Approximating the distribution of Cohen's d_p in within-subject designs

Denis Cousineau [a] ✉ ⓘ

[a]Université d'Ottawa

Abstract ■ In this document, I demonstrate an approximate distribution of Cohen's d_p in within-subject designs. The distribution follows a noncentral t distribution with degrees of freedom depending on the correlation between the measures. The result generalizes the distribution of Cohen's d_p to both between-subject and single-group designs, yielding a flexible and integrative measure for comparing effect sizes across different study designs.

Keywords ■ Standardized mean difference, Cohen's d_p, noncentral t distribution.

✉   denis.cousineau@uottawa.ca

Introduction

The standardized mean difference is a convenient measure to describe differences between two groups. Proposed by Jacob Cohen in the late sixties (Cohen, 1969), its most common version is called d_p and is given by

$$d_p = \frac{M_1 - M_2}{S_p} \tag{1}$$

in which M_1 and M_2 are the two group means and S_p is the pooled standard deviation.

Hedges (1981) provided the distribution of the d_p measure in between-subject designs, a noncentral t distribution. He also demonstrated that d_p is a biased statistic which can be unbiased using a correction term (called c in Hedges, 1981, and J in Goulet-Pelletier & Cousineau, 2018; when unbiased, it is recommended to call the statistic Hedges' g_p). From this distribution, it was possible to derive exact confidence intervals (Steiger & Fouladi, 1997; Lecoutre, 2007) as well as pseudo confidence intervals (see Cousineau & Goulet-Pelletier, 2020; Viechbauer, 2007, for reviews).

Whereas this statistic has a definite and definitive solution for the between-subject design, this is not the case in within-subject designs for which the exact distribution of the statistic is unknown. Becker (1988) provided the exact distribution for a related statistic, $d_{\mathbf{D}}$. This second statistic is relative to the standard deviation of the *difference* between scores. The two standardized difference scores, d_p

and $d_{\mathbf{D}}$, are not on the same scale and as a consequence, cannot be compared directly.

Herein, I provide an approximate distribution for d_p in within-subject designs:

$$d_p \approx \sqrt{\frac{2(1-\rho)}{n}} \times t'_\nu(\lambda) \tag{2}$$

where t' is the noncentral t distribution, $\nu = 2(n-1)/(1+\rho^2)$ are the degrees of freedom, ρ is the correlation between the measures, and $\lambda = \sqrt{\frac{n}{2(1-\rho)}} \times \frac{\Delta}{\sigma}$ is the noncentrality parameter which depends on the difference between the population means (here noted Δ) and on the population standard deviation (here noted σ). The solution is similar to the one found by Becker (1988); both use the noncentral t distribution except that here the degrees of freedom are fractional between $1 \times (n-1)$ and $2 \times (n-1)$ as a function of the correlation between the repeated measures. Also, when ρ is null, as in between-subject designs, the solution is identical to the one reported by Hedges (1981).

In what follows, it is assumed that the measures are from a bivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the folowing parameters: mean vector $\boldsymbol{\mu} = \{\mu_{\mathbf{X}}, \ \mu_{\mathbf{Y}}\}$ with $\Delta = \mu_{\mathbf{X}} - \mu_{\mathbf{Y}}$ the separation between the two means. The demonstration assumes that the variances are homogeneous so that $\boldsymbol{\Sigma}$ reduces to $\begin{pmatrix} \sigma^2 & \rho\,\sigma^2 \\ \rho\,\sigma^2 & \sigma^2 \end{pmatrix}$ in which $\rho \equiv \rho_{\mathbf{XY}}$ is the population correlation between the two measurements labeled $\mathbf{X}$ and $\mathbf{Y}$.

The following relation $\sigma_\mathbf{D}^2 = \sigma_\mathbf{X}^2 + \sigma_\mathbf{Y}^2 - 2\sigma_\mathbf{X}\sigma_\mathbf{Y}\rho$ is used to derive the variance of the difference $\mathbf{D} = \mathbf{X} - \mathbf{Y}$ (Kendall & Stuart, 1977). It is based on the population variance of the scores ($\sigma_\mathbf{X}^2$ and $\sigma_\mathbf{Y}^2$) as well as the pairwise correlation between these scores (ρ). When variances are homogeneous in the population (i.e., $\sigma_\mathbf{X}^2 = \sigma_\mathbf{Y}^2 = \sigma^2$), it simplifies to $\sigma_\mathbf{D}^2 = 2(1-\rho)\,\sigma^2$ and therefore the standard error of the difference is $\sqrt{2(1-\rho)}\,\sigma/\sqrt{n}$.

Demonstration of the result

Let us define $\overline{D} = \overline{X} - \overline{Y}$, the mean difference between the two repeated measurements, typically a pretest and a post-test. The pooled variance is defined as $S_p^2 = (S_\mathbf{X}^2 + S_\mathbf{Y}^2)/2$; as the two sets of measurements have the same number of observations n, the variance simplifies to the mean of the variances on measurements $\mathbf{X}$ and $\mathbf{Y}$.

Hereafter, $\mathcal{N}(\mu, \sigma)$ is used to denote a random variable which follows a normal distribution with mean μ and variance σ^2 and χ_ν^2 to denote a random variable following a chi-square distribution with degrees of freedom ν.

The distribution of $d_p = (\overline{X} - \overline{Y})/S_p$ is then given by

$$Pr(d_p < d) = Pr\left(\frac{\overline{D}}{S_p} < d\right) \tag{3a}$$

$$= Pr\left(\frac{\mathcal{N}\left(\Delta, \frac{\sqrt{2(1-\rho)}}{\sqrt{n}}\,\sigma\right)}{\sqrt{\sigma^2\,\chi_\nu^2/\nu}} < d\right) \tag{3b}$$

$$= Pr\left(\sqrt{\frac{2(1-\rho)}{n}} \times \frac{\mathcal{N}\left(\sqrt{\frac{n}{2(1-\rho)}}\frac{\Delta}{\sigma}, 1\right)}{\sqrt{\chi_\nu^2/\nu}} < d\right) \tag{3c}$$

$$= Pr\left(\sqrt{\frac{2(1-\rho)}{n}} \times t_\nu'\left(\sqrt{\frac{n}{2(1-\rho)}}\frac{\Delta}{\sigma}\right) < d\right) \tag{3d}$$

where t' is a noncentral t distribution with degrees of freedom $2(n-1)/(1+\rho^2)$ and noncentrality parameter $\sqrt{n/(2(1-\rho))}\,\Delta/\sigma$.

On Step (3b), I used the fact that the standard error of the difference from a pair of correlated means is given by $\sqrt{2(1-\rho)}\,\sigma/\sqrt{n}$, as indicated earlier. I also used the fact that the distribution of the pooled variance follows approximately $\sigma^2 \times \chi_\nu^2/\nu$ where $\nu = 2(n-1)/(1+\rho^2)$ (Ben, 2020). From Allaire, reported in Laurencelle (2016), it was found that the two variances from correlated bivariate data are also correlated (if the correlation at the data level is ρ, then the correlation of their variances is ρ^2).

Step (3d) follows from the definition of a noncentral t

distribution. This completes the demonstration.

An illustration

To illustrate the distribution, I generated simulated Cohen's d_p from the following: The population is bivariate normal with means $-\Delta/2$ and $+\Delta/2$, a common variance for both scores of σ^2 and a correlation of ρ. I chose the values $\Delta = 15, \sigma = 15$ and $\rho = 0.50$ along with samples of size $n = 10$. The population Cohen's d_p from these parameters is $\Delta/\sigma = 1$. From this simulated sample, I computed d_p as the difference in observed means onto the pooled standard deviation. I replicated this process five million times.

The distribution of simulated d_p is shown in Figure 1 along with the theoretical distribution (Eq. 2). As seen, even for such small samples, the fit is excellent.

Discussion

The demonstration provides an approximate distribution of Cohen's d_p in within-subject design. The exact degrees of freedom depend on the population correlation ρ^2, a result anticipated by Fitts (2020). The population ρ is unknown but many estimators have been proposed (e.g., Olkin & Pratt, 1958; Kubokawa, Marchand, & Strawderman, 2017) and will be explored in a subsequent report (Cousineau & Goulet-Pelletier, in preparation). It also depends on the noncentrality parameter given by $\sqrt{n/(2(1-\rho))}\Delta/\sigma$. Using the distribution and estimates of the degrees of freedom and of the noncentrality parameter, confidence intervals can be determined for this statistic. These will be examined in Cousineau and Goulet-Pelletier (in preparation).

The solution turns out to be very close to the solution proposed by Becker (1988) except for one difference: Becker's (1988) degrees of freedom ($n-1$) are adjusted by a factor $2/(1+\rho^2)$. This factor introduces a continuum between a between-subject design (in which ρ is null) where degree of freedom is $2(n-1)$ as usual (Hedges, 1981) and perfectly correlated situations (in which $|\rho| = 1$) which reduces to the 1-group design where the degree of freedom is $1(n-1)$.

With the distribution at hand, I believe that d_p has the potential to become the sole measure of standardized difference. Its alternative, $d_\mathbf{D}$, has an exact distribution and is commonly used in power planning. However, both measures cannot be compared. Depending on correlation, the second can be smaller or larger than d_p. Thus, when using standardized difference in within-subject designs, it is necessary to report correlation so that d_p can be converted into $d_\mathbf{D}$ or vice versa. Having two different measures sharing the same name and the same symbol may create confusions; make sure to explicitly disclose what Cohen's d was reported and how corresponding confidence intervals were computed (Goulet-Pelletier & Cousineau, 2018, and

Figure 1 ■ Simulated d_p from a bivariate normal distribution with difference in means of 15 points and standard deviation of 15 points as well so that the population Cohen's d_p is 1. Simulated samples are small ($n = 10$). The blue dashed line is the distribution derived herein; the red, full, line is (bottom) the t' distribution with $1 \times (n-1)$ degrees of freedom (from Becker, 1988); (top) the t' distribution with degrees of freedom $2 \times (n-1)$ as recommended in Goulet-Pelletier and Cousineau (2018).
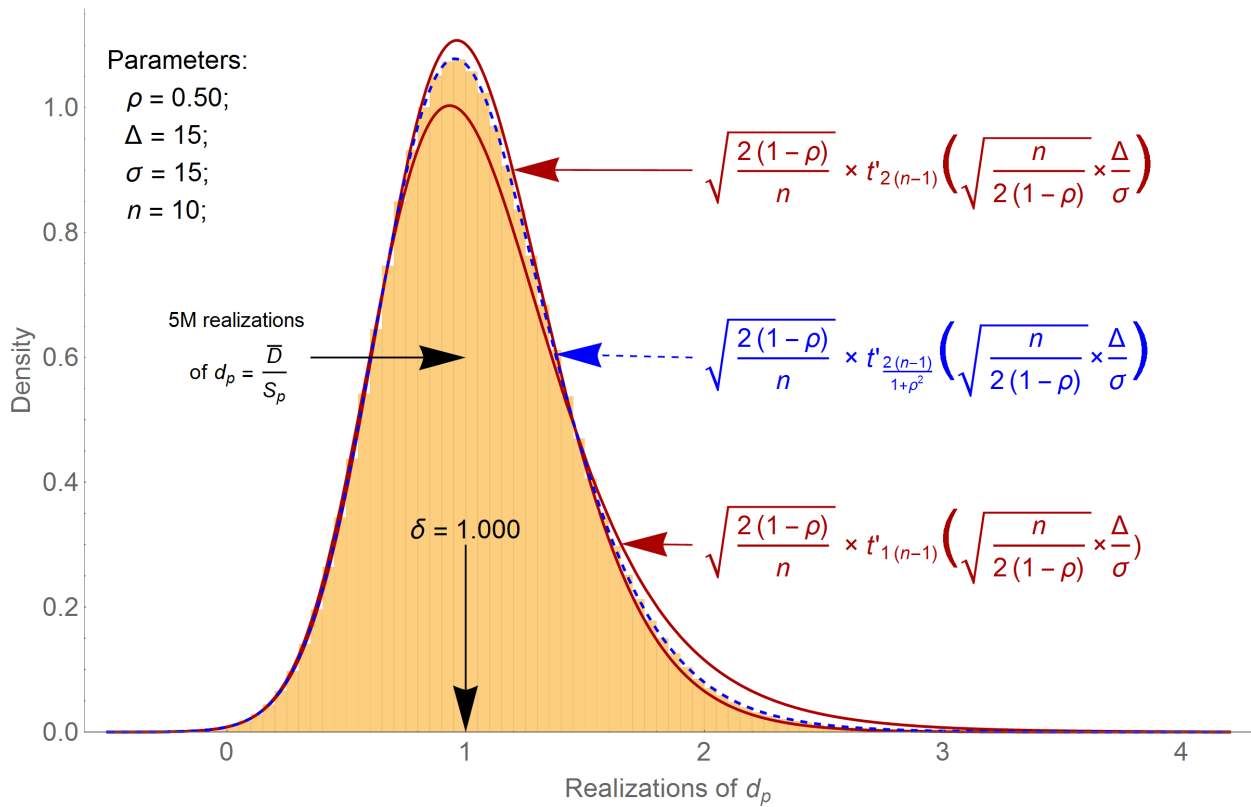
Viechbauer, 2007, identified hundreds of variantes).

Authors' note

I wish to thank Robert Calin-Jageman, Douglas Fitts, Bruno Lecoutre, Jean-Christophe Goulet-Pelletier and an anonymous reviewer for helpful comments.

References

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41,* 257–278. doi:10 . 1111 / j . 2044-8317.1988.tb00901.x

Ben. (2020). Distribution of the pooled variance in paired samples. Retrieved August 11, 2020, from https://stats.stackexchange.com/q/482118

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* Academic Press.

Cousineau, D., & Goulet-Pelletier, J.-C. (2020). Five techniques to derive confidence intervals with a special attention to the Cohen's d_p in the between-group design. *PsyArXiv,* 1–35. doi:10.31234/osf.io/s2597

Fitts, D. A. (2020). Commentary on "A review of effect sizes and their confidence intervals, Part I: The Cohen's d family": The degrees of freedom for paired samples designs. *The Quantitative Methods for Psychology,* 1–35.

Goulet-Pelletier, J.-C., & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, Part I: The Cohen's d family. *The Quantitative Methods for Psychology, 14,* 242–265. doi:10.20982/tqmp.14.4.p242

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107–128. doi:10.2307/1164588

Kendall, M. G., & Stuart, A. (1977). *The advanced theory of statistics, vol. 1.* New York: Macmillan.
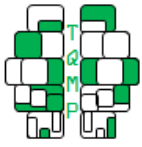
Kubokawa, T., Marchand, É., & Strawderman, W. E. (2017). A unified approach to estimation of noncentrality parameters, the multiple correlation coefficient, and mixture models. *Mathematical Methods of Statistics*, *26*, 134–148. doi:10.3103/S106653071702003X

Laurencelle, L. (2016). Le quotient de deux variances corrélées, sa distribution et son test. *The Quantitative Methods for Psychology*, *12*, 131–137. doi:10.20982/tqmp.12.2.r131

Lecoutre, B. (2007). Another look at confidence intervals from the noncentral t distribution. *Journal of Modern Applied Statistical Methods*, *6*, 107–116. doi:10.22237/jmasm/1177992600

Olkin, I., & Pratt, J. W. (1958). Unbaised estimation of certain correlation coefficients. *The Annals of Mathematical Statistics*, *29*, 201–211. doi:10.1214/aoms/1177706717.

Steiger, J. H., & Fouladi, R. T. (1997). *Noncentral interval estimation and the evaluation of statistical models*. Mahwah: Erlbaum.

Viechbauer, W. (2007). Approximate confidence intervals for standardized effect sizes in the two-independent and two-dependent samples design. *Journal of Educational and Behavioral Statistics*, *32*, 39–60. doi:10.3102/1067998606298034

Citation

Cousineau, D. (2020). Approximating the distribution of $c$ohen's d_p in within-subject designs. *The Quantitative Methods for Psychology*, *16*(4), 418–421. doi:10.20982/tqmp.16.4.p418