Journal of Modern Applied Statistical Methods

Volume 14 | Issue 2 Article 21

11-1-2015

JMASM34: Two Group Program for Cohen's d, Hedges' g, η2, Radj2, ω2, ε2, Confidence Intervals, and Power

David A. Walker
Northern Illinois University, dawalker@niu.edu

JMASM34: Two Group Program for Cohen's d, Hedges' g, η2, Radj2, ω2, ε2, Confidence Intervals, and Power

Erratum
This paper was originally published in JMASM Algorithms & Code without its enumeration, JMASM34.

Journal of Modern Applied Statistical Methods
November 2015, Vol. 14, No. 2, 282-292.

ISSN 1538 − 9472

JMASM Algorithms and Code: Two Group Program for Cohen's d, Hedges' g, η^2, R_{adj}^2, ω^2, ε^2, Confidence Intervals, and Power

David A. Walker
Northern Illinois University
DeKalb, IL

The purpose of this research is to provide an application for users interested in a SPSS syntax program to determine an array of commonly-employed effect sizes and confidence intervals not readily available in SPSS functionality, such as the standardized mean difference and r-related squared indices, for a between-group design.

Keywords: Effect size, confidence intervals, SPSS, syntax

Introduction

The purpose of this research is to provide an application for researchers and practitioners interested in a SPSS syntax program (Walker, 2015) to determine an array of commonly-employed effect sizes and confidence intervals not readily available in SPSS functionality, such as the standardized mean difference and *r*-related squared indices, for a between-group design using descriptive statistics: means, standard deviations, and sample sizes.

As a brief *précis*, in the social sciences, there has been a sustained effort by researchers, editorial boards, and professional organizations for mandatory reporting of effect sizes with statistical significance testing (American Educational Research Association, 2006; American Psychological Association [APA], 2010; Cohen, 1992; Ferguson, 2009; Levine & Hullett, 2002; Thompson, 1998; Wilkinson & The APA Task Force on Statistical Inference, 1999). Cohen (1988, p. 10) noted that an effect size, "…serves as an index of degree of departure from the null hypothesis." When reported with statistically significant results, effect sizes can provide information, for example, pertaining to the extent

Dr. Walker is a Professor of Educational Research, Technology, and Assessment in the College of Education. Email him at dawalker@niu.edu.

of the difference between means or the magnitude of a relationship in terms of the proportion of the total variance accounted for in an outcome (Cohen, 1988). Effect sizes can also be employed to indicate the functional, applied effect of an outcome (Nickerson, 2000).

Ferguson (2009) and Thompson (2009) proposed that effect sizes differentiate generally into the subsequent categories: 1) variance accounted for measures such as squared indices of *r*; 2) corrected estimates, typically employed to reduce estimation bias, such as R_{adj}^2; and 3) standardized mean differences, for example, Cohen's *d*. The current study's program will extrapolate effect sizes from all of these categories.

Cohen (1988) suggested that for *r*-related squared indices, which indicate the proportion of variance in the dependent variable accounted for by the effect of the independent variable, values of .01, .06, and .14 should serve as markers of small, medium, and large effects, respectively. Further, Cohen (1988) defined the values of effect sizes for the standardized difference between means as small = .20, medium = .50, and large = .80. However, it should be appropriately noted that it is at the discretion of the researcher to determine the context in which qualifying labels such as "small," "medium," and "large" effects are being defined when using any effect size index. This caution has been stated by Glass, McGaw, and Smith (1981) with reiteration from Cohen (1988) and Thompson (2009).

Lastly, there has been an emphasis in the literature (APA, 2010; Cohen, 1994; Sapp, 2004; Vacha-Haase & Thompson, 2004; Wilkinson & The APA Task Force on Statistical Inference, 1999) that not only should effect sizes be reported with statistically significant results, but confidence intervals ought to complement said point estimate indices for more comprehensive analysis and interpretation of outcomes. As noted by Levin and Robinson (2003, p. 235), "Reporting and interpreting effect sizes (with corresponding confidence intervals) in multiple experiment studies where the effect of interest is replicated (i.e., its direction is confirmed) may provide readers with more useful information concerning the believability and magnitude of the effect…"

Two group program

The SPSS syntax program will create an internal matrix table to assist users in determining the effects pertaining to the standardized mean difference and/or the proportion of variance in the dependent variable accounted for by the effect of the independent variable for two groups. The preponderance of the ensuing formulas are derived from Aaron, Ferron, and Kromrey (1998), Cohen (1988), Cohen and Cohen (1983), Cooper and Hedges (1994), and Richardson (1996).

The variance accounted for effect size measures include eta squared (η^2: Note equal to R^2 (Beasley & Schumacker, 1995)), which is known to be a positively-biased index, particularly with small sample sizes, and is defined as:

$$\eta^2 = \frac{d^2}{\left(d^2 + 4\right)} \tag{1}$$

where d = Cohen's d value.

Additionally, correction indices for η^2, such as omega squared (ω^2), epsilon squared (ε^2), and $R_{adj}{}^2$, all algebraically and theoretically-related measures (Cohen, 1988), are part of the program and formulated as:

$$\omega^2 = \frac{\left(t^2 - 1\right)}{\left(t^2 + N_1 + N_2 - 1\right)} \tag{2}$$

$$\varepsilon^2 = 1 - \left(1 - \eta^2\right) \times \frac{\left(N_1 + N_2 - 1\right)}{\left(N_1 + N_2 - 2\right)} \tag{3}$$

$$R_{adj}^2 = \eta^2 - \left(\left(1 - \eta^2\right) \times \left(\frac{2}{\left(N_1 + N_2 - 3\right)}\right)\right) \tag{4}$$

where t is the t value derived from the model as

$$d \times SQRT\left[\frac{\left(N_1 \times N_2\right)}{\left(N_1 + N_2\right)}\right] \tag{5}$$

M_1, SD_1, N_1 and M_2, SD_2, N_2 are the means, standard deviations, and sample sizes for Group 1 and Group 2, respectively.

Finally, Cohen's d is a measure of standardized mean difference and is defined as:

$$d = \frac{(M_1 - M_2)}{SQRT\left[\frac{(SD_1^2 + SD_2^2)}{2}\right]} \tag{6}$$

Note that Kraemer (1983) indicated the formula for *d* is optimal when both sample sizes are relatively equal and also large. Further, Cohen's *d* is recognized as a biased estimate (Hedges, 1981) and; thus, Hedges' g is a correction measure for this concern. It should be mentioned; however, that *d* and g are approximately equivalent when $n = 30$ (Hedges & Olkin, 1985). Hedges' g is defined as:

$$g = d \times \left(1 - \left(\frac{3}{(4 \times (N_1 + N_2) - 9)}\right)\right) \tag{7}$$

For the syntax program, the squared indices' estimated confidence intervals (CI) are set at 90% and based on the work of Cohen, Cohen, West, and Aiken (2003). For these estimated CIs, it is agreed that the sample size should be > 60, which, comparatively, assumes negligible error and; therefore, the absence of an adjustment for noncentrality. The error term for the approximated CI is defined as:

$$\begin{aligned} R^2E = {} & SQRT\left(4 \times \eta^2\right) \times \left(1 - \eta^2\right)^2 \times \frac{(N-1-1)^2}{(N^2-1)} \\ & \times (N+3) \ldots R^2 \pm R^2E \times 1.645 \end{aligned} \tag{8}$$

For the standardized mean difference CIs, these are set at 95%. The program's estimated CI formula is based on previous research by Grissom and Kim (2005), Hedges and Olkin (1985), and Steiger (2004). Bird (2002) found that if *d* is < 2.00, which in social science research frequently can be the circumstance with middling-sized effects (Richard, Bond, & Stokes-Zoota, 2003; Rosnow & Rosenthal, 2003), adjustment for noncentrality is not compulsory. The error term for this approximated CI is defined as:

$$d1; g1 = \frac{N}{(N_1 \times N_2)} + \frac{d^2}{(2 \times N)} \ldots d \pm d1 \times 1.96 \tag{9}$$

Note: For any CI within the program, the user can alter it by changing the *Z* value within the syntax, for example, to values such as 1.28 (80% CI), 1.645 (90% CI), 1.96 (95% CI), or 2.58 (99% CI), where Cohen (1990, p. 1310) observed "I don't think that we should routinely use 95% intervals: Our interests are often better served by more tolerant 80% intervals."

Results

As seen in Appendix A, the user would put the two-group descriptive data (M_1, SD_1, N_1 for Group 1 and M_2, SD_2, N_2 for Group 2) in the space between BEGIN DATA and END DATA along with the total sample size (*N*). Thus, these descriptive data in the example from the program are, in group order, 16.45 2.23 30 11.77 4.66 34 64 and represent continuous data for the dependent variable (Depression Score) and categories for the independent variable Group (i.e., Group 1 [Treatment] and Group 2 [Control]).

Once the program is run, the results show that the matrix produced will cluster the effect sizes by the categories noted previously: standardized means difference, squared index, and corrected squared indices. Additionally, the matrix generates an overall model *post-hoc* power value, which is predicated on alpha established at .05 and the particular sample sizes for Group 1 and Group 2.

As can be seen in the results from Table 1, the standardized mean difference effect size for Cohen's *d* was 1.256 or a "large" effect of over one standard deviation difference in Depression Score between Group 1 and Group 2 with 95% CI at (1.109, 1.403) and overall model power = .999, where power ≥ .80 is desired in social science research (Nunnally, 1978). The correction for Cohen's *d*, Hedges' *g*, was very comparable in value at 1.241 (1.094, 1.387).

Table 1. Standardized Mean Difference, Confidence Intervals, and Model Post-Hoc Power.

Cohen's *d*	$95\%_{CIL}$	$95\%_{CIU}$	Hedges' g	$95\%_{CIL}$	$95\%_{CIU}$	Power
1.256	1.109	1.403	1.241	1.094	1.387	0.999

Note. CI = Confidence Interval; L = Lower; U = Upper.

For the squared and corrected squared indices, the results in Table 2 indicated that the proportion of variance in the dependent variable accounted for by the effect of the independent variable was "large" overall for all of the various indices. As would be expected, these effect size measures ranged from a low of

25.9% for the correction R_{adj}^2 (90% CI .089, .430) to a high of 28.3% for the non-corrected η^2 (90% CI .107, .458).

Table 2. Proportion of Variance in the DV Accounted for by the Effect of the IV and Confidence Intervals.

η^2	$90\%_{CIL}$	$90\%_{CIU}$	R_{adj}^2	$90\%_{CIL}$	$90\%_{CIU}$	ω^2	$90\%_{CIL}$	$90\%_{CIU}$	ε^2	$90\%_{CIL}$	$90\%_{CIU}$
.283	.107	.458	.259	.089	.430	.274	.100	.448	.271	.098	.444

References

Aaron, B., Kromrey, J. D., & Ferron, J. M. (1998, November). *Equating r-based and d-based effect size indices: Problems with a commonly recommended formula*. Paper presented at the annual meeting of the Florida Educational Research Association, Orlando, FL.

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher, 35*, 33-40.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Beasley, T. M., & Schumacker, R. E. (1995). Multiple regression approach to analyzing contingency tables: Post hoc and planned comparison procedures. *Journal of Experimental Education, 64*, 79-93.

Bird, K. D. (2002). Confidence intervals for effect sizes in analysis of variance. *Educational and Psychological Measurement, 62*, 197–226. doi:10.1177/0013164402062002001

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

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*, 1304-1312.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

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

Cohen, J., & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah: NJ: Erlbaum.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers, *Professional Psychology: Research and Practice*, *40*(5), 532-538. doi:10.1037/a0015808

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. New York: Psychology Press.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107-128.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego: Academic Press.

Kraemer, H. C. (1983). Theory of estimation and testing of effect sizes: Use in meta-analysis. *Journal of Educational Statistics*, *8*, 93-101.

Levin, J. R., & Robinson, D. H. (2003). The trouble with interpreting statistically nonsignificant effect sizes in single-study investigations. *Journal of Modern Applied Statistical Methods*, *2*(1), 231-236. Available at http://digitalcommons.wayne.edu/jmasm/vol2/iss1/23/

Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research*, *28*(4), 612-625. doi:10.1111/j.1468-2958.2002.tb00828.x

Nickerson, R. S. (2000). Null hypothesis significance testing: A Review of an old and continuing controversy. *Psychological Methods*, *5*(2), 241-301. doi:10.1037/1082-989X.5.2.241

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*(4), 331-363. doi:10.1037/1089-2680.7.4.331

Richardson, J. T. E. (1996). Measures of effect size. *Behavior Research Methods, Instruments, & Computers*, *28*, 12-22.

Rosnow, R., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology, 57*(3), 221–237. doi:10.1037/h0087427

Sapp, M. (2004). Confidence intervals within hypnosis research. *Sleep and Hypnosis, 6*(4), 169–176.

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*(2), 164–182. doi:10.1037/1082-989X.9.2.164

Thompson, B. (1998). Statistical significance and effect size reporting: Portrait of a possible future. *Research in the Schools, 5*(2), 33-38.

Thompson, B. (2009). A brief primer on effect sizes. *Journal of Teaching in Physical Education, 28*, 251-254.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 51*(4), 473-481. doi:10.1037/0022-0167.51.4.473

Walker, D. A. (2015). Two group program for Cohen's *d*, Hedges' g, η^2, R_{adj}^2, ω^2, ε^2, confidence intervals, and power. [Computer program]. DeKalb, IL: Author.

Wilkinson, L., & The APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594-604.

Appendix A: SPSS syntax two group program for Cohen's *d*, Hedges' g, η^2, R_{adj}^2, ω^2, ε^2, confidence intervals, and power.

```
DATA LIST LIST /M1 SD1 (2F9.3) N1 (F8.0) M2 SD2 (2F9.3) N2 N (2F8.0).
***********************************************************************
Put your two-group data (M1, SD1, N1 for Group 1 and M2, SD2, N2 for Group 2) in
the space between BEGIN DATA and END DATA along with the total sample size (N)
***********************************************************************.
BEGIN DATA
16.45 2.23 30 11.77 4.66 34 64
END DATA.
COMPUTE POOLD = ((N1-1)*(SD1**2)+(N2-1)*(SD2**2))/((N1+N2)-2).
COMPUTE COHEND = ABS((M1-M2)/SQRT(POOLD)).
COMPUTE D1 = N/(N1*N2) + COHEND**2/(2*N).
COMPUTE HEDGESG = COHEND*(1-(3/(4*(N1 + N2)-9))).
COMPUTE G1 = N/(N1*N2) + HEDGESG**2/(2*N).
COMPUTE CRITICAL = 0.05.
COMPUTE K = 1.
COMPUTE H = (2*N1*N2)/(N1+N2).
COMPUTE NCP = ABS((COHEND*SQRT(H))/SQRT(2)).
COMPUTE ALPHA = IDF.T(1-CRITICAL/2,N1+N2-2).
COMPUTE POWER1 = 1-NCDF.T(ALPHA,N1+N2-2,NCP).
COMPUTE POWER2 = 1-NCDF.T(ALPHA,N1+N2-2,-NCP).
COMPUTE B = POWER1 + POWER2.
COMPUTE ETA2 = COHEND**2/(COHEND ** 2 + 4).
COMPUTE EPSILON = 1-(1-ETA2) * (N1  +  N2-1) / (N1  +  N2-2).
COMPUTE TTEST = COHEND  * SQRT((N1  *  N2) /(N1  +  N2)).
COMPUTE OMEGA = (TTEST**2-1)/(TTEST**2 + N1 + N2 -1).
COMPUTE SEETA1 = (1-ETA2)/SQRT(N1 + N2-1).
COMPUTE SEETA2 = 2/(N1 + N2 - 2).
COMPUTE SEETA3 = SQRT(SEETA2 + 4*ETA2).
COMPUTE SEETA = SEETA1 * SEETA3.
COMPUTE TTEST = COHEND  * SQRT((N1  *  N2) /(N1  +  N2)).
COMPUTE ADJR2 = ETA2 - ((1-ETA2)*(2/(N1 + N2 -3))).
COMPUTE ADJR2A = (((4*ADJR2)*(1-ADJR2)*(N-K-1)**2)).
COMPUTE ADJR2B = (N**2-1)*(N+3).
COMPUTE ADJR2C = ADJR2A/ADJR2B.
```

```
COMPUTE ADJR21 = SQRT(ADJR2C).
*********************************************************************
NOTE: Confidence Intervals can be altered below by changing the Z = value to
   either 1.96 = (95%) or 2.58 = (99%) For the squared indices, they are at 90%
*********************************************************************.
COMPUTE Z = 1.645.
COMPUTE ADJR2L = (ADJR2-(Z*ADJR21)).
COMPUTE ADJR2H = (ADJR2+(Z*ADJR21)).
COMPUTE OMEGA = (TTEST**2-1)/(TTEST**2 + N1 + N2 -1).
COMPUTE SEE1 = (1-EPSILON)/SQRT(N1  +  N2-1).
COMPUTE SEE2 = 2/(N1 + N2 - 2).
COMPUTE SEE3 = SQRT(SEE2 + 4*EPSILON).
COMPUTE SEEPSILON = SEE1 * SEE3.
COMPUTE SEO1 = (1-OMEGA)/SQRT(N1  +  N2-1).
COMPUTE SEO2 = 2/(N1 + N2 - 2).
COMPUTE SEO3 = SQRT(SEO2 + 4*OMEGA).
COMPUTE SEOMEGA = SEO1 * SEO3.
COMPUTE ETAA = (((4*ETA2)*(1-ETA2)*(N-K-1)**2)).
COMPUTE ETAB = (N**2-1)*(N+3).
COMPUTE ETAC = ETAA/ETAB.
COMPUTE ETA1 = SQRT(ETAC).
COMPUTE ETAL = (ETA2-(Z*ETA1)).
COMPUTE ETAH = (ETA2+(Z*ETA1)).
COMPUTE OMEGAA = (((4*OMEGA)*(1-OMEGA)*(N-K-1)**2)).
COMPUTE OMEGAB = (N**2-1)*(N+3).
COMPUTE OMEGAC = OMEGAA/OMEGAB.
COMPUTE OMEGA1 = SQRT(OMEGAC).
COMPUTE OMEGAL = (OMEGA-(Z*OMEGA1)).
COMPUTE OMEGAH = (OMEGA+(Z*OMEGA1)).
COMPUTE EPSILONA = (((4*EPSILON)*(1-EPSILON)*(N-K-1)**2)).
COMPUTE EPSILONB = (N**2-1)*(N+3).
COMPUTE EPSILONC = EPSILONA/EPSILONB.
COMPUTE EPSILON1 = SQRT(EPSILONC).
COMPUTE EPSILONL = (EPSILON-(Z*EPSILON1)).
COMPUTE EPSILONH = (EPSILON+(Z*EPSILON1)).
*********************************************************************
NOTE: Confidence Intervals for Cohen's d are at 95%
*********************************************************************.
```

```
COMPUTE Z = 1.96.
COMPUTE GH = (HEDGESG+(G1*Z)).
COMPUTE GL = (HEDGESG-(G1*Z)).
COMPUTE DH = (COHEND+(D1*Z)).
COMPUTE DL = (COHEND-(D1*Z)).
EXECUTE.
FORMAT POOLD to DL (F9.3).
VARIABLE LABELS COHEND 'Cohens d'/B 'Power'/ETA2 'Eta Squared'/OMEGA 'Omega
    Squared'/EPSILONL '90% CI Lower'/
EPSILONH '90% CI Upper'/OMEGAL '90% CI Lower'/ OMEGAH '90% CI Upper'/ETAL '90%
    CI Lower'/ADJR2L '90% CI Lower'/
GL '95% CI Lower'/ GH '95% CI Upper'/HEDGESG 'Hedges g'/ADJR2H '90% CI
    Upper'/ADJR2 'Adjusted R2'/DL '95% CI Lower'/
DH '95% CI Upper'/ETAH '90% CI Upper'/EPSILON 'Epsilon Squared'/.
REPORT FORMAT=LIST AUTOMATIC ALIGN(CENTER)
  /VARIABLES= COHEND DL DH HEDGESG GL GH B
  /TITLE "Standardized Mean Difference, Confidence Intervals, and Model Post-Hoc
    Power".
REPORT FORMAT=LIST AUTOMATIC ALIGN(LEFT)
MARGINS (*,150)
  /VARIABLES= ETA2 ETAL ETAH ADJR2 ADJR2L ADJR2H OMEGA OMEGAL OMEGAH EPSILON
    EPSILONL EPSILONH
  /TITLE "Proportion of Variance in the DV Accounted for by the Effect of the IV
    and Confidence Intervals".
```