DOCUMENT RESUME

ED 438 308 TM 030 619

AUTHOR	Mahadevan, Lakshmi
TITLE	The Effect Size Statistic: Overview of Various Choices.
PUB DATE	2000-01-00
NOTE	19p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Dallas, TX, January 27-29, 2000).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Data Analysis; *Effect Size; Measurement Techniques; Meta Analysis; *Statistical Significance

ABSTRACT

Over the years, methodologists have been recommending that researchers use magnitude of effect estimates in result interpretation to highlight the distinction between statistical and practical significance (cf. R. Kirk, 1996). A magnitude of effect statistic (i.e., effect size) tells to what degree the dependent variable can be controlled, predicted, or explained by the independent variable (P. Snyder and S. Lawson, 1993). There are a number of ways one can compute an effect size statistic as part of data analysis. There is no concept of "one size fits all" (B. Thompson, 1999), so it is up to the smart researcher to choose the index best suited for a particular research endeavor. It has now become necessary that such a statistic always be included to enable other researchers to carry out meta-analyses and to inform judgment regarding the practical significance of results. This paper provides a tutorial summary of some of the effect size choices so that researchers will be able to follow the recommendations of the American Psychological Association (APA) publication manual, those of the APA Task Force on Statistical Inference, and the publication requirements of some journals. (Contains 3 tables and 11 references.) (Author/SLD)

The Effect Size Statistic: Overview of Various Choices

Lakshmi Mahadevan
Texas A&M University

Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, January 27-29, 2000

Abstract

Over the years, methodologists have been recommending that researchers use magnitude of effect estimates in result interpretation to highlight the distinction between statistical and practical significance (cf. Kirk, 1996). A magnitude of effect statistic (i.e., effect size) tells us to what degree the dependent variable can be controlled, predicted or explained by the independent variable(s) (Snyder & Lawson, 1993).

There are a number of ways one can compute an effect size statistic as a part of data analysis. There is no concept of "one-size fits all" (Thompson, 1999), so it is up to the smart researcher to choose the index best suited for a particular research endeavor. However, it has now become necessary that such a statistic always be included to enable other researchers to carry out meta-analyses and to inform judgment regarding the practical significance of results.

The purpose of the present paper is to provide a tutorial summary of some of the many effect size choices, so that SERA members will be better able to follow the recommendations of the APA publication manual, the APA Task Force on Statistical Inference, and the publication requirements of some journals.

Over the years, statistical significance has been the prominent feature of data analyses in the field of education and other social sciences. However, the results of statistical significance tests do not always aid the researcher in determining whether these results are of practical significance (Kirk, 1996). Methodologists suggest that researchers use magnitude of effect estimates in result interpretation to highlight the distinction between statistical and practical significance (cf.\Shaver, 1991). A magnitude of effect statistic (e.g. effect size) tells us how much of the dependent variable can be controlled, predicted or explained by the independent variable(s)(Snyder & Lawson, 1993).

Given the criticisms of statistical significance tests (cf. Cohen, 1994; Schmidt, 1996), researchers are increasingly emphasizing effect sizes as being critical to thoughtful research practice (cf. Kirk, 1996; Thompson, 1996). Indeed, the APA Task Force on Statistical Inference recently suggested, "Always provide some effect-size estimate when reporting a *p* value" (Wilkinson & The APA Task Force on Statistical Inference, 1999, p. 599, emphasis added), and later noted that, "We must stress again that reporting and interpreting effect sizes in the context of previously reported effects is essential to good research" (p. 599, emphasis added).

Definition

The phrase "effect size", can be used to mean "the degree to which the phenomenon is present in the population," or " the degree to which the null hypothesis is false"(Cohen, 1988). Therefore, effect size is a name given to a family of indices that measure the magnitude of a treatment effect. Effect size can be measured in various ways (Kirk, 1996), but the two most common matrices are:

a) standardized differences and

b) the variance-accounted for or correlation between the independent variable and the dependent variable. This correlation is called the "effect size correlation" (Rosnow & Rosenthal, 1996). There are several choices in each of these two families.

Various Representations of the Effect Size Statistic

Effect Size Measures for Two Independent Groups

1. Standardized difference between two groups.

Cohen's *d*. Cohen (1988) suggested that when the research involves the comparison of two groups, it is common to examine the difference between the two means. This difference, however, will have little meaning apart from the particular scale of measurement involved. It is therefore useful to divide the difference between the two means by the common within-group standard deviation (a) so that the effect can be represented in σ units. These units according to Cohen (1969, 1977) can be referred to as the units of *d*. Therefore,

$$d = \frac{M_1 - M_2}{\sigma}$$

Cohen argued that the standard deviation of either group can be used when the variances of the two groups are homogenous. The *d* is a descriptive measure.

However, most researchers use the pooled standard deviation, σ_{pooled} (Rosnow & Rosenthal, 1996). The pooled standard deviation is found as the root mean square of the two standard deviations (Cohen, 1988, p. 44). That is, the pooled standard deviation is the square root of the average of the squared standard deviations. When the two standard deviations are similar the root mean square will not differ much from the simple average of the two variances, because the average of the two equal numbers equals the two numbers being averaged.

The *d* can also be computed from the value of the t test of the differences between the two groups (Rosnow & Rosenthal, 1991). In the following equation "df" is the degrees of freedom for the t test. The "n's" are the number of cases for each group. The formula without the n's should be used when the n's are equal. The formula with separate n's should be used when the n's are not equal.

$$d = 2t \big/ \sqrt{(df)}$$

or

$$d = t(n_1 + n_2) / \left[\sqrt{(df)}\sqrt{n_1 n_2}\right].$$

d can also be computed from r, the effect size correlation:

$$d = 2r \big/ \sqrt{(1 - r^2)}.$$

And *d* can also be computed from Hedges's g.

$$d = g\sqrt{(\mathrm{N} / df)}$$

<u>Interpretation of Cohen's </u>*d*. Cohen defined effect sizes as "small, *d=.2*", "medium, *d=.5*", and "large, *d=.8*". The terms "small", "large" and "medium" are relative not only to each other, but also to the area of behavioral science or even more particularly to the specific content and research method applied in the given setting (p. 25).

Effect sizes can also be thought of as the average percentile standing of the average treated (or experimental) participant relative to the average untreated (or control) participant. An effect size of 0.0 indicates that the mean of the treated group is at the 50th percentile of the untreated group. An effect size of 0.8 indicates that the mean of the treated group is at the 79th percentile of the untreated group. An effect size of 1.7 indicates that the mean of the treated group is at the 95.5 percentile of the untreated group.

Or effect sizes can be interpreted in terms of the percent of nonoverlap of the treated group's scores with those of the untreated group. An effect size of 0.0 indicates that the distribution of scores for the treated group overlaps completely with the distribution of scores for the untreated group, there is 0% of non-overlap. An ES of 0.8 indicates a non-overlap of 47.4% in the two distributions. An effect size of 1.7 indicates a non-overlap of 75.4% in the two distributions. This is indicated in Table 1.

Insert Table 1 about here.

Hedges' *g*. Hedges ' s *g* is an inferential measure. It is normally computed by using the square root of the Mean Square Error from the analysis of variance testing for differences

$$g = \mathrm{M}_1 - \mathrm{M}_2 / S_{pooled}$$

between the two groups. Hedges's *g* is named for Gene V Glass, one of the pioneers of meta-analysis:

$$S = \sqrt{\left[\Sigma(\mathrm{X}-\mathrm{M})^2 / \mathrm{N}-1\right]}$$

and

$$S_{pooled} = \sqrt{\mathrm{M}S_{within}}$$

Hedges's *g* can be computed from the value of the t test of the differences between the two groups (Rosenthal and Rosnow, 1991). The formula with separate n's should be used when the n's are not equal. The formula with the overall number of cases, N, should be used when the n's are equal:

$$g = {}^{t\sqrt{n_1} + n_2}\Big/_{\sqrt{n_1 n_2}}$$

or

$$g = 2t / \sqrt{\mathrm{N}}$$

The pooled standard deviation, σ_{pooled}, can be computed from the unbiased estimator of the pooled population value of the standard deviation, Spooled, and vice versa, using the following formula (Rosnow & Rosenthal, 1996, p. 334):

$$\sigma_{pooled} = S_{pooled}\sqrt{(df / N)},$$

Where df = the degrees of freedom for the MSerror and N the total number of cases.

Hedges's *g* can be computed from Cohen's *d*.

$$g = d / \sqrt{N/df}$$

2. <u>Correlation measures of effect size.</u> The effect size correlation can be computed directly as the point-biserial correlation between the dichotomous independent variable and the continuous dependent variable:

$$r_{Y\lambda} = r_{dv,iv}.$$

The effect size correlation can be computed from a single degree of freedom Chi Square value by taking the square root of the Chi Square value divided by the number of cases, N. This value is also known as phi:

$$r_{Y\lambda} = \Phi = \sqrt{\left(X^2(1)/N\right)}.$$

The effect size correlation can be computed from the t test value:

$$r_{Y\lambda} = \sqrt{\left[t^2 / \left(t^2 + df\right)\right]}$$

The effect size can be computed from a single degree of freedom F test value (e.g., a one-way analysis of variance with two groups):

$$r_{Y\gamma} = \sqrt{[(F(1,\_)/ F(1,\_) + dferror)]}.$$

The effect size correlation can be computed from Cohen's *d*:

$$r_{Y\lambda} = d / \sqrt{d^2 + 4}.$$

The effect size correlation can be computed from Hedges's *g*:

$$r_{Y\lambda} = \sqrt{\left\{\left(g^2 n_1 n_2\right) / \left[g^2 n_1 n_2 + \left(n_1 + n_2\right) df\right]\right\}}.$$

<u>Indices of effect sizes in relation to POWER, sample size, statistical significance, chi-square and F.</u>

Insert Table 2 about here.

The effect size associated with t is Cohen's *d* defined as:

$$d = \frac{M_1 - M_2}{\sigma}.$$

On inspection of the entries under *t* in sections A,B and C in Table 2, to achieve a modest power level of .50 we will require sample sizes of 30, 35 and 200, in each group for the three combinations of expected effect size and alpha respectively.

The effect size associated with r is r itself. The definitions of small, medium and large effects are not as consistent between *r* and *t*.

Insert Table 3 about here.

To achieve the moderate power level of .50 sample sizes of 40, 70 and 400 will be required for the three combinations of expected effect size and alpha. Comparison of sample sizes listed for t and r show the sample sizes required for .r to be consistently higher while the total sample sizes required are lower than that which is required for by t.

The difference between correlation coefficients (rl-r2) is indexed by *q*. Often, enormous sample sizes are required to detect differences and to determine *q:*

$$\frac{1}{2}\log_e \frac{1+r}{1-r}.$$

The difference between an obtained proportion (P) and .50 (*P*-.50) is referred to as *g*.

The difference between two obtained proportions (*P*1-*P*2)is indexed by *h* i.e. the difference between the arcs in transformations of the two proportions.

The effect size associated with χ^2 is called *w*. It is defined as the square root of the sum over all cells (of any size table of frequencies) of the square of the difference between the proportion expected and the proportion obtained in each cell divided by the proportion expected in that cell, or:

$$w = \sqrt{\Sigma \frac{\left(P_{expected} - P_{obtained}\right)}{P_{expected}}}.$$

The effect size associated with *F* is called *f* and is defined as the σ of the means divided by the σ within conditions. In the case of just two groups, *f* is related to *d* by *f=d/2*. More generally *f* is related to the correlation ratio eta, by

$$f = \sqrt{\frac{eta^2}{1-eta^2}}$$

Conclusion

There are a number of ways one can compute an effect size statistic as a part of data analysis. There is no concept of "one-size fits all" (Thompson, 1999), it is up to the researcher to choose the method best suited for his or her purpose. However, it has now become necessary that such a statistic be included in every study (Wilkinson & The APA Task Force on Statistical Inference, 1999) so as to enable other researchers to carry out extensive meta-analyses and possible replication of studies. The magnitude of effect estimates add high value to the research design and increased confidence in the reliability and validity of inferences drawn.

References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Earlbaum.

Cohen, J. (1994). The earth is round (p < .05). American Psychologist, 49, 997—1003.

Kirk, R. (1996). Practical significance: A concept whose time has come. Educational and Psychological Measurement, 56, 746-759.

Rosenthal, R. & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw Hill.

Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. Pyschological Methods, 1, 331-340.

Schmidt, F. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. Psychological Methods, 1, 115—129.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. Journal of Experimental Education, 61, 334-349.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. Educational Researcher, 25(2), 26-30.

Thompson, B. (1999, April). Common methodology mistakes in educational research, revisited, along with a primer on both effect sizes and the bootstrap. Invited address presented at the annual meeting of the American Educational Research Association, Montreal. (ERIC Document Reproduction Service No. ED 429 110)

http://www.apa.org/journal/amp/amp548594.html]

Wilkinson, L., & The APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. American Psychologist, 54, 594-604. [reprint available through the APA Home Page:

Wilson, S. A., Becker, L. A., & Tinker, R. H. (1995). Eye movement desensitization and reprocessing (EMDR) treatment for psychologically traumatized individuals. Journal of Consulting and Clinical Psychology, 63, 928-937.

Table 1

Percentages of non-overlap according to Cohen's Effect Size standards.

Cohen's Standard	Effect Size	Percentile Standing	Percent of Non-overlap
	2.0	97.7	81.1%
	1.9	97.1	79.4%
	1.8	96.4	77.4%
	1.7	95.5	75.4%
	1.6	94.5	73.1%
	1.5	93.3	70.7%
	1.4	91.9	68.1%
	1.3	90	65.3%
	1.2	88	62.2%
	1.1	86	58.9%
	1.0	84	55.4%
	0.9	82	51.6%
LARGE	0.8	79	47.4%
	0.7	76	43.0%
	0.6	73	38.2%
MEDIUM	0.5	69	33.0%
	0.4	66	27.4%
	0.3	62	21.3%
SMALL	0.2	58	14.7%
	0.1	54	7.7%
	0.0	50	0%

Note. Adapted from Cohen (1988).

Table 2

Multipurpose power tables.

Statistics and effect sizes							
Statistic	t	r	r_1-r_2	$P-.50$	P_1-P_2	χ^2	F
Effect Size	d	r	q	g	h	w	f
a. small	.20	.10	.10	.05	.20	.10	.10
b. medium	.50	.30	.30	.15	.50	.30	.25
c. large	.80	.50	.50	.25	.80	.50	.40
A	Sample size (rounded) required to detect medium effect at .05, two-tail						
Power	t	r	r_1-r_2	$P-.50$	P_1-P_2	χ^2 (df=1)	F(df=1 for numerator)
.15	10	10	20	<10	<10	<25	10
.20	10	15	30	10	10	<25	10
.30	20	35	50	20	20	25	20
.40	25	40	70	25	25	30	25
.50	30	55	90	30	30	45	30
.60	40	70	115	40	40	55	40
.70	50	90	150	50	50	70	50
.80	65	115	175	65	65	90	65
.90	85	140	235	85	85	120	85
definition of n:	a	b	c	d	c	d	a
B	Sample size (rounded) required to detect medium effect at .01, two-tail						
Power	t	r	r_1-r_2	$P-.50$	P_1-P_2	χ^2 (df=1)	F (df=1)
.15	20	30	55	30	20	25	20
.20	25	35	70	40	25	35	25
.30	35	45	95	50	35	45	35
.40	45	60	125	60	45	60	45
.50	55	70	150	70	55	75	55
.60	65	85	180	85	65	90	65
.70	80	100	220	100	75	110	80
.80	95	125	260	130	95	130	95
.90	120	160	330	160	120	160	120
C	Sample size (rounded) required to detect 'small' effect at .05, two-tail						
Power	t	r	r_1-r_2	$P-.50$	P_1-P_2	χ^2 (df=1)	F (df=1)
.15	45	85	170	90	40	80	45
.20	65	125	250	120	65	125	65
.30	105	200	400	200	105	200	105
.40	150	300	600	300	140	300	150
.50	200	400	800	400	200	400	200
.60	250	500	1000	500	250	500	250
.70	300	600	1250	650	300	600	300
.80	400	800	1600	800	400	800	400
.90	550	1000	2100	1000	500	1000	550

a. each group or condition
b. *n* of score pairs
c. *n* of each sample
d. total N

Note. Adapted from Cohen (1977).

Table 3

The levels of r that are equivalent to each level of *d*.

	d	Cohen's r	r equivalent to *d**
Small	.20	.10	.10
medium	.50	.30	.24
Large	.80	.50	.37

*where r is obtained from d by

$$r_{Y\lambda} = d / \sqrt{d^2 + 4}$$

Note. See Rosnow and Rosenthal (1984).

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

ERIC

TM030619

I. DOCUMENT IDENTIFICATION:

Title: THE EFFECT SIZE STATISTIC: OVERVIEW OF VARIOUS CHOICES	
Author(s): LAKSHMI MAHADEVAN	
Corporate Source:	Publication Date: 1/00

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] Sample sticker to be affixed to document

Check here

Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document [ ]

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

or here

Permitting reproduction in other than paper copy.

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: Lakshmi Mahadevan	Position: RES ASSOCIATE
Printed Name: LAKSHMI MADADEVAN	Organization: TEXAS A&M UNIVERSITY
Address: TAMU DEPT EDUC PSYC COLLEGE STATION, TX 77843-4225	Telephone Number: (409) 845-1831 Date: 1/20/00

OVER