The Standardized Difference Between Means: Much Variance In Notation. Also, Differences Between *d* and r_{pb} as Effect Size Estimates

I use "d" to stand for the population parameter which is the difference between means divided by the two populations common standard deviation. I use "*d*" to stand for the least squares estimator of d. Since most psychologists use "*d*" to stand for this estimator, I also encourage the use of "$\hat{d}$" for this estimator.

Unfortunately, the notation used by statisticians for the parameter and the estimator varies greatly, as noted by McGrath and Meyer (2006). They recommend using "δ" for the parameter, "*d*" for the maximum likelihood estimate, and "*g*" for the least squares estimate. They define the pooled standard deviation differently than do I, but their method of computing *g*, although it looks different from mine, produces values identical to those using my method. McGrath and Meyer also provide an unbiased estimator of the parameter, and also point out that the differences between the least squares estimate and the maximum likelihood estimate and the unbiased estimate are generally trivial.

The "*d*" that is commonly reported in the literature is really "*g*." I used to use "*g*," but switched to "*d*" since almost nobody knows that the proper symbol is "*g*."

An alternate measure of the size of the difference between two means is the point biserial correlation coefficient. McGrath and Meyer point out that the value of the r_{pb} is affected by the ratio of the sample sizes, n_1/n_2, but the value of *g* is not. For a fixed difference between means, the absolute value of the point biserial decreases and the ratio n_1/n_2 moves away from one.

McGrath and Meyer also note that Cohen's benchmarks for small, medium and large effects for *d* are not comparable to those for ρ (the point biserial correlation in the population. For example, when $n_1 = n_2$, a *d* of .8 (large) corresponds to a ρ_{pb} of .37, but the benchmark for "large" with ρ is .5. When the sample sizes differ, the difference between benchmark-implied size of effect for *d* and ρ increases. With great differences between n_1 and n_2, the value of *d* may indicate a large effect while the value of ρ_{pb} would indicate a trivial effect.

McGrath and Meyer discuss the relative merits of *g* and r_{pb} as estimates of effect size. Their discussion is very interested, but I shall not take time to summarize it well here. I generally prefer *d* over r_{pb}, which may reflect my training in experimental rather than nonexperimental research methods. I will note that when the goal is to estimate what proportion of the variance in an existing population is accounted for by the dichotomous predictor variable, the squared point biserial correlation is clearly the estimate of choice – but for it to be relatively unbiased the ratio of n_1 to n_2 must be the same as what it is in the population of interest. When researchers use a sample n_1/n_2 ratio closer to one than it is in the population, they increase power to detect an effect but at the same time increase bias in r_{pb} as an estimator of the size of the effect.

Check out the calculator at https://www.socscistatistics.com/effectsize/default3.aspx . It will compute the maximum likelihood *d* and the least squares *g*. Remember that *g* is what most people call *d*.

Group 1		*Group 2*	
Mean (*M*):	10	Mean (*M*):	11
Standard deviation (*s*):	1	Standard deviation (*s*):	1
Sample size (*n*):	10	Sample size (*n*):	10

Calculate Reset

Success!

Cohen's *d* = (11 - 10)/1 = 1.

Glass's *delta* = (11 - 10)/1 = 1.

Hedges' *g* = (11 - 10)/1 = 1.

Equal sample sizes and equal variances, *d* = *g*.

Group 1		*Group 2*	
Mean (*M*):	10	Mean (*M*):	11
Standard deviation (*s*):	2	Standard deviation (*s*):	.5
Sample size (*n*):	10	Sample size (*n*):	10

Calculate Reset

Success!

Cohen's *d* = (11 - 10)/1.457738 = 0.685994.

Glass's *delta* = (11 - 10)/2 = 0.5.

Hedges' *g* = (11 - 10)/1.457738 = 0.685994.

Equal sample sizes, disparate variances, *d* = *g*.

Group 1		*Group 2*	
Mean (*M*):	10	Mean (*M*):	11
Standard deviation (*s*):	1	Standard deviation (*s*):	1
Sample size (*n*):	15	Sample size (*n*):	5

Calculate Reset

Success!

Cohen's *d* = (11 - 10)/1 = 1.

Glass's *delta* = (11 - 10)/1 = 1.

Hedges' *g* = (11 - 10)/1 = 1.

Disparate sample sizes, equal variances, $d = g$.

Group 1		*Group 2*	
Mean (*M*):	10	Mean (*M*):	11
Standard deviation (*s*):	.5	Standard deviation (*s*):	1.5
Sample size (*n*):	5	Sample size (*n*):	15

Calculate Reset

Success!

Cohen's *d* = (11 - 10)/1.118034 = 0.894427.

Glass's *delta* = (11 - 10)/0.5 = 2.

Hedges' *g* = (11 - 10)/1.34371 = 0.744208.

Disparate sample sizes and variances. Variance higher in the group with the larger sample size. For computing the standardizer for g, the larger sample size in Group 2 will give it greater influence. $g < d$.

Group 1		Group 2	
Mean (*M*):	10	Mean (*M*):	11
Standard deviation (*s*):	1.5	Standard deviation (*s*):	.5
Sample size (*n*):	5	Sample size (*n*):	15

Calculate Reset

Success!

Cohen's *d* = (11 - 10)/1.118034 = 0.894427.

Glass's *delta* = (11 - 10)/1.5 = 0.666667.

Hedges' *g* = (11 - 10)/0.833333 = 1.2.

Variance smaller in the group with more scores. $g > d$.

With d, the standardizer is $s_{pooled} = \sqrt{.5(s_1^2 + s_2^2)}$. It is not affected by disparity in sample sizes. If you are confident that the variances in the two populations are identical, then it makes sense to weight the two standard deviations equally even if the sample sizes happen to differ.

With g, the standardizer is , $s_{pooled} = \sqrt{\Sigma(p_j s_j^2)}$, where for each group s^2_j is the within-group variance and $p_j = \frac{n_j}{N}$, the proportion of the total number of scores (in both groups, N) which are in that group (n_j).

Reference

McGrath, R. E. & Meyer, G. J. (2006). When effect sizes disagree: The case of *r* and *d*. *Psychological Methods*, *11*, 386-401.

Karl L. Wuensch, East Carolina University, November, 2020.

Return to Wuensch's Stat Help Page