

A Framework for Planning Sample Sizes Regarding Prediction Intervals of the Normal Mean Using R Shiny Apps

Wei-Ming Luh¹, Jiin-Huarng Guo²

[1] *Institute of Education, National Cheng Kung University, Tainan City, Taiwan ROC.* [2] *Department of Applied Mathematics, National Pingtung University, Pingtung City, Taiwan ROC.*

Methodology, 2024, Vol. 20(4), 283–303, <https://doi.org/10.5964/meth.13549>

Received: 2023-12-22 • **Accepted:** 2024-11-08 • **Published (VoR):** 2024-12-23

Handling Editor: Belén Fernández-Castilla, Universidad Nacional de Educación a Distancia, Madrid, Spain

Corresponding Author: Wei-Ming Luh, National Cheng Kung University, #1 University Road, Tainan, Taiwan, 701, ROC. E-mail: luhwei@mail.ncku.edu.tw

Abstract

Replication is a core principle for research, and the recent recognition of the importance of constructing prediction intervals for precise replications highlights the need for robust sample-size planning methodologies. However, methodological and technical complexities often hinder researchers from efficiently achieving this task. This study addresses this challenge by developing five R Shiny apps specifically tailored to determine sample sizes concerning prediction intervals for the mean of the normal distribution. Two measures of precision, absolute and relative widths, are considered. Additionally, the apps consider unequal sampling unit costs and sample size allocations to achieve optimal results by exhaustive search. Simulation results validate the proposed methodology, demonstrating favorable coverage rates. Two illustrative examples of one-sample and two-sample problems showcase these apps' versatility and user-friendly nature, providing researchers with a valid and straightforward approach for systematically planning sample sizes.

Keywords

allocation ratio, absolute width, relative width, sampling cost, optimal sample size allocation, replicability

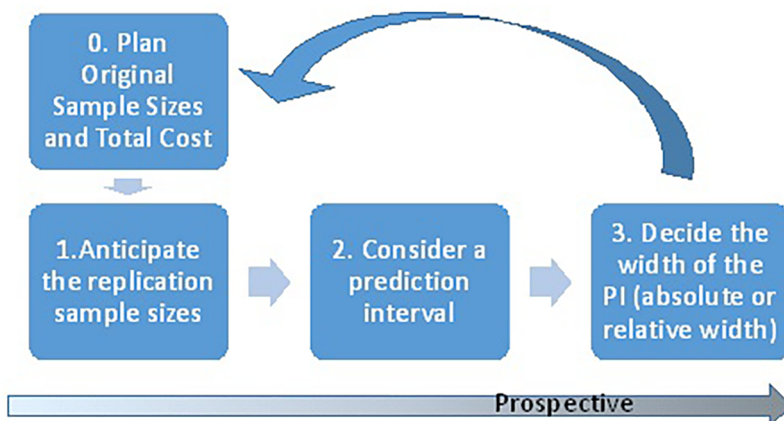
In the past, replications were seldom conducted due to a lack of incentives and were rarely published as they were not considered novel (Nosek & Lakens, 2014). However, the replication crisis has recently made replication a core principle for the future of psychology and social sciences (Anderson & Maxwell, 2016; Cumming, 2008; Higgins et al., 2009; Verhagen, 2022). A critical issue is ensuring that the outcomes of replication



studies are consistent with those of the original studies, thereby evaluating the replicability of the original findings (Patil et al., 2016; Spence & Stanley, 2024). When original study results are imprecise, they lead to wide prediction intervals (PIs). PIs, based on the results of an original (past or current) sample, predict the results of a replication (future) sample drawn repeatedly from the same population with a specified probability $1 - \alpha$ (Patel, 1989; Proschan, 1953). They are crucial in estimating the likelihood of treatment effects from sample to sample, making them invaluable in quality control, meta-analysis, and power calculations for planning replication studies (IntHout et al., 2016; Spence & Stanley, 2016). However, they are often neglected in introductory textbooks, except in the context of regression (Hartnack & Roos, 2021; Preston, 2000). Take one-sample problems as an example; a general form of a PI can be a point estimate $\pm SD t_{n-1; 1-\alpha/2} \sqrt{1/n + 1/m}$, where n and m represent the original and replication sample sizes, respectively. The computation of PIs, by considering sampling errors, can indicate whether the results of replication studies align with those of the original studies. In other words, PIs capture the uncertainty and heterogeneity associated with the estimation of the replication sample, providing a range of expected results *before* the replication study is conducted. (Calin-Jageman & Cumming, 2019; Chiolero et al., 2012; IntHout et al., 2016; Meeker et al., 2017; Roth, 2009). Note that, for constructing PI, the anticipated replication sample size should be decided beforehand (See Figure 1). That becomes a chicken-and-egg conundrum. There has been little discussion regarding this issue, although the essence of planning ahead is fundamental to experimental designs.

Figure 1

A Holistic Framework of Sample Size Planning



Sample size determination is a crucial yet enigmatic aspect of research design and replication studies. It involves probability theory, sampling distribution, sampling error, and uncertainties that form the core of statistical thinking. Despite its mandatory nature, mastering sampling size determination remains challenging, particularly in experimental fields (Anderson & Kelley, 2024; Marszalek et al., 2011; Mellis, 2018). The statistical reform has shifted the focus of sample size calculation from power to precision (i.e., the width of confidence intervals) (Calin-Jageman & Cumming, 2019; Lai & Kelley, 2012; Luh, 2022). However, this shift has primarily applied to original studies, not replication studies. In the prediction context, required sample sizes are based on precision, similar to the construction of confidence intervals. Precision can be expressed in two ways: absolute width or relative width, as Meeker and Hahn (1982) described. The absolute width is the difference between the upper and lower bounds of the prediction interval, which measures the uncertainty associated with the prediction in the same units as the data. On the other hand, the relative width is defined as the ratio of the absolute width to the limiting PI width, providing a measure of the uncertainty relative to the magnitude of the data for comparison across different scales or units. However, these approaches in terms of sample size calculation have not been extensively detailed since then.

In the context of estimating the mean for a one-sample problem, while Hahn (1970a, 1970b) and Meeker and Hahn (1982) discussed the original sample size needed based on relative width for PIs on the mean, little attention has been given to the absolute width. A holistic approach to sample size planning has seen little advancement in this context. For two-sample problems, Hahn (1977) and Niwitpong and Niwitpong (2008) have focused on constructing PIs for the difference between two means of normal distributions. Hahn's approach allows unequal variances to construct an approximate PI, while Niwitpong and Niwitpong use a known ratio of variances. However, the issue of cost-effectiveness remains unaddressed in sample size planning. Existing literature has suggested unbalanced designs for cost-effectiveness, but their application in determining sample size is limited (Hsu, 1994; Liu, 2003; Peckham et al., 2015). In summary, optimal sample size allocation is a crucial yet understudied issue for two-group problems.

One reason that sample size determination has received insufficient attention is the complexity of the calculation involved. Recent advancements in computing have sparked renewed interest in web applications (Doi et al., 2016). Shiny, developed by R Studio, stands out for its interactive, user-friendly, and publicly accessible features. It is an ideal tool for sample size planning for practitioners and applied researchers with minimal coding experience (Chang et al., 2023). To address the research gap, the present study aims to develop several R Shiny apps for sample size planning regarding PIs. Given the complexity of iteration and calculation, these apps are invaluable for practitioners and applied researchers. For the one-sample problem, the present study aims to develop an R Shiny app to determine the original sample size, given the pre-specified replication sample size. For the two-sample problem, considering the unequal unit cost for groups,

two apps were created for Hahn's and Niwitpong's methods, respectively, and the optimal allocation is proposed to utilize an exhaustive search method (Luh, 2022) to find the needed original sample sizes for achieving cost-effectiveness. Finally, another two apps were developed for a scenario when one group size (experimental or control) is fixed to determine another group's sample size (Luh & Guo, 2009; 2016). These scenarios represent a significant advancement compared to existing commercial packages.

The subsequent sections of this study are structured as follows. Section 2 delves into determining sample sizes for the one-group problem using two precision measures. Section 3 presents sample size determination using Hahn's and Niwitpong's methods. This is followed by two illustrative examples of one- and two-group problems showcasing the proposed R Shiny apps. Simulation results and tables are also provided. Finally, the last section offers conclusions from the study's findings and discusses their implications.

The One-Sample Problem

In estimating the mean in the one-sample problem, let X_i be independent and normally distributed with mean μ and variance σ^2 , $i = 1, \dots, n$ (the original sample size). Let $\bar{X}_n = \sum X_i/n$ and $S^2 = \sum (X_i - \bar{X}_n)^2/(n-1)$ be the original sample's mean and variance, respectively. For a replication study, the replication sample is independent of the original sample, and the mean of a replication sample is denoted as \bar{X}_m for a given replication sample size m from the same distribution as the original sample. The discrepancy between the original and the replication samples on the estimation of the mean can be shown by the statistic

$$T_1 = \frac{\bar{X}_n - \bar{X}_m}{S\sqrt{1/n + 1/m}},$$

which is distributed as a t distribution with $n-1$ degrees of freedom and S is the standard deviation of the original sample. The $(1-\alpha)100\%$ two-sided PI for \bar{X}_m is denoted as $[\bar{X}_n \pm St_{n-1; 1-\alpha/2}\sqrt{1/n + 1/m}]$. The absolute width of the PI, denoted as W_{PI} , is $2St_{n-1; 1-\alpha/2}\sqrt{1/n + 1/m}$ which is a random variable and related to the size of m and n . As Meeker and Hahn (1980) noted, if the replication sample size (m) increases to infinity, the prediction interval is the same as the confidence interval (CI). Generally, the width of a prediction interval (PI) is greater than that of the corresponding confidence interval (CI). The precision of PI can be expressed as absolute width (AW) and relative width (RW), and the corresponding sample sizes are demonstrated as follows.

The Absolute Width

To determine the sample size, the absolute width (W_{PI}) should be no larger than a given reasonable value of width (w) from the researcher with a designated probability $(1-\gamma)$

(say, 0.8 or 0.9), i.e., $P(W_{PI} \leq w) \geq 1 - \gamma$. Then, we can find the original sample size to satisfy

$$\begin{aligned} P(W_{PI} \leq w) &= P(2St_{n-1; 1-\alpha/2} \sqrt{1/n + 1/m} \leq w) \\ &= P\left(\frac{(n-1)S^2}{\sigma^2} \leq \frac{(n-1)w^2}{4\sigma^2 t_{n-1; 1-\alpha/2}^2 (1/n + 1/m)}\right) \geq 1 - \gamma, \end{aligned} \quad (1)$$

where the distribution of $(n-1)S^2/\sigma^2$ is a central chi-squared χ_{n-1}^2 with $n-1$ degrees of freedom. To ease the calculation, the present study developed App (I) (Luh & Guo, 2024a) for researchers and practitioners. Given an anticipated replication sample size m , using Panel 1, we can find the minimal original sample size n to meet the following condition:

$$n \geq 1 / \left(\frac{(n-1)w^2}{4\sigma^2 t_{n-1; 1-\alpha/2}^2 \chi_{n-1; 1-\gamma}^2} - \frac{1}{m} \right).$$

The Relative Width

Another measure of precision is relative width. Calculating the relative width involves the ratio (r) of the absolute width and the limiting interval of a PI. For example, if $r = 1.2$, meaning that the width of the PI is no more than twenty percent larger than the width of the limiting interval. According to Meeker and Hahn (1982), when n goes to infinity and m is fixed, the limiting interval of a PI regards as the population $(1 - \alpha)100\%$ two-sided PI for the replication sample mean \bar{X}_m , and it becomes $[\mu \pm \sigma z_{1-\alpha/2} \sqrt{1/m}]$ with a width of $2\sigma z_{1-\alpha/2} \sqrt{1/m}$, denoted as W_{LPI} , which is a decreasing function of m . That is, the larger the m , the smaller the W_{LPI} . When m is fixed, W_{LPI} is then fixed, and it can be used as a baseline to form a relative width (RW), a random variable, by dividing the absolute width of a PI by its limiting interval width as

$$RW = \frac{W_{PI}}{W_{LPI}} = \frac{St_{n-1; 1-\alpha/2} \sqrt{1/n + 1/m}}{\sigma z_{1-\alpha/2} \sqrt{1/m}} = \frac{St_{n-1; 1-\alpha/2} \sqrt{m/n + 1}}{\sigma z_{1-\alpha/2}}.$$

Note that RW is an increasing function of m , but a decreasing function of n . If n goes to infinity, the value of RW approaches 1. To determine the required original sample sizes, researchers can set the value of RW not less than an acceptable ratio of $r (>1)$ with a desired probability of $(1 - \gamma)$. Thus, for the designated value r and probability $1 - \gamma$, we can obtain the required original sample size to satisfy $P(RW \leq r) \geq 1 - \gamma$; i.e.,

$$P\left(\frac{St_{n-1; 1-\alpha/2} \sqrt{m/n + 1}}{\sigma z_{1-\alpha/2}} \leq r\right) \geq 1 - \gamma.$$

Because S^2/σ^2 is distributed as a $\chi^2_{n-1}/(n-1)$ distribution with $n-1$ degrees of freedom, then we have

$$\sqrt{\frac{\chi^2_{n-1; 1-\gamma} t_{n-1; 1-\alpha/2} \sqrt{m/n+1}}{n-1}} \leq r.$$

Thus, given m , we can find a minimal n by using App (I) (Panel 2) to satisfy

$$n \geq m / \left[\left(\frac{r z_{1-\alpha/2}}{t_{n-1; 1-\alpha/2}} \right)^2 \frac{n-1}{\chi^2_{n-1; 1-\gamma}} - 1 \right].$$

The Two-Sample Problem

In estimating the difference between two means in the two-sample problem, let X_{ij} be independent and normally distributed with mean μ_j and variance σ_j^2 , $i = 1, \dots, n_j$ (original sample sizes), $j = 1, 2$. Let $\bar{X}_{n_j} = \sum_{i=1}^{n_j} X_{ij}/n_j$ and $S_j^2 = \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{n_j})^2 / (n_j - 1)$ be the original sample means and variances for the j^{th} group. The replication samples are independent of the original samples for a replication study. Moreover, the replication and original samples in Group 1 are from the same distribution. This condition also holds for Group 2. The two replication sample means are denoted as $(\bar{X}_{m_1}, \bar{X}_{m_2})$ with the given replication sample sizes (m_1, m_2) . The present study then adopted the methods from [Hahn \(1977\)](#) and [Niwitpong and Niwitpong \(2008\)](#), respectively, and employed two precisions, i.e., the absolute and relative widths, to calculate the original sample size needed. In the following, we further consider cost constraints and use optimal sample allocations to demonstrate the features of the proposed approaches. A constraint of total cost for the original sample is $C_n = c_1 n_1 + c_2 n_2$ with the sampling unit costs (c_1, c_2) , then, $n_1 = C_n / (c_1 + k c_2)$ and $n_2 = n_1 k$ can be obtained. The allocation ratio for the original sample is $k = n_2 / n_1$. Note, if the total cost is constrained, the optimal allocation ratio is set as $(\sigma_2 / \sigma_1) \sqrt{c_1 / c_2}$ ([Luh, 2022](#); [Pentico, 1981](#)) for Hahn’s statistic and $\sqrt{\eta c_1 / c_2}$ Niwitpong’s statistic, where η is the ratio of σ_2^2 / σ_1^2 .

Hahn’s Method

The test statistic in [Hahn \(1977\)](#) is

$$T_2 = \frac{\bar{X}_{n_1} - \bar{X}_{n_2} - (\bar{X}_{m_1} - \bar{X}_{m_2})}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} + \frac{S_1^2}{m_1} + \frac{S_2^2}{m_2}}} = \frac{\bar{X}_{n_1} - \bar{X}_{n_2} - (\bar{X}_{m_1} - \bar{X}_{m_2})}{\sqrt{a_1 S_1^2 + a_2 S_2^2}},$$

where $a_1 = 1/n_1 + 1/m_1, a_2 = 1/n_2 + 1/m_2, S_1^2$ and S_2^2 are the original sample variances of Groups 1 and 2, respectively. The random variable

$$(a_1S_1^2 + a_2S_2^2)/(a_1\sigma_1^2 + a_2\sigma_2^2) \quad (2)$$

is distributed approximately as a chi-squared $\chi_{v_2}^2/v_2$, and then T_2 is distributed approximately as a t distribution with v_2 degrees of freedom as

$$v_2 = \frac{(a_1S_1^2 + a_2S_2^2)^2}{(a_1S_1^2)^2/(n_1 - 1) + (a_2S_2^2)^2/(n_2 - 1)} \quad (3)$$

(Satterthwaite, 1946; Welch, 1938). The $(1 - \alpha)100\%$ Hahn's PI of $\bar{X}_{m_1} - \bar{X}_{m_2}$ is denoted as $[\bar{X}_{n_1} - \bar{X}_{n_2} \pm t_{v_2; 1 - \alpha/2} \sqrt{a_1S_1^2 + a_2S_2^2}]$ and the absolute width of this PI is $2t_{v_2; 1 - \alpha/2} \sqrt{a_1S_1^2 + a_2S_2^2}$, denoted as W_{PI} . We discussed two precisions as follows:

Absolute Width

To consider the absolute width of the $(1 - \alpha)100\%$ Hahn's PI of $\bar{X}_{m_1} - \bar{X}_{m_2}$, we need to set a designated width w and a desired probability $(1 - \gamma)$. Then, we can find the original sample sizes to satisfy

$$\begin{aligned} P(W_{PI} \leq w) &= P(2t_{v_2; 1 - \alpha/2} \sqrt{a_1S_1^2 + a_2S_2^2} \leq w) \\ &= P(\chi^2 \leq \frac{v_2 w^2}{4(a_1\sigma_1^2 + a_2\sigma_2^2)t_{v_2; 1 - \alpha/2}^2}) \geq 1 - \gamma, \end{aligned}$$

or

$$\frac{v_2 w^2}{4(a_1\sigma_1^2 + a_2\sigma_2^2)t_{v_2; 1 - \alpha/2}^2} \geq \chi_{v_2; 1 - \gamma}^2 \quad (4)$$

where $\chi^2 = v_2(a_1S_1^2 + a_2S_2^2)/(a_1\sigma_1^2 + a_2\sigma_2^2)$ is distributed approximately as a chi-squared $\chi_{v_2}^2$ variable with degrees of freedom v_2 in Equation (3). We developed App (II) (Luh & Guo, 2024b) to ease the calculation. Given an anticipated replication sample sizes (m_1, m_2) , we can find the minimal n_1 using Panel 1 to meet

$$n_1 \geq (\sigma_1^2 + \sigma_2^2/k) / \left[\frac{v_2 w^2}{4\chi_{v_2; 1 - \gamma}^2 t_{v_2; 1 - \alpha/2}^2} - \left(\frac{\sigma_1^2}{m_1} + \frac{\sigma_2^2}{m_2} \right) \right] \text{ and then } n_2 = n_1 k,$$

where $k = n_2/n_1$.

Relative Width

It is known that if the original sample sizes increase to infinity, the limiting Hahn's PI is $[\mu_1 - \mu_2 \pm z_{1 - \alpha/2} \sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}]$ with the width of $2z_{1 - \alpha/2} \sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}$, denoted as

W_{LPI} , which is a decreasing function of (m_1, m_2) . Then, the relative width (RW_H) of the PI width to its limiting interval width is defined as

$$RW_H = \frac{t_{v_2; 1-\alpha/2} \sqrt{a_1 S_1^2 + a_2 S_2^2}}{z_{1-\alpha/2} \sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}}$$

The value of RW_H should be less than an acceptable ratio r with a desired probability $(1 - \gamma)$. We can obtain the required original sample sizes to satisfy $P(RW_H \leq r) \geq 1 - \gamma$; that is,

$$\begin{aligned} &P\left(\frac{t_{v_2; 1-\alpha/2} \sqrt{a_1 S_1^2 + a_2 S_2^2}}{z_{1-\alpha/2} \sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}} \leq r\right) \\ &= P\left(\frac{a_1 S_1^2 + a_2 S_2^2}{a_1 \sigma_1^2 + a_2 \sigma_2^2} \leq \frac{r^2 z_{1-\alpha/2}^2 (\sigma_1^2/m_1 + \sigma_2^2/m_2)}{t_{v_2; 1-\alpha/2}^2 (a_1 \sigma_1^2 + a_2 \sigma_2^2)}\right) \geq 1 - \gamma. \end{aligned} \tag{5}$$

Based on Equations (2) and (5), we have

$$\sqrt{\frac{\chi_{v_2; 1-\gamma}^2}{v_2} \frac{t_{v_2; 1-\alpha/2} \sqrt{a_1 \sigma_1^2 + a_2 \sigma_2^2}}{z_{1-\alpha/2} \sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}}} \leq r. \tag{6}$$

For easy application, App (II) can be used. Given (m_1, m_2) , we can find the minimal n_1 by using Panel 2 to meet

$$n_1 \geq \frac{\sigma_1^2 + \sigma_2^2/k}{\sigma_1^2/m_1 + \sigma_2^2/m_2} / \left[\frac{r^2 v_2}{\chi_{v_2; 1-\gamma}^2} \left(\frac{z_{1-\alpha/2}}{t_{v_2; 1-\alpha/2}} \right)^2 - 1 \right] \text{ and then } n_2 = n_1 k.$$

Determining Another Group Size When One Group Size is Fixed

In another scenario, when the original sample size for one group is fixed n_1 , the task is to determine the sample size n_2 for another group. We employed Hahn’s method and developed R Shiny App (III) (Luh & Guo, 2024c). By giving (m_1, m_2) , we can find the required size n_2 . Panel 1 is for the precision of absolute width to satisfy Equation (4), and Panel 2 is for the relative width to satisfy Equation (6).

Niwitpong’s Method

In the following, we consider another method based on Niwitpong and Niwitpong (2008), which involves a known ratio value of two variances. The test statistic is defined as

$$T_3 = \frac{\bar{X}_1 - \bar{X}_2 - (\bar{X}_{m_1} - \bar{X}_{m_2})}{\tilde{S}_p \sqrt{b_1 + b_2}},$$

where $b_1 = 1/n_1 + \eta/n_2$, $b_2 = 1/m_1 + \eta/m_2$ with a known ratio of two variances $\sigma_2^2/\sigma_1^2 = \eta$, and $\tilde{S}_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2/\eta}{n_1 + n_2 - 2}$, an unbiased estimator of σ_1^2 . Note that \tilde{S}_p^2/σ_1^2 follows a distribution as $\chi_{v_3}^2/v_3$ and T_3 follows a t distribution with $v_3 = n_1 + n_2 - 2$ degrees of freedom. The $(1 - \alpha)100\%$ Niwitpong's PI of $\bar{X}_{m_1} - \bar{X}_{m_2}$ is denoted as $[\bar{X}_{n_1} - \bar{X}_{n_2} \pm t_{v_3; 1 - \alpha/2} \tilde{S}_p \sqrt{b_1 + b_2}]$ with the absolute width $2t_{v_3; 1 - \alpha/2} \tilde{S}_p \sqrt{b_1 + b_2}$, denoted as W_{PI} . Then, we discussed two precisions as follows:

Absolute Width

For the absolute width of the $(1 - \alpha)100\%$ Niwitpong's PI of $\bar{X}_{m_1} - \bar{X}_{m_2}$ with a specified desired probability $(1 - \gamma)$ and a designated value of w for W_{PI} , we can calculate the original sample sizes needed to satisfy

$$\begin{aligned} P(W_{PI} \leq w) &= P(2t_{v_3; 1 - \alpha/2} \tilde{S}_p \sqrt{b_1 + b_2} \leq w) \\ &= P(\chi^2 \leq \frac{v_3 w^2}{4\sigma_1^2 t_{v_3; 1 - \alpha/2}^2 (b_1 + b_2)}) \geq 1 - \gamma \end{aligned}$$

or

$$\frac{v_3 w^2}{4\sigma_1^2 t_{v_3; 1 - \alpha/2}^2 (b_1 + b_2)} \geq \chi_{v_3; 1 - \gamma}^2 \quad (7)$$

where $\chi^2 = v_3 \tilde{S}_p^2 / \sigma_1^2$ is a distribution of $\chi_{v_3}^2$ with degrees of freedom $v_3 = n_1 + n_2 - 2$. We developed App (IV) (Luh & Guo, 2024d) to obtain the minimal n_1 , given (m_1, m_2) , using Panel 1 to meet

$$\begin{aligned} n_1 &\geq (1 + \eta/k) / \left(\frac{v_3 w^2}{4\sigma_1^2 \chi_{v_3; 1 - \gamma}^2 - v_3 t_{v_3; 1 - \alpha/2}^2} - b_2 \right) \text{ and then } n_2 = n_1 k, \\ &\text{where } k = n_2 / n_1. \end{aligned}$$

Relative Width

If the original sample sizes go to infinity, the limiting PI is $[\mu_1 - \mu_2 \pm z_{1 - \alpha/2} \sigma_1 \sqrt{b_2}]$ with the limiting width $2z_{1 - \alpha/2} \sigma_1 \sqrt{b_2}$, denoted as W_{LPI} . The relative width is defined as

$$RW_N = \frac{t_{v_3; 1 - \alpha/2} \tilde{S}_p \sqrt{b_1/b_2 + 1}}{z_{1 - \alpha/2} \sigma_1}.$$

For a designated ratio r and a probability $1 - \gamma$, to obtain the required original sample sizes to satisfy $P(RW_N \leq r) \geq 1 - \gamma$, we can have

$$P(RW_N \leq r) = P\left(\frac{t_{v_3; 1-\alpha/2} \tilde{S}_p \sqrt{b_1/b_2 + 1}}{z_{1-\alpha/2} \sigma_1} \leq r\right) \geq 1 - \gamma,$$

and \tilde{S}_p^2/σ_1^2 follows a distribution as $\chi_{v_3}^2/v_3$. Hence, we have

$$\sqrt{\frac{\chi_{v_3; 1-\gamma}^2 t_{v_3; 1-\alpha/2} \sqrt{b_1/b_2 + 1}}{v_3 z_{1-\alpha/2}}} \leq r. \tag{8}$$

Thus, given (m_1, m_2) , we can find the minimal n_1 using App (IV) Panel 2 to meet

$$n_1 \geq (1 + \eta/k) / \left[b_2 \left(\frac{v_3 r^2 z_{1-\alpha/2}^2}{\chi_{v_3; 1-\gamma}^2 t_{v_3; 1-\alpha/2}^2} - 1 \right) \right] \text{ and then } n_2 = n_1 k.$$

Determining Another Group Size When One Group Size is Fixed

We employed Niwitpong’s methods and developed R Shiny App (V) (Luh & Guo, 2024e) to find the required size n_2 when n_1 is fixed, by providing (m_1, m_2) . Panel 1 is for the precision of absolute width to satisfy Equation (7), and Panel 2 is for relative width to satisfy Equation (8).

Illustrative Examples and Simulation

Illustrative Examples

To enhance the application of the proposed approaches, two examples of determining the original sample size are illustrated, and the proposed apps are demonstrated as follows.

The One-Sample Problem

We used the example of Spence and Stanley (2024) regarding the estimation of hours of sleep for college students to determine the needed sample size in universities. From their study, the original sample size is $n = 50$ with a mean = 7.21 and $S = 2.2$, and a hypothetical replication sample size $m = 70$ and a 95% prediction interval can be constructed as $95\%PI = M_{original} \pm t_{0.975, (n-1)} S \sqrt{1/n + 1/m} = 7.21 \pm 2.01 \times 2.2 \sqrt{1/50 + 1/70} = 7.21 \pm 0.8186$ with a PI width of $1.6372 = (2 \times 0.8186)$. Then, we can use this information to plan an original sample size. Suppose the precision of absolute width is considered; we aim for the random variable W_{PI} to be less than or equal to $w = 1.6372$ with a probability of 0.8. Then, we can use App (I) (Panel 1) to determine the appropriate original sample size n for a given $m = 70$. After plugging in a significance level of $\alpha = .05$, a desired probability $1 - \gamma = .8$, a planning value of $\sigma^2 = 2.2^2 = 4.84$, and the designated value of width $w = 1.6372$ (see Figure 2), the output shows that the minimum original sample size n is 63, achieving the probability of 80.39% in

our simulation results (refer to Table 1). Thus, Spence and Stanley's anticipated original sample size ($n = 50$) is insufficient to achieve the desired probability; they achieve the designated probability of .80, only 52.73% of the time for the absolute width ($W_{PI} \leq 1.6372$).

Suppose researchers want to use the precision of relative width instead of absolute width; we set the limiting width as 1.031 ($= 2\sigma z_{1-\alpha/2} \sqrt{1/m} = 2 \times 2.2 \times 1.96 / \sqrt{70}$) and set that the random variable RW is less than 1.588 ($= 1.6372/1.031$), with a probability of 0.8, then how large should the original size (n) be? For a given $m = 70$, we can use App (I) (panel 2) to calculate that the most diminutive original size n should be 63 to satisfy Equation (1) and to achieve the probability of 80.32% in our simulation results (refer to Table 1). Our simulation shows that while Spence and Stanley's coverage rate is 95.36%, they achieve the designated probability of .80 only 52.62% for the relative width ($RW \leq 1.588$) for $m = 70$ and $n = 50$.

Figure 2

A Screenshot of R Shiny App (I) for the One-Sample Problem

Determining the Sample Size for the One-sample Problem

1. For absolute width 2. For relative width

Determining the Original Sample Size, Given the Replication Sample Size

alpha, significance level
0.05

1 - gamma, a desired probability
0.8

Planning value of σ^2
4.84

m, replication sample size
70

w, a designated value of PI width
1.6372

Submit

output
Original sample size n = 63

Table 1

For the One-Sample Problem, the Required Original Sample Size and the Simulation Results

Absolute Width ^a			Simulation Results	
Given Replication Sample Size <i>m</i>	Original Sample Size <i>n</i>	Coverage (%)	$W_{PI} \leq w$ (%)	
50	90	94.80	80.89	
70	63	94.95	80.39	
90	54	95.24	79.87	
Relative Width ^b			Simulation Results	
Given Replication Sample Size <i>m</i>	Original Sample Size <i>n</i>	Coverage (%)	$RW \leq r$ (%)	
50	48	95.01	80.55	
70	63	94.95	80.32	
90	78	94.96	80.17	

^a Set $w = 1.6372$ and planning value of $\sigma^2 = 4.84$ in App (I) Panel 1 and denote the absolute width of PI as W_{PI} . ^b Set $r = 1.5$ in App (I) Panel 2 and denote the relative width as RW .

The Two-Sample Problem

In the following, we applied the study by Losordo et al. (2011) and utilized the proposed App (II) for Hahn's method to illustrate sample size planning in a two-sample problem. Their study involved two groups, each comprising 56 patients. Group 1 received a single dose of mobilized autologous CD34+ cells, while Group 2, serving as the control, received an equal volume of diluent over six months. Their findings revealed a significant improvement in exercise tolerance among patients in the treatment group compared to those in the control group patients (139 ± 151 versus 69 ± 122). Without loss of generality, we assumed $c_1 = \$9$ for CD34+ and $c_2 = \$1$ for diluent. Because the information is limited and the prediction interval is not provided, we used the precision of relative width by setting $r = 1.5$ to estimate the original sample sizes (n_1, n_2) by anticipating the replication sample size $(m_1, m_2) = (54, 54)$. Based on Hahn's method, using App (II) (Panel 2), we set the significance level $\alpha = .05$, a desired probability $1 - \gamma = .8$, and planning values of variances ($151^2 = 22801$ and $122^2 = 14884$). As for the allocation ratio, setting optimal allocation ($k = 0$) can obtain $(n_1, n_2) = (45, 99)$ (refer to Figure 3). Group 2 has a larger sample size than Group 1 because it is less expensive. Additionally, for $k = 1$, the resulting $(n_1, n_2) = (56, 56)$. Finally, if the relative width is set by $r = 1.2$, other conditions hold, and the resulting original sample sizes become (160, 160), indicating that the smaller the ratio, the larger the sample size.

Figure 3

A Screenshot of R Shiny App (II) for the Two-Sample Problem

Determining Sample Sizes for PIs of Two-sample Problems Based on Hahn's Method 1. For absolute width 2. For relative width

Determining Original Sample Sizes, Given Replication Sample Sizes

α , significance level
0.05

1 - γ , a desired probability
0.8

a, ($0 < a < 1$)
0.7

b, ($b > 1$)
1.3

Planning value of σ_1^2
22801

Planning value of σ_2^2
14884

m_1 , future sample size of group 1
54

m_2 , future sample size of group 2
54

c_1 , unit cost of group 1
9

c_2 , unit cost of group 2
1

r, a designated ratio for RW, $r > 1$
1.5

k, allocation for the original sample; set k = 0 for the optimal case
0

Submit

output
Original sample size $n_1 = 45$,
original sample size $n_2 = 99$,
and obtained cost for the original study = 504

Tables and Simulations

To assess the effectiveness of the proposed apps, this section presents sample size tables and simulation results using R codes (R Core Team, 2023). To produce sample size tables, we specified $\alpha = .05$, $1 - \gamma = 0.8$ to find the needed original sample sizes. The simulation procedure, taking the one-sample problem as an example, is as follows:

1. For each simulation, first, given the needed sample sizes (m , n) obtained from the apps and a planning population variance (σ^2), generate an original sample with size n from a normal distribution. Then, obtain the original sample mean and variance, and construct a $100(1 - \alpha)\%$ prediction interval.
2. Independently generate a replication sample with size m from the same distribution and obtain the replication sample mean.
3. Check if the prediction interval covers the replication sample mean over 10,000 simulations to cumulate the coverage rate.
4. For the absolute width, calculate W_{PI} and check if the value is less than or equal to the designated value w . For the relative width, calculate RW and check if the value is less than or equal to the designated value r .
5. Finally, over 10,000 simulations, the percentages of $W_{PI} \leq w$ and $RW \leq r$ are recorded, respectively.

For the one-sample problem, Table 1 demonstrates the required minimal original sample size (n) for absolute width and relative width, respectively. The results reveal that the required original sample size is negatively related to the replication sample size in the case of absolute width but positively related in the case of relative width. Finally, the simulation results show that, given the sizes (m , n), the proposed approach can achieve a 95% coverage rate as $\alpha = .05$, signifying that approximately 95% of the random prediction intervals will encompass their corresponding replication sample means (not the population mean) in the long run. Also, the percentages of $W_{PI} \leq w$ and $RW \leq r$ all meet the desired probability of .80.

For the two-sample problem with unit costs $(c_1, c_2) = (\$9, \$1)$, Tables 2 and 3 present pairs of (n_1, n_2) for a given (m_1, m_2) and various allocation ratios for the absolute width and the relative width, respectively. Several notable findings emerge: Firstly, in optimal cases (setting $k = 0$) under the same (m_1, m_2) , the resulting total costs are minimized as expected. Secondly, compared to Hahn's method, Niwitpong's method typically requires a slightly smaller sample size and incurs lower total costs. The reason is that Niwitpong's method only requires a proportion of variances, not their actual values. Thirdly, if we compare Tables 2 and 3, when $(m_1, m_2) = (54, 54)$, it is found that the resulting original sample sizes are the same. This is because, in the case of absolute width, the given width $w = r \times W_{LPI} = r \times 2 \times z_{1-\alpha/2} \times \sqrt{151^2/54 + 122^2/54} = 155.33$ is based on $r = 1.5$, generating the same condition. Finally, simulation results indicate that, given the required sample sizes, the proposed approach can achieve 95% coverage and reach the

target (i.e., $W_{PI} \leq w$ for absolute width or $RW \leq r$ for relative width) about 80% of the time.

Table 2

For the Two-Sample Problem, the Required Original Sample Sizes, Resulting Total Costs, and Simulation Results for Absolute Width (W_{PI})

Replication Sample Sizes, (m_1, m_2)	Allocation Ratio, k	Computational Results		Simulation Results	
		Original Sample Sizes (n_1, n_2)	C_n	Coverage (%)	$W_{PI} \leq w$ (%)
Hahn's method					
(54, 54)	0	(45, 99)	504 ^a	95.64	80.43
	0.5	(77, 39)	732	95.28	79.63
	1	(56, 56)	560	95.44	80.95
	2	(46, 92)	506	95.15	80.13
(35, 70)	0	(60, 118)	658 ^a	95.06	79.34
	0.5	(97, 49)	922	95.08	80.14
	1	(72, 72)	720	94.93	80.45
	2	(60, 120)	660	94.99	80.24
Niwitpong's method					
(54, 54)	0	(39, 116)	467 ^a	94.94	80.28
	0.5	(77, 39)	732	95.28	80.94
	1	(56, 56)	560	95.45	81.65
	2	(44, 88)	484	95.30	81.77
(35, 70)	0	(49, 148)	589 ^a	95.20	80.28
	0.5	(97, 49)	922	95.08	80.33
	1	(70, 70)	700	95.11	80.17
	2	(55, 110)	605	95.05	80.56

Note. Set Planning value of $\sigma_1^2 = 22801$, $\sigma_2^2 = 14884$ for Hahn's method in App (II) Panel 1, and $\sigma_1^2 = 22801$, $\eta = 0.65278$ for Niwitpong's method in App (IV) Panel 1.

^a the optimal case.

Table 3

For the Two-Sample Problem, the Required Original Sample Sizes, Resulting Total Costs, and Simulation Results for Relative Width (RW)

Replication Sample Sizes, (m ₁ , m ₂)	Allocation Ratio, k	Original Sample Sizes(n ₁ , n ₂)	C _n	Coverage (%)	RW ≤ r (%)
Hahn's method					
(54, 54)	0	(45, 99)	504 ^a	95.64	80.43
	0.5	(77, 39)	732	95.28	79.63
	1	(56, 56)	560	95.44	80.95
	2	(46, 92)	506	95.15	79.81
(35, 70)	0	(39, 82)	433 ^a	94.96	79.54
	0.5	(64, 32)	608	94.92	80.02
	1	(47, 47)	470	95.17	79.96
	2	(40, 80)	440	95.19	81.00
Niwitpong's method					
(54, 54)	0	(39, 116)	467 ^a	94.94	80.27
	0.5	(77, 39)	732	95.28	80.94
	1	(56, 56)	560	95.45	81.65
	2	(44, 88)	484	95.30	81.77
(35, 70)	0	(32, 98)	386 ^a	95.02	80.04
	0.55	(64, 32)	608	94.93	80.55
	1	(46, 46)	460	95.34	80.10
	2	(36, 72)	396	95.17	80.07

Note. Set Planning value of σ₁² = 22801, σ₂² = 14884 for Hahn's method in App (II) Panel 2, and η = 0.65278 for Niwitpong's method in App (IV) Panel 2.

^a the optimal case.

Conclusions and Discussion

Replication research has attracted attention over the last decade, leading to the advancement of constructing prediction intervals, often neglected in introductory textbooks except in the context of regression. This study emphasizes the pivotal role of sample size determination and the often-overlooked intricacies tied to prediction intervals in research design and statistical planning in the context of replication. The existing gap in the literature on the precision of the PI and the cost-effectiveness considerations underscores the need for dedicated design and innovation. To meet this challenge, the present study introduces a pioneer contribution by developing R Shiny apps explicitly designed for calculating sample sizes for prediction intervals related to the normal distribution mean, bridging from the original sample to the replication sample. These tools tackle issues concerning designated probability associated with interval widths and optimizing allocation ratios. Notably, they advance methodological rigor while enhancing the practicality and efficiency of replication research endeavors. The proposed approaches empower researchers with unprecedented capabilities to navigate the complexities of sample size determination. Being able to prefigure original sample sizes within a holistic

framework is more coherent and systematic, enhancing the credibility of the research results.

As regards the two measures of precision, the relative width (W_{PI}/W_{LPI}) is rather hard to comprehend but easy to use to prefigure as a ratio. Because the value of relative width should be less than a designated value r with a desired probability $(1 - \gamma)$, we can regard the absolute width of the PI as less than r times the width of the limiting interval. Hence, we suggest choosing a designated value r within a reasonable interval between 1 and 2, with a smaller value requiring a larger sample size when other conditions hold constant. On the other hand, the absolute width (W_{PI}) is intuitive but not easy to apply as expected because information regarding the PI might not be available. Researchers must anticipate the replication sample size to calculate the PI width before conducting a replication study. Furthermore, choosing a designated value (w) is challenging due to various studies across disciplines. Thus, we recommend selecting a reasonable value of w within a broader range as $\max(W_{LPI})/100 < w < \max(W_{LPI})$, where $\max(W_{LPI}) = 2\sigma z_{1-\alpha/2}$ by setting a planning value of σ . Additionally, in the absolute width case, $W_{PI} = 2St_{n-1; 1-\alpha/2} \sqrt{1/n + 1/m}$, which shows that the sample sizes m and n are reciprocally related like a 2-way tug, restraining to achieve $W_{PI} \leq w$. However, for the case of relative width, if the replication sample size holding constant, the original sample size is negatively related to the ratio (r).

Several reminders regarding the use of the proposed apps: First, to avoid encountering errors in running apps, caution statements appear if the plugged-in values of parameters (w for absolute width or r for relative width) are too small or too large. Second, to achieve optimization in the two-sample problem, the apps use exhaustive algorithms and require researchers to set two parameters, i.e., $0 < a < 1 < b$. Using $a = 0.7$ and $b = 1.3$ is recommended to define a narrow search space. If a warning like 'a is too large' or 'b is too small' arises, these values must be adjusted to expand the search space. Third, the selection of replication sample sizes hinges on the budget allocated for the study. Researchers can determine the total cost of the replication study in advance, which then dictates the acquisition of (m_1, m_2) . Finally, if the cost is not the issue, researchers can set it, $c_1 = c_2 = 1$ and apps still can be used to calculate the sample size.

The proposed framework includes both an original and a replication study, presented in an abridged edition. We acknowledge that a single replication is insufficient. In an applied setting, we recommend conducting multiple replications. This involves merging the original and the first replication studies into a new "original study" and planning for the sequent replications to ensure the process continues indefinitely for robust and reliable findings. Final reminders: the sample sizes provided by the proposed apps are based on an ideal condition, but there are always nuanced variations that researchers have to deal with in the applied settings. It should also be pointed out, echoing Parker and Berman (2003), that the "right" sample size is not a singular number, after specifying the goal of precision or statistical power. Instead, it serves as a factor crucial for assessing

a study's utility, necessitating justification for the assumptions involved and for the condition constrained. The problems researchers encounter in science are not merely statistical. The scenarios and the corresponding sample sizes presented in this study can be considered best-case scenarios. Employing considerations of "how much" and "how uncertain" fosters cautious scientific judgments, promoting consistency and reliability.

Funding: This research was supported by grants from the National Science Council, Taiwan (MOST 110-2410-H-006-034-MY3).

Acknowledgments: We thank Associate Editor Belén Fernández-Castilla and the reviewers for their valuable suggestions and comments.

Competing Interests: The authors have declared that no competing interests exist.

References

- Anderson, S. F., & Kelley, K. (2024). Sample size planning for replication studies: The devil is in the design. *Psychological Methods*, 29(5), 844–867. <https://doi.org/10.1037/met0000520>
- Anderson, S. F., & Maxwell, S. E. (2016). There's more than one way to conduct a replication study: Beyond statistical significance. *Psychological Methods*, 21(1), 1–12. <https://doi.org/10.1037/met0000051>
- Calin-Jageman, R. J., & Cumming, G. (2019). The new statistics for better science: Ask how much, how uncertain, and what else is known. *American Statistician*, 73(Sup. 1), 271–280. <https://doi.org/10.1080/00031305.2018.1518266>
- Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2023). *Shiny: Web application framework for R* [R package Version 1.7.4.9002]. <https://shiny.rstudio.com/>
- Chiolero, A., Santschi, V., Burnand, B., Platt, R. W., & Paradis, G. (2012). Meta-analyses: With confidence or prediction intervals? *European Journal of Epidemiology*, 27, 823–825. <https://doi.org/10.1007/s10654-012-9738-y>
- Cumming, G. (2008). Replication and p intervals: P values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, 3(4), 286–300. <https://doi.org/10.1111/j.1745-6924.2008.00079.x>
- Doi, J. A., Potter, G. E., & Wong, J. (2016). Web application teaching tools for statistics using R and Shiny. *Technology Innovations in Statistics Education*, 9(1). <https://doi.org/10.5070/T591027492>
- Hahn, G. J. (1970a). Statistical intervals for a normal population, Part I. Tables, examples and applications. *Journal of Quality Technology*, 2(3), 115–125. <https://doi.org/10.1080/00224065.1970.11980426>

- Hahn, G. J. (1970b). Statistical intervals for a normal population, Part II. Formulas, assumptions, some derivations. *Journal of Quality Technology*, 2(4), 195–206.
<https://doi.org/10.1080/00224065.1970.11980438>
- Hahn, G. J. (1977). A prediction interval on the difference between two future sample means and its application to a claim of product superiority. *Technometrics*, 19(2), 131–134.
<https://doi.org/10.1080/00401706.1977.10489520>
- Hartnack, S., & Roos, M. (2021). Teaching confidence, prediction and tolerance intervals in scientific practice: A tutorial on binary variables. *Emerging Themes in Epidemiology*, 18, 17.
<https://doi.org/10.1186/s12982-021-00108-1>
- Higgins, J. P. T., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, 172(1), 137–159. <https://doi.org/10.1111/j.1467-985X.2008.00552.x>
- Hsu, L. M. (1994). Unbalanced designs to maximize statistical power in psychotherapy efficacy studies. *Psychotherapy Research*, 4(2), 95–106. <https://doi.org/10.1080/10503309412331333932>
- IntHout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, 6, Article e010247.
<https://doi.org/10.1136/bmjopen-2015-010247>
- Lai, K., & Kelley, K. (2012). Accuracy in parameter estimation for ANCOVA and ANOVA contrasts: Sample size planning via narrow confidence intervals. *British Journal of Mathematical & Statistical Psychology*, 65(2), 350–370. <https://doi.org/10.1111/j.2044-8317.2011.02029.x>
- Liu, X. (2003). Statistical power and optimum sample allocation ratio for treatment and control having unequal costs per unit of randomization. *Journal of Educational and Behavioral Statistics*, 28(3), 231–248. <https://doi.org/10.3102/10769986028003231>
- Losordo, D. W., Henry, T. D., Davidson, C., Lee, J. S., Costa, M. A., Bass, T., Mendelsohn, F., Fortuin, F. D., Pepine, C. J., Traverse, J. H., Amrani, D., Ewenstein, B. M., Riedel, N., Story, K., Barker, K., Povsic, T. J., Harrington, R. A., Schatz, R. A., & the ACT34-CMI Investigators. (2011). Intramyocardial, autologous CD34+ cell therapy for refractory angina. *Circulation Research*, 109, 428–436. <https://doi.org/10.1161/CIRCRESAHA.111.245993>
- Luh, W. M. (2022). Probabilistic thinking is the name of the game: Integrating test and confidence intervals to plan sample sizes. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 18(2), 80–98. <https://doi.org/10.5964/meth.6863>
- Luh, W. M., & Guo, J. H. (2009). The sample size needed for the trimmed t test when one group size is fixed. *Journal of Experimental Education*, 78(1), 14–25.
<https://doi.org/10.1080/00220970903224578>
- Luh, W. M., & Guo, J. H. (2016). Sample size planning for the noninferiority or equivalence of a linear contrast with cost considerations. *Psychological Methods*, 21(1), 13–34.
<https://doi.org/10.1037/met0000039>
- Luh, W. M., & Guo, J. H. (2024a). *Determining the sample size for the one-sample problem* [Shiny App I]. <https://prediction-interval-means.shinyapps.io/Normal-1-Mean-PI/>

- Luh, W. M., & Guo, J. H. (2024b). *Determining sample sizes for PIs of two-sample problems based on Hahn's method* [Shiny App II].
<https://prediction-interval-means.shinyapps.io/Normal-2-Means-PI-Hahn/>
- Luh, W. M., & Guo, J. H. (2024c). *Determining another group's sample size when one group size is fixed based on Hahn's method* [Shiny App III].
<https://prediction-interval-means.shinyapps.io/Normal-2-Means-PI-fixed-n1-Hahn/>
- Luh, W. M., & Guo, J. H. (2024d). *Determining sample size for PIs of two-sample problems based on Niwitpong's method* [Shiny App IV].
<https://prediction-interval-means.shinyapps.io/Normal-2-Means-PI-Niwitpong/>
- Luh, W. M., & Guo, J. H. (2024e). *Determining another group's sample size when one group size is fixed based on Niwitpong's method* [Shiny App V].
<https://prediction-interval-means.shinyapps.io/Normal-2-Means-PI-fixed-n1-Niwitpong/>
- Marszalek, J. M., Barber, C., Kohlhart, J., & Holmes, C. B. (2011). Sample size in psychological research over the original 30 years. *Perceptual and Motor Skills*, *112*(2), 331–348.
<https://doi.org/10.2466/03.11.PMS.112.2.331-348>
- Meeker, W. Q., & Hahn, G. J. (1980). Prediction intervals for the ratios of normal distribution sample variances and exponential distribution sample means. *Technometrics*, *22*(3), 357–366.
<https://doi.org/10.1080/00401706.1980.10486167>
- Meeker, W. Q., & Hahn, G. J. (1982). Sample sizes for prediction intervals. *Journal of Quality Technology*, *14*(4), 201–206. <https://doi.org/10.1080/00224065.1982.11978821>
- Meeker, W. Q., Hahn, G. J., & Escobar, L. (2017). *Statistical intervals. A guide for practitioners and researchers* (pp. 177, 403). Wiley.
- Mellis, C. (2018). Lies, damned lies and statistics: Clinical importance versus statistical significance in research. *Paediatric Respiratory Reviews*, *25*, 88–93. <https://doi.org/10.1016/j.prrv.2017.02.002>
- Niwitpong, S., & Niwitpong, S. A. (2008). Prediction interval on the difference between two future sample means with a known ratio of variances. *International Journal of Intelligent Technology and Applied Statistics*, *1*, 75–86. <https://doi.org/10.6148/IJTAS.2008.0102.05>
- Nosek, B. A., & Lakens, D. (2014). A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137–141. <https://doi.org/10.1027/1864-9335/a000192>
- Parker, R. A., & Berman, N. G. (2003). Sample size: More than calculations. *American Statistician*, *57*, 166–170. <https://doi.org/10.1198/0003130031919>
- Patel, J. K. (1989). Prediction intervals—A review. *Communications in Statistics. Theory and Methods*, *18*(7), 2393–2465. <https://doi.org/10.1080/03610928908830043>
- Patil, P., Peng, R. D., & Leek, J. T. (2016). What should researchers expect when they replicate studies? A statistical view of replicability in psychological science. *Perspectives on Psychological Science*, *11*(4), 539–544. <https://doi.org/10.1177/1745691616646366>
- Peckham, E., Brabyn, S., Cook, L., Devlin, T., Dumville, J., & Torgerson, D. J. (2015). The use of unequal randomisation in clinical trials—An update. *Contemporary Clinical Trials*, *45*(Pt A), 113–122. <https://doi.org/10.1016/j.cct.2015.05.017>

- Pentico, D. W. (1981). On the determination and use of optimal sample sizes for estimating the difference in means. *American Statistician*, 35(1), 40–42.
<https://doi.org/10.1080/00031305.1981.10479301>
- Preston, S. (2000). Teaching prediction intervals. *Journal of Statistics Education : An International Journal on the Teaching and Learning of Statistics*, 8(3).
<https://doi.org/10.1080/10691898.2000.12131297>
- Proschan, F. (1953). Confidence and tolerance intervals for the normal distribution. *Journal of the American Statistical Association*, 48(263), 550–564.
<https://doi.org/10.1080/01621459.1953.10483493>
- R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Roth, J. V. (2009). Prediction interval analysis is underutilized and can be more helpful than just confidence interval analysis. *Journal of Clinical Monitoring and Computing*, 23, 181–183.
<https://doi.org/10.1007/s10877-009-9165-0>
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2(6), 110–114. <https://doi.org/10.2307/3002019>
- Spence, J. R., & Stanley, D. J. (2016). Prediction interval: What to expect when you're expecting . . . A replication. *PLoS One*, 11(9), Article e0162874. <https://doi.org/10.1371/journal.pone.0162874>
- Spence, J. R., & Stanley, D. J. (2024). Tempered expectations: A tutorial for calculating and interpreting prediction intervals in the context of replications. *Advances in Methods and Practices in Psychological Science*, 7(1). Advance online publication.
<https://doi.org/10.1177/25152459231217932>
- Verhagen, M. D. (2022). A pragmatist's guide to using prediction in the social sciences. *Socius: Sociological Research for a Dynamic World*, 8. <https://doi.org/10.1177/23780231221081702>
- Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29(3–4), 350–362. <https://doi.org/10.1093/biomet/29.3-4.350>



Methodology is the official journal of the European Association of Methodology (EAM).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.