thumb and effect sizes have to be evaluated within a substantive area. Cohen (7) defined the d effect size as the difference between two means divided by the standard deviation or some form of variability, and Cohen provided the following rough rules of thumb for interpreting the d effect sizes: d=.2 small effect size, d=.5 medium effect size, d=.8 large effect size.

Confidence intervals around correlations are the same processes described within this article within the section on confidence intervals around validity indices; however, confidence intervals around d effect sizes is more complex

Table 1. Highest Average Effect Size Measures for Traditional Cognitive-Behavioral Orientations

Orientation	Effect size d	Effect size r	Number of studies
Behavioral therapy	1.06	.48	214
Hypnotherapy	1.82	.68	475
Systematic Desensitization	1.05	.47	475
Implosive therapy	.68	.33	475
Behavior modification	.73	.35	475
Cognitive-behavioral therapy	1.13	.50	475
Cognitive therapy	1.00	.46	214
Self-control strategies	1.01	.46	214
Biofeedback	.91	.42	214
Covert-behavioral	1.52	.61	214
Flooding	1.12	.50	214
Relaxation therapy	.90	.42	214
Reinforcement	.97	.45	214
Modeling	1.43	.59	214

and require, like coefficient alpha, non-centralized distributions. In addition, readers who are interested in the software for these distributions can consult Bird (8) and Smithson (9), and Professor Geoff Cummings at La Trobe University in Australia, has developed software that runs under the program Excel, and it can be downloaded from the following website: http://www.latrobe.edu.au/psy/esci.

The highest average effect sizes for traditional cognitive-behavioral orientations or therapies are found in Table 1. Cognitive-behavioral therapies are eclectic groups of techniques that combine strategies from cognitive and behavioral psychology. Albert Ellis is the grandfather of cognitive-behavioral therapy within the area of clinical psychology; whereas, Aaron T. Beck is a prominent figure within the area psychiatry. Traditional forms of cognitive-behavioral therapies developed from academic psychology or were embraced by academic psychology. As previously stated, traditional cognitive-behavioral orientations are found in Table 1.

Behavioral therapy is the collection of techniques to change behavior. Hypnotherapy is the therapeutic use of hypnosis to change behavior. Moreover, hypnosis parallels guided imagery, biofeedback, and progressive relaxation. In essence, it is a therapeutic relationship between a therapist and client in which the client receives suggestions that have psychophysiological effects. In addition, hypnosis can lead to behavioral changes and there are several styles or approaches to hypnosis. Systematic desensitization is an imagination technique that pairs relaxation and anxiety-evoking stimuli.

Implosive therapy is also an imagination-based procedure that is similar to flooding, but employs psychoanalytic imagery scenes. Behavior modification is the use of behavioral techniques to modify or change behavior. As previously stated, cognitive-behavioral therapy is a blend of cognitive and behavioral psychology, and it states that clients' problems are the results of faulty belief systems. Cognitive therapy is a form of cognitive-therapy that was developed by Aaron Beck, and the goals are to change clients distorted cognitions.

Self-control strategies are techniques that clients implement themselves, such as self-monitoring, self-reinforcement, self-punishment, bibliotherapy, self-hypnosis, and so on. Biofeedback is the use of computers and other technologies to provide clients with feedback about physiological processes, such as body temperature, blood pressure, and heart rate. Covert behavioral therapy is another

Table 2. Highest Average Effect Size Measures for Nontraditional Cognitive-Behavioral Orientations

Orientation	Effect size d	Effect size r	Number of studies
Adlerian	.71	.34	375
Transactional analysis	.67	.33	475
Reality therapy	.75	.35	21

imagery-based technique that uses aversive stimuli to desensitize clients to anxiety. Flooding is yet another imagination procedure or in vivo, real-life-based procedure that leads to the extinction or unlearning of behavior.

Relaxation therapy is the implementation of progressive relaxation or muscle-tension exercises or guided imagery to produce relaxation. Reinforcement is a technique that leads to an increase in behavior. Finally, modeling is a social learning technique that employs models to model and modify certain behaviors.

Within Table 1, the reader should notice that cognitive-behavioral therapy has a d effect size of 1.13, which is a large effect size, and 475 studies supported this statistic. Interestingly, hypnotherapy has a d effect size of 1.82-again a large effect size measure. Likewise, cognitive therapy has a large effect size of 1.00 with 214 studies to support the overall effect. Moreover, implosive therapy has a medium effect size of .68. In contrast, flooding, a technique related to implosive therapy, has a large effect size of 1.12.

Effect sizes of nontraditional cognitive-behavioral orientations are found in Table 2. Adlerian therapy, a forerunner of traditional cognitive-behavioral therapies, is an analytical form of psychotherapy. Transactional analysis is a form of cognitive-behavioral therapy that did not develop from academic or traditional behavior therapy. Within this form of psychotherapy, usually there are not references to classical or operant conditioning, but an analysis of clients ego states, which is a behavioral way of operationally defining Freud's constructs of personality-id, ego, superego; however, transactional analysis refers to these as the child, adult, and parent ego states, respectively. In contrast to Freud's psychoanalysis, transactional analysis is a humanistic-existential form of psychotherapy that stresses clients can choose and change even when experiencing harsh social and environmental forces.

Reality therapy did not develop from academic psychology either, and it has humanistic-existential and behavior modification aspects. Reality therapy teaches clients to take responsibility for their behaviors and actions and to do things in order for change to occur.

Table 2 has the highest average effect size measures for nontraditional cognitive-behavioral therapies. Adlerian therapy had a d effect size of .71, which is a medium effect size. Transactional analysis had a d effect size of .67. which is another medium effect size measure. Finally, reality therapy had a d effect size of .75, which rounded, is approximately .80-indicating a large effect size; however, there are only 21 studies used in calculating this effect size.

Table 3 has the 95% confidence intervals for the highest average d effect sizes for the various forms of psychotherapy along with the power values, and these intervals were calculated with software found Smithson (10), and this table provides more information than Tables 1 and 2. For example, clearly short-term dynamic therapy and reality therapy both have lower power values than the other therapies. Stevens defined power values greater than .70 as adequate and values greater than .90 as excellent. Both short-term dynamic therapy and reality therapy have lower power values because of their small sample sizes. Power is not usually an issue with sample sizes of about 100. The readers should also notice that short-term dynamic therapy and reality therapy both have zeros within their confidence intervals. This

Table 3. Confidence Intervals for Highest Effect Sizes for Various Forms of Psychotherapy with Power Values

Orientation	Effect size d	Estimated t-values	95% Confidence interval for d	Power values	Number of studies
Short-Term Dynamic Therapy	0.71	1.81	-.0451, .7483	0.41	26
Psychodynamic	0.69	7.52	.2523, .4374	1.00	475
Dynamic Eclectic	0.89	9.69	.3505, .5392	1.00	474
Adlerian	0.71	6.87	.2502, .4589	1.00	375
Dynamic/Humanistic	0.40	2.93	.0648, .3354	0.83	214
Person-Centered	0.63	6.10	.2111, .4185	1.00	375
Gestalt	0.64	6.97	.2274, .4118	1.00	475
Transactional Analysis	0.67	7.30	.2424, .4272	1.00	475
Behavioral Therapy	1.06	7.75	.3861, .6723	1.00	214
Hypnotherapy	1.82	19.83	.8025, 1.0163	1.00	475
Systematic Desensitization	1.05	11.44	.4287, .6206	1.00	475
Implosive Therapy	0.68	7.41	.2474, .4323	1.00	475
Behavior Modification	0.73	7.95	.2717, .4575	1.00	475
Cognitive-Behavioral Therapy	1.13	12.31	.4677, .6614	1.00	475
Cognitive Therapy	1.00	7.31	.3571, .6413	1.00	214
Self-Control	1.01	7.39	.3624, .6469	1.00	214
Biofeedback	0.91	6.66	.3140, .5956	1.00	214
Covert-Behavioral	1.52	11.12	.6072, .9116	1.00	214
Flooding	1.12	8.19	.4151, .7034	1.00	214
Relaxation Therapy	0.90	6.58	.3087, .5899	1.00	214
Reinforcement	0.97	7.09	.3424, .6258	1.00	214
Modeling	1.43	10.46	.5642, .8645	1.00	214
Reality Therapy	0.75	1.72	-.0722, .8083	0.38	21

d: population effect size for the t-test for independent two groups and $t = (\sqrt{N}/2)d$ and $d = 2t/\sqrt{N}$
According to Stevens (2002), power > .70 is adequate and > .90 is excellent.

suggest the null hypothesis was not rejected and negative values within these intervals also suggest that both therapies can produce negative effects or harm; however, clearly any form of psychotherapy can produce negative effects, especially treatments with large effect sizes.

In summary, both short-dynamic therapy and reality therapy have confidence intervals that include zero, and confidence intervals around d provide more information than the single point estimates of d.

Cognitive-behavioral therapy had a lower limit of .4677(about a medium effect size) and an upper limit of .6614, again a value reflecting a medium effect size). Hypnotherapy had a lower limit of .8025(large effect size) and an upper limit of 1.0163(large effect size).

Dynamic/humanistic therapy, person-centered therapy, Gestalt therapy all had lower confidence intervals that are within the lower ranges, with dynamic/humanistic therapy with a lower limit of .0648-which is the lowest of the humanistic therapies.

In terms of cognitive-behavioral therapies with upper limits of values that are .80 or larger, only hypnotherapy, covert-behavioral therapy, modeling, and reality therapy had upper limits with high d effect sizes using confidence intervals.

Clearly, Table 3 provides more information than Tables 1 and 2, and the confidence intervals suggest that true population parameters vary from the d effect size point estimates and most forms of cognitive-behavioral therapies have upper confidence intervals within the medium effect size ranges, and not the larger ranges as suggested by the point estimates of d.

DISCUSSION

Confidence intervals are necessary for

hypnosis research to advance. Specifically, it is known that hypnosis can increase the effect sizes of cognitive-behavioral and dynamically oriented psychotherapies (11). What is needed within hypnosis research is the reporting of effect sizes and confidence interval around effect size measures. Moreover, from a measurement point of view, hypnosis researchers need to report reliability measures for all items used within studies and confidence intervals around reliability. Likewise, confidence intervals are also needed for validity measures used within hypnosis research. When hypnosis researchers start routinely reporting confidence intervals within studies, this will aid in the meta-analytical thinking. And as this article demonstrated confidence intervals provide more information than point estimates. Finally, confidence intervals, especially ones around reliability, can allow researchers to test specific hypotheses about reliability as opposed to testing the null hypothesis. In summary, with the new software that exist within the social sciences, hypnosis researchers can improve the area of hypnosis research by providing confidence intervals and improving quantitative

REFERENCES

1. *Thompson, B. What future quantitative social science research could look like: Confidance intervals for effect sizes.*
2. *Sapp M. Test anxiety: Applied research, assessment, and treatment interventions (2nd ed.) Lanham, MD: University Press of America, 1999.*
3. *Sapp M. (2002). Psychological and educational test scores. Spring field, ILL: Charles C. Thomas Publisher.*
4. *Fan X, Thompson B. Confidence intervals about score reliability coefficients, please: An EPM guidelines editorial. Educational and Psychological Measurement 2001;61:517-531.*
5. *Sapp M. Counseling and psychotherapy: Theories, associated research, and issues. Lanham, MD: University Press of America, 1997.*
6. *Sapp M. Cognitive-behavioral theories of counseling: Traditional and nontraditional approaches. Springfield, IL: Charles C. Thomas Publisher, (2004).*
7. *Cohen J. Statistical power analysis for the behavioral sciences. New York: Academic Press, 1977.*
8. *Bird KD. Confidence intervals for effect sizes in analysis of variance. Educational and Psychological Measurement 2002;62:197-226.*
9. *Smithson M. Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentralized distributions in computing intervals. Educational and Psychological Measurement 2001;61:605-632.*
10. *Smithson M. Confidence intervals. Thousand Oaks, CA: Sage Publications, 2003.*
11. *Kirsch, I., Montgomery, G. and Sapirstein, G. (1995). Hypnosis as an adjunct to cognitive behavioral psychotherapy: A meta-analysis. Journal of Consulting and Clinical Psycholog 63:214-220*