Journal of Statistical Planning and Inference ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Comparing the overlapping of two independent confidence intervals with a single confidence interval for two normal population parameters

Saeed Maghsoodloo *, Ching-Ying Huang

Department of Industrial and Systems Engineering, Auburn University, Auburn, AL 36849, USA

ARTICLE INFO

Article history:
Received 4 August 2009
Received in revised form
27 April 2010
Accepted 27 April 2010

Keywords:
Overlap type I error
Percent overlap
Overlap confidence intervals on means and variances
Small sample sizes
Overlap type II error rate
Relative Efficiency
Noncentral *t*

ABSTRACT

Two overlapping confidence intervals have been used in the past to conduct statistical inferences about two population means and proportions. Several authors have examined the shortcomings of Overlap procedure and have determined that such a method distorts the significance level of testing the null hypothesis of two population means and reduces the statistical power of the test. Nearly all results for small samples in Overlap literature have been obtained either by simulation or by formulas that may need refinement for small sample sizes, but accurate large sample information exists. Nevertheless, there are aspects of Overlap that have not been presented and compared against the standard statistical procedure. This article will present exact formulas for the maximum % overlap of two independent confidence intervals below which the null hypothesis of equality of two normal population means or variances must still be rejected for any sample sizes. Further, the impact of Overlap on the power of testing the null hypothesis of equality of two normal variances will be assessed. Finally, the noncentral *t*-distribution is used to assess the Overlap impact on type II error probability when testing equality of means for sample sizes larger than 1.

1. Introduction

When testing equality of two normal population means, the sampling distribution (SMD) of the difference of two sample means must be used to conduct statistical inference (i.e., estimation and test of hypothesis) about the corresponding populations mean difference $\mu_x - \mu_y$. An interesting problem arises as to whether the same conclusions will be reached if the SMD of individual sample means is used to construct separate confidence intervals (CIs) for μ_x and μ_y and examine the amount of overlap of the individual CIs in order to make statistical inferences about $\mu_x - \mu_y$. Asymptotic relationships are given by Schenker and Gentleman (2001) about the changes in the type I and II error probabilities (Prs) if the overlapping of two confidence intervals are used to make inferences about the difference in two population quantities Q_1 and Q_2 (such as two population proportions, two means, etc), where the authors made no assumptions about the two underlying populations. The authors used the geometry in their Fig. 1 (p. 183) to show that the total length of two overlapping intervals is longer than that of the corresponding CI from the Standard procedure. Further, in section 3 (p. 184) they proved the asymptotic deficiencies of Overlap relative to the Standard method for both type I and II error probabilities. We will use the restricted assumption of normal underlying populations and an analytical procedure to

* Corresponding author.
E-mail address: maghsood@eng.auburn.edu (S. Maghsoodloo).

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.04.057

Nomenclature

$N(\mu, \sigma^2)$	a normal universe with population mean μ and variance σ^2
Z	$N(0, 1)$
CIL	confidence interval length
$L(\mu_x)$	lower CI limit
SE	standard error
$\mathrm{SE}(\bar{x}-\bar{y})$	$=\sqrt{\sigma_x^2/n_x+\sigma_y^2/n_y}$
$se(\bar{x}-\bar{y})$	$=\sqrt{S_x^2/n_x+S_y^2/n_y}$
PWF	power function
ARE	asymptotic relative efficiency
β	type II error rate
k	sample *se ratio*$=(S_x/\sqrt{n_x})/(S_y/\sqrt{n_y})$

obtain similar results at a level of significance, α, and will verify that in order to attain a nominal type I error rate of 5%, the corresponding two confidence levels must be set at 83.4224%, which is consistent with the simulated 85% value reported by Payton et al. (2000, p. 547).

When the process variances are unknown and sample sizes are not large (i.e., the real-life encountered cases), this paper will obtain exact formulas for the Overlap type I Pr, and also for Overlap type II error probability at a specified standardized difference. Further, the computation of type II error probability (when testing H_0: $\mu_x-\mu_y=0$) requires the use of noncentral t-distribution, although Schenker and Gentleman (2001) provide the impact of Overlap on the Power Function (PWF$=1-\beta$) for the limiting case in terms of n_x and n_y (which also includes the known-variances case).

We will use the noncentral t-distribution to obtain the PWF of testing H_0: $\mu_x-\mu_y=0$ (in the unknown variances case, which has been available in statistical literature for many years) and also the Overlap PWF for sample sizes n_x & $n_y \geq 2$. Even if the underlying distributions are not Laplace–Gaussian, the t-distribution can still be used for statistical inferences about two process means for moderate and large sample sizes, because the application of t-distribution requires the assumption that only sample means be approximately normally distributed (due to the Central Limit Theorem).

Although Payton et al. (2000), Schenker and Gentleman (2001), Payton et al. (2003), and others have somewhat rectified the Overlap problem and have pointed out the misconceptions therein, there are still some details to be worked out. Thus, the objective of this paper is to investigate the exact differences between the Overlap and the Standard [a term coined by Schenker and Gentleman (2001)] methods for testing the null hypotheses H_0: $\sigma_x=\sigma_y$ and H_0: $\mu_x=\mu_y$. We will also use the last two authors' terminology "Standard" to imply the exact correct statistical procedure. Schenker and Gentleman (2001) report results for the impact of Overlap on type I and II error Prs in testing H_0: $Q_1=Q_2$ for the case of large sample sizes, where they refer to Q_1 and Q_2 as quantities (or parameters) of any two, not necessarily normal, populations. Therefore, this work will investigate the same only for underlying normal populations, and other aspects of Overlap but for all sample sizes ≥ 2. To be on the conservative side, we refer to $n \leq 20$ as small, $20 < n \leq 50$ as moderate, and $n > 50$ as large, although some statisticians prefer $n > 60$ as large because for $n > 60$, $t_{\alpha,\nu} \cong Z_\alpha$ to one decimal place, where the degrees of freedom $(df)\nu=n-1$, and Z_α represents the $(1-\alpha)$ quantile of a standard normal deviate.

In summary, the primary objectives of this article are: (1) To quantify the impact of the Overlap procedure on type I error probability (Pr) for a LOS α when testing equality of two normal process variances, or two normal population means for unknown process variances and sample sizes ≥ 2. Payton et al. (2000) obtained results for the latter objective, but used simulation to obtain their Table 1, p. 551; further, the former objective has not been investigated. (2) To determine the maximum % overlap of two individual CIs below which the null hypothesis (either H_0: $\sigma_x=\sigma_y$, or H_0: $\mu_x=\mu_y$) must still be rejected at a given LOS α. (3) To examine the impact of Overlap procedure on type II error rate for sample sizes ≥ 2 and unknown normal population variances.

Note that all primed symbols in this article pertain to the Overlap procedure, and we denote $\bar{x}$ as the larger of the two sample means.

2. The Overlap against Standard method for difference in means of two normal populations with known variances

Consider random samples of sizes n_x and n_y from two independent normal populations $N(\mu_x$, known $\sigma_x^2)$ and $N(\mu_y$, known $\sigma_y^2)$. It is widely known that the $(1-\alpha)\times 100\%$ confidence interval lengths are $\mathrm{CIL}(\mu_x)=2Z_{\alpha/2}\times\sigma_x/\sqrt{n_x}$, and $\mathrm{CIL}(\mu_y)=2Z_{\alpha/2}\times\sigma_y/\sqrt{n_y}$. Suppose the two CIs for μ_x and μ_y are disjoint; then, it follows that either $L(\mu_x) > U(\mu_y)$, or $L(\mu_y) > U(\mu_x)$, where $L(\mu_x)$ and $U(\mu_x)$ represent the lower and upper CI limits for μ_x, respectively. These two possibilities lead to the mutually exclusive requirements that either $\bar{x}-Z_{\alpha/2}\sigma_x/\sqrt{n_x} > \bar{y}+Z_{\alpha/2}\times\sigma_y/\sqrt{n_y}$, or $\bar{y}-Z_{\alpha/2}\sigma_y/\sqrt{n_y} > \bar{x}+Z_{\alpha/2}\sigma_x/\sqrt{n_x}$. Combining these two conditions leads to the Overlap rejection of H_0: $\mu_x=\mu_y$ iff $|\bar{x}-\bar{y}| > Z_{\alpha/2}\times(\sigma_x/\sqrt{n_x}+\sigma_y/\sqrt{n_y})$. If α is set at the nominal rate of 5%, this last inequality will lead to the same asymptotic condition (4) of Schenker and Gentleman (2001, p. 183). In the balanced case of $\sigma_x=\sigma_y=\sigma$, because statistical theory dictates that n_x/n_y should equal σ_x/σ_y, then $n_x=n_y=n$, and the above rejection condition reduces to $|\bar{x}-\bar{y}| > 2Z_{\alpha/2}\sigma/\sqrt{n}$ at the significance level α based on the Overlap method.

Let $K=(\sigma_x/\sqrt{n_x})/(\sigma_y/\sqrt{n_y})=\mathrm{SE}(\bar{x})/\mathrm{SE}(\bar{y})\geq 0$ represent the ratio of two independent normal population standard errors (SEs) for any sample sizes n_x and n_y. It can be shown (all omitted proofs are available on request from the first author) that if the Standard type I error rate is α, but we reject H_0 when the two independent CIs are disjoint, then the Overlap type I

Please cite this article as: Maghsoodloo, S., Huang, C.-Y., Comparing the overlapping of two independent confidence intervals with a single confidence interval for two.... J. Statist. Plann. Inference (2010), doi:10.1016/j.jspi.2010.04.057

error Pr is given by

$$\alpha' = 2 \times \Pr[\bar{x}-\bar{y} > Z_{\alpha/2}\sigma_y(1+K)/\sqrt{n_y}] = 2 \times \Pr[Z > Z_{\alpha/2}(1+K)/\sqrt{1+K^2}], \tag{1a}$$

which is identical to that of asymptotic Eq. (7) provided by Schenker and Gentleman (2001, p. 184) when their standardized difference, d, is set equal to 0. Eq. (1a) shows that as $K \to 0$ or ∞, the value of α' slowly (from below) approaches the exact type I error probability α [consistent with Table 3, p. 3 of Payton et al. (2003)]. Further, since $(1+K)/\sqrt{1+K^2} \geq 1$ and $Z_{\alpha/2}(1+K)/\sqrt{1+K^2} \geq Z_{\alpha/2}$, then $\alpha' = 2 \times \Pr[Z > Z_{\alpha/2}(1+K)/\sqrt{1+K^2}]$ is smaller than $\alpha = 2 \times \Pr(Z > Z_{\alpha/2})$, which means that Overlap always leads (except in the limiting case of $K=0$ or ∞) to a smaller type I error Pr than that of the Standard method, consistent with Figure 3 of Schenker and Gentleman (2001, p. 184). With the aid of calculus, we can show that the minimum value of α' occurs when the SE ratio $K=1$. For the case of $K=1$ (which includes the special case of $\sigma_x=\sigma_y=\sigma$ and $n_x=n_y=n$), the Overlap type I error Pr from Eq. (1a), [using Φ as the cumulative distribution function (*cdf*) of a N(0, 1) density], reduces to

$$\alpha' = 2 \times \Pr(Z > Z_{\alpha/2}\sqrt{2}) = 2 \times \Phi(-Z_{\alpha/2}\sqrt{2}) \tag{1b}$$

- Setting α at 0.01, Eq. (1b) leads to the Overlap LOS of $\alpha' = 0.00026971696 \ll 0.01$. The % relative error in the Overlap type I Pr is $[(0.01-0.00026971696)/0.01] \times 100\% = 97.303\%$.
- For the nominal significance level $\alpha = 0.05$, from (1b) the value of $\alpha' = 0.0055746 \ll 0.05$.

This last value is consistent with the limiting value of 0.006 provided by Payton et al. (2003, p. 2) in their Eq. (6). The % relative error of Overlap to the Standard method is 88.851%. As a result, the larger the LOS α is, the smaller the % relative error becomes. Payton et al. (2000) provide simulation results, for run sizes of 10,000 pairs from the N(0, 1) distribution in their column 3 of Table 1, p. 551, where the value of α' ranges from 0.0039 at $n=5$ to 0.0055 at $n=50$ (n incremented by 5). Eq. (1b) shows that in the case of known equal variances and equal sample sizes, the value of Overlap type I error Pr, α', does not depend on n. However, in a later article Payton et al. (2003) provide more accurate results in their Table 4, page 3, again through simulation run sizes of 10,000 pairs from a $N(0, 1)$ distribution.

At $K=2$ or 1/2, $\alpha=0.05$, Eq. (1a) gives $\alpha'=0.0085494$; at $K=5$, or 1/5, $\alpha'=0.021095$, and at $K=10$, or 0.10, $\alpha'=0.031932$. Clearly, Eq. (1a) shows that as $K \to \infty$ or 0, then $\alpha' \to \alpha$. If the alternative H_1 is one-sided, say H_1: $\mu_x - \mu_y > 0$, then from the Overlap standpoint, H_0 must be rejected only if both conditions $\bar{x}-\bar{y} > 0$ and $L(\mu_x) > U(\mu_y)$ hold, and as a result [for details see Huang (2008, Chapter 3)] the overlap type I error Pr for the case of $\sigma_x=\sigma_y$ and $n_x=n_y$ reduces to

$$\alpha'_1 = \Pr[Z > Z_{\alpha/2}(\sigma_x+\sigma_y)/\sqrt{\sigma_x^2+\sigma_y^2}] = \Pr(Z > Z_{\alpha/2}\sqrt{2})$$

which is equal to ½ of the 2-sided α' from Eq. (1b). Thus, the impact of Overlap on type I error Pr is even greater for an one-sided alternative than the 2-sided case.

Let O represent the amount of overlap length between two individual CIs. From Fig. 1(a and b), the value of O will be zero if either $L(\mu_x) > U(\mu_y)$ or $L(\mu_y) > U(\mu_x)$, in which case H_0: $\mu_x=\mu_y$ is rejected at the LOS $< \alpha$. Thus, O is larger than 0 when $U(\mu_x) > U(\mu_y) > L(\mu_x)$, or $U(\mu_y) > U(\mu_x) > L(\mu_y)$. The overlap is 100% if $U(\mu_x) \geq U(\mu_y) > L(\mu_y) \geq L(\mu_x)$, or if $U(\mu_y) \geq U(\mu_x) > L(\mu_x) \geq L(\mu_y)$. Without loss of generality, X denotes the sample for which $\bar{x}-\bar{y} \geq 0$, and Fig. 1(a) shows that

$$O = U(\mu_y) - L(\mu_x) = Z_{\alpha/2}(\sigma_x/\sqrt{n_x} + \sigma_y/\sqrt{n_y}) - (\bar{x}-\bar{y}) \tag{2}$$

Let O_r be the critical value of O at which H_0 is barely rejected at an α-level. Substituting the Standard critical rejection limit of $\bar{x}-\bar{y} = Z_{\alpha/2}\sigma_y\sqrt{1+K^2}/\sqrt{n_y}$ into (2) results in

$$O_r = Z_{\alpha/2}\sigma_y \times (1+K-\sqrt{1+K^2})/\sqrt{n_y} \tag{3a}$$

Eq. (3a) indicates that H_0: $\mu_x - \mu_y = 0$ must be rejected at the LOS $\leq \alpha$ if the amount of overlap $O \leq O_r$. Further, the span of the two individual CIs, assuming $\bar{x} \geq \bar{y}$, is

$$U(\mu_x) - L(\mu_y) = Z_{\alpha/2}(1+K)\sigma_y/\sqrt{n_y} + (\bar{x}-\bar{y}) \tag{3b}$$

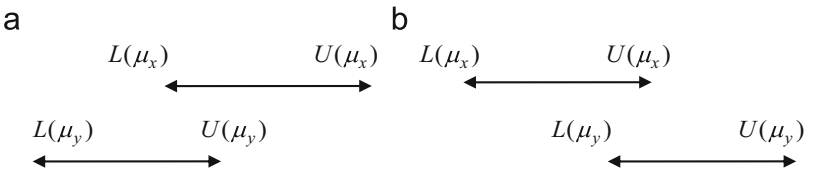

Fig. 1.

Please cite this article as: Maghsoodloo, S., Huang, C.-Y., Comparing the overlapping of two independent confidence intervals with a single confidence interval for two.... J. Statist. Plann. Inference (2010), doi:10.1016/j.jspi.2010.04.057

Combining Eqs. (2) and (3b), the exact percent α-Overlap is given by

$$\omega = \frac{Z_{\alpha/2}\sigma_y(1+K)/\sqrt{n_y} - (\bar{x}-\bar{y})}{Z_{\alpha/2}\sigma_y(1+K)/\sqrt{n_y} + (\bar{x}-\bar{y})} \times 100\% \tag{3c}$$

As $\bar{x}-\bar{y} \geq 0$ increases, the *P-value* of the Z-test decreases, and Eq. (3c) shows that the % overlap also decreases. Because H_0: $\mu_x - \mu_y = 0$ must be rejected at the LOS α iff $|\bar{x}-\bar{y}| \geq Z_{\alpha/2}\sigma_y\sqrt{1+K^2}/\sqrt{n_y}$, the maximum % overlap above which H_0 cannot be rejected at an α-level, from (3c), is given by

$$\omega_r(K) = \frac{1+K-\sqrt{1+K^2}}{1+K+\sqrt{1+K^2}} \times 100\% \tag{3d}$$

Eq. (3d) shows that the maximum percent overlap, below which H_0 must be rejected, does not depend on α and reduces to 17.1573% at $K=1$, which includes the special case of $\sigma_x=\sigma_y=\sigma$ and $n_x=n_y=n$. Again, calculus shows that [for details see Huang (2008)] $K=1$ maximizes $\omega_r(K)$, and as $K\to 0$ or ∞, $\omega_r(K)\to 0$, and Overlap very gradually approaches an exact α-level test [consistent with Table 3 of Payton et al. (2003, p.3)], i.e., Overlap gradually becomes less deficient as K departs from 1.

Next, what should the individual confidence levels, $(1-\gamma)$, be so that the comparisons of individual CIs will lead to an exact $\alpha-$level test? It can be proven that

$$1-\gamma = 1-2\times\Phi[-Z_{\alpha/2}\sqrt{1+K^2}/(1+K)] \tag{4}$$

Eq. (4) shows that the level of each CI must be set at $1-2\times\Phi[-Z_{\alpha/2}\sqrt{1+K^2}/(1+K)]$ in order to reject H_0 at the α LOS iff the two CIs are disjoint, which is in agreement with Eq. (8) of Payton et al. (2003, p. 2). Calculus will show that $K=1$ maximizes γ, and Eq. (4) shows that $1-\gamma$ approaches $1-\alpha$ as K departs from 1. If $\alpha=0.05$ and $K=1$, then $\gamma = 2\Phi(-Z_{0.025}/\sqrt{2}) = 0.165776273$, $1-\gamma=0.834223727$, which implies that the confidence level of each individual interval must be set at 83.4224% in order to reject H_0 at the 5% level iff the two CIs are disjoint. This assertion is in complete agreement with the value of 83.4% in Table 3 at $K=1$ of Payton et al. (2003, p.3). Further, at $\alpha=0.05$, $K=2$ or 1/2, the value of $1-\gamma=0.85595$; at $K=3$, or 1/3, $1-\gamma=0.87874$; at $K=4$, or 1/4, $1-\gamma=0.89395$, while at $K=5$, or 0.20, $1-\gamma=0.90422$, all of which are in agreement with row 2 in Table 3 of Payton et al. (2003, p.3). At $\alpha=0.05$, $K=20$ or 0.05, $1-\gamma=0.93837$, showing that $1-\gamma$ very slowly approaches 95% as $K\to\infty$ or 0, i.e., the Overlap method becomes less deficient as K departs from 1.

Finally, the impact of Overlap on type II error Pr for the known-variance normal case is investigated. It can be proven that the Overlap type II error Pr, is given by

$$\beta' = \Phi\left(Z_{\alpha/2}\frac{(1+K)}{\sqrt{1+K^2}} - d\right) - \Phi\left(-Z_{\alpha/2}\frac{(1+K)}{\sqrt{1+K^2}} - d\right) \tag{5a}$$

where $d = (\mu_x-\mu_y)/\sqrt{\sigma_x^2/n_x+\sigma_y^2/n_y} = \delta\sqrt{n_y}/(\sigma_y\sqrt{1+K^2})$, $K^2 = V(\bar{x})/V(\bar{y})$, and $\delta=\mu_x-\mu_y\geq 0$. Eq. (5a) at $K=0$ (or ∞) gives the type II error Pr, β, from the Standard method. Thus, the PWF of the Overlap procedure in the case of known-variances is

$$1-\beta' = \Phi\left(d - Z_{\alpha/2}\frac{(1+K)}{\sqrt{1+K^2}}\right) + \Phi\left(-Z_{\alpha/2}\frac{(1+K)}{\sqrt{1+K^2}} - d\right) \tag{5b}$$

Eq. (5b) is identical to the asymptotic PWF given in Eq. (7) of Schenker and Gentleman (2001, p. 184), where they provide the expression for $1-\beta'$ at the nominal value of $\alpha=0.05$. Because the function $(1+K)/\sqrt{1+K^2}$ lies within the interval $[1,\ \sqrt{2}]$ for all $K\geq 0$, $Z_{\alpha/2}>0$ for $0<\alpha<0.50$, and Φ is a monotonically increasing function of its argument, it follows from (5b), that the PWF, $1-\beta'$, attains its maximum when $K=0$ or ∞, which represents the case of Standard procedure.

If the alternative is one-sided, say H_1: $\mu_x-\mu_y>0$, clearly the expression for $1-\beta'$ given in Eq. (5b) stays intact, but the Standard method type II error Pr becomes $\beta_1 = \Phi(Z_\alpha - \delta\sqrt{n/2}/\sigma)$, so that the impact of Overlap on type II error Pr for the one-sided H_1 is greater than that of the two-sided case.

3. Comparing the overlap of two independent CIs with a single CI for the ratio of two normal population variances

Because there are two different *t*-tests (the pooled and two-independent-sample *t*-tests) to compare independent normal means when variances are unknown, it is prudent to pretest H_0:$\sigma_x^2=\sigma_y^2$ at an $\alpha-$level. Consider random samples of sizes n_x and n_y from normal universes $N(\mu_x,\ \sigma_x^2)$ and $N(\mu_y,\ \sigma_y^2)$. It is widely known that the confidence interval length $\mathrm{CIL}(\sigma_x^2) = U(\sigma_x^2) - L(\sigma_x^2) = \nu_x S_x^2 \times [(\chi^2_{1-\alpha/2,\nu_x})^{-1} - (\chi^2_{\alpha/2,\nu_x})^{-1}]$, where $\chi^2_{1-\alpha/2,\nu_x}$ represents the $\alpha/2$ quantile of chi-square with $\nu_x=n_x-1$ *df*, and similar expression for $\mathrm{CIL}(\sigma_y^2)$. Based on the Standard method, H_0: $\sigma_x^2/\sigma_y^2=1$ must be rejected at the α-level if either $F_0 = S_x^2/S_y^2 < F_{1-\alpha/2,\nu_x,\nu_y}$ or $F_0 = S_x^2/S_y^2 > F_{\alpha/2,\nu_x,\nu_y}$. The type I error Pr for the Standard method using the null SMD of S_x^2/S_y^2, which is Fisher's F_{ν_x,ν_y}, is α. It can be proven that the Overlap type I error Pr from the two disjoint CIs is given by

$$\alpha'(\text{two disjoint CIs}) = \Pr\left(F_{\nu_x,\nu_y} < \frac{\nu_y}{\nu_x}\times C_{1-\alpha/2,\nu_x,\nu_y}\right) + \Pr\left(F_{\nu_x,\nu_y} > \frac{\nu_y}{\nu_x}\times C_{\alpha/2,\nu_x,\nu_y}\right), \tag{6}$$

where $C_{\alpha/2,,\nu_x,\nu_y} = \chi^2_{\alpha/2,\nu_x}/\chi^2_{1-\alpha/2,\nu_y}$. Thus, based on Overlap we reject H_0: $\sigma_x^2=\sigma_y^2$ if either $F_0 = \frac{S_x^2}{S_y^2} < \frac{\nu_y}{\nu_x}\times C_{1-\alpha/2,\nu_x,\nu_y}$, or $F_0 = \frac{S_x^2}{S_y^2} > \frac{\nu_y}{\nu_x}\times\times C_{\alpha/2,\nu_x,\nu_y}$. Eq. (6) verifies that at $\alpha=0.05$, as $\nu_x=\nu_y$ increase, α' decreases toward 0.0055746, similar to the

Please cite this article as: Maghsoodloo, S., Huang, C.-Y., Comparing the overlapping of two independent confidence intervals with a single confidence interval for two.... J. Statist. Plann. Inference (2010), doi:10.1016/j.jspi.2010.04.057

overlapping of CIs for two normal means, while at $v_x = v_y = 1$, $\alpha' = 0.0178$. However, as $v_x / v_y \to 0$ or ∞, the value of α' very slowly approaches α.

Let ω_r be the maximum percent overlap below which $H_0: \sigma_x^2 = \sigma_y^2$ must be rejected at an α-level. It can be proven that

$$\omega_r = \left(\frac{v_y \times C_{\alpha/2,v_x,v_y} \times \chi^2_{\alpha/2,v_y} - v_x \times F_{\alpha/2,v_x,v_y} \times \chi^2_{\alpha/2,v_y}}{v_x \times C_{\alpha/2,v_x} \times F_{\alpha/2,v_x,v_y} \times \chi^2_{\alpha/2,v_y} - v_y \times \chi^2_{\alpha/2,v_x}} \right) \times 100\% \tag{7a}$$

Eq. (7a) shows that H_0: $\sigma_x^2 = \sigma_y^2$ must not be rejected at an α-level if the percent overlap exceeds ω_r. For the balanced case of $n_x = n_y = n$, the percent overlap in Eq. (7a) reduces to

$$\omega_r = \left(\frac{C_{\alpha/2,n-1} - F_{\alpha/2,n-1,n-1}}{C_{\alpha/2,n-1} \times F_{\alpha/2,n-1,n-1} - 1} \right) \times 100\% \tag{7b}$$

Eq. (7b) shows that the rejection percent overlap between the two CIs for the ratio of two variances will increase as n increases, and unlike normal population means, is α-dependent for finite sample sizes. Matlab shows that at $v = n - 1 = 7{,}819{,}285$ *df* , the 0.05-level overlap is 17.157261356%, which is very close to the overlap for two independent normal population means discussed in Section 2. Further, for sample sizes within the interval [2, 3], the variance-Overlap method is almost an α-level test, like the case of CIs for population means when K is far away from 1.

Third, what should each individual confidence level, $1-\gamma$, be so that the two independent CIs lead to an exact α-level test on $H_0: \sigma_x^2 = \sigma_y^2$? It can be proven that the value of γ can be obtained from

$$v_x F_{\alpha/2,v_x,v_y} / v_y = C_{\gamma/2,v_x,v_y} \tag{8a}$$

Eq. (8a) clearly shows that the value of $1-\gamma$, similar to Eq. (4), depends only on the LOS α of testing $H_0: \sigma_x^2 = \sigma_y^2$ and the sample sizes n_x and n_y. For example, when $\alpha = 0.05$, $n_x = 21$ & $n_y = 11$, Eq. (8a) reduces to $6.8371 = \chi^2_{\gamma/2,20} / \chi^2_{1-\gamma/2,10}$. Through trial & error, the solution to this last equality is $\gamma/2 = 0.07119$ (and $\gamma = 0.14238$). In the case of balanced sampling, Eq. (8a) reduces to

$$F_{\alpha/2,n-1,n-1} = C_{\gamma/2,n-1} \tag{8b}$$

where $C_{\gamma/2,n-1} = \chi^2_{\gamma/2,n-1} / \chi^2_{1-\gamma/2,n-1} > 1$ for all finite n. For a specified α, the value of $\gamma/2$ from (8b) is an increasing function of n. For a 0.05-level test, at $n = 4$, the solution is $\gamma/2 = 0.07127$; for moderate sample sizes $10 < n \le 30$, the approximate solution is $\gamma/2 = 0.08$. As $n \to \infty$, $\gamma \to 0.1657760$, similar to overlapping of CIs for means.

Finally, the impact of Overlap on the type II error Pr is investigated. The OC (Operating Characteristic) curve of the Fisher F-test on equality of two normal process variances has been numerously documented in statistical literature, and at the α-level, it is repeated below

$$\beta(\lambda) = cdfF_{v_x,v_y}(F_{\alpha/2,v_x,v_y} / \lambda^2) - cdfF_{v_x,v_y}(F_{1-\alpha/2,v_x,v_y} / \lambda^2), \tag{9a}$$

where $\lambda = \sigma_x / \sigma_y$. It can be shown [see Huang (2008, Chapter 5)] that the type II error Pr for Overlap of two variance-ratio CIs is given by

$$\beta'(\lambda) = cdfF_{v_x,v_y}\left(\frac{v_y}{v_x} \times C_{\alpha/2,v_x,v_y} / \lambda^2 \right) - cdfF_{v_x,v_y}\left(\frac{v_y}{v_x} \times C_{1-\alpha/2,v_x,v_y} / \lambda^2 \right), \tag{9b}$$

To evaluate the approximate RELEFF (Relative Efficiency) of Overlap to the Standard method at $\alpha = 0.05$, we make use of Eq. (10), listed below, and determine n'_x and n'_y by equating the approximate value of $\beta'(\lambda)$ to that of $\beta(\lambda)$, i.e.,

$$cdfF_{v'_x,v'_y}\left(\frac{v'_y}{v'_x} \times C_{0.025,v'_x,v'_y} / \lambda^2 \right) \cong cdfF_{v_x,v_y}(F_{0.025,\ v_x,v_y} / \lambda^2) \tag{10}$$

It is impossible to find a general closed-form solution from (10) for n'_x and n'_y, whose values depend on n_x, n_y and λ. Accordingly, we used MS Excel and Matlab to ascertain some knowledge about the Overlap RELEFF. Our findings are as follows:

- As λ increases, the RELEFF increases. For example, at $n_x = n_y = 20$ and $\lambda = 1.20$, the RELEFF is 26.32%, while at $\lambda = 1.6$ the same RELEFF is equal to 45.60%.
- Matlab shows that the asymptotic RELEFF is 100% as n_x & $n_y \to \infty$. The farther λ is from 1, the more rapidly the ARE approaches 100%.

4. The impact of Overlap on type I error PR of testing H_0: $\mu_x = \mu_y$ for unknown normal population variances and sample sizes ≥ 2

4.1. The case of $H_0: \sigma_x = \sigma_y = \sigma$ not rejected leading to the pooled t-test

In practice, a preliminary test on H_0: $\sigma_x^2 = \sigma_y^2 = \sigma^2$ is advisable before deciding whether to use the pooled t-test in preference to the two-independent samples t-test. For the sake of being conservative and cautious, we will use the pooled

Please cite this article as: Maghsoodloo, S., Huang, C.-Y., Comparing the overlapping of two independent confidence intervals with a single confidence interval for two.... J. Statist. Plann. Inference (2010), doi:10.1016/j.jspi.2010.04.057

t-test if the *P-value* of the pretest on H_0:$\sigma_x=\sigma_y=\sigma$ exceeds 20%, although some authors, such as Browne (1979, Sec. 3, p. 658), recommend just the nominal significance of 5% level for the pretest. Further, according to Devore (2008, p. 340) "the *F test of equal variances is quite sensitive to the assumption of normal population distributions much more so than t procedures.*" Accordingly, if n_x and n_y both are less than or near 10, pooling should be avoided unless the *P-value* of testing H_0:$\sigma_x=\sigma_y=\sigma$ exceeds 40%. When H_0:$\sigma_x=\sigma_y=\sigma$ is tenable, it is well-known that the CIL for $\mu_x-\mu_y$ is $2t_{\alpha/2,\nu}\times S_p\sqrt{1/n_x+1/n_y}$, i.e., H_0: $\mu_x=\mu_y$ must be rejected at the Standard LOS α if $|\bar{x}-\bar{y}|\geq t_{\alpha/2,\nu}\times S_p\sqrt{1/n_x+1/n_y}$, where $S_p^2=[(n_x-1)S_x^2+(n_y-1)S_y^2]/\nu$ and $\nu=(n_x+n_y-2)$. However, for the individual two t-CIs, the Overlap type I error Pr for testing H_0: $\mu_x=\mu_y$, bearing in mind that $t_\nu^2=F_{1,\nu}$, is given by [see Huang (2008), Chapter 6]

$$\alpha'=\Pr\left\{F_{1,n_x+n_y-2}>\nu(k\times t_{\alpha/2,\nu_x}+t_{\alpha/2,\nu_y})^2/[(\nu_y+F_0\nu_x)(1+R_n)]\right\} \tag{11a}$$

where $R_n=n_y/n_x$, and $k=\sqrt{R_nF_0}=(s_x\sqrt{n_y})/(s_y\sqrt{n_x})$ is the sample *se ratio*. For the pooled t-test, in the most common case of balanced sampling design, Eq. (11a) reduces to

$$\alpha'=\Pr\left[F_{1,2(n-1)}>F_{\alpha,1,n-1}\left(1+\sqrt{F_0}\right)^2/(1+F_0)\right], \tag{11b}$$

where the pretest statistic $F_0=S_x^2/S_y^2$ must range within the interval $(F_{0.90,n-1,n-1}, F_{0.10,n-1,n-1})$. The random function $(1+\sqrt{F_0})^2/(1+F_0)$ inside the argument of the RHS of (11b) attains its maximum at $F_0=1$ and its minimum at $F_{0.10,n-1,n-1}$, or $F_{0.90,n-1,n-1}$. As a result, the minimum value of α' occurs at $F_0=1$ and its maximum occurs at either $F_{0.10,n-1,n-1}$ or $F_{0.90,n-1,n-1}$. At the same F_0, α' in (11b) is a monotonically increasing function of n. As $n\to\infty$, α' in Eq. (11b) approaches 0.0055746, which is very close to the known-variance case of testing H_0: $\mu_x=\mu_y$. Eq. (11b) for Overlap type I error Pr is different from $1-\Pr(A)$ atop p. 549 of Payton et al. (2000) because theirs pertains to the general two-independent samples t-statistic, discussed in the next section, while (11b) pertains only to the pooled t-test. As before, it can be shown [for details see Huang (2008, Chapter 6)] that $\alpha'<\alpha$ for all finite $n>1$.

4.2. The case of H_0: $\sigma_x=\sigma_y$ rejected leading to the two-independent sample t-test

In this section, H_0:$\sigma_x=\sigma_y$ is rejected at the 20% level leading to the assumption that the F-statistic $F_0=S_x^2/S_y^2$ is outside the interval $(F_{0.90,n_x-1,n_y-1}, F_{0.10,n_x-1,n_y-1})$. It has been shown in statistical literature, Winer (1971), that if the assumption $\sigma_x=\sigma_y$ is not tenable, the random variable $[(\bar{x}-\bar{y})-(\mu_x-\mu_y)]/\sqrt{(S_x^2/n_x)+(S_y^2/n_y)}$ has an approximate Student's t-distribution with *df*

$$\nu=(S_x^2/n_x+S_y^2/n_y)^2\Big/\left[\frac{(S_x^2/n_x)^2}{n_x-1}+\frac{(S_y^2/n_y)^2}{n_y-1}\right]=\frac{\nu_y\nu_x(k^2+1)^2}{\nu_yk^4+\nu_x}, \tag{12}$$

the B.L. Welch approximation improving as sample sizes increase. Eq. (12) shows that ν depends only on n_x, n_y, and the sample *se ratio* $k=\sqrt{F_0R_n}=(S_x\sqrt{n_y})/(S_y\sqrt{n_x})$ and will always lie within the interval $\mathrm{Min}(\nu_x,\nu_y)<\nu\leq\nu_x+\nu_y$. When H_0:$\sigma_x=\sigma_y$ is rejected at the 20% level, the approximate LOS of testing H_0:$\mu_x-\mu_y=0$ from the Standard method is given by

$$\alpha\cong\Pr(t_\nu^2>t_{\alpha/2,\nu}^2\,|\,\mu_x-\mu_y=0)=\Pr(F_{1,\nu}>F_{\alpha,1,\nu}\,|\,\delta=0) \tag{13}$$

Note that because the sample mean and variance from an underlying normal population are independent, the LOS in (13) of testing H_0:$\mu_x-\mu_y=0$ is not altered based on the decision of pre-testing H_0:$\sigma_x=\sigma_y$. For disjoint t-CIs, it can be proven that

$$\alpha'\approx\Pr[F_{1,\nu}>(k\times t_{\alpha/2,\nu_x}+t_{\alpha/2,\nu_y})^2/(1+k^2)] \tag{14a}$$

When $n_x=n_y=n$, $R_n=1$, $k=S_x/S_y$, $k^2=F_0$, then Eq. (14a) reduces to

$$\alpha'\approx\Pr[F_{1,\nu}>F_{\alpha,1,n-1}\times(1+\sqrt{F_0})^2/(1+F_0)], \tag{14b}$$

and Eq. (12) simplifies to $\nu=(n-1)(1+F_0)^2/(1+F_0^2)$. Note that this last formula for ν reduces to $2(n-1)$ at $F_0=1$, which is the *df* of the pooled t-test, as it should because the unlikely realization $F_0=1$ (for which the *P-value*=100%) is in perfect agreement with H_0: $\sigma_x^2=\sigma_y^2$. Eq. (14a) shows that α' does not depend on the specific values of S_x^2 *and* S_y^2 but only on their ratio through $k=\sqrt{F_0R_n}$. For Payton et al.'s (2000) example of $n_1=n_2=n=10$, $S_1=0.80$, $S_2=1.60$, $F_0=(0.8/1.6)^2=0.25$, $\nu=(10-1)(1+0.25)^2/(1+0.25^2)=13.2353$, and at $\alpha=0.05$, the use of Eq. (14b) shows that the value of $\alpha'\approx0.00940573$, which is different from 0.0149 reported by Payton et al. (2000, p. 549). The *df* used by them was 9 ($=n-1$), which resulted in an imprecise value of $\alpha'=0.0149$, also partially due to small sample sizes.

4.3. Comparing the paired t-CI with two independent t-CIs from normal populations

Unlike the two independent samples t-CI for $\mu_x-\mu_y$ for a completely randomized design (CRD), the paired t-CI must be formed only for a randomized complete block design (RCBD), where the rvs X and Y are paired observations, or a random vector $[\boldsymbol{x}\quad\boldsymbol{y}]^T$, from a single bivariate normal population. The objective here is merely to assess the impact of correlation

Please cite this article as: Maghsoodloo, S., Huang, C.-Y., Comparing the overlapping of two independent confidence intervals with a single confidence interval for two.... J. Statist. Plann. Inference (2010), doi:10.1016/j.jspi.2010.04.057

on Overlap, because the paired and two-independent t-tests occur from two completely different sampling designs. It can be proven that for the two individual non-overlapping CIs

$$\alpha' = \Pr\left[F_{1,n-1} > F_{\alpha,1,n-1}\left(1+\sqrt{F_0}\right)^2 / \left(1+F_0-2r\sqrt{F_0}\right)\right] \tag{15}$$

Eq. (15) shows that the effect of negative correlation is to increase α' toward α, i.e., as $r \to -1$, $\alpha' \to \alpha$ for all n and F_0 so that Overlap becomes less and less deficient. When $r=0$, α' in Eq. (15) becomes similar to Eq. (14b), as it should because zero correlation implies independence in the case of underlying bivariate normal populations. Further, Eq. (15) shows that the impact of positive correlation is to reduce α' , i.e., Overlap becomes more and more deficient (or the power of Overlap $\to 0$ as $r \to 1$), consistent with Section (5.2) of Schenker and Gentleman (2001, p. 185).

5. The percent overlap that leads to rejection of H_0: $\mu_x = \mu_y$

5.1. The case of unknown normal populations with $\sigma_x = \sigma_y = \sigma$ and sample sizes ≥ 2

Throughout this section, it is understood that a pretest on H_0: $\sigma_x = \sigma_y = \sigma$ has yielded a *P-value* > 0.20 so that the null hypothesis H_0: $\sigma_x = \sigma_y = \sigma$ is tenable leading to a pooled t-test on H_0:$\mu_x = \mu_y$.

It can be proven that the exact % α-Overlap is given by

$$\omega = \left[\frac{(t_{\alpha/2,v_x}S_x/\sqrt{n_x} + t_{\alpha/2,v_y}S_y/\sqrt{n_y}) - (\bar{x}-\bar{y})}{(t_{\alpha/2,v_x}S_x/\sqrt{n_x} + t_{\alpha/2,v_y}S_y/\sqrt{n_y}) + (\bar{x}-\bar{y})}\right] \times 100\% \tag{16a}$$

As $\bar{x}-\bar{y} \geq 0$ increases, the *P-value* of the t-test decreases (i.e., H_0: $\mu_x = \mu_y$ must be rejected more strongly) and the statistic ω in Eq. (16a) decreases. Because H_0: $\mu_x = \mu_y$ must be rejected at the α-level iff $|\bar{x}-\bar{y}| \geq t_{\alpha/2,v} \times S_p\sqrt{1/n_x + 1/n_y}$, where $v = n_x + n_y - 2$, then on substitution of this last borderline value into (16a), H_0: $\mu_x - \mu_y = 0$ must be rejected at the level $\leq \alpha$ iff

$$\omega \leq \frac{k \times t_{\alpha/2,v_x} + t_{\alpha/2,v_y} - t_{\alpha/2,v}\sqrt{(1+R_n)(v_xF_0+v_y)/v}}{k \times t_{\alpha/2,v_x} + t_{\alpha/2,v_y} + t_{\alpha/2,v}\sqrt{(1+R_n)(v_xF_0+v_y)/v}} \times 100\% \tag{16b}$$

where $R_n = n_y/n_x$, $F_0 = S_x^2/S_y^2$ and $k = \sqrt{R_nF_0} = se(\bar{x})/se(\bar{y})$. Thus, the maximum percent overlap at or below which H_0 must be rejected at the α-level is given by

$$\omega_r = \left[\frac{k \times t_{\alpha/2,v_x} + t_{\alpha/2,v_y} - t_{\alpha/2,v}\sqrt{(1+R_n)(v_xF_0+v_y)/v}}{k \times t_{\alpha/2,v_x} + t_{\alpha/2,v_y} + t_{\alpha/2,v}\sqrt{(1+R_n)(v_xF_0+v_y)/v}}\right] \times 100\% \tag{16c}$$

Notice that unlike the maximum percent overlap in (16c) above which H_0 cannot be rejected, Browne (1979, p. 658) defines a measure, D, that gives the proportion of separation of the shorter to the longer interval. His Table 1, p. 660, provides the minimum values of D above which H_0: $\mu_x = \mu_y$ must be rejected at the 5% level. Further, Browne's measure D is not in general equal to $1-\omega_r$. For the case of balanced CRD, the maximum overlap at the rejection limits in (16c) reduces to

$$\omega_r = \left[\frac{t_{\alpha/2,n-1}(1+\sqrt{F_0}) - t_{\alpha/2,2(n-1)}\sqrt{1+F_0}}{t_{\alpha/2,n-1}(1+\sqrt{F_0}) + t_{\alpha/2,2(n-1)}\sqrt{1+F_0}}\right] \times 100\% \tag{16d}$$

As $n \to \infty$ and per force $F_0 \to 1$, $t_{\alpha/2} \to Z_{\alpha/2}$, and Eq. (16d) yields $\omega_r = [(2-\sqrt{2})/(2+\sqrt{2})] \times 100\% = 17.1573\%$, which is identical to the known-&-equal variances case given in Eq. (3d) at $K=1$. Note that unlike the case of known-variances, ω_r in (16c) is α-dependent unless n_x & $n_y > 150$.

It can be proven [see Huang (2008, Chapter 7)] that the exact value of γ for all sample sizes n_x & $n_y \geq 2$ can be obtained from the following equation:

$$t_{\gamma/2,v_y} + k \times t_{\gamma/2,v_x} = t_{\alpha/2,v} \times \sqrt{(R_n+1)(v_xF_0+v_y)/v} \tag{17a}$$

where $v = n_x + n_y - 2$ and $k = (S_x\sqrt{n_y})/(S_y\sqrt{n_x})$. When sampling is balanced, Eq. (17a) reduces to

$$F_{\gamma,1,n-1} = F_{\alpha,1,2(n-1)} \times (1+k^2)/(1+k)^2 \tag{17b}$$

For a 0.05-level t-test, $n_x = n_y = n$, and $F_0 = k^2 = 1$, the values of γ from (17b) range from 0.2021653 at $n-1=1$ down to 0.166305 at $n-1=100$. In order to obtain the limiting value of γ , we let $n \to \infty$ in (17a), and per force $F_0 \to 1$ because the pooled t-test requires that $\sigma_x^2 = \sigma_y^2$, resulting in the limit $t_{\gamma/2,n-1}(n \to \infty) = 1.959964/\sqrt{2} = 1.385904 \to$ Limit γ (as $n \to \infty$) $= \Pr(|Z| \geq 1.385904) = 0.16578$, which is identical to the known-&-equal population variances case obtained from Eq. (4) at $K=1$.

Please cite this article as: Maghsoodloo, S., Huang, C.-Y., Comparing the overlapping of two independent confidence intervals with a single confidence interval for two.... J. Statist. Plann. Inference (2010), doi:10.1016/j.jspi.2010.04.057

5.2. The case of H_0:$\sigma_x=\sigma_y$ rejected leading to Welch's Approximate two-sample t-test

Without proofs we provide all the pertinent formulas below.

$$\omega=\frac{(t_{\alpha/2,v_x}S_x/\sqrt{n_x}+t_{\alpha/2,v_y}S_y/\sqrt{n_y})-(\bar{x}-\bar{y})}{(t_{\alpha/2,v_x}S_x/\sqrt{n_x}+t_{\alpha/2,v_y}S_y/\sqrt{n_y})+(\bar{x}-\bar{y})}\times 100\% \tag{18a}$$

$$\omega_r=\left[\frac{(kt_{\alpha/2,v_x}+t_{\alpha/2,v_y})-t_{\alpha/2,v}\times\sqrt{1+k^2}}{(kt_{\alpha/2,v_x}+t_{\alpha/2,v_y})+t_{\alpha/2,v}\times\sqrt{1+k^2}}\right]\times 100\% \tag{18b}$$

where v is computed from Eq. (12) and H_0: $\mu_x=\mu_y$ must not be rejected iff % overlap exceeds (18b). When the sampling design from the two independent normal populations is balanced, the maximum % overlap in Eq. (18b) that still leads to the rejection of H_0: $\mu_x=\mu_y$ at the α-level reduces to

$$\omega_r=\left[\frac{t_{\alpha/2,n-1}(1+\sqrt{F_0})-t_{\alpha/2,v}\times\sqrt{1+F_0}}{t_{\alpha/2,n-1}(1+\sqrt{F_0})+t_{\alpha/2,v}\times\sqrt{1+F_0}}\right]\times 100\%, \tag{18c}$$

where in the balanced case $v=(n-1)(S_x^2+S_y^2)^2/(S_x^4+S_y^4)=(n-1)(1+F_0)^2/(1+F_0^2)$. When $n_x=n_y$, the value of ω_r in (18c) lies within the interval (0, 20.548028%), where zero pertains to the limiting value as $k=\sqrt{F_0}\to 0$ or ∞, and the upper limit pertains to $k^2=F_{0.90,3,3}$ or $F_{0.10,3,3}$. The limiting (as either $R_n=n_y/n_x$ or $k\to 0$ or ∞) value of (18b) is 0, so that the Overlap approaches an α-level test.

For the case of rejected H_0:$\sigma_x=\sigma_y$ at the 20% level, the value of γ for Welch's approximate t-test is obtained [see Huang (2008, Chapter 7)] from the following Eq. (19a):

$$t_{\gamma/2,v_y}+k\times t_{\gamma/2,v_x}=t_{\alpha/2,v}\times\sqrt{k^2+1} \tag{19a}$$

For the case of a balanced sampling design, (19a) reduces to

$$F_{\gamma,1,n-1}=F_{\alpha,1,v}\times(1+k^2)(1+\sqrt{F_0})^2 \tag{19b}$$

For example, using (19b) at $\alpha=0.05$, $n_x=n_y=n=10$, $k=\sqrt{F_0}=0.5$, $F_0=k^2=0.25$, from Eq. (12) the value of $v=13.2353$, and $F_{0.05,1,13.2353}=4.6503672$ results in $F_{\gamma,1,9}=2.58353732\to\gamma=\Pr(F_{1,9}\geq 2.583537)=0.142442$. Payton et al. (2000) report this last value as 0.1262 because the denominator df of $F_{\gamma,1,v}$ on the far RHS in the formula atop their page 550 should be $v=13.2353$, not 9 ($=n-1$) as reported. However, for larger n values, their formula becomes more accurate. The limiting value of γ in Eq. (19a), as $n\to\infty$, can easily be obtained from $Z_{\gamma/2}=Z_{\alpha/2}\times\sqrt{\sigma_x^2+\sigma_y^2}/(\sigma_x+\sigma_y)$. This last result is consistent with the case of $\sigma_x=\sigma_y$ of Eq. (4) because its solution at $\alpha=0.05$ is $\gamma=0.16578$.

5.3. Comparing paired t-CI with two independent t-CIs for underlying normal populations

For the paired t-test, the % α-Overlap is given by

$$\omega=\frac{t_{\alpha/2,n-1}(S_x+S_y)/\sqrt{n}-(\bar{x}-\bar{y})}{t_{\alpha/2,n-1}(S_x+S_y)/\sqrt{n}+(\bar{x}-\bar{y})}\times 100\% \tag{20a}$$

where the X-variate has the larger or equal mean. Because H_0: $\mu_d=0$ must be rejected at the level $\leq\alpha$ iff $\bar{x}-\bar{y}\geq t_{\alpha/2,n-1}\times S_d/\sqrt{n}$, then from (20a) H_0 must be rejected at the $\alpha\times 100\%$ level or less if

$$\omega\leq\left[\frac{t_{\alpha/2,n-1}(S_x+S_y)/\sqrt{n}-t_{\alpha/2,n-1}\times S_d/\sqrt{n}}{t_{\alpha/2,n-1}(S_x+S_y)/\sqrt{n}+t_{\alpha/2,n-1}\times S_d/\sqrt{n}}\right]\times 100\% \tag{20b}$$

Thus, the maximum % overlap below which H_0: $\mu_d=0$ must be rejected at the α-level is given by

$$\omega_r=\left[\frac{1+\sqrt{F_0}-\sqrt{1+F_0-2r\sqrt{F_0}}}{1+\sqrt{F_0}+\sqrt{1+F_0-2r\sqrt{F_0}}}\right]\times 100\% \tag{20c}$$

The % overlap in (20c) depends only on the correlation coefficient r and the ratio of the two observed standard deviations, i.e., it does not depend on α and specific values of S_x and S_y. It is interesting to note that when $r=0$ (i.e., X and Y are independent) and $F_0=1$, then Eq. (20c) reduces to 17.1573%.

Next, what should the individual confidence levels, $1-\gamma$, be so that the two independent CIs lead to an exact α-level test on H_0: $\mu_d=0$? It can be shown [see Huang (2008, Chapter 7)] that the value of γ is obtained from:

$$t_{\gamma/2,n-1}=t_{\alpha/2,n-1}\times\sqrt{1+F_0-2r\sqrt{F_0}}/(1+\sqrt{F_0}) \tag{21}$$

As $r\to -1$, Eq. (21) shows that $\gamma\to\alpha$, and consequently, the Overlap procedure becomes an exact α-level test, while if $r=1$, the RHS of (21) attains its minimum leading to maximum value of γ. When $r=0$ (i.e., independent X & Y), for very large or very small values of F_0 ($=k^2$), in the limit Overlap becomes an α-level test.

Please cite this article as: Maghsoodloo, S., Huang, C.-Y., Comparing the overlapping of two independent confidence intervals with a single confidence interval for two.... J. Statist. Plann. Inference (2010), doi:10.1016/j.jspi.2010.04.057

6. The impact of Overlap on β for unknown variances and $n \geq 2$

6.1. The case of H_0: $\sigma_x=\sigma_y=\sigma$ not rejected leading to the pooled t-test

If the assumption $\sigma_x=\sigma_y=\sigma$ is tenable and because statistical theory dictates that the total resources, $N=n_x+n_y$, should be allocated according to $n_x=\sigma_x N/(\sigma_x+\sigma_y)=N/2=n_y$, then the most common applications of the pooled t-test occur under equal sample sizes. Because type II error Pr can be computed only when a LOS is specified and $\alpha=0.05$ is nominal for most applications, throughout this section all computations will be performed for a 0.05-level test. However, our results are applicable to any LOS α by replacing 0.025 with $\alpha/2$.

It is well known [see Johnson et al. (1995), Chapter 31] that the Standard type II error rate, for underlying normal populations, is given by (a proof is also available on request)

$$\beta = \Pr\left\{-t_{0.025,\nu} \leq t'_{n_x+n_y-2}\left[\delta\sqrt{n_x n_y/(n_x+n_y)}/\sigma\right] \leq t_{0.025}\right\} \tag{22a}$$

where $t'_{n_x+n_y-2}[\delta\sqrt{n_x n_y/(n_x+n_y)}/\sigma]$ represents a noncentral $t-rv$ with $\nu=n_x+n_y-2$ df and noncentrality parameter $\xi=\delta\sqrt{n_x n_y/(n_x+n_y)}/\sigma$; further, $t_{0.025}=t_{0.025,\nu}$ only for notational convenience. Note that only when $\delta=0$, the $t'_{n_x+n_y-2}$ in (22a) becomes the central t and β becomes equal to $1-\alpha$. When the sampling design is balanced, the OC function of (22a) reduces to

$$\beta = \Pr\left[-t_{0.025} \leq t'_{2(n-1)}(\delta\sqrt{n/2})/\sigma \leq t_{0.025}\right] \tag{22b}$$

As an example, suppose we draw a random sample $n_x=9$ from a $N(\mu_x$, unknown $\sigma^2)$ and one of size $n_y=9$ from another $N(\mu_y, \sigma^2)$ with the objective of testing H_0:$\mu_x-\mu_y=0$ at the nominal significance level of 5% versus the 2-sided alternative H_1:$\mu_x-\mu_y\neq 0$. We wish to answer the question "what is the Pr of accepting H_0 if the true mean difference $\delta=\mu_x-\mu_y$ were not zero but were equal to 0.80σ?", i.e., we wish to compute the type II error Pr at $\delta=0.80\sigma$. Then, the value of the noncentrality parameter is equal to $\xi=(\delta/\sigma)\sqrt{n_x n_y/(n_x+n_y)}=(0.80\sigma/\sigma)\sqrt{81/18}=1.69706$ and the corresponding type II error Pr from Eq. (22b) is equal to β=Pr$(-t_{0.025,16}\leq t'_{16}(1.69706)\leq t_{0.025,16})$=$cdf$[of $t'_{16}(1.69706)$ at 2.119905]$-cdf$[of $t'_{16}(1.69706)$ at (-2.119905)]. Fortunately, both Minitab and Matlab provide the cdf of the noncentral t-distribution. Using Minitab, we obtain β(at $\delta=0.80\sigma$)=0.64205 so that the power of the pooled t-test at $\delta=0.80\sigma$ is equal to $1-\beta=0.35795$.

It can be proven that the corresponding type II error Pr for Overlap is given below.

$$\beta' = \Pr\left\{-A_P \leq t'_{n_x+n_y-2}\left[\delta\sqrt{n_x n_y/(n_x+n_y)}/\sigma\right] \leq A_P\right\}, \tag{23a}$$

where $A_p=(t_{\alpha/2,\nu_x}S_x/\sqrt{n_x}+t_{\alpha/2,\nu_y}S_y/\sqrt{n_y})/(S_p\sqrt{1/n_x+1/n_y})$. Note that if $\delta=0$, Eq. (23a) reduces to $1-\alpha'$ obtained from Eq. (11a). In the case of balanced design, Eq. (23a) reduces to

$$\beta' = \Pr\left[-\frac{t_{\alpha/2,n-1}(1+\sqrt{F_0})}{\sqrt{1+F_0}} \leq t'_{2(n-1)}\left(\frac{\delta}{\sigma}\sqrt{\frac{n}{2}}\right)) \leq \frac{t_{\alpha/2,n-1}(1+\sqrt{F_0})}{\sqrt{1+F_0}}\right] \tag{23b}$$

Unfortunately, unlike the Standard β from Eq. (22a), the Overlap type II error Pr in Eq. (23b) not only depends on the specified value of δ/σ but also on the realized value of F_0. Thus, it is impossible to compute a priori Pr of Overlap type II error, unless some rough value of F_0 is available; this implies that there are an uncountable number of Overlap OC curves at $\alpha=0.05$, each valid for a realized value of F_0.

As an example, suppose samples of sizes $n_x=n_y=9$ are drawn from two independent normal populations with unknown but equal population variances. We wish to compute the Pr of accepting H_0:$\mu_x-\mu_y=0$ at $\alpha=0.05$ if $\mu_x-\mu_y=0.80\sigma$ and the sample statistics are available as $S_x=0.650$ and $S_y=0.540$. Note that when $n_x=n_y=n$, it is sufficient to provide the statistic $F_0=S_x^2/S_y^2$ instead of the specific values of S_x and S_y. Because $t_{\alpha/2,n-1}=t_{0.025,8}=2.306004$, the noncentrality parameter $\xi=(0.80\sigma/\sigma)\sqrt{81/18}=1.6970563$, Eq. (23b) yields β'(at $\delta=\mu_x-\mu_y=0.80\sigma$, $n_x=9$, $n_y=9$)=Pr$[-3.247338\leq t'_{16}(1.6970563)\leq 3.247338]$=0.904231. This last value of β' is much larger than $\beta=0.64205$ from the Standard method, as expected. The % relative error depends on n_x, n_y and δ/σ.

6.2. The case of H_0:$\sigma_x=\sigma_y$ rejected leading to Welch's Approximate two-sample t-test

The formulas for degrees of freedom in Eq. (12) rarely lead to an integer and is generally rounded down to make the test on H_0:$\mu_x-\mu_y=0$ conservative, i.e., the rounding down ν in (12) increases the P-$value$ of t-test. However, programs like Matlab and Minitab provide the cdf and quantiles of the t-distribution for non-integer values of ν. It has been verified by the authors that ν in Eq. (12) attains its minimum when the smaller sample has much larger variance and vice a versa. Even then, it is for certain that $\text{Min}(\nu_x, \nu_y)<\nu\leq\nu_{x+}\nu_y$, and hence the two-sample approximate t-test, as expected and well known, is less powerful than the pooled t-test. When H_0:$\sigma_x=\sigma_y$ is rejected at the 20% level, the type II error Pr of a 5%-level test, for underlying independent normal populations, is approximately given by

$$\beta \approx \Pr(-t_{0.025,\nu} \leq t_0 \leq t_{0.025,\nu} \,|\, \mu_x-\mu_y=\delta), \tag{24}$$

Please cite this article as: Maghsoodloo, S., Huang, C.-Y., Comparing the overlapping of two independent confidence intervals with a single confidence interval for two.... J. Statist. Plann. Inference (2010), doi:10.1016/j.jspi.2010.04.057

where $t_0=(\bar{x}-\bar{y})/\sqrt{(S_x^2/n_x)+(S_y^2/n_y)}$ is approximately central t distributed when H_0 is true with df, v, given in Eq. (12). When H_0 is false, the authors have also verified that the exact SMD of the statistic $t_0=(\bar{x}-\bar{y})/\sqrt{(S_x^2/n_x)+(S_y^2/n_y)}$ under the alternative H_1: $\mu_x-\mu_y=\delta\neq 0$, unlike the case of $\sigma_x=\sigma_y=\sigma$, is intractable using a central χ_v^2, due to the fact that t_0 is not a likelihood ratio test statistic. As far as we know, the exact PWF of the t-Prime test (or the two-independent samples t-test) has not yet been derived in statistical literature. The following, results existing in statistical literature, is only an approximation because, to our knowledge, there does not exist an exact equation for type II error Pr of testing H_0:$\mu_x-\mu_y=0$ when the two underlying normal population variances are unknown and also unequal. Consequently,

$$\beta\approx \Pr(-t_{0.025}-\Delta\le t_v\le t_{0.025}-\Delta), \tag{25}$$

where the Studentized true mean difference $\Delta=\delta/\sqrt{(S_x^2/n_x)+(S_y^2/n_y)}$, and v is computed from (12). Unfortunately, the approximate expression for β in Eq. (25) still depends on the sample $se(\bar{x}-\bar{y})=\sqrt{(S_x^2/n_x)+(S_y^2/n_y)}$, and therefore, the approximation in (25) can be carried out if δ is specified in terms of $se(\bar{x}-\bar{y})$, or in units of $\mu_x-\mu_y$, in which case the realized values of S_x^2 and S_y^2 have to be used posteriori in order to approximate a priori type II error probability.

For example, suppose samples of sizes $n_x=11$ and $n_y=7$ are drawn from two independent normal populations with unknown and unequal variances. We wish to compute the Pr of accepting H_0:$\mu_x-\mu_y=0$ at $\alpha=0.05$ if $\mu_x-\mu_y=\delta=0.4$, the sample statistics are $S_x=0.880$, $S_y=0.540$, $F_0=2.65569$ (note that this is not significant at the 20% level but sample sizes are too small), and the sample se ratio $k=1.30000$. Eq. (12) gives $v=60(1.3000^2+1)^2/(6\times 1.3000^4+10)=15.99928$ df, $t_{0.025}=t_{0.025,15.99928}=2.119913$, $\Delta=(\mu_x-\mu_y)/\sqrt{(S_x^2/n_x)+(S_y^2/n_y)}=1.194924$, so that $-t_{0.025}-\Delta=-3.314837$, $t_{0.025}-\Delta=0.92499$, and β (at $\delta=0.40$) $\approx \Pr(-3.314837\le t_{15.99928}\le 0.92499)=0.8156432-0.00219022=0.813453$. Note that this last approximate type II error Pr would be in precise agreement with what UCLA's Statistics Department Power Calculator may have on their website (www.stat.ucla.edu), as stated by Devore (2008, pp. 340-341), because Eq. (25) exactly checked against Devore's (2008) examples atop his p. 341, to 4 decimals, which used the Power Calculator.

The type II error Pr from the Overlap is given below [for details see Huang (2008), Chapter 8].

$$\beta'\approx \Pr[(-A-\delta)/se(\bar{x}-\bar{y})\le t_v\le (A-\delta)/se(\bar{x}-\bar{y})], \tag{26}$$

where $A=t_{\alpha/2,v_x}S_x/\sqrt{n_x}+t_{\alpha/2,v_y}S_y/\sqrt{n_y}$ and v is given by Eq. (12). For example, if $\delta=0.40$, $n_x=11$, and $n_y=7$, $S_x=0.880$ and $S_y=0.540$, $A=1.090609$, $v=15.9993$ as before, $se(\bar{x}-\bar{y})=0.33475$, and Eq. (26) now gives β'(at$\delta=0.40$) $\approx$ $\Pr[-1.49061/se(\bar{x}-\bar{y})\le t_{15.9993}\le 0.69061/0.33475]=0.971937$, as compared to the Standard value of β(at $\delta=0.40$)$=0.813453$, resulting in a % relative error of 19.483% in type II error Pr.

6.3. The impact of Overlap on type II error probability for correlated samples

Consider a 5%-level test of H_0:$\mu_x-\mu_y=\mu_d=0$ versus the 2-sided alternative H_1: $\mu_d\neq 0$, where the paired response (x, y) comes from a bivariate normal universe so that X and Y are correlated random variables with unknown correlation coefficient ρ. It is widely known that for a 5%-level test

$$\beta=\Pr[-t_{0.025,n-1}\le t'_{n-1}(\xi)\le t_{0.025,n-1}] \tag{27}$$

Eq. (27) shows that the exact SMD of $t_0=\bar{d}\sqrt{n}/S_d$ under the alternative H_1: $\mu_d\neq 0$ is the noncentral t with $v=n-1$ df and noncentrality parameter $\xi=\mu_d\sqrt{n}/\sigma_d$ (a proof is also available on request), while the null SMD of t_0 is the central $t_{n-1}=t'_{n-1}(0)$. For example, suppose we wish to compute the type II error Pr when testing H_0: $\mu_d=0$ at the 5% level with a random sample of size $n=10$ blocks from a bivariate normal population versus the alternative H_1: $\mu_d=0.50\sigma_d$. From Eq. (27), β(at $\mu_d=0.50\sigma_d$)$=\Pr(-t_{0.025,9}\le t'_{n-1}(\xi)\le t_{0.025,9})$, where $\xi=0.50\sigma_d\sqrt{10}/\sigma_d=1.581139$. Consequently using Matlab, we obtain β (at $\mu_d=0.50\sigma_d$, $n=10$)$=\Pr(-2.262157\le t'_9(1.581139)\le 2.262157)=0.7068244$.

It can be proven [see Huang (2008, Chapter 8)] that the type II error Pr for the paired t-test using the Overlap is given by

$$\beta'=\Pr\left[-t_{\alpha/2,n-1}\left(\frac{1+\sqrt{F_0}}{\sqrt{1+F_0-2r\sqrt{F_0}}}\right)\le t'_{n-1}(\xi)\le t_{\alpha/2,n-1}\left(\frac{1+\sqrt{F_0}}{\sqrt{1+F_0-2r\sqrt{F_0}}}\right)\right] \tag{28}$$

where $\xi=\mu_d\sqrt{n}/\sigma_d$ as before. Eq. (28) shows that the impact of negative r is to reduce the Overlap type II error Pr while the impact of positive correlation is to increase β'. As $r\to 0$, β' in (28) does not approach β' in (23b) because (28) was derived from a RCBD, while (23b) was derived from a CRD.

7. Summary and conclusions

Section 2 used the normal underlying populations with known variances to prove asymptotic results that already existed in Overlap literature, some of which were also obtained only through simulation. It was proven that for the nominal significance level $\alpha=0.05$, the corresponding 95% overlapping CIs provide a much smaller LOS $\alpha'=0.0055746$. Although, the Overlap literature has never considered the one-sided alternative, we showed that the Overlap LOS is ½ of

Please cite this article as: Maghsoodloo, S., Huang, C.-Y., Comparing the overlapping of two independent confidence intervals with a single confidence interval for two.... J. Statist. Plann. Inference (2010), doi:10.1016/j.jspi.2010.04.057

the corresponding two-sided alternative (i.e., the Overlap procedure becomes even more conservative for a one-sided case). Second, a concept that had not been discussed in Overlap literature is the maximum % overlap that two independent CIs can have below which the null hypothesis of equality of means or variances must still be rejected at a pre-assigned LOS α. It was proven that this maximum % overlap depends only on the standard error ratio, K, and in the case of known-variances is equal to 17.1573% only at $K=1$, and uniformly diminishes to zero as $K\to 0$ or ∞, i.e., the Overlap slowly converges to an exact α-level test for the limiting values of K.

Section 3 examined the overlapping CIs for the ratio of two normal population variances against the Standard method that uses Fisher's F_{v_x,v_y} distribution. Statistical literature had not investigated the Overlap procedure for the ratio of two variances. As in the case of process means, Overlap greatly reduces the LOS of the test, and the limiting value (i.e., as $n_x=n_y\to\infty$) of α' at $\alpha=0.05$ is also 0.0055746, while at $n_x=n_y=2$, $\alpha'=0.0178$. Further, as $n_x/n_y\to\infty$ or 0, $\alpha'\to\alpha$.

Section 4 examined the impact of Overlap on type I error Pr for underlying normal populations with unknown variances, but sample sizes n_x & $n_y\geq 2$, using the pooled t and two-independent sample t statistics, and also examined the effect of negative and positive correlations on the Overlap procedure. We also highlighted some inexact results reported in Overlap literature for small sample sizes.

Section 5 derived the % overlap and the individual confidence levels of two independent CIs when variances are unknown for all sample sizes ≥ 2.

Section 6 used the noncentral t-distribution to obtain formulas for the OC curve (and also power function) of Overlap in the case of underlying normal populations with unknown variances and sample sizes n_x and $n_y\geq 2$, which also holds approximately true when the underlying distributions are non-normal but both n_x & $n_y>60$.

References

Browne, R.H., September 1979. On visual assessment of the significance of a mean difference. Biometrics 35, 657–775.

Devore, J.L., 2008. In: Probability and Statistics. Thomson Brooks/Cole, Canada.

Huang, C.-Y., 2008. Comparing the overlapping of two independent confidence intervals with single confidence interval for two normal population parameters. Ph.D. Dissertation, Industrial Engineering Department, Auburn University, Alabama.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1995. In: Continuous Univariate Distributions 2nd edition John Wiley&Sons, Inc.

Payton, M.E., Miller, A.E., Raun, W.R., 2000. Testing statistical hypotheses using standard error bars and confidence intervals. Communication in Soil Science and Plant Analysis 31, 547–552.

Payton, M.E., Greenstone, M.H., Schenker, N., 2003. Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance? The Journal of Insect Science 3, 34–39

Schenker, N., Gentleman, J.E., 2001. On judging the significance of differences by examining the overlap between confidence intervals. The American Statistician 55, 182–186.

Winer, B.J., 1971. Statistical Principals in Experimental Design. McGraw Hill, New York.

Please cite this article as: Maghsoodloo, S., Huang, C.-Y., Comparing the overlapping of two independent confidence intervals with a single confidence interval for two.... J. Statist. Plann. Inference (2010), doi:10.1016/j.jspi.2010.04.057