

The APP procedure for estimating the Cohen's effect size

Cohen's effect size

Xiangfei Chen

*Mathematical Sciences, New Mexico State University,
Las Cruces, New Mexico, USA*

David Trafimow

*Department of Psychology, New Mexico State University,
Las Cruces, New Mexico, USA*

Tonghui Wang and Tingting Tong

*Mathematical Sciences, New Mexico State University,
Las Cruces, New Mexico, USA, and*

Cong Wang

Mathematics, University of Nebraska Omaha, Omaha, Nebraska, USA

289

Received 4 August 2021
Revised 16 August 2021
Accepted 16 August 2021

Abstract

Purpose – The authors derive the necessary mathematics, provide computer simulations, provide links to free and user-friendly computer programs, and analyze real data sets.

Design/methodology/approach – Cohen's d , which indexes the difference in means in standard deviation units, is the most popular effect size measure in the social sciences and economics. Not surprisingly, researchers have developed statistical procedures for estimating sample sizes needed to have a desirable probability of rejecting the null hypothesis given assumed values for Cohen's d , or for estimating sample sizes needed to have a desirable probability of obtaining a confidence interval of a specified width. However, for researchers interested in using the sample Cohen's d to estimate the population value, these are insufficient. Therefore, it would be useful to have a procedure for obtaining sample sizes needed to be confident that the sample Cohen's d to be obtained is close to the population parameter the researcher wishes to estimate, an expansion of the *a priori* procedure (APP). The authors derive the necessary mathematics, provide computer simulations and links to free and user-friendly computer programs, and analyze real data sets for illustration of our main results.

Findings – In this paper, the authors answered the following two questions: The precision question: How close do I want my sample Cohen's d to be to the population value? The confidence question: What probability do I want to have of being within the specified distance?

Originality/value – To the best of the authors' knowledge, this is the first paper for estimating Cohen's effect size, using the APP method. It is convenient for researchers and practitioners to use the online computing packages.

Keywords Effect size, Cohen's d , Non-central t distribution, Minimum sample size

Paper type Research paper

1. Introduction

Cohen (e.g. 1988) famously argued that researchers should be concerned not only with whether an effect is present but with the size of the effect too. Cohen discussed a variety of different effect size indices, and other researchers have added to the effect size toolbox. Nevertheless, for typical studies, where economic data for two groups are compared, economic data for two countries are compared, or where an experimental group is compared



© Xiangfei Chen, David Trafimow, Tonghui Wang, Tingting Tong and Cong Wang. Published in *Asian Journal of Economics and Banking*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>.

Asian Journal of Economics and
Banking
Vol. 5 No. 3, 2021
pp. 289-306
Emerald Publishing Limited
e-ISSN: 2633-7991
p-ISSN: 2615-9821
DOI 10.1108/AJEB-08-2021-0095

to a control group, Cohen's d remains by far the most popular effect size index. As will be explained in more detail in the subsequent section, Cohen's d denotes the difference in means divided by the standard deviation. Thus, Cohen's d provides valuable information about how much the means differ in standard deviation units.

One reason that scientists in the social and economic sciences have found Cohen's d useful is that many of the dependent measures these sciences do not have intrinsic meaning. For instance, whereas it might be reasonably clear what a dollar means, the meaning of a unit on an attitude scale might be less clear. If the mean attitude in the experimental condition is 2 and the mean attitude in the control condition is 1, is this a small difference or a large one? By converting the difference in means to Cohen's d , the researcher can gain an idea of the size of the difference in standard deviation units, even when the scale units are not themselves intrinsically meaningful. Another advantage of Cohen's d is that it facilitates comparisons across studies. Even if scale units are different for different studies, thereby rendering them seemingly impossible to compare, researchers still can compare in terms of standard deviation units. Many researchers have taken advantage of this, particularly in meta-analytic research. Despite the popularity of Cohen's d and its obvious usefulness, there remains an important limitation. Specifically, the Cohen's d that a researcher obtains in a particular experiment is a sample statistic. It is not a population value. Typically, researchers are not interested in sample statistics for their own sake, but because they provide useful estimates of population values. Thus, there is an important question that has not been properly addressed: how well does Cohen's d estimate the population effect size? Although researchers have long known how to compute traditional confidence intervals for Cohen's d , traditional confidence intervals do not properly address the question. This is because, for example, although 95% of 95% confidence intervals surround the population parameter, it is not the case that the population parameter has a 95% chance of being within a 95% confidence interval. This last is unknown. In addition, [Trafimow and Uhalt \(2020\)](#) have shown that sample confidence intervals tend to be inaccurate representations of population confidence intervals unless sample sizes are much larger than those typically employed. An alternative way to address the issue is to use the *a priori* procedure (APP) that has been employed previously in a variety of ways not pertaining to Cohen's d (e.g. [Li et al., 2020](#); [Trafimow, 2017, 2019](#); [Trafimow and MacDonald, 2017](#); [Trafimow et al., 2020a](#); [Wang et al., 2020, 2021](#); [Wei et al., 2020](#)). Although the APP uses confidence intervals, it does so in a way that deviates importantly from traditional confidence intervals. To use APP thinking to address Cohen's d , the researcher would ask the two bullet-pointed questions below.

RQ1. The precision question: How close do I want my sample Cohen's d to be to the population value?

RQ2. The confidence question: What probability do I want to have of being within the specified distance?

For example, the research might wish to have a 95% probability of obtaining Cohen's d within a tenth of a standard deviation of the population value. The present goal is to determine the sample size the researcher needs to collect to meet the precision and confidence specifications in the contexts of independent and matched samples experimental designs.

This paper is organized as follows. Definitions of Cohen's effect sizes for both populations and samples are given in [Section 2](#), together with properties of noncentral t distributions. In [Section 3](#), the APP methods are applied for estimating population effect size θ in independent case and θ_D in dependent case. In [Section 4](#), the simulation study, the coverage rate, and real data examples are provided in supporting our main results given in [Section 3](#). Conclusion remarks are given in [Sections 5](#).

2. Preliminaries

Effect size is a statistical concept that measures the strength of the relationship between two variables on a numeric scale. For example, in medical education research studies that compare different educational interventions, effect size is the magnitude of the difference between groups. The absolute effect size is the difference between the average, or mean, outcomes in two different intervention groups. The standard deviation of the effect size is of critical importance, since it indicates how much uncertainty is included in the measurement. For more details and applications, see, [Sullivan and Feinn \(2012\)](#), [Schafer and Schwarz \(2019\)](#), [Bhandari \(2020\)](#).

Cohen's d is one of the most common ways to measure effect size, which is known as the difference of two population means and it is divided by the standard deviation from the data. Mathematically, Cohen's effect size is denoted by:

$$\theta = \frac{\mu_1 - \mu_2}{\sigma}, \quad (2.1)$$

where μ_1 and μ_2 are means of two populations, and σ is the standard deviation based on either or both populations.

Cohen's d is defined as the difference between two means divided by a standard deviation for the data obtained from both populations:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S}, \quad (2.2)$$

where \bar{X}_1 and \bar{X}_2 are sample means and S , defined by Jacob Cohen, is the pooled standard deviation (for two independent samples)

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}},$$

and n_1 , S_1 , n_2 , S_2 are sample sizes and sample variances of two independent samples, respectively.

Note that confidence intervals of standardized effect sizes, especially Cohen's d , rely on the calculation of confidence intervals of noncentrality parameters. In order to find the minimum sample sizes for estimating the Cohen's effect size θ given in (2.1) by Cohen's d given in (2.2), we need the following definition.

Definition 2.1. Let Z and U be independent random variables, $Z \sim N(\lambda, 1)$, the normal distribution with mean λ and standard deviation 1, and $U \sim \chi_m^2$, the chi-square distribution with m degrees of freedom. The random variable T given by

$$T = \frac{Z}{\sqrt{U/m}} \quad (2.3)$$

is said to have a noncentral t -distribution with m degrees of freedom and the noncentrality parameter λ , denoted by $T \sim t_m(\lambda)$.

It is easy to obtain the following properties of $T \sim t_m(\lambda)$ (see [Nguyen and Wang, 2008](#)).

(1) The probability density function (pdf) of T is given by

$$f_T(t; \lambda) = \frac{1}{2^{(m+1)/2} \Gamma(\frac{m}{2}) \sqrt{m\pi}} \int_0^\infty x^{\frac{m-1}{2}} \exp\left\{-\frac{1}{2} \left[x + \left(t\sqrt{\frac{x}{m}} - \lambda\right)^2\right]\right\} dx.$$

(2) The mean and variance of T are

$$E(T) = \lambda \sqrt{\frac{m}{2}} \frac{\Gamma((m-1)/2)}{\Gamma(m/2)}, \quad \text{for } m > 1$$

and

$$Var(T) = \frac{m(1+\lambda^2)}{m-2} - \frac{\lambda^2 m}{2} \left(\frac{\Gamma((m-1)/2)}{\Gamma(m/2)} \right)^2 \quad \text{for } m > 2,$$

respectively. For convenience, if we use the correction factor $J(m)$ given by

$$J(m) = \sqrt{\frac{m}{2}} \frac{\Gamma((m-1)/2)}{\Gamma(m/2)}, \tag{2.4}$$

then the mean and the variance of T are

$$E(T) = J(m) \lambda, \quad \text{for } m > 1, \tag{2.5}$$

and

$$Var(T) = \frac{m(1+\lambda^2)}{m-2} - (J(m) \lambda)^2, \quad \text{for } m > 2. \tag{2.6}$$

We will use the following results in the proofs of our main results to be given in next section.

Proposition 2.1. Let X_{11}, \dots, X_{1n_1} be a random sample of size n_1 from $N(\mu_1, \sigma^2)$, X_{21}, \dots, X_{2n_2} be a random sample of size n_2 from $N(\mu_2, \sigma^2)$. Assume that two samples are independent. Let θ and d be the Cohen's effect sizes given in (2.1) and (2.2), respectively. Then

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}} d \sim t_{n_1+n_2-2}(\lambda) \quad \text{with } \lambda = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \theta.$$

Proof. From the basic statistics, we know that.

- (1) $\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, (n_1^{-1} + n_2^{-1})\sigma^2)$ so that $Z \equiv \sqrt{\frac{n_1 n_2}{n_1 + n_2}} d \sim N(\lambda, 1)$.
- (2) $\frac{(n_1-1)S_1^2}{\sigma^2} \sim \chi_{n_1-1}^2$ and $\frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi_{n_2-1}^2$ so that $\frac{(n_1+n_2-2)S^2}{\sigma^2} \sim \chi_{n_1+n_2-2}^2$ as S_1^2 and S_2^2 are independent.

Now by Definition 2.1, $T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} d \sim t_{n_1+n_2-2}(\lambda)$, the desired result follows. \square

For the dependent case, we have the following result.

Proposition 2.2. Let $(X_{11}, X_{21})', \dots, (X_{1n}, X_{2n})'$ be a random sample of size n from a bivariate normal population with mean vector μ and covariance matrix Σ , where

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Let $D_i = X_{1i} - X_{2i}$, $i = 1, \dots, n$ and $\theta_D = \frac{\mu_1 - \mu_2}{\sigma}$ and $d_D = \frac{\bar{D}}{S_D}$ be the Cohen's sizes of the population and matched sample, respectively, where

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i \quad \text{and} \quad S_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2$$

are the mean and variance of D_i 's, respectively. Then $\sqrt{n} d_D \sim t_{n-1}(\lambda_D)$, where the noncentrality parameter $\lambda_D = \sqrt{\frac{n}{2(1-\rho)}} \theta_D$.

Proof. Note that $\bar{D} \sim N(\mu_1 - \mu_2, \frac{2(1-\rho)\sigma^2}{n})$ so that $Z \equiv \frac{\sqrt{n}\bar{D}}{\sqrt{2(1-\rho)}\sigma} \sim N(\lambda_D, 1)$, where $\lambda_D = \sqrt{\frac{n}{2(1-\rho)}} \theta_D$. Also it is easy to know that

$$U \equiv \frac{(n-1)S_D^2}{2(1-\rho)\sigma^2} \sim \chi_{n-1}^2,$$

and \bar{D} and S_D^2 are independent. Therefore, by Definition 2.1,

$$T = \frac{Z}{\sqrt{U/(n-1)}} = \frac{\frac{\sqrt{n}\bar{D}}{\sqrt{2(1-\rho)}\sigma}}{\sqrt{(n-1)\frac{S_D^2}{2(1-\rho)\sigma^2}/(n-1)}} = \sqrt{n} d_D \sim t_{n-1}(\lambda_D). \quad (2.7)$$

so that the desired result follows. \square

Remark 2.1. The graphs of density curves of $t_m(\lambda)$ with different mean values $J(m)\lambda$ and different degrees of freedom m are given in Figures 1 and 2, respectively. From graphs, we know that (i) density curves are symmetric about mean $J(m)\lambda$, which is a function of m , so that the equal tailed confidence intervals should be the best choice, and (ii) density curves tend to $N(\lambda, 1)$ as m to ∞ .

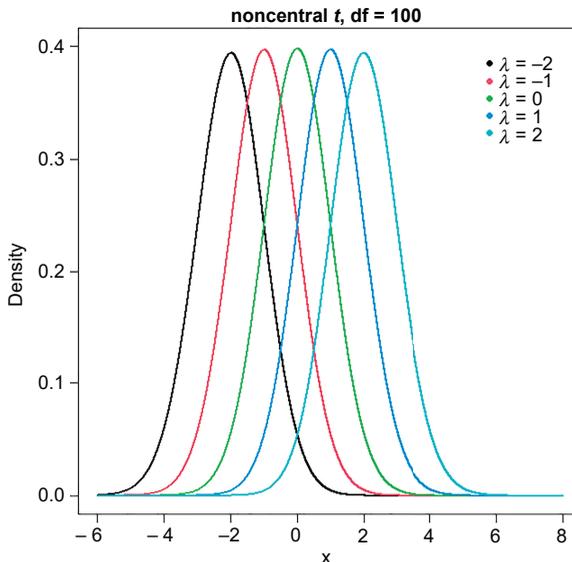


Figure 1. The density curve of noncentral t -distribution with degrees of freedom 100 and different values of noncentral parameter λ

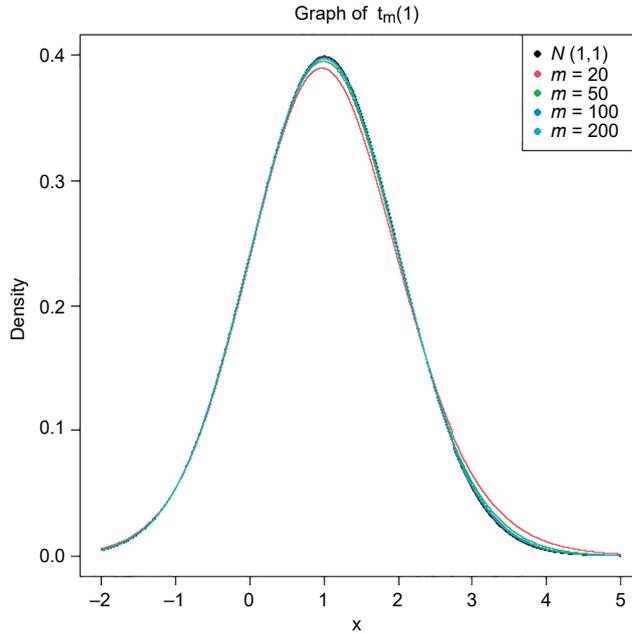


Figure 2.
The density curve of noncentral t -distribution with noncentral parameter $\lambda = 1$ and different values of degrees of freedom

Remark 2.2. For a given the confidence level $c = 1 - \alpha$, the c 100% confidence intervals of λ based on T_1 with m_1 degrees of freedom and T_2 with m_2 with $m_1 < m_2$ degrees of freedom have the following relationship:

$$c = P(|T_1 - \lambda| \leq t_{m_1, (1-c)/2}) \geq P(|T_2 - \lambda| \leq t_{m_1, (1-c)/2}), \quad (2.8)$$

where $t_{m, (1-c)/2}$ is the critical value of the t -distribution with m degrees of freedom. More details for the confidence interval of Cohen's effect size θ will be given in next section.

3. The APP methods for estimating both θ and θ_D

In this section, we will apply the APP methods for estimating Cohen's effect size θ in independent samples' case and θ_D in matched sample case.

3.1 The minimum sample size required for estimating θ

First, we consider two independent samples from two normal populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ with equal unknown variances: $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Theorem 3.1. Let X_{11}, \dots, X_{1n_1} be a random sample of size n_1 from $N(\mu_1, \sigma^2)$, X_{21}, \dots, X_{2n_2} be a random sample of size n_2 from $N(\mu_2, \sigma^2)$. Assume that two samples are independent. Let c be the confidence level and f be the precision which satisfies

$$P(-f\sigma_{d_1} \leq d - E(d) \leq f\sigma_{d_1}) = c, \quad (3.1)$$

where $E(d) = J(n_1 + n_2 - 2)\theta$, $d_1 = \frac{X_{11} - X_{21}}{S}$ and

$$\text{Var}(d_1) \equiv \sigma_{d_1}^2 = \frac{n_1 + n_2 - 2}{n_1 + n_2 - 4} (2 + \theta^2) - (J(n_1 + n_2 - 2)\theta)^2,$$

$J(n_1 + n_2 - 2)$ is the correction factor given in (2.4). Both θ and d are given in (2.1) and (2.2). Let $n = \min\{n_1, n_2\}$ and $f_T(\cdot)$ be the density of t -distribution with degrees of freedom $2(n - 1)$ and noncentrality parameter $\lambda_* = \sqrt{n/2}\theta$. Then the required sample size n can be obtained by solving

$$\int_L^U f_{H_*}(h) dt = c, \tag{3.2}$$

where $d_* = (\bar{X}_{1*} - \bar{X}_{2*})/S_*$, \bar{X}_{1*} , S_{1*}^2 and \bar{X}_{2*} , S_{2*}^2 are sample means and variances of independent samples with same sample size n , respectively. Here $S_*^2 = (S_{1*}^2 + S_{2*}^2)/2$, $H_* = [d_* - E(d_*)]/\sigma_{d_*}$, which has density given by

$$f_{H_*}(h) = \sqrt{\frac{n}{2}} \sigma_{d_*} f_T \left(\sqrt{\frac{n}{2}} \sigma_{d_*} h + J(2n - 2) \lambda_* \right), \tag{3.3}$$

where

$$\begin{aligned} E(d_*) &= J(2n - 2) \theta, & \text{Var}(d_*) &= \sigma_{d_*}^2 = \frac{2}{n} \left[\frac{n-1}{n-2} (1 + \lambda_*^2) - [J(2n - 2)\lambda_*]^2 \right], \\ L &= \frac{-f \sigma_{d_1*}}{\sigma_{d_*}}, & U &= -L, & \sigma_{d_{1*}}^2 &= \frac{n-1}{n-2} (2 + \theta^2) - (J(2n - 2)\theta)^2. \end{aligned} \tag{3.4}$$

Proof. By Proposition 2.1, we know that $T_1 = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} d \sim t_{n_1 + n_2 - 2}(\lambda)$, where $\lambda = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \theta$. Thus, the mean and variance of d , are given, respectively, by

$$E(d) = E \left[\sqrt{\frac{n_1 + n_2}{n_1 n_2}} T_1 \right] = J(n_1 + n_2 - 2) \theta \tag{3.5}$$

and

$$\text{Var}(d) \equiv \sigma_d^2 = \frac{n_1 + n_2}{n_1 n_2} \left[\frac{n_1 + n_2 - 2}{n_1 + n_2 - 4} (1 + \lambda^2) - (J(n_1 + n_2 - 2) \lambda)^2 \right]. \tag{3.6}$$

Now it is easy to obtain the density of $H = \frac{d - E(d)}{\sigma_d}$, which is symmetric about 0 and given by

$$f_H(t) = \sqrt{\text{Var}(T_1)} f_{T_1} \left(\sqrt{\text{Var}(T_1)} t + J(n_1 + n_2 - 2) \lambda \right), \tag{3.7}$$

where f_{T_1} is the density of T_1 . Note that if we have two independent random samples of sizes n_1 and n_2 , we can construct the $100c\%$ confidence interval for θ based on (3.7). Now, let $T_2 = \frac{d_1}{\sqrt{2}} = \frac{X_1 - Y_1}{\sqrt{2S}}$, which has a noncentral t distribution with $n_1 + n_2 - 2$ degrees of freedom and noncentrality parameter $\theta/\sqrt{2}$. Thus, the unbiased estimator of θ is $\sqrt{2} T_2$. Note that the variance of d_1 is

$$\sigma_{d_1}^2 = 2 V(T_2) = \frac{n_1 + n_2 - 2}{n_1 + n_2 - 4} (2 + \theta^2) - [J(n_1 + n_2 - 2)]^2 \theta^2. \tag{3.8}$$

Therefore, we can setup the confidence interval for given confidence level c and precision f , which is given in (3.1). Since there are two unknowns n_1 and n_2 , there are no solutions using Equation (3.1) so we need to modify this equation. Let $n = \min\{n_1, n_2\}$ so that the degrees of freedom $n_1 + n_2 - 2 \geq 2n - 2$. Suppose that we have two independent samples of size n in both; then, $S_*^2 = \frac{S_1^2 + S_2^2}{2}$, and the distribution of $d_* = (\bar{X}_1 - \bar{X}_2)/S_*$ is the noncentral t with $2(n - 1)$ degrees of freedom and the noncentrality parameter $\lambda_* = \frac{\theta}{2}$ so that its mean and variance are

$$E(d_*) = J(2n - 2) \theta \quad \text{and} \quad \text{Var}(d_*) = \sigma_{d_*}^2 = \frac{2}{n} \left[\frac{n - 1}{n - 2} (1 + \lambda_*)^2 - [J(2n - 2)\lambda_*]^2 \right].$$

Thus, the standardized random variable $H_* = \frac{d_* - E(d_*)}{S_*}$ has the density given in Equation (3.3). Therefore, the required n can be obtained by solving Equation (3.2). Similarly to Remark 2.2, we know that $H_{2(n-1), (1-c)/2} \geq H_{n_1+n_2-2, (1-c)/2}$ where $H_{m, (1-c)/2}$ is the critical value of the distribution H . It is easy to see that

$$c = P(-H_{2(n-1), (1-c)/2} \leq H_* \leq H_{2(n-1), (1-c)/2}) \geq P(-H_{2(n-1), (1-c)/2} \leq H \leq H_{2(n-1), (1-c)/2}) \tag{3.9}$$

so that the desired results follows. \square

Remark 3.1. The required n obtained in Theorem 3.1 is unique. Also, if the conditions in Theorem 3.1 are satisfied, we can construct a $c \times 100\%$ confidence interval for $\theta = \frac{\mu_1 - \mu_2}{\sigma}$

$$\left[\frac{d_* - \sigma_{d_*} f}{J(2n - 2)}, \frac{d_* + \sigma_{d_*} f}{J(2n - 2)} \right], \tag{3.10}$$

where σ_{d_*} is given in (3.4) and $\theta = \theta_0$, which can be obtained from the previous data, otherwise the default $\theta = 0$.

Remark 3.2. In order to see that Equation (3.9) holds numerically, we provide probabilities of the $c = 95\%$ confidence intervals for different sample sizes n_1 and n_2 , which is given in Table 1.

θ	n_1	n_2	df	$H_{2n-2, (1-c)/2}$	$P(H < H_{2n-2, (1-c)/2})$
0	100	100	198	13.87366	0.95
		120	218		0.9395
		150	248		0.9267
0.2	100	100	198	13.87366	0.9512
		120	218		0.9408
		150	248		0.9281
0.5	100	100	198	14.04887	0.9528
		120	218		0.9427
		150	248		0.9302
0.8	100	100	198	14.14443	0.9547
		120	218		0.9445
		150	248		0.9322

Table 1. Probabilities of the $c = 95\%$ confidence intervals for different sample sizes n_1 and n_2

Remark 3.3. Researchers can access at the following website <https://appcohensd.shinyapps.io/independent/> to obtain the required sample size. The input variables are the value of θ_0 from the previous data by previous data or $\theta_0 = 0$, precision f , and confidence level c . For convenience, the output variable is the required sample size $n = \min \{n_1, n_2\}$. The required sample sizes for different values of precision $f = 0.1, 0.15, 0.2, 0.25$, confidence levels $c = 0.95, 0.90$ and $\theta_0 = 0, 0.1, \dots, 1$ for independent case are given in [Table 2](#). The relationship between required sample n and parameter θ for different values of precision f is given in [Figure 3](#).

θ		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$f = 0.1$	$c = 0.95$	382	384	386	389	392	396	405	411	415	423	436
	$c = 0.90$	268	269	271	273	276	279	283	287	294	300	304
$f = 0.15$	$c = 0.95$	169	170	171	173	174	176	180	183	186	191	194
	$c = 0.90$	118	119	120	121	122	124	125	127	130	133	135
$f = 0.2$	$c = 0.95$	94	95	96	97	98	99	100	102	104	106	110
	$c = 0.90$	66	66	67	67	68	69	70	71	73	74	76
$f = 0.25$	$c = 0.95$	60	60	61	62	63	63	64	65	67	68	70
	$c = 0.90$	41	41	42	42	43	44	44	45	46	47	49

Table 2. The desired sample sizes n for $f = 0.1, 0.15, 0.2, 0.25, \theta = 0, 0.1, \dots, 1$ and $c = 0.95, 0.9$ in the independent case

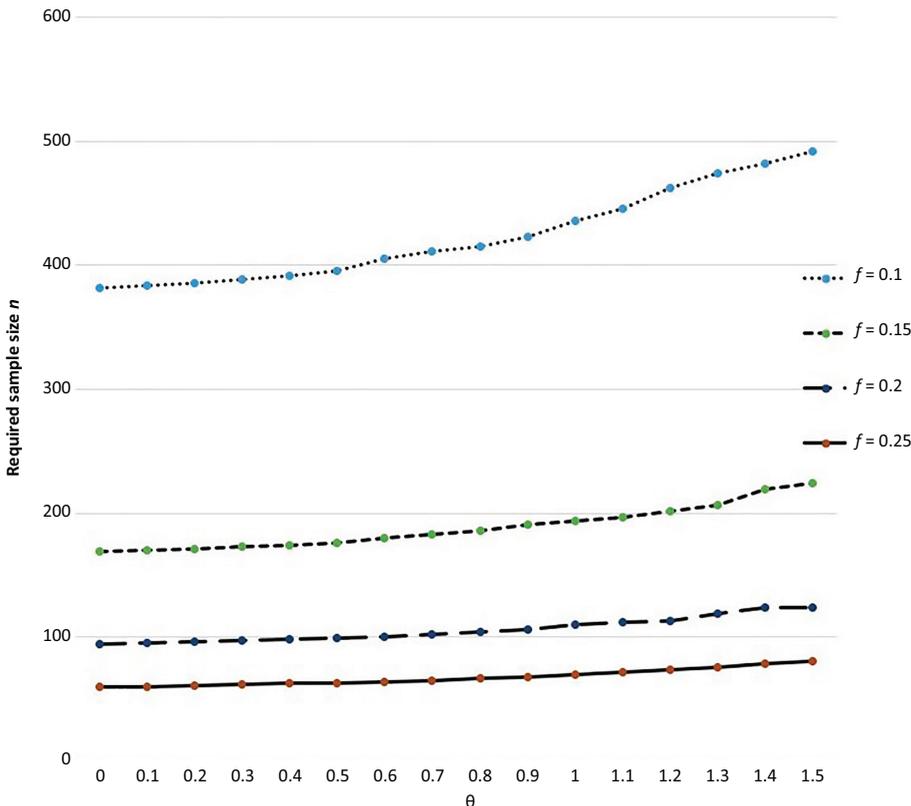


Figure 3. The relationship between required sample n and parameter θ for $c = 0.90$ and different values of precision f

Theorem 3.2. Let $(X_{11}, X_{21})', \dots, (X_{1n}, X_{2n})'$ be a random sample of size n from a bivariate normal population with mean vector μ and covariance matrix Σ , where

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Denote $D_i = X_{1i} - X_{2i}$, $i = 1, \dots, n$. Let $\theta_D = \frac{\mu_1 - \mu_2}{\sigma}$ and $d_D = \frac{\bar{D}}{S_D}$ be the Cohen's effect sizes of the population and matched sample, respectively, where

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i \quad \text{and} \quad S_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2$$

are the mean and variance of D_i 's, respectively. The density function of $\sqrt{n} d_D$ is f_T , the density noncentral t -distribution with $n - 1$ degrees of freedom and noncentrality parameter $\lambda_D = \sqrt{\frac{n}{2(1-\rho)}} \theta_D$. Also, the mean and variance of d_D are given by

$$E(d_D) = J(n-1) \lambda_D / \sqrt{n} \quad \text{and} \quad \sigma_{d_D}^2 = \frac{1}{n} \left[\frac{n-1}{n-3} (1 + \lambda_D^2) - (J(n-1) \lambda_D)^2 \right].$$

Let c be the confidence level and f be the precision which satisfies

$$P(-f \sigma_{d_D} \leq d_D - E(d_D) \leq f \sigma_{d_D}) = c, \tag{3.11}$$

where $d_D^1 = \frac{D_i}{S_D}$ and

$$E(d_D^1) = \frac{J(n-1) \theta_D}{\sqrt{2(1-\rho)}}, \quad \text{Var}(d_D^1) = \sigma_{d_D^1}^2 = \frac{n-1}{n-3} \left(1 + \frac{\theta_D^2}{2(1-\rho)} \right) - \frac{[J(n-1) \theta_D]^2}{2(1-\rho)}.$$

Then the required sample size n can be obtained by

$$\int_L^U f_H(t) dt = c, \tag{3.12}$$

where f_H is the density of $H = [d_D - E(d_D)] / \sigma_{d_D}$ given by

$$f_H(h) = \sqrt{n} \sigma_{d_D} f_T(\sqrt{n} \sigma_{d_D} h + J(n-1) \lambda_D) \tag{3.13}$$

and

$$L = -f \frac{\sigma_{d_D^1}}{\sigma_{d_D}} \quad U = -L. \tag{3.14}$$

Proof. By Proposition 2.2, we know that $T = \sqrt{n} d_D \sim t_{n-1}(\lambda_D)$, where $\lambda_D = \sqrt{\frac{n}{2(1-\rho)}} \theta_D$.

It is easy to see that the mean and variance of d_D are given by

$$E(d_D) = J(n - 1) \lambda_D / \sqrt{n} \quad \text{and} \quad \sigma_{d_D}^2 = \frac{1}{n} \left[\frac{n-1}{n-3} (1 + \lambda_D^2) - (J(n-1) \lambda_D)^2 \right].$$

Thus, the density of H is given in Equation (3.13). Now, we know that $d_D^1 = D_1/S_D \sim t_{n-1}(\lambda_D^1)$, where $\lambda_D^1 = \frac{\theta_D}{\sqrt{2(1-\rho)}}$. Then, the variance of d_D^1 is

$$\sigma_{d_D^1}^2 = \frac{n-1}{n-3} (1 + (\lambda_D^1)^2) - (J(n-1) \lambda_D^1)^2.$$

From the distribution of H , we can obtain the required sample size n which is given in Equations (3.12) - (3.14) so that the desired result follows. \square

Remark 3.4. The value of n obtained is unique with f . Also, if the conditions in Theorem 3.2 are satisfied, we can construct a $c \times 100\%$ confidence interval for θ_D given by

$$\left[\frac{(d_D - \sigma_{d_D^1} f) \sqrt{2(1-\rho)}}{J(n-1)}, \frac{(d_D + \sigma_{d_D^1} f) \sqrt{2(1-\rho)}}{J(n-1)} \right],$$

where $\theta_D = \theta_{D_0}$, which can be obtained from the previous data, otherwise the default $\theta_D = 0$.

Remark 3.5. Researchers can access at the following website: <https://appcohensd.shinyapps.io/matched/>

The input variables are the confidence level c , precision f , correlation coefficient ρ and $\theta_D = \theta_{D_0}$ obtained from previous data information. The default value of $\theta_D = 0$. The output is the desired sample size n required. Table 3 provides the n for $c = 0.90, 0.95$ for different values of f and different values of θ_{D_0} 's. Also the relationship between n and θ_D is given in Figure 4.

θ_D			0	0.2	0.5	0.8
$f = 0.1$	$\rho = 0.2$	$c = 0.95$	382	389	418	470
		$c = 0.90$	267	273	293	330
	$\rho = 0.5$	$c = 0.95$	382	392	436	517
		$c = 0.90$	267	276	304	363
	$\rho = 0.8$	$c = 0.95$	382	407	514	705
		$c = 0.90$	267	284	361	490
$f = 0.15$	$\rho = 0.2$	$c = 0.95$	168	173	186	207
		$c = 0.90$	117	120	130	144
	$\rho = 0.5$	$c = 0.95$	168	175	198	231
		$c = 0.90$	117	122	135	163
	$\rho = 0.8$	$c = 0.95$	168	180	234	325
		$c = 0.90$	117	126	162	216
$f = 0.25$	$\rho = 0.2$	$c = 0.95$	59	61	67	76
		$c = 0.90$	40	42	46	51
	$\rho = 0.5$	$c = 0.95$	59	62	71	81
		$c = 0.90$	40	42	48	58
	$\rho = 0.8$	$c = 0.95$	59	65	82	118
		$c = 0.90$	40	44	58	77

Table 3.
The desired sample sizes n for $f = 0.1, 0.15, 0.25, \theta_D = 0, 0.2, 0.5, 0.8$ and $c = 0.95, 0.9$ in the case of $\rho = 0.2, 0.5, 0.8$

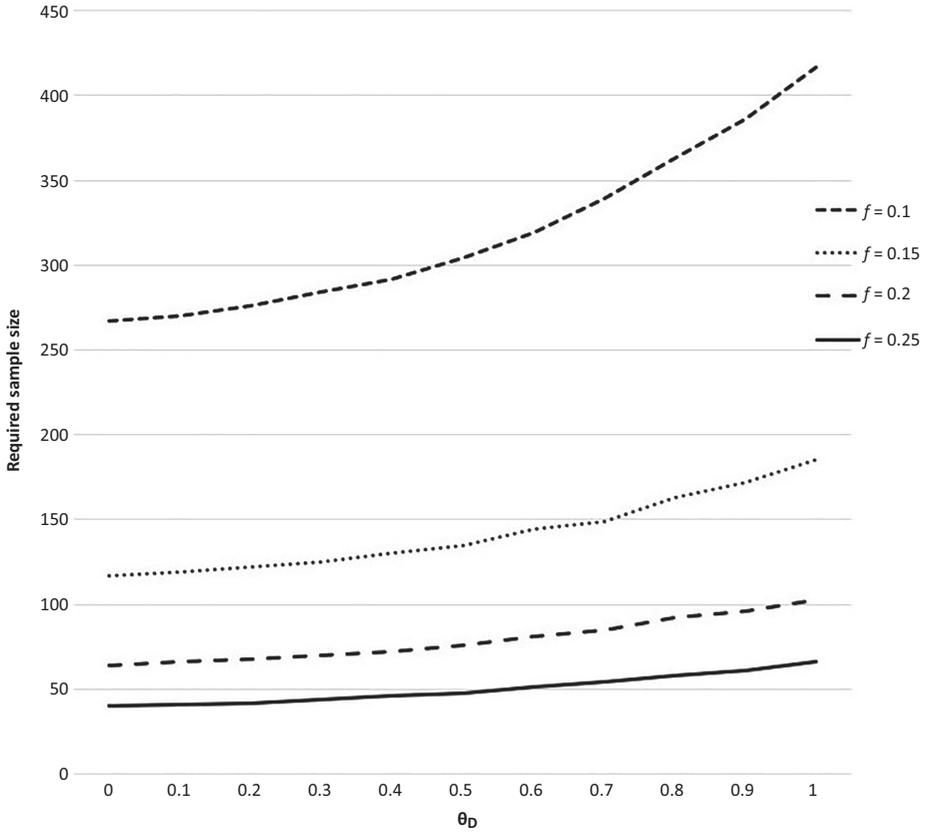


Figure 4. The relationship between required sample n and parameter θ_D for $\rho = 0.5$, $c = 0.90$ and different values of precision f

4. Simulation and real data examples

In this section we will provide some simulation results and two real data examples to support our main results. Based on $M = 100,000$ runs, coverage rates of the confidence intervals and the corresponding point estimates of θ in independent case and θ_D for matched data are given in Tables 4 and 5. From these two tables, we can see that the performances of our APP procedures work very well and the biases are really small.

To evaluate our results, we provide a real data example for independent and match sample cases, respectively.

Example 4.1. Consider the salary data (in 2011) of all professors from College of Arts and Sciences and College of Engineering (population 1 with size 85) and other colleges (population 2 with size 93) of New Mexico State University, which are publicly available https://riograndefoundation.org/downloads/rgf_pr_nmsu.pdf

The estimated distributions based on the data set are approximately $N(8.0868, 1.1099^2)$ for population 1 and $N(9.5273, 1.4115^2)$ for population 2 (with unit \$10,000) (see Figures 5 and 7). The Q–Q plots of the data sets are given in Figures 6 and 8, showing that the scatters lie close to the line with no obvious pattern coming away from the line in both Q–Q plots. Now we consider the 95% confidence interval of θ with precision $f = 0.25$ and default $\theta = 0$.

f	θ	$c = 0.95$			$c = 0.90$		
		n	$\widehat{\theta}$	Coverage rate	n	$\widehat{\theta}$	Coverage rate
0.1	0	382	0.0001	0.9489	268	0.0001	0.8971
	0.2	386	0.1999	0.9509	271	0.2004	0.8976
	0.5	396	0.5002	0.9499	279	0.5007	0.8988
	0.7	411	0.7007	0.9509	287	0.7012	0.9008
	1	436	1.0007	0.9509	304	1.0012	0.8997
0.15	0	169	-0.0002	0.9489	118	-0.0003	0.8976
	0.2	171	0.2008	0.9499	120	0.2006	0.8998
	0.5	176	0.5007	0.9503	124	0.5009	0.9006
	0.7	183	0.7013	0.9511	127	0.7023	0.9011
	1	194	1.0019	0.9508	135	1.0031	0.8987
0.2	0	94	-0.0004	0.9473	66	0.0005	0.8981
	0.2	96	0.2013	0.948	67	0.1998	0.8978
	0.5	99	0.5009	0.9497	69	0.5045	0.898
	0.7	102	0.7023	0.952	71	0.7029	0.8985
	1	110	1.0025	0.9469	76	1.0051	0.9007
0.25	0	60	0.0002	0.9469	41	-0.0005	0.8913
	0.2	61	0.2009	0.9472	42	0.2026	0.8951
	0.5	63	0.5028	0.9469	44	0.5049	0.898
	0.7	65	0.7046	0.9489	45	0.7058	0.8974
	1	70	1.0054	0.9522	48	1.0085	0.8983

Table 4. The corresponding point estimates of θ and the coverage rates when n satisfies the required minimum sample size for $f = 0.1, 0.15, 0.2, 0.25, \theta = 0, 0.2, 0.5, 0.7, 1$ and $c = 0.95, 0.9$ in the independent case

f	θ_D	$c = 0.95, \rho = 0.2$			$c = 0.95, \rho = 0.5$			$c = 0.95, \rho = 0.8$		
		n	$\widehat{\theta}_D$	Coverage rate	n	$\widehat{\theta}_D$	Coverage rate	n	$\widehat{\theta}_D$	Coverage rate
0.1	0	382	0.0001	0.9503	382	-0.0001	0.9488	382	0.0001	0.9499
	0.2	389	0.2002	0.9496	392	0.2004	0.9502	407	0.2006	0.9501
	0.5	418	0.5007	0.9514	436	0.5007	0.9505	514	0.5008	0.9535
	0.7	447	0.7008	0.9506	493	0.7011	0.9529	625	0.7009	0.9513
	1	514	1.0015	0.9517	594	1.0011	0.9539	892	1.0008	0.9542
0.15	0	168	-0.0003	0.9468	168	-0.0001	0.947	168	-0.0001	0.9486
	0.2	173	0.2013	0.9499	175	0.2009	0.9495	173	0.2007	0.9509
	0.5	186	0.5024	0.9498	198	0.5019	0.9529	207	0.5014	0.9544
	0.7	199	0.7027	0.9515	222	0.7024	0.9544	291	0.7019	0.9564
	1	234	1.0033	0.9545	257	1.0032	0.9509	391	1.0018	0.9517
0.2	0	93	0.0001	0.9448	93	-0.0003	0.9444	93	0	0.9452
	0.2	97	0.2016	0.9486	98	0.2013	0.949	102	0.2014	0.9511
	0.5	104	0.5038	0.9485	110	0.5039	0.9509	129	0.5026	0.9524
	0.7	115	0.704	0.9546	125	0.7043	0.9531	159	0.7034	0.9513
	1	129	1.0058	0.9527	154	1.0044	0.9575	228	1.0034	0.956
0.25	0	59	0.0001	0.9448	59	0.0001	0.9451	59	-0.0003	0.9438
	0.2	61	0.2025	0.9486	62	0.2022	0.9489	65	0.2019	0.9491
	0.5	67	0.5062	0.9497	71	0.5057	0.9514	82	0.5042	0.9532
	0.7	73	0.7071	0.9526	80	0.7066	0.9538	105	0.7051	0.9562
	1	82	1.0098	0.9529	97	1.0075	0.9563	146	1.0055	0.956

Table 5. The corresponding point estimates of θ_D and the coverage rates when n satisfies the required minimum sample size for $f = 0.1, 0.15, 0.2, 0.25, \theta = 0, 0.2, 0.5, 0.7, 1, \rho = 0.2, 0.5, 0.8$ and $c = 0.95$ in the matched case

From Table 2, the minimum sample sizes n needed is 60, so we randomly choose samples of size 60 from both populations and obtain the value of Cohen's d is $d = -1.1430$. Then by Remark 3.1, the 95% confidence interval of θ is $[-1.4900, -0.7814]$, which includes the true value of population effect size $\theta = -1.1284$.

Figure 5.
Histogram and
estimated
density curve

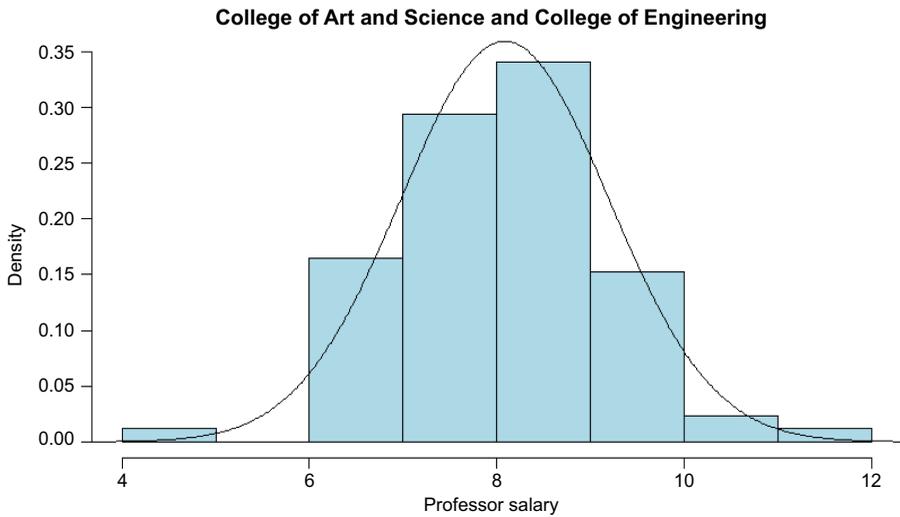
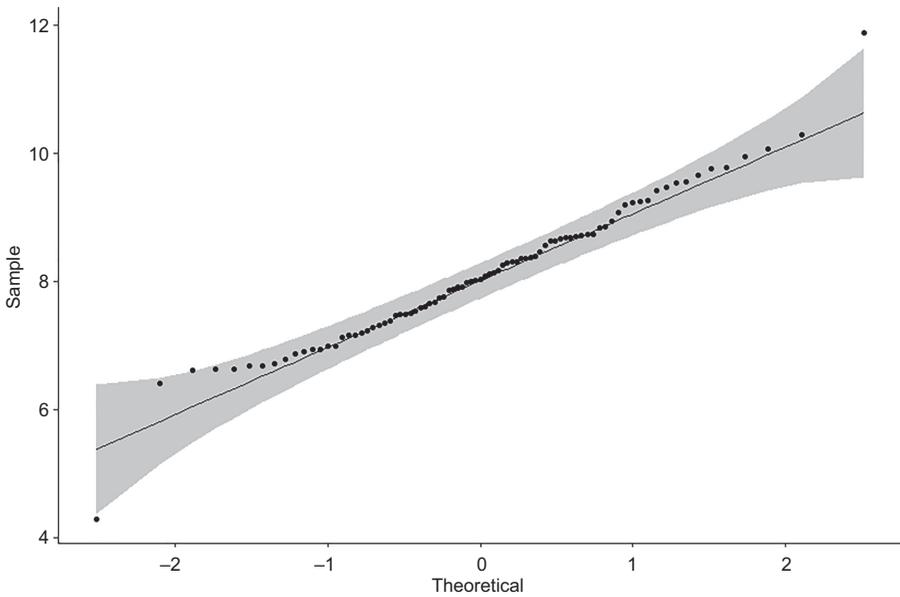
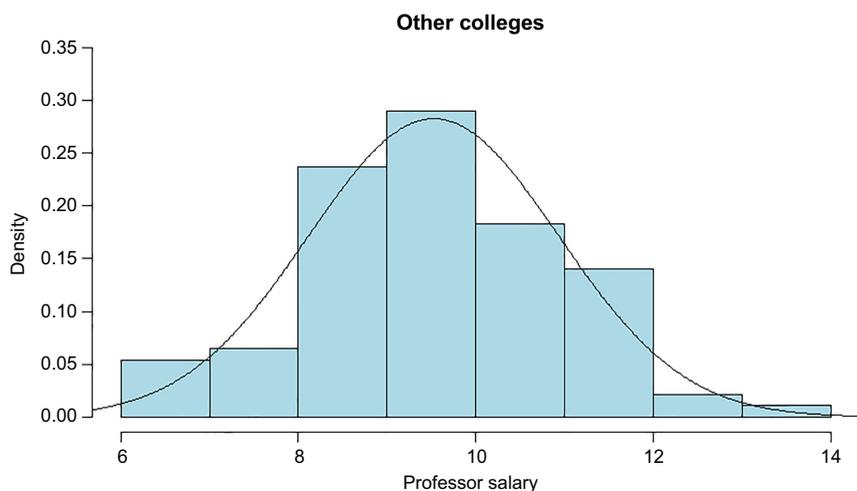


Figure 6.
Q-Q plot



Example 4.2. For matched case, we consider the data set named “Rugby” from Package “PairedDate” in R by [Champely \(2018\)](#). This data set provides the ratings on a continuous ten-point scale of two experts about 93 actions during several rugby union matches. Let D be the difference between ratings of two experts. The histogram and estimated density curve of D are given in [Figure 9](#). From the data, we obtain $\bar{D} = -0.3011$ and $S_D = 1.4872$. The pattern of the points of the scatter plot shown in [Figure 10](#) shows a positive



Cohen's effect size

303

Figure 7. Histogram and estimated density curve

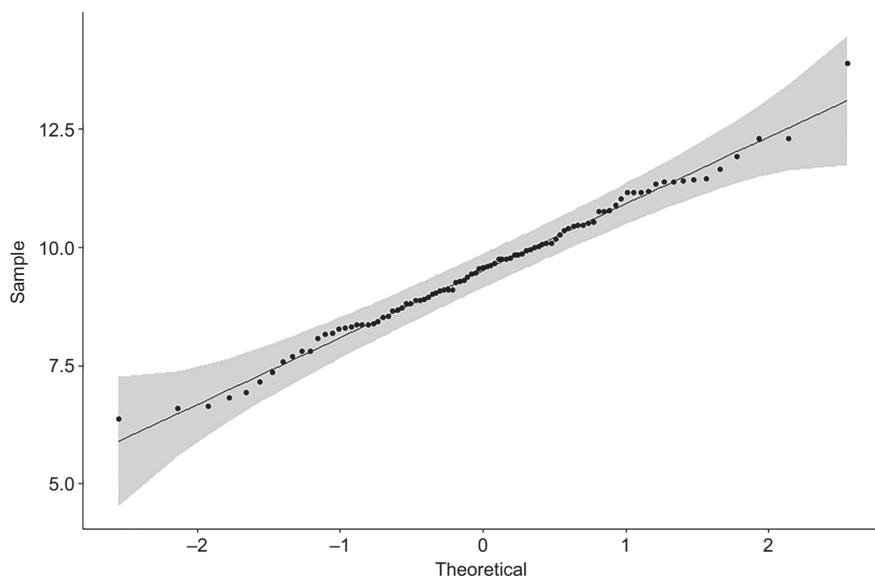


Figure 8. Q-Q plot

linear relationship between these two variables. After calculation, we get the correlation coefficient is $\rho = 0.85$. Now we consider the 95% confidence interval of population effect size θ_D with precision $f = 0.25$ and default $\theta = 0$. By the shinyApp provided in Section 4 for matched data, the minimum sample sizes n needed is 59. Randomly select a sample with 59 paired data; then by Remark 3.2, we have the 95% confidence interval of θ_D is $[-0.3122, -0.0371]$, which includes the true value of population effect size -0.1122 .

Figure 9.
Histogram and
estimated
density curve

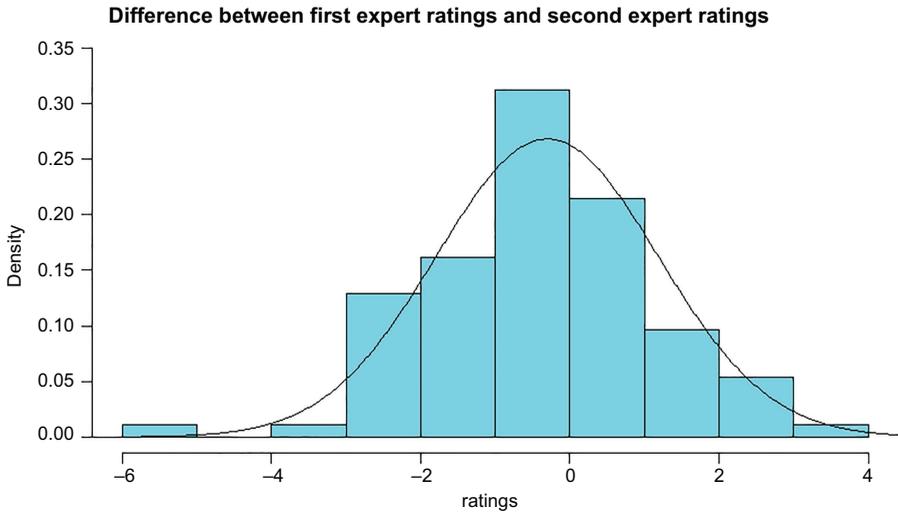
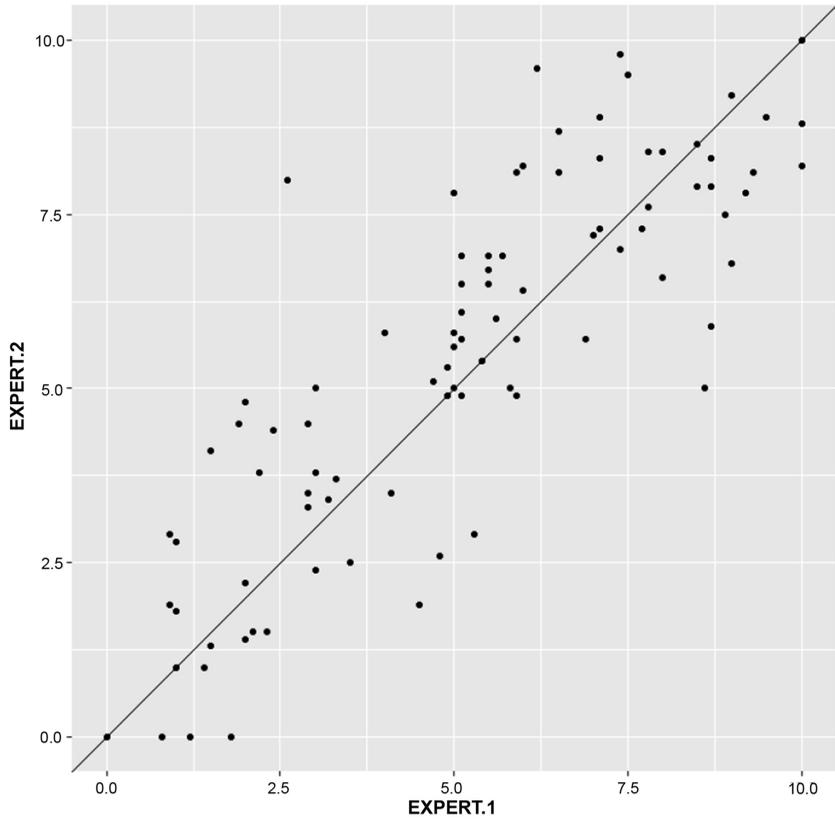


Figure 10.
Correlation



5. Conclusion remarks

Our goal was to derive ways to perform the APP with respect to Cohen's d for independent and matched samples. The present mathematics provide those derivations. In turn, computer simulations support the mathematical derivations. We also provide links to free and user-friendly programs to facilitate researchers performing the APP to determine sample sizes to meet their specifications for precision and confidence. An advantage of the programs is that even researchers who are unsophisticated in mathematics nevertheless can avail themselves of APP advantages.

In addition to the obvious benefit of aiding researchers who wish to compute Cohen's d determine the samples sizes they need the present mathematics provide an additional benefit. Specifically, the famous article in Science by the [Open Science Collaboratio \(2015\)](#) included replications of studies in top psychology journals. They found that the average effect size in the replication cohort of studies was less than half that in original cohort of studies. Thus, effect sizes tend not to replicate across study cohorts. Our suspicion is that one reason for irreproducibility is that sample sizes are too small and traditional power analyses are insufficient because they do not address the precision issue ([Trafimow and Myüz, 2019](#); [Trafimow et al., 2020b](#)), though significance testing doubtless plays a role too. The present mathematics, along with the links to computer programs, provide a solution. We hope and expect that researchers who wish to use Cohen's d to index their effect sizes will be better able to determine appropriate sample sizes, and thereby increase reproducibility in the social sciences.

References

- Bhandari, P. (2020), *Effect Size in Statistics-Scribbr*, available at: <https://www.scribbr.com/statistics/effect-size/>.
- Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed., Erlbaum, Hillsdale, NJ.
- Champely, S. (2018), "PairedData: paired data analysis. R package version: 1.1.1", available at: <https://CRAN.R-project.org/package=PairedData>.
- Li, H., Trafimow, D., Wang, T., Wang, C. and Hu, L. (2020), "User-friendly computer programs so econometricians can run the a priori procedure", *Frontiers in Management and Business*, Vol. 1 No. 1, pp. 2-6, doi: [10.25082/FMB.2020.01.002](https://doi.org/10.25082/FMB.2020.01.002).
- Nguyen, H.T. and Wang, T. (2008), *A Graduate Course in Probability and Statistics*, Vol. 2, Tsinghua University Press, Beijing.
- Open Science Collaboration (2015), "Estimating the reproducibility of psychological science", *Science*, Vol. 349 No. 6251, p. aac4716, doi: [10.1126/science.aac4716](https://doi.org/10.1126/science.aac4716).
- Schafer, T. and Schwarz, M. (2019), "The meaningfulness of effect sizes in psychological research: differences between sub-disciplines and the impact of potential biases", *Frontiers in Psychology*, Vol. 10, p. 813. doi: [10.3389/fpsyg.2019.00813](https://doi.org/10.3389/fpsyg.2019.00813).
- Sullivan, G.M. and Feinn, R. (2012), "Using effect size-or why the p value is not enough", *Journal of Graduate Medical Education*, Vol. 4 No. 3, pp. 279-282, doi: [10.4300/JGME-D-12-00156.1](https://doi.org/10.4300/JGME-D-12-00156.1).
- Trafimow, D. (2017), "Using the coefficient of confidence to make the philosophical switch from a posteriori to a priori inferential statistics", *Educational and Psychological Measurement*, Vol. 77 No. 5, pp. 831-854, doi: [10.1177/0013164416667977](https://doi.org/10.1177/0013164416667977).
- Trafimow, D. (2019), "A frequentist alternative to significance testing, p -values, and confidence intervals", *Econometrics*, Vol. 7 No. 2, pp. 1-14, available at: <https://www.mdpi.com/2225-1146/7/2/26>.
- Trafimow, D. and MacDonald, J.A. (2017), "Performing inferential statistics prior to data collection", *Educational and Psychological Measurement*, Vol. 77 No. 2, pp. 204-219, doi: [10.1177/0013164416659745](https://doi.org/10.1177/0013164416659745).

- Trafimow, D. and Myüz, H.A. (2019), "The sampling precision of research in five major areas of psychology", *Behavior Research Methods*, Vol. 51 No. 5, pp. 2039-2058, doi: [10.3758/s13428-018-1173-x](https://doi.org/10.3758/s13428-018-1173-x).
- Trafimow, D. and Uhalt, J. (2020), "The inaccuracy of sample-based confidence intervals to estimate a priori ones", *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, Vol. 16 No. 2, pp. 112-126, doi: [10.5964/meth.2807](https://doi.org/10.5964/meth.2807).
- Trafimow, D., Wang, C. and Wang, T. (2020a), "Making the a priori procedure (APP) work for differences between means", *Educational and Psychological Measurement*, Vol. 80 No. 1, pp. 186-198, doi: [10.1177/0013164419847509](https://doi.org/10.1177/0013164419847509).
- Trafimow, D.T., Hyman, M.R. and Kostyk, A. (2020b), "The (im)precision of scholarly consumer behavior research", *Journal of Business Research*, Vol. 114, pp. 93-101, doi: [10.1016/j.jbusres.2020.04.008](https://doi.org/10.1016/j.jbusres.2020.04.008).
- Wang, C., Wang, T., Trafimow, D. and Talordphop, K. (2020), "Extending the a priori procedure to one-way analysis of variance model with skew normal random effects", *Asian Journal of Economics and Banking*, Vol. 4 No. 2, pp. 77-90.
- Wang, C., Wang, T., Trafimow, D., Li, H., Hu, L. and Rodriguez, A. (2021), "Extending the A Priori procedure (APP) to address correlation coefficients", in Ngoc Thach, N., Kreinovich, V. and Trung, N.D. (Eds), *Data Science For Financial Econometrics*, Springer. doi: [10.1007/978-3-030-48853-610](https://doi.org/10.1007/978-3-030-48853-610).
- Wei, Z., Wang, T., Trafimow, D. and Talordphop, K. (2020), "Extending the a priori procedure to normal Bayes models", *International Journal of Intelligent Technologies and Applied Statistics*, Vol. 13 No. 2, pp. 169-183, doi: [10.6148/IJITAS.202006-13\(2\).0004](https://doi.org/10.6148/IJITAS.202006-13(2).0004).

Further reading

- Baguley, T. (2009), "Standardized or simple effect size: what should be reported?", *British Journal of Psychology*, Vol. 100 No. Pt 3, pp. 603-617.
- Bobko, P., Roth, P.L. and Bobko, C. (2001), "Correcting the effect size of d for range restriction and unreliability", *Organizational Research Methods*, Vol. 4 No. 1, pp. 46-61.
- Cohen, J. (2013), *Statistical Power Analysis for the Behavioral Sciences*, Elsevier Science, United Kingdom.
- Demidenko, E. (2016), "The p-value you can't buy", *The American Statistician*, Vol. 70 No. 1, pp. 33-38, doi: [10.1080/00031305.2015.1069760](https://doi.org/10.1080/00031305.2015.1069760).
- Gillett, R. (2003), "The metric comparability of meta-analytic effect-size estimators from factorial designs", *Psychological Methods*, Vol. 8 No. 4, pp. 419-433.
- Hedges, L.V. (1981), "Distribution theory for Glass's estimator of effect size and related estimators", *Journal of Educational Statistics*, Vol. 6 No. 2, p. 107, doi: [10.2307/1164588](https://doi.org/10.2307/1164588).
- Wang, C., Wang, T., Trafimow, D. and Chen, J. (2019), "Extending a priori procedure to two independent samples under skew normal settings", *Asian Journal of Economics and Banking*, Vol. 3 No. 2, pp. 29-40.
- Wang, C., Wang, T., Trafimow, D. and Xu, Z. (2021), "A priori procedure (APP) for estimating the scale parameter in gamma populations for known shape parameter", in Sriboonchitta, S., Kreinovich, V. and Yamaka, W. (Eds), *Quantum Computing in Econometrics and Quantum Economics*, Springer-Verlag.

Corresponding author

Tonghui Wang can be contacted at: twang@nmsu.edu

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com