DOCUMENT RESUME

ED 253 567 TM 850 107

AUTHOR Reynolds, Sharon; Day, Jim
TITLE Monte Carlo Studies of Effect Size Estimates and Their Approximations in Meta-Analysis.
PUB DATE 24 Aug 84
NOTE 19p.; Paper presented at the Annual Meeting of the American Psychological Association (Toronto, ON, August 24-28, 1984).
PUB TYPE Speeches/Conterence Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Computer Simulation; *Effect Size; *Estimation (Mathematics); *Meta Analysis; *Monte Carlo Methods; Research Problems; Sampling; Statistical Studies

ABSTRACT

Monte Carlo studies explored the sampling characteristics of Cohen's d and three approximations to Cohen's d when used as average effect size measures in meta-analysis. Reviews of 10, 100, and 500 studies (M) were simulated, with degrees of freedom (df) varied in seven steps from 8 to 58. In a two independent groups design, samples were obtained from populations whose mean differences represented a zero, small, medium, or large effect size. One thousand replications of studies within each of the 84 combinations of effect size, df, and M were conducted, and a mean and standard error were obtained for each combination of conditions. As expected, d was a positively biased estimator of effect size, overestimating by as much as 13 percent even with Hedges' correction factor. Surprisingly, the most unbiased estimator of effect size and the highest relative efficiency was obtained with the approximation to d computed from the obtained t and corrected according to Hedges. The approximations to d from the nonparametric statistic and the obtained conventional significance levels were not consistent estimators of effect size. These simulation results suggest that additional study of the behavior of effect size estimators should precede a more widespread application of meta-analysis. (Author)

Monte Carlo Studies of Effect Size Estimates and Their Approximations in Meta-Analysis

Sharon Reynolds

Texas Christian University

and

Jim Day

Texas Woman's University

Paper presented at the Annual Convention of the American Psychological Association, Toronto, CAN, August 24, 1984

Abstract

Monte carlo studies explored the sampling characteristics of Cohen's d and three approximations to Cohen's d when used as average effect size measures in meta-analysis. Reviews of 10, 100, and 500 studies (M) were simulated, with degrees of freedom (df) varied in seven steps from 8 to 58. In a two independent groups design, samples were obtained from populations whose mean differences represented a zero, small, medium, or large effect size (Δ). One thousand replications of studies within each of the 84 combinations of Δ, df, and M were conducted, and a mean and standard error were obtained for each combination of conditions. As expected, d was a positively biased estimator of Δ, overestimating Δ by as much as 13% even with Hedges' correction factor. Surprisingly, the most unbiased estimator of Δ and the highest relative efficiency was obtained with the approximation to d computed from the obtained t and corrected according to Hedges. The approximations to d from the nonparametric statistic and the obtained conventional significance levels were not consistent estimators of Δ. The results of these simulations suggest that additional study of the behavior of effect size estimators should precede a more widespread application of meta-analysis.

Introduction

Recent scientific endeavor is characterized by an exponential increase in the number of studies constituting a body of literature. Since the data items of such a literature are often disparately published and contain inconsistent findings, the extrication of relationships among explanatory variables is often difficult to achieve. The integrative summary of the literature has consequently emerged as a more important facet of the scientific process. Recently, methods of secondary analysis, including cluster and probability approaches, and meta-analytic methods have been proposed to add a quantitative dimension to such literature reviews.

The most common meta-analytic approach currently utilized seems to be the one proposed by Glass (1976) in which an average effect size is computed across studies in a variable domain. The most widely used estimate of effect size is the standardized mean difference, d, where $d = (\bar{X}_e - \bar{X}_c)/S_c$, described by Cohen in 1969. Hedges (1981), however, has formally demonstrated that d is a biased estimate of Δ, where $\Delta = (\mu_e - \mu_c)\sigma_c$ and represents the true effect size. Hedges proposed a correction factor, $1-(3/(4 * df) -1)$, which, when multiplied by each d before averaging, corrects the bias.

Since the means and standard deviations necessary to compute d are often not available to the reviewer, approximations to d based on parametric and nonparametric statistics and reported significance levels have been suggested. Before the widespread adoption of these estimates of d occurs, however, their sampling characteristics should be understood. Thus the purpose to the Monte Carlo studies reported here was to empirically examine the behavior of d and the approximations of d, and their corrections according to Hedges, as estimators of Δ.

Method

The studies simulated an experimental situation with two independent groups represented as two populations consisting of normally distributed, computer-generated random numbers having a variance of unity. The population designated as simulating the control group had a mean of 10.0; the experimental populations had means of 10.0, 10.2, 10.5, and 10.8, the first simulating a situation in which the null hypothesis is true and the other three corresponding to effect sizes that were considered to represent small, medium, and large effects, respectively. To examine the influence of the number of effect sizes entering the analysis, reviews were simulated with 10, 100, and 500 studies (M). In order to determine the effects of varying degrees of freedom (df), each review was conducted on equal-n "studies" with sample sizes of 5, 10, 20, and 30 and unequal-n studies simulated by the pairs 10-15, 10-20, and 10-25.

A FORTRAN program generated 1000 reviews for each unique combination of , M and df. For each study within each review, the following statistics were computed: d, a parametric approximation to d, $d1$, a nonparametric approximation $d2$, and an approximation based on significance level, $d3$. The approximations were computed as $t\sqrt{1/n_1 + 1/n_2}$, where t was the value of the obtained t, the value of t at the significance level associated with the obtained U, and where t was the value of t associated with the conventional level of significance exceeded by the obtained t (.05 or .01) respectively. In addition, each of these effect size estimates was multiplied by Hedges' correction factor to produce d', $d1'$, $d2'$, and $d3'$, respectively. A mean corrected and uncorrected estimated effect size was computed for each review, and, subsequently means and standard erors were determined for the 1000 reviews within each of the 84 combinations of conditions.

Results

As expected from the work of Hedges, for all positive values of Δ, d overestimated Δ, and the magnitude of this bias was directly proportional to the value of Δ (see Tables). Although the bias of the corrected d was less than that of d, d' was still positively baised; in fact, d' overestimated Δ by as much as 13% for large values of Δ and small values of M and df. Surprisingly, the parametric approximation to d was a less biased estimator of Δ than d, d1' deviating no more than 1.5% from Δ in any condition. Both d2 and d3 overestimated Δ for Δ = 0.0 and 0.2 and underestimated Δ for effect sizes of 0.5 and 0.8. The magnitude of the bias of these latter two approximations was very large for the very small and very large effect sizes, and Hedges' correction factor exacerbated this bias for both measures. Although d and its parametric approximation showed consistency across df and M, neither d2 nor d3 appeared to be a consistent estimator of Δ. All estimators showed very poor relative efficiency for small values of df and M, and in all cases, the corrected estimate had only a slightly smaller standard error than the uncorrected one. The parametric approximation, furthermore, was a more efficient estimator than d.

Conclusions

Several conclusions seem warranted by these findings. (a) In the long run, the use of a mean _d_, corrected or otherwise, in integrative literature reviews will overestimate the magnitude of the relationship between explanatory variables. (b) The corrected parametric approximation to _d_ is the least biased of the heretofore suggested estimators of the true effect size. (c) The nonparametric approximation to _d_ and the approximation based on the achieved conventional significance level are not consistent estimators of the true effect size and thus have less utility than the other effect size estimators. (d) Due to the poor efficiency of all the estimators, the meta-analyst should exercise extreme caution when reviewing a small number of studies characterized by small _df_. In general, the results of these simulations suggest that additional study of the behavior of effect size estimates should precede a more widespread application of meta-analysis.

Deviations of dl' from Δ = .0

DF	M=10	M=100	M=500
8	0.0057	0.0034	0.0017
18	0.0023	0.0020	-0.0005
23	-0.0010	-0.0008	-0.0001
28	0.0015	-0.0014	-0.0010
33	0.0001	0.0011	0.0005
38	-0.0019	0.0001	-0.0002
58	0.0029	0.0012	0.0001

Deviations of d1' from Δ = .2

DF	M=10	M=100	M=500
8	0.0042	0.0032	0.0018
18	0.0018	0.0020	-0.0004
23	-0.0012	-0.0008	0.0000
28	0.0015	-0.0014	-0.0010
33	0.0000	0.0011	0.0006
38	-0.0019	0.0001	-0.0002
58	0.0029	0.0012	0.0001

Deviations of dl' from Δ = .5

DF	M=10	M=100	M=500
8	0.0021	0.0028	0.0019
18	0.0010	0.0018	-0.0003
23	-0.0016	-0.0010	0.0001
28	-0.0014	-0.0014	-0.0005
33	-0.0001	0.0012	0.0005
38	-0.0019	0.0001	-0.0001
58	0.0029	0.0012	0.0002

Deviations of dl' from Δ = .8

DF	M=10	M=100	M=500
8	0.0000	0.0024	0.0020
18	0.0003	0.0017	-0.0003
23	0.0019	-0.0011	0.0002
28	0.0014	-0.0014	-0.0008
33	-0.0003	0.0012	0.0007
38	-0.0019	0.0000	0.0000
58	0.0029	0.0012	0.0002

Deviations of d2' from $\underline{\Lambda}$ = .0

DF	M=10	M=100	M=500
8	0.3601	0.3668	0.3688
18	0.2910	0.2887	0.2877
23	0.2673	0.2697	0.2704
28	0.2551	0.2543	0.2525
33	0.2495	0.2476	0.2471
38	0.2151	0.2130	0.2128
58	0.1752	0.176	0.1763

Deviations of d2' from λ = .5

DF	M=10	M=100	M=500
8	-0.0548	-0.0517	-0.0509
18	-0.1149	-0.1147	-0.1155
23	-0.1309	-0.1298	-0.1302
28	-0.1415	-0.1419	-0.1420
33	-0.1446	-0.1457	-0.1458
38	-0.1665	-0.1642	-0.1650
58	-0.1796	-0.1809	-0.1812

Deviations of d2' from $\Delta = .8$

DF	M=10	M=100	M=500
8	-0.2438	-0.2422	-0.2415
18	-0.2762	-0.2791	-0.2805
23	-0.2950	-0.2937	-0.2932
28	-0.3009	-0.2997	-0.2932
33	-0.3039	-0.3008	-0.3009
38	-0.3104	-0.3086	-0.3087
58	-0.3169	-0.3172	-0.3178

Deviations of d3' from $\Delta = .0$

DF	M=10	M=100	M=500
8	0.3067	0.8124	0.8168
18	0.2475	0.5850	0.5880
23	0.2102	0.5330	0.5381
28	0.2041	0.5084	0.5120
33	0.1990	0.4924	0.4948
38	0.1763	0.4155	0.4186
58	0.1372	0.3406	0.3423

Deviations of d3' from Δ = .2

DF	M=10	M=100	M=500
8	0.1542	0.6172	0.6194
18	0.0977	0.3915	0.3924
23	0.0925	0.3434	0.3431
28	0.0981	0.3168	0.3163
33	0.0918	0.3001	0.3000
38	0.0675	0.2239	0.2237
58	0.0554	0.1484	0.1484

Deviations of d3' from $\Delta = .5$

DF	M=10	M=100	M=500
8	0.0002	0.3286	0.3285
18	0.1099	0.1030	0.1029
23	0.0055	0.0547	0.0543
28	0.0285	0.0285	0.0281
33	-0.0104	0.0124	0.0122
38	-0.0697	-0.0632	-0.0632
58	-0.1369	-0.1374	-0.1375

Deviations of d3' from λ = .8

DF	M=10	M=100	M=500
8	-0.2378	0.0375	0.3382
18	-0.1120	-0.1835	-0.1837
23	-0.0677	-0.1835	-0.2316
28	-0.0445	-0.2569	-0.2569
33	-0.0271	-0.2726	-0.2726
38	-0.0467	-0.3467	-0.3467
58	-0.1203	-0.4203	-0.4204