COMPARING THE OVERLAPPING OF TWO INDEPENDENT CONFIDENCE INTERVALS WITH A SINGLE CONFIDENCE INTERVAL FOR TWO NORMAL POPULATION PARAMETERS

Except where reference is made to the work of others, the work described in this dissertation is my own or was done in collaboration with my advisory committee. This dissertation does not include proprietary or classified information.

Ching Ying Huang

Certificate of Approval:

Alice E. Smith
Professor
Industrial and Systems Engineering

Saeed Maghsoodloo, Chair
Professor
Industrial and Systems Engineering

Kevin T. Phelps
Professor
Mathematics and Statistics

George T. Flowers
Dean
Graduate School

COMPARING THE OVERLAPPING OF TWO INDEPENDENT CONFIDENCE INTERVALS WITH A SINGLE CONFIDENCE INTERVAL FOR TWO NORMAL POPULATION PARAMETERS

Ching-Ying Huang

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Doctor of Philosophy

Auburn, Alabama
December 19, 2008

COMPARING THE OVERLAPPING OF TWO INDEPENDENT CONFIDENCE INTERVALS WITH A SINGLE CONFIDENCE INTERVAL FOR TWO NORMAL POPULATION PARAMETERS

Ching Ying Huang

____
Signature of Author

____
Date of Graduation

VITA

Ching-Ying Huang, daughter of Shuh-Peir Huang and Yueh-E Lin, was born April 14, 1979, in Taipei, Taiwan. She entered Chang Gung University at Taoyuan, Taiwan in September, 1998 and received the Bachelor of Science degree in Business Administration June, 2002. She started her graduate program in Industrial and Systems Engineering at Auburn University August 2003. She married Chih-Wei Chiang on December 30, 2006.

DISSERTATION ABSTRACT

COMPARING THE OVERLAPPING OF TWO INDEPENDENT CONFIDENCE INTERVALS WITH A SINGLE CONFIDENCE INTERVAL FOR TWO NORMAL POPULATION PARAMETERS

Ching-Ying Huang

Doctor of Philosophy, December 19, 2008
(B.S., Chang Gung University, 2002)

156 Typed Pages

Directed by Saeed Maghsoodloo

Two overlapping confidence intervals have been used in many sources in the past 30 years to conduct statistical inferences about two normal population means (μ_x and μ_y). Several authors have examined the shortcomings of Overlap procedure in the past 13 years and have determined that such a method completely distorts the significance level of testing the null hypothesis H_0: $\mu_x = \mu_y$ and reduces the statistical power of the test. Nearly all results for small sample sizes in Overlap literature have been obtained either by simulation or by somewhat inaccurate formulas, and only large-sample (or known-variance) exact information has been provided. Nevertheless, there are many aspects of Overlap that have not yet been presented in the literature and compared against the standard statistical procedure. This paper will present exact formulas for the % overlap,

ranging in the interval (0, 61.3626%] for a 0.05-level test, that two independent confidence intervals (CIs) can have, but the null hypothesis of equality of two population means must still be rejected at a pre-assigned level of significance α for sample sizes ≥ 2.

The exact impact of Overlap on the α-level and the power of pooled-t test will also be presented. Further, the impact of Overlap on the power of the F-statistic in testing the null hypothesis of equality of two normal process variances will be assessed. Finally, we will use the noncentral t distribution, which has never been applied in Overlap literature, to assess the Overlap impact on type II error probability when testing H_0: $\mu_x = \mu_y$ for sample sizes n_x and $n_y \geq 2$.

ACKNOWLEDGEMENT

The author would like to thank her parents, Shuh-Peir Huang and Yueh E Lin, and two sisters, Yu-Hsin and Yu-Li Huang, for their positive attitude and encouragement. Especially, she dedicates this work to her husband, Chih-Wei Chiang, whose patience and moral support have made it possible.

The author expresses her gratitude to Professor Saeed Maghsoodloo for his experienced knowledge and assistance on this research and Professor Alice E. Smith for her support during the author's graduate program. The author is also thankful to Professor Kevin Phelps for helpful suggestions.

Computer software used: MS Excel

MS Word

MathType

Matlab

TABLE OF CONTENTS

LIST OF TABLES

LIST OF NOTATION

SMD	sampling distribution
μ_x	mean of population X
μ_y	mean of population Y
$\bar{x}$	mean of sample X
$\bar{y}$	mean of sample Y
S_x^2	variance of sample X = $\sum_{i=1}^{n_x}(x_i-\bar{x})^2/(n_x-1)$
S_y^2	variance of sample Y
σ_x^2	variance of population X
σ_y^2	variance of population Y
H_0	null hypothesis
H_1	alternative hypothesis
CI	confidence interval
CIL	confidence interval length
$L(\mu)$	lower $(1-\alpha)$% CI limit for μ
$U(\mu)$	upper $(1-\alpha)$% CI limit for μ
A_L	lower bound for acceptance interval

A_U	upper bound for acceptance interval
Pr	probability
LOS	level of significance
α	type I error Pr = Pr(reject H_0 | H_0 is true) by the Standard Method
α'	type I error Pr from two overlapping CIs
α'_1	type I error Pr from two overlapping CIs for the one-sided alternative
α'_B	type I error Pr using the Bonferroni procedure
β	type Π error Pr = Pr(not rejecting H_0 | H_0 is false)
β'	type Π error Pr from overlapping CIs
β_1	type Π error Pr for the one-sided alternative
β'_B	Bonferroni type Π error Pr
SE	population standard error
se	sample standard error
K	Standard Error ratio for populations, K= $(\sigma_X / \sqrt{n_X})/(\sigma_Y / \sqrt{n_Y})$
k	standard error ratio for samples, k = $(S_x / \sqrt{n_x})/(S_y / \sqrt{n_y})$
df	degrees of freedom; the symbol ν will denote df
OC Curve	operating characteristic curve
O	amount of overlap length between the two individual CIs
O_r	borderline value of Δ at which H_0 is barely rejected at the LOS α .
ω	the exact percentage of the overlap
ω_r	maximum percent overlap below which H_0 must be rejected at or below α

$N(\mu, \sigma^2)$	a normal Pr density function (pdf) with population mean μ and population variance σ^2
$\Phi(z)$	the cumulative distribution function (cdf) of the standard normal density at point z.
PWF	power Function (The graph of $1-\beta$ versus the parameter under H_0)
F_0	the ratio of two sample variances, S_x^2 / S_y^2
R_n	the ratio of two sample sizes, n_y / n_x
δ	$\mu_x - \mu_y$
ξ	equals $(\mu_x - \mu_y) / \sqrt{\sigma^2 / n_x + \sigma^2 / n_y}$, which represents the noncentrality parameter of the pooled-t statistic
Δ	studentized $\delta = \mu_x - \mu_y$ when $\sigma_x^2 \neq \sigma_y^2$, i.e., $\Delta = (\mu_x - \mu_y) / \sqrt{(S_x^2 / n_x) + (S_y^2 / n_y)}$
LUB	least upper bound
GLB	greatest lower bound
RELEFF	relative efficiency
ARE	asymptotic RELEFF

1.0 Introduction

When testing the equality of means of two processes, the sampling distribution (SMD) of the difference of two sample means must be used to conduct statistical inference (confidence intervals and test of hypothesis) about the corresponding processes' 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 are used to construct separate confidence intervals for μ_x and μ_y and examine the amount of overlap of the individual confidence intervals in order to make statistical inferences about $\mu_x - \mu_y$. If the underlying distributions are normal with known variances, exact relationships are given by Schenker and Gentleman, (2001) about the changes in the type I and II error probabilities if the overlapping of individual confidence intervals are used to make inferences about $\mu_x - \mu_y$ at the 5% level. Because there is no mention of proof in the above article, we will use the normal theory to generalize their formulas in chapter 3 for any LOS 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 exactly at 83.42237%, which is nearly consistent with the 85% reported by Payton *et al.* (2000).

When the process variances are unknown and sample sizes are small (i.e., the real-life encountered cases), this dissertation will obtain exact formulas for type I and II error probabilities whose values can be obtained once the unbiased estimators, S_x^2 and S_y^2, of process variances are realized. However, this dissertation will verify that in general

Using individual confidence intervals diminishes type I error rate, depending on sample sizes n_x and n_y, and increases type II error probability. Assessment of type Π error probability (β) for the general unknown variances case has not been investigated in the literature because the computation of type II error probability requires the use of noncentral t-distribution, although Schenker and Gentleman (2001) provide the impact of Overlap on the Power Function (PWF =1 − β) only for the limiting case in terms of n_x and n_y (or the known-variances case). The noncentral t-distribution has wide-spread applications when testing a hypothesis about one or two normal means. Specifically both the OC (Operating Characteristic) and Power Function (PWF =1 − β) for testing H_0: $\mu = \mu_0$ and H_0: $\mu_x - \mu_y = \delta_0$ (in the unknown variance cases) are constructed using the noncentral t-distribution. We will use the noncentral t-distribution to obtain the PWF of testing H_0: $\mu_x - \mu_y = 0$ (in the unknown variance cases) both using the SMD of $\bar{x} - \bar{y}$ (i.e., the Standard method which has been available in statistical literature for well over 50 years) and also the Overlap for sample sizes ≥ 2. It will be determined that the type II error rate always increases if individual confidence intervals are used to make inferences about $\mu_x - \mu_y$. 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 the t-distribution requires the

---

* Kendal and Stuart (1963, Vol.1, p.135) report that "The description of the distribution as the "normal," due to Karl Pearson (who is known for the definition of product-moment correlation coefficient and Pearson System of Statistical distributions), is now almost universal among English writers. Continental writers refer to it variously as the second law of Laplace, the Laplace distribution, the Gauss distribution, the Laplace-Gauss distribution and the Gauss-Laplace distribution. As an approximation to the binomial it was reached by DeMoivre in 1738 but he did not discuss its properties."

assumption that only sample means be approximately normally distributed (due to the Central Limit Theorem).

Investigation of the overlapping CIs is worthy because as Schenker and Gentleman (2001) mentioned in the article "On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals" that there are many articles, such as Mancuso (2001), that still use the Overlap method for testing the equality of two population quantities. Although we found some articles, such as Payton *et al.* (2000) entitled *"Testing Statistical Hypotheses Using the Standard Error Bars and Confidence Intervals"* that 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 is to investigate the exact differences between the Overlap method and the Standard [a term coined by Schenker and Gentleman (2001)] method for testing the null hypotheses H_0: $\sigma_x = \sigma_y$ and H_0: $\mu_x = \mu_y$ under different assumptions. The former hypothesis has never been investigated with the Overlap method. The statistical literature reports results for the impact of Overlap on type I and II error probabilities in testing H_0: $\mu_x = \mu_y$ only for the case of large sample sizes (i.e., the limiting case where n_x and $n_y \to \infty$). Therefore, this work will investigate the same and other aspects of Overlap but for small sample sizes (i.e., $n \le 20$, which also will hold true for moderate and large sample sizes). To be on the conservative side, we refer to $n \le 20$ as small, $20 < n \le 50$ as moderate and $n > 50$ as large in this dissertation, although some statisticians prefer $n > 60$ as large because for $n > 60$, $t_{\alpha,\nu} \cong Z_\alpha$ to one decimal place, where Z_α represents the $(1 - \alpha)$ quantile of a standard normal deviate.

The contents of different chapters are as follows: In chapter 2, an extensive literature survey and the results thus far are provided. In chapters 3.1 and 3.2, the known-variance case is discussed and compared with what has been reported without proof in the literature for the limiting case. In chapter 4, the Bonferroni method is compared against the overlap. In chapter 5, the statistical inference on the ratio of two process variances (σ_x^2 / σ_y^2) from the Overlap is compared against the Standard method. Chapter 6 discusses the impact of Overlap on type I error probability. Chapter 7 discusses the amount and % overlap required to reject H_0: $\mu_x = \mu_y$ at the α-level of significance when process variances are unknown and samples sizes are small and moderate. Similarly, chapter 8 considers the impact of Overlap on type II error Pr when process variances are unknown for n_x and $n_y \leq 50$. Finally, chapter 9 summarizes the dissertation findings.

In summary, the primary objectives of this dissertation are: (1) To examine the impact of Overlap procedure on type I error probability (Pr) when testing equality of two process variances or two population means for unknown process variances and sample sizes ≥ 2. Payton *et al.* (2000) obtained results for the latter objective but there are inaccuracies (for $n < 50$) in their development; further, the former objective has not been investigated. Moreover, the Overlap literature has not considered the case of pooled t-test and little has been mentioned by Schenker and Gentleman (2001) about the paired t-test. (2) To determine the maximum % overlap of two individual confidence intervals (CIs) below which the null hypothesis (either H_0: $\sigma_x = \sigma_y$ or H_0: $\mu_x = \mu_y$) cannot still be rejected at a given level of significance (LOS) α. This Objective has not yet been investigated. (3) To examine the impact of Overlap procedure on type II error Pr for sample sizes n_x and $n_y \geq 2$. Schenker and Gentleman (2001) carried out this last objective only for limiting case (i.e., as n_x and $n_y \rightarrow \infty$, and or known σ_x and σ_y).

The above objectives are worthy of further investigation because there are many researchers who still use the overlapping CIs to test hypotheses, especially in the biology and medical papers (see the references mentioned in chapter 3.1). Furthermore, some statistical software, such as Minitab, still exhibit overlapping CIs that may lead users to wrong conclusions. Although the CI for two population quantities is the common method to make decisions regarding H_0: $\sigma_x = \sigma_y$ or H_0: $\mu_x = \mu_y$, our objective is to ascertain the exact relationship between the overlapping of two individual CIs and the corresponding single CI. Most former researches have only discussed the limiting case (i.e., as n_x and $n_y \rightarrow \infty$). In the real-life situations, sufficient resources may not be available to gather very large samples. Thus, the case of small ($n \leq 20$) to moderate sample sizes ($20 < n \leq 50$) is a major contribution of this dissertation.

The reader should bear in mind that all primed symbols in this dissertation pertain to the Overlap method.

2.0 Literature Review

It has been well known that when two underlying populations are normal, the null hypothesis H_0: $\mu_x = \mu_y$ is tested, in case of known variances, using the sampling distribution of $\overline{X} - \overline{Y}$, which is also the Laplace-Gaussian $N(\mu_x - \mu_y,\ \sigma_x^2 / n_x + \sigma_y^2 / n_y)$. However, in practice, rarely the population variances are known. Thus, the equality of two process variances should first be tested with an F-statistic. If H_0: $\sigma_x^2 / \sigma_y^2 = 1$ is not rejected (and *P-value* > 0.20), the two-sample pooled-t procedure will be applied for testing H_0: $\mu_x = \mu_y$. Otherwise, the two-independent sample t-statistic has to be used to perform statistical inferences about $\mu_x - \mu_y$. In case of related samples (or paired observations), the paired t-statistic has to be used to conduct statistical inferences about $\mu_x - \mu_y$.

The above rules are the formal (or Standard) procedures for testing H_0: $\mu_x = \mu_y$. What if we discuss this question with two individual relevant intervals? Cole *et al.* (1999) mentioned that using two individual CIs to test the null hypothesis H_0: $\mu_x = \mu_y$ would lead to a smaller type I Error and larger type Π error rate than the formal procedures. Payton *et al.* (2000) pointed out that many researchers use the standard error bars (sample mean $\pm$ standard error of the mean) to test the equality of two population means. Therefore, if the two individual standard error bars fail to overlap, they will conclude that two sample means are significantly different. Actually, these researchers are making a test of hypothesis with an approximate Pr(type I error) = $\alpha = 0.16$ not $\alpha = 0.05$. Payton *et al.* (2000) also derived a formula for the probability of the overlap from two individual CIs. Payton *et al*. (2000) defined A to be the event that confidence intervals computed individually for two population means to overlap. Thus, if the sample sizes are equal ($n_1 = n_2 = n$), and population variances are unknown, they deduced that $\Pr(A) = \Pr$ $\left[\frac{n(\overline{Y_1} - \overline{Y_2})^2}{S_1^2 + S_2^2} < F_{\alpha,1,n-1}\frac{(S_1 + S_2)^2}{S_1^2 + S_2^2}\right]$. They state that the random variable $\frac{n(\overline{Y_1} - \overline{Y_2})^2}{S_1^2 + S_2^2}$ has the F-distribution with numerator degrees of freedom (df) $\nu_1 = 1$ and denominator df $\nu_2 = (n - 1)$ if the two samples are from the same normal population. It will be shown in Chapter 6 that their above statement is inaccurate. The two samples need not originate from the same population, and that the denominator df of the F-distribution is not (n − 1) but rather in the case of $n_1 = n_2 = n$ it is given by $\nu = \frac{(n-1)(S_1^2 + S_2^2)^2}{S_1^4 + S_2^4}$, where $(n - 1) < \nu < 2(n - 1)$. Further, they state that if the two samples are from two different normal populations with the same mean but unequal variances, the quantity $\frac{n(\overline{Y_1} - \overline{Y_2})^2}{S_1^2 + S_2^2}$ is still approximately F-distributed with $\nu_1 = 1$ and $\nu_2 = (n - 1)$ df, where their value of $\nu_2 = (n - 1)$ df is accurate only in the limiting case. Therefore, they conclude that

$$\Pr(A) = \Pr(Intervals\_overlap) \cong \Pr\left[F_{1,n-1} < F_{\alpha,1,n-1}\left(1 + \frac{2S_1S_2}{S_1^2 + S_2^2}\right)\right].$$

Payton *et al*. (2000) further state that for the 95% CIs, $n_x = n_y = n = 10$, S_1 (= S_x) = 0.80 and $S_2 = 1.60$, $1 - \Pr(A) = 1 - \Pr(Intervals\_overlap) \cong 1 - \Pr(F_{1,9} < F_{0.05,1,9} \times 1.8) = 1 -$

0.9859 = 0.0141 (which was misprinted as 0.0149). It will be shown in Chapter 6 that this last Overlap Pr should be revised to $\alpha' = 0.00608057$, i.e., their result has a relative error of 56.8754%

Moreover, Payton *et al*. (2000) used SAS Version 6.11 to simulate from a N(0, 1) when sample sizes varied from n = 5 to n = 50 in order to ascertain the accuracy of the above formula. In this article, the authors do not give information about the known variances case. For the unknown variances case, they only consider the case when the sample sizes are equal. The largest sample size Payton *et al*. (2000) considered was n = 50. Furthermore, Schenker and Gentleman (2001) found more than 60 articles in the health sciences for testing the equality of two population means by using the Overlap method. Schenker and Gentleman (2001) state that the Overlap method will fail to reject H_0 when the Standard method would reject it. In other words, the Overlap will lead to less statistical power than the Standard method. The authors considered three population quantities Q_1, Q_2 and $Q_1 - Q_2$. They state that Brownlee (1965) provided the 95% confidence intervals for the three quantities as $\widehat{Q}_1 \pm 1.96\widehat{SE}_1$, $\widehat{Q}_2 \pm 1.96\widehat{SE}_2$ and $(\widehat{Q}_1 - \widehat{Q}_2) \pm 1.96\sqrt{\widehat{SE}_1^2 + \widehat{SE}_2^2}$. However, using the Overlap method, the null hypothesis will not be rejected if and only if $(\widehat{Q}_1 - \widehat{Q}_2) \pm 1.96(\widehat{SE}_1 + \widehat{SE}_2)$ contains zero. Schenker and Gentleman (2001) defined k as the limiting *SE* (standard error) ratio, i.e., either SE_1/SE_2 or SE_2/SE_1, and considered only ratios that are greater than or equal to 1. For a limiting *SE* ratio of k and a standardized difference of $d = (Q_1 - Q_2)/\sqrt{SE_1^2 + SE_2^2}$, they reported that the asymptotic power for the standard method is $\Phi(-1.96 + d) + \Phi(-1.96 - d)$, where Φ represents the cdf of N(0, 1), and the asymptotic power for the

Overlap method is $\Phi(\frac{-1.96(1+k)}{\sqrt{1+k^2}}+d)+\Phi(\frac{-1.96(1+k)}{\sqrt{1+k^2}}-d)$. Note here, in this dissertation, we use different definition for k from Schenker's definition. Schenker and Gentleman (2001) use small case k as the standard error ratio for the limiting sample sizes or known variances cases but we use k as the standard error ratio for small to moderate sample sizes or unknown variances cases. It means k=$(S_x/\sqrt{n_x})/(S_y/\sqrt{n_y})$ in this dissertation. Therefore, for distinguishing, let K = $(\sigma_x/\sqrt{n_x})/(\sigma_y/\sqrt{n_y})$ represent the limiting sample sizes or known variances cases. In this chapter, we still use Schenker's symbol to represent their work. Schenker and Gentleman (2001) also stated that for the Pr of type I error, just simply let d = 0 in the above formulas. Then, the authors concluded that the Overlap method will lead to smaller α and larger β.

Furthermore, Schenker and Gentleman (2001) state that when SE_1 is nearly equal to SE_2, the Overlap method is expected to be more deficient (i.e., smaller type I error Pr and larger type II error Pr) relative to the Standard method. In this article, the authors did not give specific values of type I error and type Π error probabilities for different k and d values. Their results pertain only to large sample sizes so that the need for using the t-distribution in the case of small and moderate sample sizes was not discussed.

Payton *et al.* (2003) continued to provide the formula Pr(*Intervals_Overlap*), which they had also obtained in the year 2000 as follows:

$$\Pr(A)=\Pr(Intervals\_overlap)\cong\Pr\left[F_{1,n-1}<F_{\alpha,1,n-1}\left(1+\frac{2S_1S_2}{S_1^2+S_2^2}\right)\right].$$

Payton *et al*. (2003) state that a large-sample version of the above statement can be derived (assuming the two populations are identical):

$\Pr(A) = \Pr(Intervals\_overlap) \cong \Pr\left[|Z| < z_{\alpha/2}\sqrt{2}\right] = \Phi(\sqrt{2}\,Z_{\alpha/2}) - \Phi(-\sqrt{2}\,Z_{\alpha/2})$, where

$\Phi(z)$ (almost) universally represents the cumulative of the standardized normal density function at point z. The authors set α at the nominal value of 5%, generated 95% confidence intervals and gave the approximate probability of overlap as

$$\Pr(Intervals\_overlap) \cong \Pr\left[-2.77 < Z < 2.77\right] = 0.994.$$

Thus, the authors concluded that "the 95% CIs will overlap over 99% of the time". They also mentioned that Schenker and Gentleman (2001) showed, for large sample sizes, that the probability of type I error when comparing the overlap of 100(1-γ)% confidence intervals is $2\Pr[Z < -z_{\gamma/2}(1+k)/\sqrt{1+k^2}]$ (k is the ratio of limiting standard errors here). Replacing k with 1 will yield a multiplier for the z value in the above probability statements of $\sqrt{2}$, which is the same as the formula of Pr(*Intervals_overlap*) by Payton *et al.* (2003). Therefore, the authors made the conclusion that when one uses 95% confidence intervals to test the equality of two population means, one should set the confidence coefficient of each confidence interval equal to roughly 83 to 84%. Payton *et al.* (2003) also used SAS to verify that if α is set at 0.16, what proportion of the times the two individual CIs from the same normal N(0, 1) population would overlap. After a simulation run of 10,000 trials, the proportion of the trials that the two 84% confidence intervals overlapped (at n = 10) was 0.949 (very close to 0.95). In this article, the authors didn't discuss the unequal sample sizes and no details about the effects from the unequal variances or the case of the unknown variances were provided. Further, no article discussed the maximum % overlap that two individual CIs can have but H_0 should still be rejected at a specified LOS α. All the articles up to now have concentrated on the limiting results and have not provided specific results for type II error Pr for the overlap case as sample sizes differ and the LOS vary.

Kelton (2004) states in *Simulation With Arena* that "looking at whether (individual) confidence intervals do or do not overlap (in order to make a decision about H_0: $\mu_x = \mu_y$) is not quite the right procedure; to do the comparison the right way, we'll use the input analyzer, as discussed next." The input-analyzer used the paired-t statistic to H_0: $\mu_x = \mu_y$. But, the authors do not explain the reason why we should not use the two individual CIs to make statistical inferences regarding the process mean difference $\mu_x - \mu_y$, nether do they justify the use of a block design over a completely randomized design.

3.0 Comparing Two Normal Population Means for the Known Variances Case, or the Limiting Unknown Variances Case Where Both Sample Sizes Approach Infinity

Consider a random sample of size n_x from the normal universe $N(\mu_x, \sigma_x^2)$; then the statistic sample mean, $\bar{x}$, is also normally distributed with expected-value equal to the population mean and variance $V(\bar{x}) = \sigma_x^2/n_x$ as depicted in figure 1.

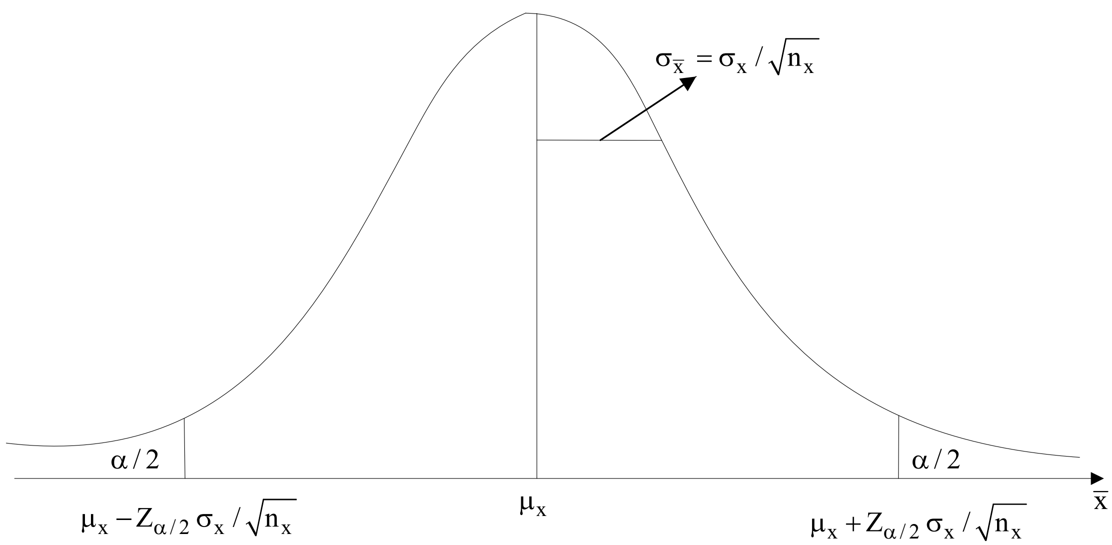


Figure 1 clearly shows that

$$\Pr(\mu_x - Z_{\alpha/2}\sigma_x/\sqrt{n_x} \le \bar{x} \le \mu_x + Z_{\alpha/2}\sigma_x/\sqrt{n_x}) = 1-\alpha$$

Rearranging the above $(1-\alpha)\times 100$ % Pr statement results in the $(1 - \alpha)$ CI for μ_x:

$$\Pr(\bar{x} - Z_{\alpha/2}\sigma_x/\sqrt{n_x} \le \mu_x \le \bar{x} + Z_{\alpha/2}\sigma_x/\sqrt{n_x}) = 1-\alpha$$

Hence, the lower CI limit for μ_x is $L(\mu_x) = \bar{x} - Z_{\alpha/2}\sigma_x/\sqrt{n_x}$ and the corresponding upper limit is $U(\mu_x) = \bar{x} + Z_{\alpha/2}\sigma_x/\sqrt{n_x}$. It resulting in the CIL (confidence interval length) of

μ_x is CIL(μ_x) = $2\times Z_{\alpha/2}\,\sigma_x/\sqrt{n_x}$. Similar procedure as above leads to L(μ_y) = $\bar{y} - Z_{\alpha/2}\times \sigma_y/\sqrt{n_y}$, U($\mu_y$) = $\bar{y} + Z_{\alpha/2}\,\sigma_y/\sqrt{n_y}$, and the corresponding CIL(μ_y) = $2Z_{\alpha/2}\,\sigma_y/\sqrt{n_y}$. Note that Figure 1 will roughly hold if the underlying distributions were non-normal and variances were unknown but both n_x and $n_y > 60$ and σ_x and σ_y are replaced by their biased estimates S_x and S_y, respectively.

3.1 The Case of $\sigma_x = \sigma_y = \sigma$

Statistical theory suggests that the total resources N = $n_x + n_y$ be allocated according to $n_x = N\sigma_x/(\sigma_x + \sigma_y)$, and hence the allocation $n_x = n_y = n = N/2$ is recommended. Suppose that the two CIs for μ_x and μ_y are disjoint; then it follows that either L(μ_x) > U(μ_y), or L(μ_y) > U(μ_x). These two possibilities lead to the condition either $\bar{x} - Z_{\alpha/2}\sigma_x/\sqrt{n_x} > \bar{y} + Z_{\alpha/2}\sigma_y/\sqrt{n_y}$, or $\bar{y} - Z_{\alpha/2}\sigma_y/\sqrt{n_y} > \bar{x} + Z_{\alpha/2}\times \sigma_x/\sqrt{n_x}$, respectively. Combining the two conditions leads to rejecting H_0: $\mu_x = \mu_y$ iff $|\bar{x} - \bar{y}| > Z_{\alpha/2}\,(\sigma_x/\sqrt{n_x} + \sigma_y/\sqrt{n_y})$; for the case of $\sigma_x = \sigma_y = \sigma$ and thus $n_x = n_y = n$, this last condition reduces to $|\bar{x} - \bar{y}| > 2\,Z_{\alpha/2}\,\sigma/\sqrt{n}$ at the level of significance α based on the Overlap method. If α is set at the nominal value of 5%, this last inequality will lead to the same condition as that of Schenker *et al*. (2001) who stated that the two intervals overlap if and only if the interval $(\widehat{Q_1} - \widehat{Q_2}) \pm 1.96(\widehat{SE_1} + \widehat{SE_2})$ contains 0.

Sometimes, it is then concluded that the null hypothesis H_0: $\mu_x = \mu_y$ must be rejected in favor of H_1: $\mu_x \neq \mu_y$ at the LOS α, such as Djordjevic *et al.* (2000), Tersmettte *et al*. (2001) and Sont *et al*. (2001) who used this concept to test H_0: $\mu_x = \mu_y$. In fact, Schenker and Gentleman (2001) state that they found more than 60 articles where the Overlap method was used either formally or informally to demonstrate visual significant difference between $\bar{x}$ and $\bar{y}$. This procedure is not accurate because the correct (1 − α)×100% CI for the difference in means of two independent normal universes must be obtained from the SMD (sampling distribution) of the statistic $\bar{x}-\bar{y}$, which is also Gaussian with $E(\bar{x}-\bar{y}) = \mu_x - \mu_y$ and $V(\bar{x}-\bar{y}) = V(\bar{x}) + V(\bar{y}) = \sigma_x^2/n_x + \sigma_y^2/n_y$ = $2\sigma^2/n$, assuming $\sigma_x = \sigma_y = \sigma$. Thus, the correct (1−α)×100% CI on $\mu_x - \mu_y$ is given by

$$\bar{x}-\bar{y} - Z_{\alpha/2}\sqrt{\sigma_x^2/n_x + \sigma_y^2/n_y} \le \mu_x - \mu_y \le \bar{x}-\bar{y} + Z_{\alpha/2}\sqrt{\sigma_x^2/n_x + \sigma_y^2/n_y} \quad \text{(1a)}$$

For the balanced design case and $\sigma_x = \sigma_y = \sigma$, Eq. (1a) reduces to

$$\bar{x}-\bar{y} - \sqrt{2}Z_{\alpha/2}\sigma/\sqrt{n} \le \mu_x - \mu_y \le \bar{x}-\bar{y} + \sqrt{2}Z_{\alpha/2}\sigma/\sqrt{n} \quad \text{(1b)}$$

The length of the above exact (1 − α)×100% CI for a balanced design is $2\sqrt{2}Z_{\alpha/2}\sigma/\sqrt{n}$. Thus, H_0: $\mu_x - \mu_y = 0$ must be rejected at the LOS α iff (i.e., it is necessary and sufficient) that

$$|\bar{x}-\bar{y}| > Z_{\alpha/2}\sqrt{(\sigma_x^2+\sigma_y^2)/n} = \sqrt{2}\,Z_{\alpha/2}\,\sigma/\sqrt{n}. \quad \text{(1c)}$$

However, requiring that the two separate CIs to be disjoint leads to rejection of H_0 iff $|\bar{x}-\bar{y}| > Z_{\alpha/2}(\sigma_x/\sqrt{n_x} + \sigma_y/\sqrt{n_y}) = 2Z_{\alpha/2}\times\sigma/\sqrt{n}$. It is clear that the requirement for rejecting H_0 of two disjoint CIs is more stringent (or more conservative) than that of the Standard method because, in the case of $\sigma_x = \sigma_y = \sigma$ and $n_x = n_y = n$, $2Z_{\alpha/2}\sigma/\sqrt{n} > Z_{\alpha/2}\sqrt{\sigma_x^2/n_x + \sigma_y^2/n_y} = \sqrt{2}\,Z_{\alpha/2}\,\sigma/\sqrt{n}$. Further, the more stringent requirement to reject H_0 (based on two independent separate CIs) leads to a smaller type I error Pr than the specified α. The correct value of α using the Standard method is given by

$$\alpha = \Pr(\bar{x} - \bar{y} < A_L, \text{ or } \bar{x} - \bar{y} > A_U \mid \mu_x - \mu_y = 0)$$

$$= \Pr[|\bar{x} - \bar{y}| > Z_{\alpha/2}\sqrt{\sigma_x^2 / n_x + \sigma_y^2 / n_y} \mid \mu_x - \mu_y = 0] = \Pr(|Z| > Z_{\alpha/2}) = \alpha,$$

where A_L and A_U denote the lower and upper α-acceptance limits, respectively.

On the other hand, if we require that the two individual CIs must be disjoint in order to reject $H_0 : \mu_x - \mu_y = 0$, then the type I error Pr from Overlap is given by

$$\alpha' = \Pr(\bar{y} - Z_{\alpha/2}\sigma_y / \sqrt{n} > \bar{x} + Z_{\alpha/2}\sigma_x / \sqrt{n}) + \Pr(\bar{x} - Z_{\alpha/2}\sigma_x / \sqrt{n} > \bar{y} + Z_{\alpha/2}\sigma_y / \sqrt{n})$$

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

$$= 2\times\Pr(Z > \sqrt{2}\, Z_{\alpha/2}) = 2\times\Phi(-\sqrt{2}\, Z_{\alpha/2}), \text{ assuming } \sigma_x = \sigma_y = \sigma \qquad (2)$$

- Setting α at 0.01 leads to the Overlap LOS of $\alpha' = 0.00026971696 << 0.01$. The % relative error, $[(\alpha - \alpha')/\alpha]\times 100\%$, in the LOS $\alpha = 0.01$ is $[(0.01 - 0.0026971696)/0.01]\times 100\% = 97.303\%$.
- For the nominal value of $\alpha = 0.05$, Eq. (2) gives $\alpha' = 0.00557459668 <<$ 0.05. The value of $\alpha' = 0.00557459668$ is consistent with the limiting value of 0.006 provided by Payton *et al*. (2003, p.36) in their equation (6). The % relative error 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 of run sizes 10,000 from two independent N(0, 1) populations in their column 3 of TABE 1, p. 551, that claim the value of α' ranges from 0.0039 at n = 5 to 0.0055 at n = 50 (n incremented by 5). Our Eq. (2) shows that in the case of

known equal variances and sample sizes the value of Overlap type I error Pr does not depend on n at all. However, their simulation inaccuracies were rectified by Payton *et. al* (2003, Table 4) again through simulation run sizes of 10,000 independent pairs from N(0, 1).

- Setting α at the maximum widely-accepted LOS of 10%, Eq. (2) shows that α' = 0.020009254 << 0.10 and the % relative error is [(0.10 − 0.020009254)/0.10] ×100% = 79.99%.

Regardless of the value of LOS α the same conclusion made by Cole *et al*. (1999) will be reached that the Overlap method will lead to a much smaller type I error rate.

If the alternative H_1 is one-sided, say H_1: $\mu_x - \mu_y > 0$, then from the Overlap standpoint H_0 should be rejected only if both conditions $\bar{x} - \bar{y} > 0$ and $L(\mu_x) - U(\mu_y) > 0$ [or $L(\mu_x) > U(\mu_y)$] hold, and as a result the Overlap type I error Pr reduces to

$$\alpha_1' = \Pr[\bar{x} - Z_{0.025}\sigma_x/\sqrt{n} > \bar{y} + Z_{0.025}\sigma_y/\sqrt{n}]$$

$$= \Pr[\bar{x} - \bar{y} > Z_{0.025}\sigma_x/\sqrt{n} + Z_{0.025}\sigma_y/\sqrt{n}] = \Pr[\bar{x} - \bar{y} > Z_{0.025}(\sigma_x + \sigma_y)/\sqrt{n}]$$

$$= \Pr(Z > Z_{0.025}(\sigma_x + \sigma_y)/\sqrt{(\sigma_x^2 + \sigma_y^2)}) = \Pr(Z > \sqrt{2}\, Z_{0.025});$$

assuming $\sigma_x = \sigma_y = \sigma$, $\alpha_1' = 0.0027873 << 0.05$. Thus the impact of Overlap on type I error Pr is even greater for a one-sided alternative than for the 2-sided one. Note that when $L(\mu_y) > U(\mu_x)$, the two CIs are disjoint but such an occurrence is congruent with H_0: $\mu_x - \mu_y \le 0$ rather than H_1: $\mu_x - \mu_y > 0$. Thus for the one-sided alternative the type I error Pr from the Overlap is exactly half of the 2-sided alternative, which was equal to 0.00557459668. Henceforth, unless specified otherwise, the alternative is two-sided.

Now, let O represent the amount of overlap length between the two individual CIs, a variable that has not been considered in the Overlap literature. From Figures (1a &b), 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 < α. 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 $U(\mu_y) \geq U(\mu_x) > L(\mu_x) \geq L(\mu_y)$. Because both conditions $U(\mu_x) > U(\mu_y) > L(\mu_x)$ and $U(\mu_y) > U(\mu_x) > L(\mu_y)$ will lead to the same result, only the case of $U(\mu_x) > U(\mu_y) > L(\mu_x)$ [Figure 2(a)] for which $\bar{x} - \bar{y} \geq 0$ is discussed here. See the illustration in Figure 2(a&b).

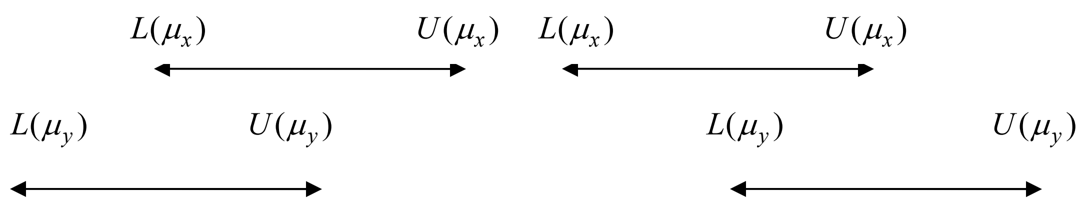


Figure 2(a) **Figure 2(b)**

That is, for the known-variance case, the larger sample mean will be denoted by $\bar{x}$. Thus for the equal-sample-size &-variance case,

$$O = U(\mu_y) - L(\mu_x) = (\bar{y} + Z_{\alpha/2} \times \sigma/\sqrt{n}) - (\bar{x} - Z_{\alpha/2} \times \sigma/\sqrt{n})$$

$$= 2Z_{\alpha/2}\sigma/\sqrt{n} - (\bar{x} - \bar{y}) \quad (3a)$$

On the other hand, the span of the two individual CIs (assuming $\bar{x} > \bar{y}$) is given by

$$U(\mu_x) - L(\mu_y) = \bar{x} + Z_{\alpha/2} \times \sigma/\sqrt{n} - (\bar{y} - Z_{\alpha/2} \times \sigma/\sqrt{n})$$

$$= 2Z_{\alpha/2}\sigma/\sqrt{n} + (\bar{x} - \bar{y}) \quad (3b)$$

Combining equations (3a & 3b) gives the exact % overlap as

$$\omega = \frac{2Z_{\alpha/2}\sigma/\sqrt{n} - (\bar{x} - \bar{y})}{2Z_{\alpha/2}\sigma/\sqrt{n} + (\bar{x} - \bar{y})} \times 100\% \qquad (3c)$$

Let O_r be the borderline value of O at which H_0 is barely rejected at the LOS α. From Eq. (1c), H_0: $\mu_x = \mu_y$ should be rejected iff $|\bar{x} - \bar{y}| \geq \sqrt{2}Z_{\alpha/2}\sigma/\sqrt{n}$. Therefore, from Eq.(3a) the value of O at which H_0 should be rejected at the α-level or less is given by $O \leq 2Z_{\alpha/2}\sigma/\sqrt{n} - \sqrt{2}Z_{\alpha/2}\sigma/\sqrt{n}$, and the exact amount of overlap that leads to an α-level test is given by

$$O_r = (2-\sqrt{2})Z_{\alpha/2}\sigma/\sqrt{n} \qquad (3d)$$

Eq. (3d) implies that H_0 must be rejected at the LOS α or less iff $O \leq (2-\sqrt{2})Z_{\alpha/2}\sigma/\sqrt{n}$.

Inserting the borderline rejection condition, $\bar{x} - \bar{y} = \sqrt{2}Z_{\alpha/2}\sigma/\sqrt{n}$, into Eq. (3b) yields

$$U(\mu_x) - L(\mu_y) = \sqrt{2}Z_{\alpha/2}\sigma/\sqrt{n} + 2Z_{\alpha/2}\sigma/\sqrt{n} = (2+\sqrt{2})Z_{\alpha/2}\sigma/\sqrt{n}. \qquad (3e)$$

Eq. (3e) implies that if the two CIs span larger than $(2+\sqrt{2})Z_{\alpha/2}\sigma/\sqrt{n}$, then H_0 must be rejected at the LOS less than α. The percent overlap in Eq. (3c) ranges from zero (occurring when $\bar{x} - \bar{y} = 2Z_{\alpha/2}\sigma/\sqrt{n}$) to 100% (occurring when $\bar{x} - \bar{y} = 0$). Inserting the borderline value of $\bar{x} - \bar{y} = \sqrt{2}Z_{\alpha/2}\sigma/\sqrt{n}$ at which H_0 must be rejected into Eq. (3c) results in

$$\omega_r = \frac{(2-\sqrt{2})Z_{\alpha/2}\sigma/\sqrt{n}}{(2+\sqrt{2})Z_{\alpha/2}\sigma/\sqrt{n}} \times 100\% = \frac{2-\sqrt{2}}{2+\sqrt{2}} \times 100\% = 17.1573\% \qquad (3f)$$

which means that H_0: $\mu_x = \mu_y$ must be rejected at the LOS α or less if the percent overlap between the two individual CIs is less than or equal to 17.1573%. It seems that the percent overlap at which H_0 should barely be rejected is 17.1573% regardless of the LOS α, but the amount of overlap from (3a) does depend on α. Further, as $|\bar{x} - \bar{y}|$ increases, the *P-value* of testing H$_0$: $\mu_x = \mu_y$ decreases and so does the value of % overlap in Eq. (3c). As $|\bar{x} - \bar{y}| \rightarrow 2Z_{\alpha/2}\sigma/\sqrt{n}$, ω → 0. Thus, in the case of known $\sigma_x = \sigma_y = \sigma$, once the % overlap exceeds 17.1573%, then H$_0$: $\mu_x = \mu_y$ must not be rejected at any α level.

If the alternative is one-sided, H$_1$: $\mu_x - \mu_y > 0$, it can be argued that the maximum percent overlap is given by

$$\frac{U(\mu_Y) - L(\mu_X)}{U(\mu_X) - L(\mu_Y)} = \frac{2Z_{\alpha/2} - Z_\alpha\sqrt{2}}{2Z_{\alpha/2} + Z_\alpha\sqrt{2}} \tag{3g}$$

and for a 5%-level test Eq. (3g) reduces to $\frac{(2Z_{0.025} - Z_{0.05}\sqrt{2})}{(2Z_{0.025} + Z_{0.05}\sqrt{2})}$ = 25.51597%, which implies that H$_0$ can be rejected at less than 5% level if the percent overlap between the two individual CIs is smaller than 25.51597%. Thus, the impact of overlap on ω_r is greater for a one-sided alternative because for the 2-sided alternative the value of ω_r = 17.15729%. Further, for the one-sided alternative the % overlap does depend on α. As an example, for a 10%-level one-sided test the value of ω_r increases to 28.96%.

The question now is what individual confidence levels, (1− γ), should be used that will lead to an exact α–level test? Clearly, the overlap amount for a (1− γ)×100% CI is given by $U'(\mu_y) - L'(\mu_x) = (\bar{y} + Z_{\gamma/2}\sigma/\sqrt{n}) - (\bar{x} - Z_{\gamma/2}\sigma/\sqrt{n})$

$$= 2\,Z_{\gamma/2}\sigma/\sqrt{n} - (\bar{x} - \bar{y}) \tag{4}$$

Because H$_0$: $\mu_x = \mu_y$ must be rejected iff $|\bar{x} - \bar{y}| \geq \sqrt{2}Z_{\alpha/2}\sigma/\sqrt{n}$ and the overlap must become zero or less in order to reject H$_0$, Eq. (4) shows that $2\,Z_{\gamma/2}\sigma/\sqrt{n}$ =

$$\sqrt{2}Z_{\alpha/2}\sigma/\sqrt{n} \rightarrow Z_{\gamma/2} = Z_{\alpha/2}/\sqrt{2} \rightarrow \gamma/2 = \Phi(-Z_{\alpha/2}/\sqrt{2}) \tag{5}$$

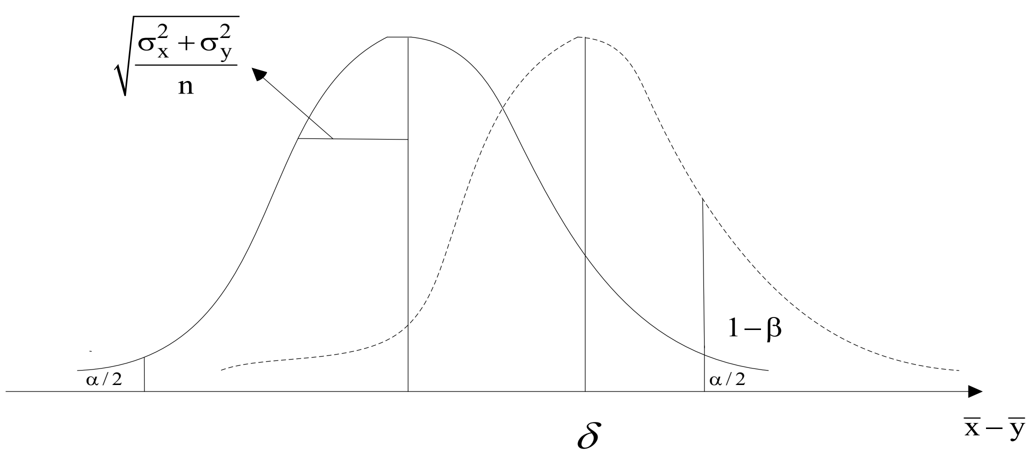
Eq. (5) shows that the confidence level for each individual interval must be set at $(1-\gamma) = 1 - 2 \times \Phi(-Z_{\alpha/2} / \sqrt{2})$ in order to reject H_0 at the LOS α iff the two CIs are disjoint. The value of $(1-\gamma)$ can also be obtained by equating the span of the two independent CIs, $2\,Z_{\gamma/2}\sigma / \sqrt{n} + (\bar{x} - \bar{y})$, to the length of the CI from the Standard method given by $2\sqrt{2}Z_{\alpha/2}\,\sigma / \sqrt{n}$, and invoking the rejection condition $\bar{x} - \bar{y}$ = $\sqrt{2}Z_{\alpha/2}\sigma / \sqrt{n}$.

- If α is set at 0.01 in Eq.(5), then γ = 0.068548146, $1-\gamma$ = 0.931451854, which implies that the confidence level of each individual interval must be set at 0.931451854 in order to reject H_0 at the 1% level iff the two CIs are disjoint.
- If α = 0.05 is substituted in Eq.(5), then γ = 0.165776273, $1-\gamma$ = 0.834223727, which implies that the confidence level of each individual interval must be set at 83.4223727% in order to reject H_0 at the 5% level iff the two CIs are disjoint. This assertion is in fair agreement with the simulation results given in TABLE 1 of Payton *et al*. (2000, p. 551) for $15 \le n \le 50$. Their TABLE 1, although inaccurate at n = 5 & 10, clearly shows that as n increases toward n = 50, the size of adjusted CIs is equal to 83.835%, which is very close to the exact $1-\gamma$ = 83.422372710%.
- Further, when the confidence level 1−α = 0.90, then $1-\gamma$ = 0.755205856. The first and third $1-\gamma$ values have not been reported in Overlap literature.

If the alternative is one-sided, H_1: $\mu_x - \mu_y > 0$, it can be argued that the value of $1-\gamma$ is given by $1-\gamma = 1-2\Phi(-Z_\alpha / \sqrt{2})$. If α = 0.05 is substituted into this last equation, then the one-sided $1-\gamma = 0.75520585634665$, which implies that the confidence level of each individual interval must be set at 0.75520585634665 in order to reject H_0 at the 5% level iff the two CIs are disjoint, while for the 2-sided alternative $1-\gamma$ was equal to 0.83422372710. Again, the impact of Overlap on individual confidence levels is greater for the one-sided alternative than that of the 2-sided one.

Lastly, since rejecting $H_0: \mu_x = \mu_y$ using the two independent CIs is more stringent than the SMD of $\bar{x}-\bar{y}$, therefore, it will lead to many more type II errors (or much less statistical power) in testing H_0: $\mu_x - \mu_y = 0$, as shown below.

In Figure 3, the solid line represents the null distribution of $\bar{x}-\bar{y}$, and the dotted line curve represents the distribution of $\bar{x}-\bar{y}$ under H_1, where $\delta = \mu_x - \mu_y > 0$ is the amount of specified shift in $\mu_x - \mu_y = \delta$ from zero, which in Figure 3 exceeds one standard error of $\bar{x}-\bar{y}$. Figure 3 clearly shows that the acceptance interval (AI) for the



sample mean difference, $\bar{x} - \bar{y}$, when testing H_0: $\mu_x - \mu_y = 0$ at the LOS α is given by

AI = $(A_L, A_U) = [-Z_{\alpha/2}\sqrt{(\sigma_x^2 + \sigma_y^2)/n}\ ,\ Z_{\alpha/2}\sqrt{(\sigma_x^2 + \sigma_y^2)/n}\ ]$, i.e., in the case of $\sigma_x = \sigma_y = \sigma$ and $n_x = n_y = n$ we cannot reject H_0 at the significance level α if our test statistic $\bar{x} - \bar{y}$ falls inside the AI = $(A_L, A_U) = (-\sqrt{2}Z_{\alpha/2}\ \sigma/\sqrt{n}\ ,\ \sqrt{2}Z_{\alpha/2}\ \sigma/\sqrt{n}\ )$. Thus, the Pr of committing a type II error as shown in Figure 3 is given by

$$\beta = \Pr[A_L \le \bar{x} - \bar{y} \le A_U \mid \mu_x - \mu_y = \delta]$$

$$= \Pr[-Z_{\alpha/2}\sqrt{(\sigma_x^2 + \sigma_y^2)/n} \le \bar{x} - \bar{y} \le +Z_{\alpha/2}\sqrt{(\sigma_x^2 + \sigma_y^2)/n} \mid \mu_x - \mu_y = \delta]$$

$$= \Pr[\bar{x} - \bar{y} \le Z_{\alpha/2}\sqrt{2\sigma^2/n} \mid \delta > 0] - \Pr[\bar{x} - \bar{y} \le -Z_{\alpha/2}\sqrt{2\sigma^2/n} \mid \delta\ ] \qquad (6a)$$

$$= \Phi\left(Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma\sqrt{2}}\right) - \Phi\left(-Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma\sqrt{2}}\right), \text{ where } \delta = \mu_x - \mu_y . \qquad (6b)$$

At α = 0.05 if the specified value of $\mu_x - \mu_y = \delta$ exceeds $0.5\sqrt{\sigma_x^2 + \sigma_y^2}$, then the value of standard normal cdf $\Phi(-Z_{0.025} - 0.5\sqrt{n}) < 0.001$ for sample sizes n ≥ 6, i.e., the last term on the RHS of equation (6a), becomes less than 0.001 once n ≥ 6. Hence, Eq. (6b) for the nominal value of α = 5% approximately reduces to

$$\beta \cong \Phi(Z_{0.025} - \delta\sqrt{n/2}/\sigma) \qquad (6c)$$

where (6c) is accurate to at least 3 decimals for n ≥ 6 and $\delta > 0.5\sqrt{\sigma_x^2 + \sigma_y^2} = 0.5\sigma\sqrt{2}$.

When the null hypothesis H_0: $\mu_x - \mu_y = 0$ is not rejected at the LOS α iff the two individual CIs ($\bar{x} - Z_{\alpha/2}\sigma_x/\sqrt{n} \le \mu_x \le \bar{x} + Z_{\alpha/2}\ \sigma_x/\sqrt{n}$) and ($\bar{y} - Z_{\alpha/2}\ \sigma_y/\sqrt{n} \le \mu_y \le \bar{y} + Z_{\alpha/2}\ \sigma_y/\sqrt{n}$) are overlapping, then the Pr of a type II error (assuming $\mu_x > \mu_y$)

from the Overlap Method is given by

$$\beta' = \Pr(\text{Overlap} | \delta > 0) = \Pr\{[L(\mu_x) \le U(\mu_y)] \cap [L(\mu_y) \le U(\mu_x)] | \delta > 0\}$$

$$= \Pr\{[\bar{x} - Z_{\alpha/2}\,\sigma_x/\sqrt{n} \le \bar{y} + Z_{\alpha/2}\,\sigma_y/\sqrt{n}] \cap [\bar{y} - Z_{\alpha/2}\frac{\sigma_y}{\sqrt{n}} \le \bar{x} + Z_{\alpha/2}\,\sigma_x/\sqrt{n}] | \delta > 0\}$$

$$= \Pr\{[\bar{x} - \bar{y} \le Z_{\alpha/2}\,\sigma_x/\sqrt{n} + Z_{\alpha/2}\,\sigma_y/\sqrt{n}] \cap$$

$$[-Z_{\alpha/2}\frac{\sigma_y}{\sqrt{n}} - Z_{\alpha/2}\,\sigma_x/\sqrt{n} \le \bar{x} - \bar{y}] | \delta > 0\}$$

When $\sigma_x = \sigma_y = \sigma$, and $n_x = n_y = n$, the $SE(\bar{x} - \bar{y}) = \sigma\sqrt{2/n}$ and as a result

$$\beta' = \Pr\{[\bar{x} - \bar{y} \le 2Z_{\alpha/2}\,\sigma/\sqrt{n}] \cap [-2Z_{\alpha/2}\,\sigma/\sqrt{n} \le \bar{x} - \bar{y}] | \delta > 0\}$$

$$= \Pr\{[-2Z_{\alpha/2}\,\sigma/\sqrt{n} \le \bar{x} - \bar{y} \le 2Z_{\alpha/2}\,\sigma/\sqrt{n}] | \delta > 0\} \quad (7a)$$

$$= \Phi(\sqrt{2}\,Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma\sqrt{2}}) - \Phi(-\sqrt{2}\,Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma\sqrt{2}}) \quad (7b)$$

Since the cdf of the standard normal density $\Phi(z)$ is a monotonically increasing function of z, comparing Eq. (6b) with Eq. (7b) shows that

$$\Phi(\sqrt{2}Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma\sqrt{2}}) > \Phi(Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sqrt{2}\sigma}) \;\&\; \Phi(-\sqrt{2}Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma\sqrt{2}}) < \Phi(-Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sqrt{2}\sigma}).$$

The above two conditions lead to

$$\beta' = \Phi(\sqrt{2}Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma\sqrt{2}}) - \Phi(-\sqrt{2}Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma\sqrt{2}}) >$$

$$\Phi(Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma\sqrt{2}}) - \Phi(-Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma\sqrt{2}}) = \beta.$$

and as a result $1-\beta' < 1-\beta$, i.e., using individual CIs loses statistical power as illustrated in Table 1 (for n = 10, 20, 40, 60 and 80 at $\alpha = 0.05$). Table 1 clearly shows that the Pr

Table 1. The Relative Power of Overlap as Compared to the Standard Method for Different Sample Sizes n and $\delta/(\sigma\sqrt{2})$ Combinations

n	$\frac{\delta}{\sigma\sqrt{2}}$	$1-\beta$	$1-\beta'$	$[\frac{\beta'-\beta}{1-\beta}]100\%$	n	$\frac{\delta}{\sigma\sqrt{2}}$	$1-\beta$	$1-\beta'$	$[\frac{\beta'-\beta}{1-\beta}]100\%$
10	0	0.050000	0.005575	88.850807	10	0.2	0.096935	0.016535	82.941952
10	0.2	0.096935	0.016535	82.941952	30	0.2	0.194775	0.046889	75.926797
10	0.4	0.244141	0.065946	72.988712	50	0.2	0.292989	0.087310	70.200085
10	0.6	0.475101	0.190941	59.810521	70	0.2	0.387332	0.136000	64.887986
10	0.8	0.715617	0.404396	43.489888	90	0.2	0.475101	0.190941	59.810521
10	1	0.885379	0.651905	26.369907	110	0.2	0.554768	0.250096	54.918821
10	1.2	0.966730	0.846828	12.402799	130	0.2	0.625674	0.311552	50.205361
10	1.4	0.993192	0.951076	4.240407	150	0.2	0.687770	0.373606	45.678672
10	1.6	0.999031	0.988926	1.011467	170	0.2	0.741418	0.434816	41.353531
10	1.8	0.999905	0.998251	0.165374	190	0.2	0.787231	0.494017	37.246245
10	2	0.999994	0.999809	0.018425	210	0.2	0.825958	0.550319	33.372049
20	0	0.050000	0.005575	88.850807	230	0.2	0.858407	0.603086	29.743564
20	0.2	0.145473	0.030356	79.132788	250	0.2	0.885379	0.651905	26.369907
20	0.4	0.432158	0.162818	62.324449	270	0.2	0.907642	0.696558	23.256232
20	0.6	0.765259	0.464729	39.271657	290	0.2	0.925899	0.736982	20.403620
20	0.8	0.947141	0.789850	16.606937	310	0.2	0.940785	0.773239	17.809209
20	1	0.994000	0.955465	3.876765	330	0.2	0.952858	0.805484	15.466533
20	1.2	0.999671	0.995267	0.440547	350	0.2	0.962600	0.833939	13.365991
20	1.4	0.999991	0.999758	0.023375	400	0.2	0.979327	0.890313	9.089308
20	1.6	1.000000	0.999994	0.000573	450	0.2	0.988775	0.929332	6.011824
20	1.8	1.000000	1.000000	0.000006	500	0.2	0.994000	0.955465	3.876765
20	2	1.000000	1.000000	0.000000	600	0.2	0.998354	0.983297	1.508145
40	0	0.050000	0.005575	88.850807	700	0.2	0.999568	0.994127	0.544334
40	0.2	0.244141	0.065946	72.988712	800	0.2	0.999891	0.998043	0.184785
40	0.4	0.715617	0.404396	43.489888	900	0.2	0.999973	0.999377	0.059617
40	0.6	0.966730	0.846828	12.402799	1100	0.2	0.999999	0.999944	0.005488
40	0.8	0.999031	0.988926	1.011467	1300	0.2	1.000000	0.999995	0.000444
40	1	0.999994	0.999809	0.018425	1500	0.2	1.000000	1.000000	0.000032
40	1.2	1.000000	0.999999	0.000072	20	0.5	0.608779	0.296070	51.366706
60	0	0.050000	0.005575	88.850807	40	0.5	0.885379	0.651905	26.369907
60	0.2	0.340845	0.110745	67.508566	60	0.5	0.972127	0.864590	11.062062
60	0.4	0.872528	0.628007	28.024464	80	0.5	0.994000	0.955465	3.876765
60	0.6	0.996402	0.969657	2.684165	100	0.5	0.998817	0.987066	1.176501
60	0.8	0.999989	0.999693	0.029611	120	0.5	0.999782	0.996589	0.319361
60	1	1.000000	1.000000	0.000032	140	0.5	0.999962	0.999167	0.079444
80	0	0.050000	0.005575	88.850807	160	0.5	0.999994	0.999809	0.018425
80	0.2	0.432158	0.162818	62.324449	180	0.5	0.999999	0.999959	0.004033
80	0.4	0.947141	0.789850	16.606937	200	0.5	1.000000	0.999991	0.000841
80	0.6	0.999671	0.995267	0.440547	220	0.5	1.000000	0.999998	0.000168
80	0.8	1.000000	0.999994	0.000573	240	0.5	1.000000	1.000000	0.000032
80	1	1.000000	1.000000	0.000000	260	0.5	1.000000	1.000000	0.000006

of type Π error from two individual CIs is always larger than the Pr of type Π error from the Standard method (i.e., $\beta' > \beta$). Thus, the statistical power of Overlap method is less than that of the standard method ($1-\beta' < 1-\beta$). And, for fixed n, both $1-\beta$ and $1-\beta'$ increase as $\delta/\sqrt{2}\sigma$ increases. Further, the power of the Overlap procedure very slowly approaches that of the standard method as n increases. The difference in percent relative power is obtained from $\{[(1-\beta)-(1-\beta')]/(1-\beta)\}100\% = [(\beta'-\beta)/(1-\beta)]100\%$. Thus, from Table 1, we can also conclude that if $\delta/(\sqrt{2}\sigma)$ is fixed, as sample size increases, the difference in percent relative power decreases. Figure 4 shows that the difference of the

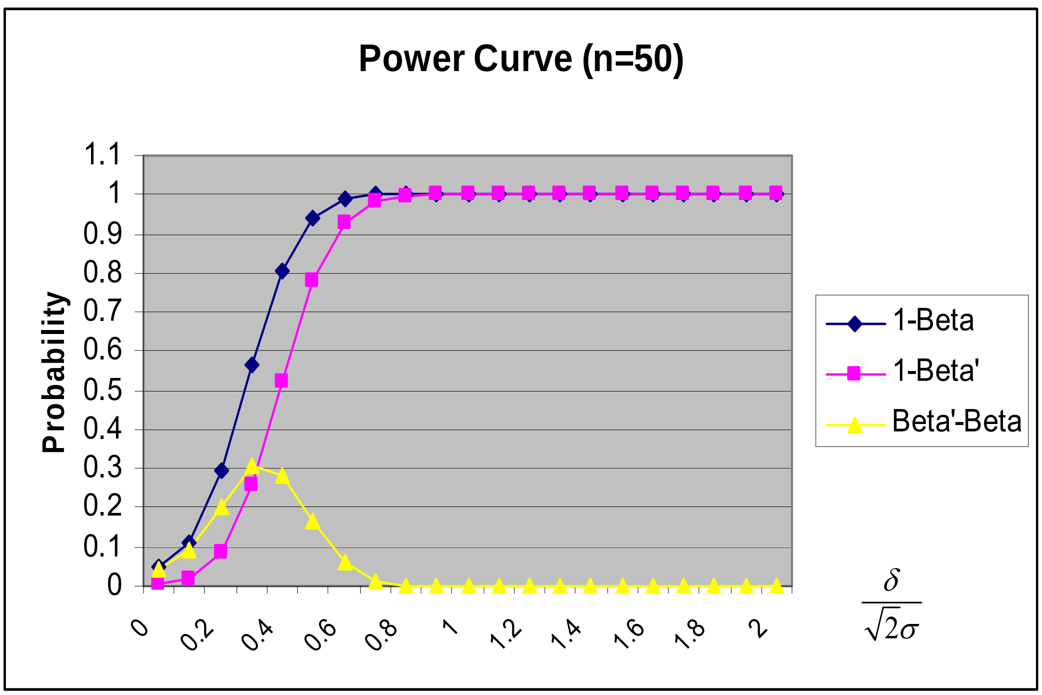


Figure 4

power ($(1-\beta)-(1-\beta') = \beta'-\beta$) increases first then decreases as $\delta/\sqrt{2}\sigma$ increases at fixed n = 50. See the summary conclusion in Table 2. To verify that the formulas for β

Table 2. Summary Conclusion of β and β'

n	$\frac{\delta}{\sigma\sqrt{2}}$	$\beta' - \beta$	$[\frac{\beta' - \beta}{1-\beta}]100\%$
fixed	increasing	increasing then decreasing	decreasing
increasing	fixed	increasing then decreasing	decreasing

(Eq.(6a)) and β' (Eq.(7)) are correct, we use the fact that if α = 0.05, then $1-\gamma$ = 0.83422372710 (γ = 0.165776273), which implies that the confidence level of each individual interval must be set at 0.8342237271 in order to reject H_0 at the 5% level iff the two CIs are disjoint. Table 3 shows that the value of β at α = 0.05 from the Standard method and the corresponding β' from Overlap at γ = 0.165776273 are exactly equal.

Table 3. Type Π Error Prs at α = 0.05 from the Standard Method and at α = 0.16578 from the Overlap Method

n	$\frac{\delta}{n\sqrt{2}}$	β at α = 0.05	β' at γ = 0.165776273	n	$\frac{\delta}{n\sqrt{2}}$	β at α = 0.05	β' at γ = 0.165776273
10	0	0.95	0.95	60	0	0.95	0.95
10	0.2	0.9030645532	0.9030645532	60	0.2	0.6591548759	0.6591548759
10	0.4	0.7558587930	0.7558587930	60	0.4	0.1274718000	0.1274718000
10	0.6	0.5248991294	0.5248991294	60	0.6	0.0035982047	0.0035982047
10	0.8	0.2843833932	0.2843833932	60	0.8	0.0000113359	0.0000113359
10	1	0.1146208592	0.1146208592	60	1	0.0000000036	0.0000000036
20	0	0.9500000000	0.9500000000	80	0	0.9500000000	0.9500000000
20	0.2	0.8545274876	0.8545274876	80	0.2	0.5678423724	0.5678423724
20	0.4	0.5678423724	0.5678423724	80	0.4	0.0528587903	0.0528587903
20	0.6	0.2347406798	0.2347406798	80	0.6	0.0003288883	0.0003288883
20	0.8	0.0528587903	0.0528587903	80	0.8	0.0000001021	0.0000001021
20	1	0.0059995300	0.0059995300	100	1	0.0000000000	0.0000000000
40	0	0.9500000000	0.9500000000	100	0	0.9500000000	0.9500000000
40	0.2	0.7558587930	0.7558587930	100	0.2	0.4839947260	0.4839947260
40	0.4	0.2843833932	0.2843833932	100	0.4	0.0206733681	0.0206733681
40	0.6	0.0332699421	0.0332699421	100	0.6	0.0000267215	0.0000267215
40	0.8	0.0009686482	0.0009686482	100	0.8	0.0000000008	0.0000000008
40	1	0.0000063680	0.0000063680	100	1	0.0000000000	0.0000000000

If the alternative is one-sided, H_1: $\mu_x - \mu_y > 0$, clearly the expression for β' given in Eq. (7b) stays in tact but the Standard method type II error Pr becomes $\beta_1 = \Phi(Z_\alpha - \delta\sqrt{n/2}/\sigma)$, where $\delta = \mu_x - \mu_y$. Because for $\delta > 0$, $\beta' = \Phi(\sqrt{2}\,Z_{\alpha/2} - \delta\sqrt{n/2}/\sigma) - \Phi(-\sqrt{2}\,Z_{\alpha/2} - \delta\sqrt{n/2}/\sigma) > \beta = \Phi(Z_{\alpha/2} - \delta\sqrt{n/2}/\sigma) - \Phi(-Z_{\alpha/2} - \delta\sqrt{n/2}/\sigma) > \beta_1 = \Phi(Z_\alpha - \delta\sqrt{n/2}/\sigma)$, it follows that the impact of Overlap on type II error Pr for the one-sided alternative is greater that that of the two-sided alternative. Note that β_1 becomes equal to β only at $\delta = 0$.

3.2 The Case of Known but Unequal Variances

If variances of the two independent processes are known but not equal, then statistical theory dictates that the two sample sizes should be allocated according to

$$n_x = \frac{\sigma_x \times N}{\sigma_x + \sigma_y}, \qquad n_y = \frac{\sigma_y \times N}{\sigma_x + \sigma_y}, \tag{8}$$

where $N = n_x + n_y$ = the total recourses available to the experimenter. The sample size allocations given in equations (8) lead to the minimum $SE(\bar{x} - \bar{y}) = (\sigma_x + \sigma_y)/\sqrt{N}$. Schenker and Gentleman (2001) make similar statement as above but did not use equation (8) to set the values of n_x and n_y. They use notational procedure by letting $k = \dfrac{\sigma_x/\sqrt{n_x}}{\sigma_y/\sqrt{n_y}}$. Note that Schenker and Gentleman (2001) refer to the limiting value of small-letter k as the *SE* ratio because they investigated the impact of Overlap on type I and II error rates only when n_x and $n_y \to \infty$. Since we discuss both limiting case (populations) and small to moderate sample size cases in this dissertation, the small k refers to the standard error ration for samples, ie, $k = \dfrac{S_x / \sqrt{n_x}}{S_y / \sqrt{n_y}}$ and K refers to the SE ratio for populations, ie, K=$\dfrac{\sigma_x / \sqrt{n_x}}{\sigma_y / \sqrt{n_y}} = \mathrm{SE}(\bar{x}) / \mathrm{SE}(\bar{y})$. Clearly,

$$\mathrm{SE}(\bar{x} - \bar{y}) = \sqrt{\sigma_x^2 / n_x + \sigma_y^2 / n_y} = \sqrt{\mathrm{K}^2 \sigma_y^2 / n_y + \sigma_y^2 / n_y}$$

$$= \sigma_y \sqrt{1 + \mathrm{K}^2} / \sqrt{n_y} = \mathrm{SE}(\bar{y}) \sqrt{1 + \mathrm{K}^2} \tag{9a}$$

Substituting equations (9a) into the Standard $(1 - \alpha)\times 100\%$ CI: $\bar{x} - \bar{y} \pm Z_{\alpha/2} \times \mathrm{SE}(\bar{x} - \bar{y})$ leads to

$$\bar{x} - \bar{y} - Z_{\alpha/2} \sigma_y \sqrt{1 + \mathrm{K}^2} / \sqrt{n_y} \le \mu_x - \mu_y \le \bar{x} - \bar{y} + Z_{\alpha/2} \sigma_y \sqrt{1 + \mathrm{K}^2} / \sqrt{n_y} \tag{9b}$$

Thus, the Standard CIL equals to $2 Z_{\alpha/2} \sigma_y \sqrt{1 + \mathrm{K}^2} / \sqrt{n_y}$. Equation (9b) shows that the null hypothesis H_0: $\mu_x - \mu_y = 0$ must be rejected at the LOS α iff the CI in equation (9b) excludes zero, or iff

$$|\bar{x} - \bar{y}| > Z_{\alpha/2} \sigma_y \sqrt{1 + \mathrm{K}^2} / \sqrt{n_y} . \tag{9c}$$

However, requiring that the two independent CIs must not overlap in order to reject H_0: $\mu_x - \mu_y = 0$ at the LOS α, is equivalent to requiring that either $L(\mu_x) > U(\mu_y)$ or $L(\mu_y) > U(\mu_x)$. These two inequalities lead to the Overlap rejection of H_0: $\mu_x - \mu_y = 0$ iff

$$|\bar{x} - \bar{y}| > Z_{\alpha/2} \sigma_x / \sqrt{n_x} + Z_{\alpha/2} \sigma_y / \sqrt{n_Y} = Z_{\alpha/2} \sigma_y (1 + \mathrm{K}) / \sqrt{n_y} \tag{10}$$

Therefore, if the exact Pr of type I error is α but we reject H_0 when the two independent CIs are disjoint, the Overlap type I error Pr reduces to

$$\alpha' = \Pr[\bar{y} - Z_{\alpha/2} \sigma_y / \sqrt{n_y} > \bar{x} + Z_{\alpha/2} \sigma_x / \sqrt{n_x}] + \Pr[\bar{x} - Z_{\alpha/2} \sigma_x / \sqrt{n_x} > \bar{y} + Z_{\alpha/2} \sigma_y / \sqrt{n_y}]$$

$$= 2\times\Pr(\ \bar{x} - \bar{y} > Z_{\alpha/2}\sigma_y(1+K)/\sqrt{n_y}\ ) = 2\times\Pr(Z > Z_{\alpha/2}(1+K)/\sqrt{1+K^2}\ ) \qquad (11)$$

which is identical to that of equation (7) provided by Schenker and Gentleman (2001, p. 184) when their standardized difference, d, is set equal to 0. Eq.(11) shows that as $K \to 0$ or ∞, the value of α' slowly approaches the exact type I error probability α [consistent with Table 3 on p. 3 of Payton *et al.* (2003)]. Further, since $(1+K)/\sqrt{1+K^2} > 1$ and $Z_{\alpha/2}(1+K)/\sqrt{1+K^2} > Z_{\alpha/2}$, 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 the Overlap always leads to a smaller type I error Pr than that of the Standard method, consistent with Figure 3 of Schenker and Gentleman (2001, p. 184). Table 4 shows the value of α' at $\alpha = 0.01$ and 0.05 for different K values. Note that Table 4 values are valid for either Gaussian underlying distributions or for the limiting values of n_x and n_y. Figure 5(a) and 5(b) show that as K increases, the value of α' slowly approaches the exact type I error probability α.

To determine the minimum value of α' from Eq.(11), let g(K)= $Z_{\alpha/2}(1+K)/\sqrt{1+K^2}$. The first derivative of g(K) is $g'(K) = Z_{\alpha/2}[\frac{1}{\sqrt{1+K^2}} - \frac{K(1+K)}{\sqrt{(1+K^2)^3}}]$. Setting $g'(K) = 0$ will lead to K = 1. To ascertain whether K=1 is a point of minimum or maximum, the second differentiation yields:

$$g''(K) = Z_{\alpha/2}[\frac{-2K}{(1+K^2)^{3/2}} + \frac{3(1+K)K^2}{(1+K^2)^{5/2}} - \frac{1+K}{(1+K^2)^{3/2}}]$$

Substituting K =1 and $Z_{0.025} = 1.959964$ into the above equation results in $g''(1) = -0.353553391 < 0$, which shows that K =1 maximize g(K). Thus, α' has the minimum value at K =1, as shown in Table 4.

Table 4. The Type I Error Pr of Two Individual CIs with Different K at $\alpha = 0.05$ and 0.01

K	$\alpha'(\alpha=0.05)$	K	$\alpha'(\alpha=0.05)$	K	$\alpha'(\alpha=0.01)$	K	$\alpha'(\alpha=0.01)$
1	0.005574597	6	0.024101169	1	0.000269717	6	0.003034255
1.2	0.005772632	7	0.026592621	1.2	0.000285833	7	0.003565806
1.4	0.006255214	8	0.028674519	1.4	0.000326631	8	0.004034767
1.6	0.006916773	9	0.030432273	1.6	0.000385984	9	0.004447733
1.8	0.007695183	10	0.031932004	1.8	0.000460718	10	0.004812093
2	0.008549353	11	0.033224353	2	0.000548586	11	0.005134764
2.2	0.009450168	12	0.034348214	2.2	0.000647644	12	0.005421799
2.4	0.010376313	13	0.035333699	2.4	0.000756080	13	0.005678348
2.6	0.011312004	14	0.036204361	2.6	0.000872186	14	0.005908733
2.8	0.012245574	15	0.036978819	2.8	0.000994382	15	0.006116572
3	0.013168478	16	0.037671953	3	0.001121233	16	0.006304887
3.2	0.014074567	17	0.038295773	3.2	0.001251466	17	0.006476214
3.4	0.014959516	18	0.038860068	3.4	0.001383967	18	0.006632687
3.6	0.015820399	19	0.039372889	3.6	0.001517779	19	0.006776110
3.8	0.016655348	20	0.039840912	3.8	0.001652089	20	0.006908016
4	0.017463296	21	0.040269717	4	0.001786217	21	0.007029710
4.2	0.018243773	22	0.040664002	4.2	0.001919599	22	0.007142316
4.4	0.018996754	23	0.041027747	4.4	0.002051774	23	0.007246798
4.6	0.019722537	24	0.041364351	4.6	0.002182369	24	0.007343994
4.8	0.020421655	25	0.041676725	4.8	0.002311086	25	0.007434630
5	0.021094804	26	0.041967382	5	0.002437694	26	0.007519342

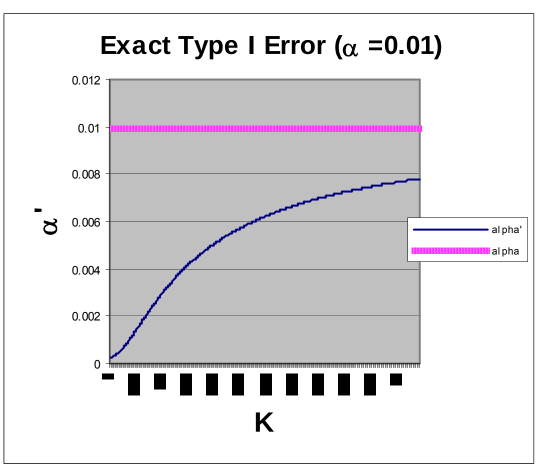


Figure 5(a)

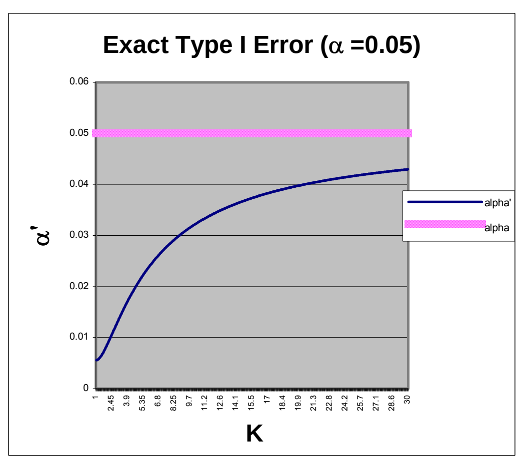


Figure 5(b)

As before, let O represent the amount of overlap length between the two individual CIs. Similar procedure as in previous section yields

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

$$= Z_{\alpha/2}(\sigma_x / \sqrt{n_x} + \sigma_y / \sqrt{n_y}) - (\bar{x} - \bar{y}) \qquad (12a)$$

Let O_r be the borderline value of O at which H_0 is barely rejected at an α-level. From Eq. (9c), H_0: $\mu_x = \mu_y$ must be rejected iff $|\bar{x} - \bar{y}| > Z_{\alpha/2}\sigma_y \sqrt{1+K^2} / \sqrt{n_y}$, which upon substitution into (12a) results in

$$O_r = Z_{\alpha/2}(\sigma_x / \sqrt{n_x} + \sigma_y / \sqrt{n_y}) - (\bar{x} - \bar{y})$$

$$= Z_{\alpha/2}(\sigma_x / \sqrt{n_x} + \sigma_y / \sqrt{n_y}) - Z_{\alpha/2}\sigma_y \sqrt{1+K^2} / \sqrt{n_y} .$$

Substituting $\sigma_x / \sqrt{n_x} = K\sigma_y / \sqrt{n_y}$ in the above equation yields

$$O_r = Z_{\alpha/2}(\sigma_y / \sqrt{n_y}) \times [1 + K - \sqrt{1+K^2}] \qquad (12b)$$

Eq. (12b) indicates that H_0 must be rejected at the LOS α or less iff $O \le Z_{\alpha/2}(\sigma_y / \sqrt{n_y}) \times$ $[1 + K - \sqrt{1+K^2}]$. Further, the span of the two individual CIs is

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

$$= Z_{\alpha/2}(\sigma_x / \sqrt{n_x} + \sigma_y / \sqrt{n_y}) + (\bar{x} - \bar{y}) \qquad (12c)$$

Thus, the exact percent α-overlap is given by

$$\omega = \frac{Z_{\alpha/2}(\sigma_x / \sqrt{n_x} + \sigma_y / \sqrt{n_y}) - (\bar{x} - \bar{y})}{Z_{\alpha/2}(\sigma_x / \sqrt{n_x} + \sigma_y / \sqrt{n_y}) + (\bar{x} - \bar{y})} \times 100\%$$

$$\rightarrow \quad \omega = \frac{Z_{\alpha/2}\sigma_y(1+\mathrm{K})/\sqrt{n_y} - (\bar{x}-\bar{y})}{Z_{\alpha/2}\sigma_y(1+\mathrm{K})/\sqrt{n_y} + (\bar{x}-\bar{y})} \times 100\% \qquad (12d)$$

As before, ω lies in the closed interval [0, 100%]. The % overlap in Eq. (12d) clearly shows that as $\bar{x}-\bar{y} > 0$ increases, the P-value of the test decreases, and Eq.(12d) shows that the % overlap also decreases. Because H_0 must be rejected at the LOS α or less iff $|\bar{x}-\bar{y}| \geq Z_{\alpha/2}\sigma_y\sqrt{1+\mathrm{K}^2}/\sqrt{n_y}$, the maximum % overlap above which H_0 cannot be rejected at an α-level is given by

$$\omega_r(k) = \frac{Z_{\alpha/2}\sigma_y(1+\mathrm{K})/\sqrt{n_y} - Z_{\alpha/2}\sigma_y\sqrt{1+\mathrm{K}^2}/\sqrt{n_y}}{Z_{\alpha/2}\sigma_y(1+\mathrm{K})/\sqrt{n_y} + Z_{\alpha/2}\sigma_y\sqrt{1+\mathrm{K}^2}/\sqrt{n_y}} \times 100\%$$

$$= \frac{1+\mathrm{K}-\sqrt{1+\mathrm{K}^2}}{1+\mathrm{K}+\sqrt{1+\mathrm{K}^2}} \times 100\% \qquad (12e)$$

Eq. (12e) shows that the maximum prevent overlap doesn't depend on α and reduces to 17.1573% when $\mathrm{K} = \frac{\sigma_x/\sqrt{n_x}}{\sigma_y/\sqrt{n_y}} = 1$. It can be verified that the 1st derivative of $\omega_r(\mathrm{K})$ is

$\omega_r'(\mathrm{K}) = \frac{2-2\mathrm{K}}{\sqrt{1+\mathrm{K}^2}\times(1+\mathrm{K}+\sqrt{1+\mathrm{K}^2})^2}$ whose root is K = 1. Moreover, the value of the 2nd derivative of $\omega_r(\mathrm{K})$ at K = 1 is -0.121320344, which means that K = 1 maximizes the % overlap and the null hypothesis H_0: $\mu_x = \mu_y$ must be rejected at any α if the overlap does not exceed 17.1573%. The farther K is from 1, the smaller the amount of allowable overlap becomes (i.e., the Overlap procedure becomes less deficient). For example, at K = 2 or 0.50, the % overlap reduces to 14.5898%. This implies that when the limiting SE ratio is K = 2 or 0.50, the two individual CIs can overlap up to 14.5898% and H_0: $\mu_x = \mu_y$ must still be rejected at the LOS α or less. At K = 3 or 1/3, the % overlap reduces to 11.696312% below which H_0 must be rejected at α or less level; at K = 10, it reduces to 0.04513682. As $K \to 0$ or ∞, $\omega_r \to 0$ so that the Overlap procedure very gradually approaches an exact α-level test [consistent with Table 3 of Payton *et al*. (2003, p. 3)]

Furthermore, what should the individual confidence level, $(1-\gamma)$, be so that comparisons of individual CIs will lead to the exact α–level test? From Eq.(12c), the corresponding span of two individual CIs at confidence level $(1-\gamma)$ is $U'(\mu_x) - L'(\mu_y) = Z_{\gamma/2}\sigma_y(1+K)/\sqrt{n_y} + (\bar{x}-\bar{y})$. From Eq.(9c), H_0: $\mu_x - \mu_y = 0$ must be rejected at the LOS α iff $|\bar{x} - \bar{y}| > Z_{\alpha/2}\sigma_y\sqrt{1+K^2}/\sqrt{n_y}$. Substituting the critical limit $\bar{x}-\bar{y} = Z_{\alpha/2}(\sigma_y/\sqrt{n_y}) \times\sqrt{1+K^2}$ into $U'(\mu_x) - L'(\mu_y)$ results in $U'(\mu_x) - L'(\mu_y) = Z_{\gamma/2}\sigma_y(1+K)/\sqrt{n_y} + Z_{\alpha/2}\sigma_y\sqrt{1+K^2}/\sqrt{n_y}$. Furthermore, the $(1-\alpha)\times100\%$ CIL from the Standard method is equal to $2Z_{\alpha/2}\sigma_y\sqrt{1+K^2}/\sqrt{n_y}$. Thus, the individual confidence levels, $(1-\gamma)$, should be set as follows which in turn leads individual CIs to an exact α–level test.

$$\rightarrow\ Z_{\gamma/2}\sigma_y(1+K)/\sqrt{n_y} + Z_{\alpha/2}\sigma_y\sqrt{1+K^2}/\sqrt{n_y} = 2Z_{\alpha/2}\sigma_y\sqrt{1+K^2}/\sqrt{n_y}$$

$$\rightarrow\ Z_{\gamma/2}\sigma_y(1+K)/\sqrt{n_y} = Z_{\alpha/2}\sigma_y\sqrt{1+K^2}/\sqrt{n_y}$$

$$\rightarrow\ \gamma = 2\times\Phi[-Z_{\alpha/2}\sqrt{1+K^2}/(1+K)] \tag{13}$$

Eq.(13) shows that the level of each CI must be set at $(1-\gamma) = 1-2\times\Phi[-Z_{\alpha/2}\sqrt{1+K^2}/(1+K)]$ in order to reject H_0 at α LOS iff the two CIs are disjoint, which is in agreement with Eq.(8) of Payton *et al*. (2003, p.2). To verify this assertion, let q(K) =

$-Z_{\alpha/2}\sqrt{1+\mathrm{K}^2}\,/(1+\mathrm{K})$. The 1st derivative of q(K) is given by $q'(\mathrm{K})=$

$\dfrac{\mathrm{K}}{\sqrt{1+\mathrm{K}^2}\,(1+\mathrm{K})}-\dfrac{\sqrt{1+\mathrm{K}^2}}{(1+\mathrm{K})^2}$. Setting $q'(\mathrm{K})=0$ results in K =1. Moreover, the 2nd

derivative of q(K) is $\left.\dfrac{d^2q(\mathrm{K})}{d\mathrm{K}^2}\right|_{\mathrm{K}=1}=-0.346476<0$, which implies K =1 maximizes q(k) and in turn also maximizes γ. Table 5 shows that as K increases toward 1, γ also increases to reach its maximum and then decreases for the fixed α as K departs from 1.

Table5. Values of γ Versus K at α = 0.05 and α = 0.01

α = 0.05				α = 0.01			
K	γ	**K**	γ	**K**	γ	**K**	γ
0.2	0.095783	3.5	0.112872	0.2	0.028594	3.5	0.037197
0.4	0.131601	4	0.106045	0.4	0.047523	4	0.033663
0.6	0.153132	4.5	0.100440	0.6	0.060458	4.5	0.030857
0.8	0.163187	5	0.095783	0.8	0.066863	5	0.028594
1	0.165776	6	0.088541	1	0.068548	6	0.025201
1.2	0.164038	7	0.083206	1.2	0.067415	7	0.022802
1.4	0.160015	8	0.079131	1.4	0.064818	8	0.021030
1.6	0.154931	9	0.075927	1.6	0.061587	9	0.019674
1.8	0.149483	10	0.073346	1.8	0.058189	10	0.018606
2	0.144051	20	0.061628	2	0.054869	20	0.014040
2.2	0.138834	30	0.057723	2.2	0.051746	30	0.012627
2.5	0.131601	40	0.055779	2.5	0.047523	40	0.011944
3	0.121265	50	0.054616	3	0.041713	50	0.011543

Lastly, the impact of Overlap on type II error probabilities for the known variance normal case is investigated. Comparing Eq.(9c) with Eq.(10), it clearly shows that the RHS of Eq. (10) is larger than that of Eq. (9c)→

$$Z_{\alpha/2}\sigma_y(1+\mathrm{K})/\sqrt{n_y}\;-\;Z_{\alpha/2}\sigma_y\sqrt{1+\mathrm{K}^2}\,/\sqrt{n_y}=Z_{\alpha/2}(\sigma_y/\sqrt{n_y})(1+\mathrm{K}-\sqrt{1+\mathrm{K}^2})>0$$

because $1+K > \sqrt{1+K^2}$. Thus rejecting H_0 when the two separate CIs are disjoint is more stringent than using the SMD of $\bar{x} - \bar{y}$ and will always lead to much less statistical power.

The Standard method Pr of committing a type II error (assuming $\mu_x > \mu_y$), using Figure 3 is given by

$$\beta = \Pr[-Z_{\alpha/2}\sqrt{\sigma_x^2/n_x + \sigma_y^2/n_y} \le \bar{x} - \bar{y} \le Z_{\alpha/2}\sqrt{\sigma_x^2/n_x + \sigma_y^2/n_y} \;|\delta] \tag{14a}$$

$$= \Pr[\frac{-Z_{\alpha/2}\sqrt{\sigma_x^2/n_x+\sigma_y^2/n_y}-\delta}{\sqrt{\sigma_x^2/n_x+\sigma_y^2/n_y}} \le \frac{(\bar{x}-\bar{y})-\delta}{\sqrt{\sigma_x^2/n_x+\sigma_y^2/n_y}} \le \frac{Z_{\alpha/2}\sqrt{\sigma_x^2/n_x+\sigma_y^2/n_y}-\delta}{\sqrt{\sigma_x^2/n_x+\sigma_y^2/n_y}}]$$

$$= \Pr[-Z_{\alpha/2} - \delta/\sqrt{\sigma_x^2/n_x+\sigma_y^2/n_y} \le Z \le Z_{\alpha/2} - \delta/\sqrt{\sigma_x^2/n_x+\sigma_y^2/n_y}] \tag{14b}$$

$$= \Pr\left[-Z_{\alpha/2} - \frac{\delta\sqrt{n_x}/\sigma_x}{\sqrt{1+K^2}} \le Z \le Z_{\alpha/2} - \frac{\delta\sqrt{n_x}/\sigma_x}{\sqrt{1+K^2}}\right] \tag{14c}$$

As in Schenker *et al.* (2001), let d represent a standardized difference, i.e., d = $\frac{\delta}{\sqrt{\sigma_x^2/n_x+\sigma_y^2/n_y}} = \frac{\delta\sqrt{n_y}/\sigma_y}{\sqrt{1+K^2}}$. Thus, the above equation results in the following form:

$$\beta = \Phi(Z_{\alpha/2} - d) - \Phi(-Z_{\alpha/2} - d) \tag{14d}$$

Eq. (14d) is the same result as Schenker *et al*. (2001, p.184) in their formula (6), except that they provide the equation for 1−β. Furthermore, when the null hypothesis H_0: $\mu_x - \mu_y = 0$ is not rejected at LOS α iff the two independent CIs ($\bar{x} - Z_{\alpha/2}\sigma_x/\sqrt{n_x} \le \mu_x \le \bar{x} + Z_{\alpha/2}\sigma_x/\sqrt{n_x}$) and ($\bar{y} - Z_{\alpha/2}\sigma_y/\sqrt{n_y} \le \mu_y \le \bar{y} + Z_{\alpha/2}\sigma_y/\sqrt{n_y}$) are overlapping, the Pr of a type II error (assuming $\mu_x > \mu_y$) from the Overlap method is given by

$$\beta' = \Pr(\text{Overlap} | \delta > 0) = \Pr\{[L(\mu_x) \le U(\mu_y)] \cap [L(\mu_y) \le U(\mu_x)] | \mu_x - \mu_y > 0\}$$

$$= \Pr\{[\bar{x} - Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n_x}} \le \bar{y} + Z_{\alpha/2}\frac{\sigma_y}{\sqrt{n_y}}] \cap [\bar{y} - Z_{\alpha/2}\frac{\sigma_y}{\sqrt{n_y}} \le \bar{x} + Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n_x}}] | \delta > 0\}$$

$$= \Pr\{[\bar{x} - \bar{y} \le Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n_x}} + Z_{\alpha/2}\frac{\sigma_y}{\sqrt{n_y}}] \cap [-Z_{\alpha/2}\frac{\sigma_y}{\sqrt{n_y}} - Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n_x}} \le \bar{x} - \bar{y}] | \delta > 0\}$$

$$= \Pr\{[-Z_{\alpha/2}\frac{\sigma_y}{\sqrt{n_y}} - Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n_x}} \le \bar{x} - \bar{y} \le Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n_x}} + Z_{\alpha/2}\frac{\sigma_y}{\sqrt{n_y}} | \delta > 0] \qquad (15a)$$

$$= \Pr[\frac{-Z_{\alpha/2}(\frac{\sigma_x}{\sqrt{n_x}} + \frac{\sigma_y}{\sqrt{n_y}}) - \delta}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \le \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \le \frac{Z_{\alpha/2}(\frac{\sigma_x}{\sqrt{n_x}} + \frac{\sigma_y}{\sqrt{n_y}}) - \delta}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}]$$

$$= \Pr[-Z_{\alpha/2}\frac{(1+\mathrm{K})}{\sqrt{1+\mathrm{K}^2}} - \frac{\delta}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \le Z \le Z_{\alpha/2}\frac{(1+\mathrm{K})}{\sqrt{1+\mathrm{K}^2}} - \frac{\delta}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}]$$

$$= \Phi(Z_{\alpha/2}\frac{(1+\mathrm{K})}{\sqrt{1+\mathrm{K}^2}} - \mathrm{d}) - \Phi(-Z_{\alpha/2}\frac{(1+\mathrm{K})}{\sqrt{1+\mathrm{K}^2}} - \mathrm{d}) \qquad (15b)$$

where $\mathrm{K}^2 = \mathrm{V}(\bar{x}) / \mathrm{V}(\bar{y})$, and $(1+\mathrm{K})/\sqrt{1+\mathrm{K}^2} = (\sigma_x/\sqrt{n_x} + \sigma_y/\sqrt{n_y})/$ $\sqrt{\sigma_x^2/n_x + \sigma_y^2/n_y}$. Thus, the PWF (power function) of the Overlap procedure in the case of known-Variances is

$$1-\beta' = \Phi(\mathrm{d} - Z_{\alpha/2}\frac{(1+\mathrm{K})}{\sqrt{1+\mathrm{K}^2}}) + \Phi(-Z_{\alpha/2}\frac{(1+\mathrm{K})}{\sqrt{1+\mathrm{K}^2}} - \mathrm{d}) \qquad (15c)$$

The result in Eq. (15c) is the same as that of Schenker *et al.* (2001, p. 184) in their Eq.(7), as they also provide the expression for $1 - \beta'$. Schenker *et al.* (2001) just provide both power functions without any explanation. The step by step derivations provided above have not been presented in statistical literature. For the comparison of β and β' see Table 6A. In Table 6A, the comparison is done for both $\alpha = 0.05$ and $\alpha = 0.01$. As the table shows, if d is fixed, as k increases, the type II error Pr increases. If k is fixed, the type II error rate decreases as d increases. Thus, as Table 6A shows, the probability of type II error based-on Overlap is larger than that of the Standard method, i.e., the Overlap method will lead to smaller statistical power. Secondly, when k is fixed, as d increases, $\beta' - \beta$ is not necessarily increasing or decreasing, this is consistent with figure 4 of Schenker and Gentleman (2001). Furthermore, Table 6A shows that at a fixed K the difference in percent relative power decreases as the standardized difference d increases for both α = 0.05 and $\alpha = 0.01$.

By definition, for an α-level test the relative efficiency of the Overlap to the Standard method, assuming the same statistical power, is given by

$$\text{RELEFF(Overlap to Standard)} = \text{RELEFF(O, ST)} = (n_x + n_y)/(n'_x + n'_y) \qquad (15d)$$

where the type II error Pr of the Standard method is given by Eq. (14) and n′ is the Overlap sample size for which $\beta' = \beta$. The exact solution to n′ is obtained by setting the first argument in Eq. (14b) to that of (15b), i.e.,

$$Z_{0.025}(1+K')/\sqrt{1+K'^2} - \delta/\sqrt{\sigma_x^2/n'_x + \sigma_y^2/n'_y} = Z_{0.025} - \delta/\sqrt{\sigma_x^2/n_x + \sigma_y^2/n_y} \qquad (15e)$$

Schenker and Gentleman (2001) state in their section 3 that the minimum ARE is ½ which clearly occurs at their limiting SE ratio of $k = 1$. We could obtain their value if we equate the argument of β on the RHS of Eq. (14a), $Z_{\alpha/2}\sqrt{\sigma_x^2/n_x + \sigma_y^2/n_y}$, to that of

Table 6A. The Relative Power of Overlap with the Standard Method for Different Sample Sizes n and $\delta/(\sigma\sqrt{2})$ Combinations for the Case of Known but Unequal Variances

$\alpha = 0.05$						$\alpha = 0.01$					
$\frac{\delta\sqrt{n_x}}{\sigma_x}$	**d**	**k**	$1-\beta$	$1-\beta'$	$(\frac{\beta'-\beta}{1-\beta})100\%$	$\frac{\delta\sqrt{n_x}}{\sigma_x}$	**d**	**k**	$1-\beta$	$1-\beta'$	$(\frac{\beta'-\beta}{1-\beta})100\%$
0.2	0.141	1	0.06898	0.00853	87.6361	0.1	0.071	1	0.01224	0.00035	97.10660
0.4	0.283	1	0.09352	0.01281	86.3005	0.3	0.212	1	0.01809	0.00060	96.67198
0.8	0.566	1	0.16323	0.02738	83.2293	0.5	0.354	1	0.02626	0.00100	96.17487
1	0.707	1	0.21026	0.03895	81.4745	1	0.707	1	0.06166	0.00333	94.60226
1.5	1.061	1	0.36849	0.08705	76.3756	1.5	1.061	1	0.12973	0.00982	92.43060
2	1.414	1	0.58524	0.17459	70.1672	2	1.414	1	0.24539	0.02584	89.46857
0.2	0.111	2	0.06445	0.00913	85.8305	0.1	0.055	1.5	0.01172	0.00044	96.27097
0.4	0.222	2	0.08220	0.01256	84.7235	0.3	0.166	1.5	0.01598	0.00066	95.86846
0.8	0.444	2	0.12947	0.02295	82.2715	0.5	0.277	1.5	0.02153	0.00099	95.42442
1	0.555	2	0.15994	0.03052	80.9184	1	0.555	1.5	0.04327	0.00255	94.10605
1.5	0.832	2	0.25936	0.05930	77.1341	1.5	0.832	1.5	0.08120	0.00614	92.43297
2	1.109	2	0.39501	0.10771	72.7329	2	1.109	1.5	0.14253	0.01379	90.32345
0.2	0.089	2	0.06141	0.01108	81.9557	0.1	0.045	2	0.01137	0.00065	94.30995
0.4	0.179	2	0.07490	0.01426	80.9631	0.3	0.134	2	0.01462	0.00089	93.87954
0.8	0.358	2	0.10911	0.02310	78.8304	0.5	0.224	2	0.01866	0.00123	93.41816
1	0.447	2	0.13034	0.02908	77.6870	1	0.447	2	0.03329	0.00262	92.11581
1.5	0.671	2	0.19735	0.05014	74.5919	1.5	0.671	2	0.05678	0.00535	90.57311
2	0.894	2	0.28663	0.08272	71.1422	2	0.894	2	0.09268	0.01042	88.75241
0.2	0.074	3	0.05934	0.01338	77.4461	0.1	0.037	2.5	0.01113	0.00093	91.64804
0.4	0.149	3	0.07008	0.01643	76.5492	0.3	0.111	2.5	0.01372	0.00121	91.19269
0.8	0.297	3	0.09634	0.02441	74.6610	0.5	0.186	2.5	0.01684	0.00156	90.71390
1	0.371	3	0.11216	0.02953	73.6684	1	0.371	2.5	0.02749	0.00291	89.40730
1.5	0.557	3	0.16065	0.04652	71.0407	1.5	0.557	2.5	0.04351	0.00525	87.93005
2	0.743	3	0.22353	0.07109	68.1980	2	0.743	2.5	0.06680	0.00918	86.26369
0.2	0.063	3	0.05787	0.01569	72.8768	0.1	0.032	3	0.01095	0.00125	88.56141
0.4	0.126	3	0.06673	0.01864	72.0702	0.3	0.095	3	0.01310	0.00156	88.09595
0.8	0.253	3	0.08783	0.02600	70.3948	0.5	0.158	3	0.01562	0.00193	87.61275
1	0.316	3	0.10023	0.03054	69.5255	1	0.316	3	0.02385	0.00326	86.32330
1.5	0.474	3	0.13738	0.04498	67.2582	1.5	0.474	3	0.03560	0.00537	84.91009
2	0.632	3	0.18434	0.06479	64.8547	2	0.632	3	0.05197	0.00865	83.36362
0.2	0.055	4	0.05678	0.01788	68.5051	0.1	0.027	3.5	0.01082	0.00159	85.26635
0.4	0.110	4	0.06430	0.02072	67.7825	0.3	0.082	3.5	0.01265	0.00192	84.80442
0.8	0.220	4	0.08183	0.02758	66.2952	0.5	0.137	3.5	0.01475	0.00231	84.32904
1	0.275	4	0.09194	0.03169	65.5304	1	0.275	3.5	0.02139	0.00362	83.07963
1.5	0.412	4	0.12165	0.04433	63.5559	1.5	0.412	3.5	0.03048	0.00557	81.73909
2	0.549	4	0.15839	0.06099	61.4915	2	0.549	3.5	0.04273	0.00842	80.30232

(15a), namely $Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n'_x}}+Z_{\alpha/2}\frac{\sigma_y}{\sqrt{n'_y}}$, and letting $\sigma_x = \sigma_y$ and $n_x = n_y = n$, but this seems to ignore the true mean difference $\delta = \mu_x - \mu_y$.

There are a numerous solutions for the Overlap sample sizes n'_x and n'_y from Eq. (15e) that must be at least as large as n_x and n_y in order to make the Overlap attain the same statistical power as the Standard method. Fortunately, an exact solution can be obtained only when $\sigma_x = \sigma_y$ and $n_x = n_y$ because it will be shown below that optimum efficiency will be achieved if $n'_x = n'_y$ and as a result the above equation reduces to

$$Z_{0.025}\sqrt{2} - (\delta/\sigma)\sqrt{n'/2} = Z_{0.025} - (\delta/\sigma)\sqrt{n/2}$$

The solution to this last equation is

$$\sqrt{n'} = Z_{0.025}(2-\sqrt{2})/(\delta/\sigma) + \sqrt{n} \qquad (15f)$$

Eq. (15f) clearly shows that as δ/σ increases, the value of n' decreases. Further, as n increases, the RELEFF of Overlap to the Standard method (n/n') increases. In fact, the larger δ/σ is, the faster the RELEFF(O, ST) approaches 100% as $n \to \infty$.

To obtain a rough approximation to (15e), we compare the 1st statement for β with the 4th statement for β' and equate $\sqrt{\sigma_x^2/n_x + \sigma_y^2/n_y}$ to $\sigma_x/\sqrt{n'_x} + \sigma_y/\sqrt{n'_y}$. Dividing both sides of the last equality by $\sigma_x/\sqrt{n_x}$ yields $\sqrt{1+K^2} = \sqrt{n_x/n'_x} + K\sqrt{n_y/n'_y}$, where $K = \frac{\sigma_y/\sqrt{n_y}}{\sigma_x/\sqrt{n_x}}$ is called the SE ratio. The equation $\sqrt{1+K^2} = \sqrt{n_x/n'_x} + K\sqrt{n_y/n'_y}$ shows that the solutions n'_x and n'_y do not depend on the specific values of σ_x and σ_y but rather only on their ratio σ_x/σ_y. Unfortunately, the same cannot be said about the ratio

$R_n = n_y/n_x$, i.e., n'_x and n'_y do depend on the specific values of n_x and n_y and not just on their ratio R_n. Further, the equation $\sqrt{1+K^2} = \sqrt{n_x/n'_x} + K\sqrt{n_y/n'_y}$ clearly shows that when $n'_x = n_x$ and $n'_y = n_y$, the RHS reduces to 1+ K which obviously exceeds the LHS $\sqrt{1+K^2}$ for all k. As $K \to \infty$, this last equation also shows that $n'_x \to n_x$ and $n'_y \to n_y$ so that the Overlap becomes an exact α-level test. When K > 1, the minimum $n'_x + n'_y$ occurs (i.e., the Overlap achieves its maximum relative efficiency) when $n'_x \geq n'_y$ and vice a versa when K < 1. It seems that we have a constrained optimization problem where $(n_x+n_y)/(n'_x+n'_y)$ is to be maximized subject to the nonlinear constraint $\sqrt{n_x/n'_x} + K\sqrt{n_y/n'_y} = \sqrt{1+K^2}$. The solution to this optimization can be obtained through the use of Lagrangian multipliers as shown below.

The objective is to maximize $f(n'_x, n'_y) = (n_x+n_y)/(n'_x+n'_y)$ subject to $\sqrt{1+K^2} - \sqrt{n_x/n'_x} - K\sqrt{n_y/n'_y} = 0$ and hence it is sufficient to maximize $f(n'_x, n'_y) = N/(n'_x+n'_y) + \lambda(\sqrt{1+K^2} - \sqrt{n_x/n'_x} - K\sqrt{n_y/n'_y})$, where $N = n_x + n_y$ and λ is an arbitrary constant. Taking the partial derivatives of $f(n'_x, n'_y)$ with respective to n'_x & n'_y and setting them equal to zero yields:

$$\partial f/\partial n'_x = -N(n'_x+n'_y)^{-2} + \lambda(\sqrt{n_x}/2)(n'_x)^{-3/2} \xrightarrow{\text{Set to}} 0$$

$$\partial f/\partial n'_y = -N(n'_x+n'_y)^{-2} + \lambda(K\sqrt{n_y}/2)(n'_y)^{-3/2} \xrightarrow{\text{Set to}} 0$$

Because λ is a completely arbitrary constant, the above system is satisfied as soon as we equate $\sqrt{n_x}\,(n'_x)^{-3/2}$ to $K\sqrt{n_y}\,(n'_y)^{-3/2}$, i.e., $\sqrt{n_x}\,(n'_x)^{-3/2} = K\sqrt{n_y}\,(n'_y)^{-3/2} \rightarrow$

$n_x (n'_x)^{-3} = K^2 n_y (n'_y)^{-3} \rightarrow K^2 n_y / n_x = (n'_x)^{-3} / (n'_y)^{-3} \rightarrow K^2 n_y / n_x = (n'_y / n'_x)^3 \rightarrow$ $n'_y / n'_x = (K^2 n_y / n_x)^{1/3}$; thus the optimum solution is obtained if we select n'_x and n'_y in such a manner that their ratio n'_y / n'_x is close to $(K^2 n_y / n_x)^{1/3}$. Table 6B provides the RELEFF of Overlap to the Standard for various values of δ/σ only for the case of $n_x = n_y = n$ and $\sigma_x = \sigma_y = \sigma$ for which K = 1. When $\sigma_x \neq \sigma_y$, there are uncountable ways that K can equal 1, and therefore, the procedure is to solve n'_x and n'_y from (15f) and to compute the RELEFF from the ratio of $(n_x+n_y)/(n'_x + n'_y)$.

The results of this chapter verifies what has been reported in Overlap literature for the limiting case (i.e., large sample sizes) by Goldstein H. & Healy MJR (1995), Payton *et al.* (2000), Schenker N. & Gentleman J. F (2001), and Payton et al. (2003). Payton *et al.* (2000) report some approximate Overlap results for smaller sample sizes ($n_x = n_y = n = 5(5)\ 50$) but used simulation to obtain them instead of the exact normal theory as applied here in Chapter 3. Further, it must be emphasized that the results reported in this chapter will also apply to non-normal underlying populations only if both n_x & $n_y > 60$. This is due the Central Limit Theorem (CLT) that states the sample mean distribution from non-normal population approaches normality as $n \rightarrow \infty$. In practice, the rate of approach to normality depends only on skewness and kurtosis of the underlying distributions. It is well known that both the skewness and kurtosis of a normal universe are zero. The closer the skewness and kurtosis of the parent populations are to zero, the more rapidly the means ($\bar{x}$ and $\bar{y}$) approach normality. For example, because the skewness of a uniform distribution is zero and its kurtosis is −1.20, only samples of size at least 6 are needed for the corresponding sample mean to be approximately normally

Table 6B. RELEFF of Overlap to the Standard Method at α = 0.05 and K=1

0.2				0.4				0.6			
n	RELEFF	n	RELEFF	n	RELEFF	n	RELEFF	n	RELEFF	n	RELEFF
4	0.06676	25	0.21671	4	0.16864	25	0.40361	4	0.26117	25	0.52305
5	0.07858	30	0.23840	5	0.19175	30	0.43053	5	0.29037	30	0.54922
6	0.08945	35	0.25758	6	0.21201	35	0.45337	6	0.31519	35	0.57094
8	0.10895	40	0.27479	8	0.24634	40	0.47312	8	0.35577	40	0.58940
10	0.12617	50	0.30462	10	0.27479	50	0.50592	10	0.38814	50	0.61940
12	0.14163	60	0.32987	12	0.29907	60	0.53236	12	0.41495	60	0.64305
14	0.15571	70	0.35174	14	0.32023	70	0.55438	14	0.43776	70	0.66237
16	0.16864	80	0.37098	16	0.33898	80	0.57313	16	0.45754	80	0.67859
18	0.18060	90	0.38814	18	0.35577	90	0.58940	18	0.47495	90	0.69248
20	0.19175	100	0.40361	20	0.37098	100	0.60370	20	0.49048	100	0.70456
0.8				1				1.5			
n	RELEFF	n	RELEFF	n	RELEFF	n	RELEFF	n	RELEFF	n	RELEFF
4	0.33898	25	0.60370	4	0.40361	25	0.66139	4	0.52305	25	0.75211
5	0.37098	30	0.62787	5	0.43658	30	0.68345	5	0.55501	30	0.76981
6	0.39760	35	0.64766	6	0.46358	35	0.70136	6	0.58052	35	0.78401
8	0.44009	40	0.66431	8	0.50592	40	0.71632	8	0.61940	40	0.79574
10	0.47312	50	0.69103	10	0.53823	50	0.74014	10	0.64822	50	0.81419
12	0.49994	60	0.71180	12	0.56411	60	0.75849	12	0.67081	60	0.82823
14	0.52240	70	0.72860	14	0.58553	70	0.77323	14	0.68919	70	0.83939
16	0.54162	80	0.74258	16	0.60370	80	0.78542	16	0.70456	80	0.84855
18	0.55836	90	0.75447	18	0.61940	90	0.79574	18	0.71769	90	0.85626
20	0.57313	100	0.76474	20	0.63317	100	0.80463	20	0.72908	100	0.86286
2				2.5				3			
n	RELEFF	n	RELEFF	n	RELEFF	n	RELEFF	n	RELEFF	n	RELEFF
4	0.60370	25	0.80463	4	0.66139	25	0.83883	4	0.70456	25	0.86286
5	0.63317	30	0.81927	5	0.68826	30	0.85126	5	0.72908	30	0.87365
6	0.65632	35	0.83092	6	0.70916	35	0.86112	6	0.74800	35	0.88217
8	0.69103	40	0.84050	8	0.90471	40	0.86919	8	0.77584	40	0.88914
10	0.71632	50	0.85546	10	0.76246	50	0.88175	10	0.79574	50	0.89995
12	0.73589	60	0.86677	12	0.77959	60	0.89119	12	0.81092	60	0.90805
14	0.75166	70	0.87571	14	0.79331	70	0.89864	14	0.82303	70	0.91443
16	0.76474	80	0.88302	16	0.80463	80	0.90471	16	0.83298	80	0.91962
18	0.77584	90	0.88914	18	0.81419	90	0.90978	18	0.84136	90	0.92395
20	0.78542	100	0.89437	20	0.82242	100	0.91411	20	0.84855	100	0.92764

distributed. This is due to the fact that the skewness of the 6-fold convolution of a U(0, 1) is zero (due to symmetry) while its kurtosis is $-1.20/6 = -0.20$. It can be shown that the kurtosis of an n-fold convolution of the U(0, 1) is exactly equal to $-1.20/n$ (see Appendix A). Further, our experience indicates [Hool J. N. and Maghsoodloo S. (1980) and Maghsoodloo S. and Hool J. N. (1981)] that the 3rd moment (skewness) plays a more important role in normal approximation of a linear combination than the 4th moment (kurtosis).

4.0 Bonferroni Intervals for Comparing Two Sample Means

The two independent 95% confidence intervals for each of the two population means have a joint Pr of 0.95^2 of containing μ_x and μ_y. Although, this concept of joint Pr has not been considered in the Overlap literature, we consider it here to investigate its impact on type I & II error rates from the Overlap method. In order to compare two 95% CIs against a single 95% CI for $\mu_x - \mu_y$, it may be best to use the Bonferroni concept so that the overall confidence Pr (regardless of the correlation structure) of the two CIs is raised from $0.95^2 = 0.9025$ to 0.95. This is accomplished by setting individual CI coefficient at $1 - \alpha = \sqrt{0.95} = 0.9746794345$ so that the joint confidence level will equal to $(\sqrt{0.95})^2$. To this end, let $(1 - \alpha_B) = \sqrt{0.95}$ (the subscript B stands for Bonferroni) = 0.9746794345; thus, $\alpha_B = 0.02532056552$, which results in $\alpha_B/2 = 0.01266028276$ and $Z_{0.0126603} = 2.23647664456$. Thus, the 97.468% confidence Pr statement for μ_x is $\Pr(\bar{x} - Z_{0.0126603}\sigma_x/\sqrt{n_x} \le \mu_x \le \bar{x} + Z_{0.0126603}\sigma_x/\sqrt{n_x}) = 0.97468$. As a result, the lower 97.468% Bonferroni CI limit for μ_x is $L(\mu_x) = \bar{x} - Z_{0.0126603}\sigma_x/\sqrt{n_x}$ and the corresponding upper limit is $U(\mu_x) = \bar{x} + Z_{0.0126603}\sigma_x/\sqrt{n_x}$ resulting in the Bonferroni CIL (confidence interval length) of $CIL(\mu_x) = 2\times Z_{0.0126603}\sigma_x/\sqrt{n_x}$. Following the same procedure, the 97.468% CI for μ_y will be: $L(\mu_y) = \bar{y} - Z_{0.0126603}\sigma_y/\sqrt{n_y}$, $U(\mu_y) = \bar{y} + Z_{0.01266}\sigma_y/\sqrt{n_y}$, and the corresponding Bonferroni $CIL(\mu_y) = 2Z_{0.0126603}\sigma_y/\sqrt{n_y}$.

The Bonferroni confidence Intervals for μ_x and μ_y will not change the 95% CI for $\mu_x - \mu_y$, i.e., the 95% CI for $\mu_x - \mu_y$ is still the same as in Eq. (9b), as shown below:

$$\bar{x} - \bar{y} - Z_{0.025}(\sigma_x / \sqrt{n_x}) \times \sqrt{1+K^2} \le \mu_x - \mu_y \le \bar{x} - \bar{y} + Z_{0.025}(\sigma_x / \sqrt{n_x}) \times \sqrt{1+K^2} .$$

The 95% CI in Eq.(9c) shows that the $H_0 : \mu_x - \mu_y = 0$ must be rejected at the 5% level of significance iff $|\bar{x} - \bar{y}| > Z_{0.025}(\sigma_x / \sqrt{n_x}) \times \sqrt{1+K^2}$. However, requiring that the two separate independent CIs must be disjoint in order to reject $H_0 : \mu_x - \mu_y = 0$ at the 5% level, is equivalent to either $L(\mu_x) > U(\mu_y)$, or $L(\mu_y) > U(\mu_x)$. These two possibilities lead to either $\bar{x} - Z_{0.01266}\sigma_x / \sqrt{n_x} > \bar{y} + Z_{0.01266}\sigma_y / \sqrt{n_y}$, or $\bar{y} - Z_{0.01266}\sigma_y / \sqrt{n_y} > \bar{x} + Z_{0.01266}\sigma_x / \sqrt{n_x}$, respectively. Inserting $\sigma_y / \sqrt{n_y} = K\sigma_x / \sqrt{n_x}$ into this last inequality leads to the rejection of H_0 iff

$$|\bar{x} - \bar{y}| > Z_{0.0126603}\sigma_x / \sqrt{n_x} + Z_{0.0126603}\sigma_y / \sqrt{n_y} = Z_{0.0126603}(1+K)\sigma_x / \sqrt{n_x} \qquad (16a)$$

Thus the Bonferroni CIL for the Eq.(16a) is $2Z_{0.0126603}(1+K)\sigma_x / \sqrt{n_x}$. (16b)

Using the same procedures as in chapter 3, if we set the exact type I error at 5% and reject H_0 when the two independent CIs do not overlap; then the Bonferroni type I error Pr reduces to

$$\alpha'_B$$

$$= \Pr(\bar{x} + Z_{0.0126603}\frac{\sigma_x}{\sqrt{n_x}} < \bar{y} - Z_{0.0126603}\frac{\sigma_y}{\sqrt{n_y}}) + \Pr(\bar{x} - Z_{0.0126603}\frac{\sigma_x}{\sqrt{n_x}} > \bar{y} + Z_{0.0126603}\frac{\sigma_y}{\sqrt{n_y}})$$

$$= 2 \times \Pr\left[\bar{x} - \bar{y} > Z_{0.0126603}\sigma_x(1+K)/\sqrt{n_x}\right] = 2 \times \Pr\left[Z > \frac{Z_{0.0126603}\sigma_x(1+K)/\sqrt{n_x}}{\sqrt{\sigma_x^2/n_x + \sigma_y^2/n_y}}\right]$$

$$= 2 \times \Pr\left[Z > \frac{Z_{0.0126603}\sigma_x (1+K)/\sqrt{n_x}}{\sqrt{\sigma_x^2/n_x + K^2\sigma_x^2/n_x}} \right] = 2\times\Pr[Z > Z_{0.0126603}\ (1+K)/\sqrt{1+K^2}\ ] \qquad (17)$$

Eq. (11) leads to $\alpha' = 2\times\Pr[Z > Z_{0.025}(1+K)/\sqrt{1+K^2}\ ]$. Comparing Eq.(17) with Eq.(11),

clearly, since $Z_{0.0126603} > Z_{0.025} \rightarrow \frac{Z_{0.0126603}\ (1+K)}{\sqrt{1+K^2}} > Z_{0.025}\frac{(1+K)}{\sqrt{1+K^2}}$, then $\alpha'_B < \alpha'$.

Thus, the Bonferroni intervals lead to an even smaller type I error Pr than both α and α', i.e., $\alpha'_B < \alpha' < \alpha$. Using the same logic as before, the minimum α'_B occurs at K = 1. Figure 6 shows that from k = 0.1 to 10 and at α = 0.05 → $\alpha'_B < \alpha' - \alpha'_B < \alpha' < \alpha = 0.05$. Moreover, Figure 6 shows that the minimum α'_B occurs when K =1 (or see Table 7).

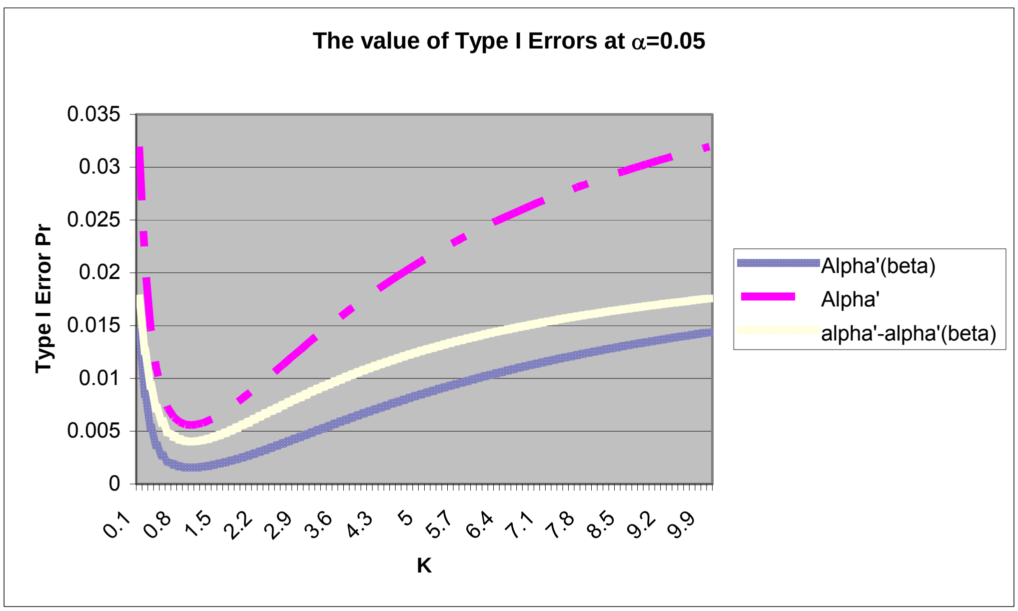


Figure 6

As before, let O represent the amount of overlap length between the two individual Bonferroni CIs. Using the same procedure as the former section →

Table7. Type I Errors for Overlap and Bonferroni Methods at α =0.05

K	α'_B	α'	K	α'_B	α'
1	0.0015623	0.0055746	6	0.0100611	0.0241012
1.2	0.0016335	0.0057726	6.2	0.0103455	0.0246370
1.4	0.0018096	0.0062552	6.4	0.0106211	0.0251532
1.6	0.0020571	0.0069168	6.6	0.0108879	0.0256507
1.8	0.0023567	0.0076952	6.8	0.0111465	0.0261302
2	0.0026949	0.0085494	7	0.0113970	0.0265926
2.2	0.0030617	0.0094502	7.2	0.0116398	0.0270387
2.4	0.0034487	0.0103763	7.4	0.0118751	0.0274693
2.6	0.0038493	0.0113120	7.6	0.0121032	0.0278850
2.8	0.0042579	0.0122456	7.8	0.0123243	0.0282865
3	0.0046702	0.0131685	8	0.0125388	0.0286745
3.2	0.0050825	0.0140746	8.2	0.0127469	0.0290496
3.4	0.0054921	0.0149595	8.4	0.0129488	0.0294124
3.6	0.0058968	0.0158204	8.6	0.0131448	0.0297634
3.8	0.0062950	0.0166553	8.8	0.0133350	0.0301032
4	0.0066853	0.0174633	9	0.0135198	0.0304323
4.2	0.0070670	0.0182438	9.2	0.0136993	0.0307510
4.4	0.0074393	0.0189968	9.4	0.0138738	0.0310600
4.6	0.0078019	0.0197225	9.6	0.0140433	0.0313595
4.8	0.0081545	0.0204217	9.8	0.0142082	0.0316501
5	0.0084970	0.0210948	10	0.0143686	0.0319320
5.2	0.0088295	0.0217428	10.2	0.0145246	0.0322057
5.4	0.0091520	0.0223665	10.4	0.0146764	0.0324714
5.6	0.0094646	0.0229668	10.6	0.0148242	0.0327296
5.8	0.0097676	0.0235447	10.8	0.0149681	0.0329805

$$O = U(\mu_y) - L(\mu_x) = (\bar{y} + Z_{0.0126603} \times \sigma_y / \sqrt{n_y}) - (\bar{x} - Z_{0.0126603} \times \sigma_x / \sqrt{n_x})$$

$$= Z_{0.0126603}\left(\sigma_x / \sqrt{n_x} + \sigma_y / \sqrt{n_y}\right) - (\bar{x} - \bar{y}) \qquad \text{(18a)}$$

Let O_r be the border line value of O at which H_0 is barely rejected at the 5% level.

From Eq. (9c), $H_0{:}\mu_x = \mu_y$ should be rejected iff $|\bar{x} - \bar{y}| > Z_{0.025}\sigma_x\sqrt{1+K^2} / \sqrt{n_x}$.

Therefore, the value of

$$O_r = Z_{0.0126603}\left(\sigma_x / \sqrt{n_x} + \sigma_y / \sqrt{n_y}\right) - Z_{0.025}\sigma_x\sqrt{1+K^2} / \sqrt{n_x}$$

$$= (\sigma_x / \sqrt{n_x}) \times \left[Z_{0.0126603}(1+\mathrm{K}) - Z_{0.025}\sqrt{1+\mathrm{K}^2} \right] \qquad (18b)$$

Eq. (18b) indicates that H_0 must be rejected at the 5% or less level iff $O \le (\sigma_x / \sqrt{n_x}) \times \left[Z_{0.0126603}(1+\mathrm{K}) - Z_{0.025}\sqrt{1+\mathrm{K}^2} \right]$. Further, the span of the two individual CIs is

$U(\mu_x) - L(\mu_y) = (\bar{x} + Z_{0.0126603} \times \sigma_x / \sqrt{n_x}) - (\bar{y} - Z_{0.0126603} \times \sigma_y / \sqrt{n_y})$

$$= (\sigma_x / \sqrt{n_x}) \times \left[Z_{0.0126603}(1+\mathrm{K}) + Z_{0.025}\sqrt{1+\mathrm{K}^2} \right] \qquad (18c)$$

Thus, the percentage of the overlap length at the borderline condition for the Bonferroni case is given by $\dfrac{U(\mu_Y) - L(\mu_X)}{U(\mu_X) - L(\mu_Y)} \times 100\%$

$$= [\frac{Z_{0.0126603}(1+\mathrm{K}) - Z_{0.025}\sqrt{1+\mathrm{K}^2}}{Z_{0.0126603}(1+\mathrm{K}) + Z_{0.025}\sqrt{1+\mathrm{K}^2}}] \times 100\% \qquad (18d)$$

Let h(K)= $[\dfrac{Z_{0.0126603}(1+\mathrm{K}) - Z_{0.025}\sqrt{1+\mathrm{K}^2}}{Z_{0.0126603}(1+\mathrm{K}) + Z_{0.025}\sqrt{1+\mathrm{K}^2}}]$. From Maple, $h'(\mathrm{K}) =$

$$\frac{Z_B - Z_{0.025}\mathrm{K}/\sqrt{1+\mathrm{K}^2}}{Z_B + Z_B\mathrm{K} + Z_{0.025}\sqrt{1+\mathrm{K}^2}} - \frac{(Z_B + Z_B\mathrm{K} - Z_{.025}\sqrt{1+\mathrm{K}^2})(Z_B + Z_{0.025}\mathrm{K}/\sqrt{1+\mathrm{K}^2})}{(Z_B + Z_B\mathrm{K} + Z_{0.025}\sqrt{1+\mathrm{K}^2})^2} \qquad (18e)$$

Plugging K = 1 into Eq.(18e), result in $h'(\mathrm{K})|_{\mathrm{K}=1} = 0.117405174 - 0.117405174 = 0$ and

$h''(\mathrm{K})|_{\mathrm{K}=1} = -0.095648707 - 0.425286147 + 0.117405174 - 0.022459306 = -0.425988985 < 0$,

which implies that K =1 maximizes h(K). Thus, the maximum overlap occurs when K=1 as before. Table 8 shows that, at the same K, the amount of overlap based on Bonferroni concept is larger than that of two individual CIs at LOS of 0.05. As K increases, the difference in overlap and Bonferroni overlap monotonically and slowly

Table 8. The Impact of Bonferroni on Percent Overlap at Different K

K	Bonferroni Overlap(%)	Overlap (%)	Difference (%)	K	Bonferroni Overlap(%)	Overlap (%)	Difference (%)
1	23.481035	17.157288	6.323747	3.1	17.907989	11.453938	6.454051
1.1	23.427525	17.102324	6.325201	3.2	17.678328	11.219816	6.458512
1.2	23.286523	16.957512	6.329012	3.3	17.456447	10.993694	6.462753
1.3	23.082143	16.747656	6.334487	3.4	17.242089	10.775302	6.466787
1.4	22.832795	16.491705	6.341089	3.5	17.034990	10.564364	6.470625
1.5	22.552471	16.204060	6.348410	3.6	16.834880	10.360601	6.474279
1.6	22.251757	15.895613	6.356143	3.7	16.641492	10.163735	6.477758
1.7	21.938629	15.574565	6.364064	3.8	16.454562	9.973490	6.481072
1.8	21.619071	15.247062	6.372009	3.9	16.273829	9.789598	6.484232
1.9	21.297543	14.917681	6.379862	4	16.099042	9.611797	6.487245
2	20.977344	14.589803	6.387541	4.1	15.929955	9.439834	6.490121
2.1	20.660894	14.265901	6.394993	4.2	15.766332	9.273465	6.492867
2.2	20.349935	13.947753	6.402182	4.3	15.607946	9.112454	6.495491
2.3	20.045701	13.636613	6.409088	4.4	15.454576	8.956577	6.497999
2.4	19.749033	13.333333	6.415700	4.5	15.306015	8.805616	6.500399
2.5	19.460480	13.038464	6.422016	4.6	15.162061	8.659366	6.502695
2.6	19.180364	12.752325	6.428039	4.7	15.022524	8.517630	6.504894
2.7	18.908840	12.475065	6.433775	4.8	14.887219	8.380218	6.507001
2.8	18.645939	12.206705	6.439233	4.9	14.755973	8.246951	6.509022
2.9	18.391594	11.947169	6.444424	5	14.628619	8.117658	6.510960
3	18.145672	11.696312	6.449360	5.1	14.504998	7.992177	6.512821

increases toward the limit of 0.065892072.

Finally, the same conclusion as before can be reached that separate CIs lead to larger type Π error when Bonferroni concept is applied. The exact Pr of type Π error is the same as Eq.(14) → $\beta = \Phi(Z_{\alpha/2} - d) - \Phi(-Z_{\alpha/2} - d)$.

For β'_B (B stands for Bonferroni) will be changed to

$$\beta'_B = \Pr(\text{Overlap} | \mu_x - \mu_y = \delta) = \Pr\{[L(\mu_x) \le U(\mu_y)] \cap [L(\mu_y) \le U(\mu_x)] | \delta\}$$

$$= \Phi\left(Z_{0.0126603}\frac{1+K}{\sqrt{1+K^2}} - d\right) - \Phi\left(-Z_{0.0126603}\frac{1+K}{\sqrt{1+K^2}} - d\right) \qquad (19)$$

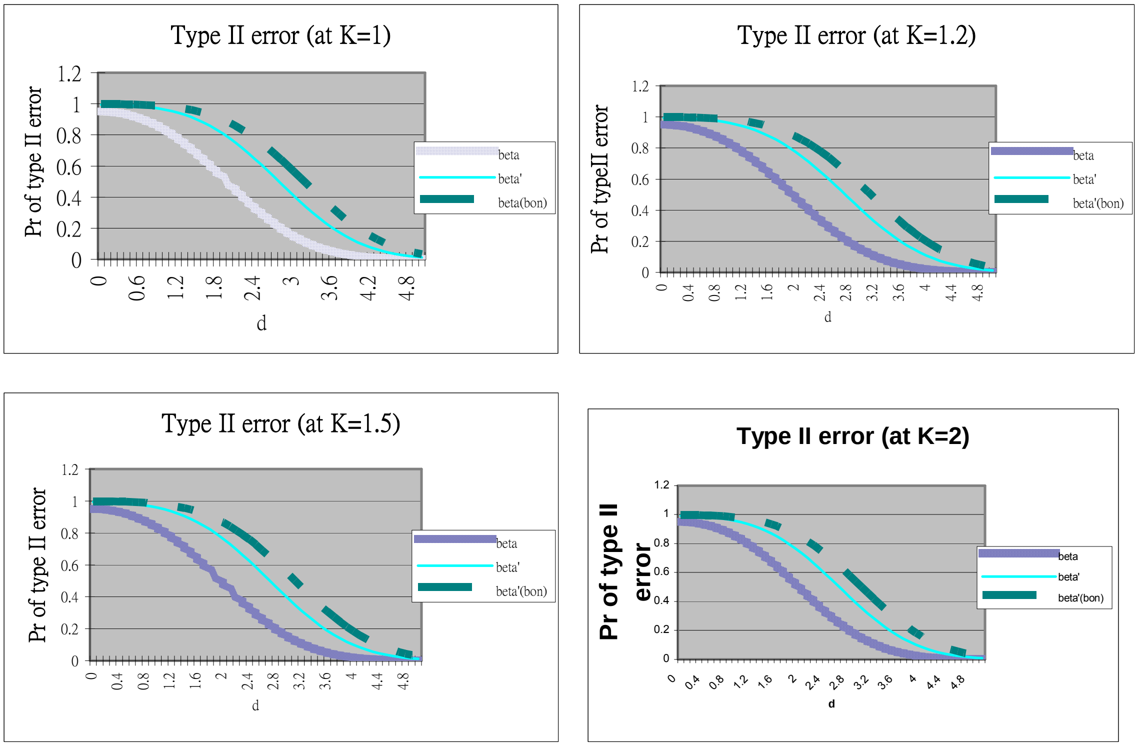
Table 9 clearly shows that the Bonferroni concept leads to the largest type Π error Pr than other two methods, i.e., $\beta'_B > \beta' > \beta$. Because the Bonferroni CIs always have larger

Table 9. Type II Error Pr for the Standard, Overlap, and Bonferroni Methods with Different K and d Combinations

K	d	β	β'	β'_B	K	d	β	β'	β'_B
1	0	0.95	0.994425	0.998438	1.8	0	0.95	0.992305	0.997643
1	0.2	0.921586	0.993461	0.998090	1.8	0.2	0.921586	0.991068	0.997157
1	0.4	0.881232	0.990392	0.996952	1.8	0.4	0.881232	0.987161	0.995579
1	0.6	0.826159	0.984692	0.994725	1.8	0.6	0.826159	0.979999	0.992544
1	0.8	0.753937	0.975507	0.990896	1.8	0.8	0.753937	0.968656	0.987431
1	1	0.662927	0.961706	0.984708	1.8	1	0.662927	0.951936	0.979356
1.2	0	0.95	0.994227	0.998367	2	0	0.95	0.991451	0.997305
1.2	0.2	0.921586	0.993237	0.998006	2	0.2	0.921586	0.990111	0.996763
1.2	0.4	0.881232	0.990085	0.996826	2	0.4	0.881232	0.985887	0.995010
1.2	0.6	0.826159	0.984241	0.994523	2	0.6	0.826159	0.97818	0.991656
1.2	0.8	0.753937	0.974842	0.990571	2	0.8	0.753937	0.96604	0.986044
1.2	1	0.662927	0.960747	0.984200	2	1	0.662927	0.948262	0.977248
1.4	0	0.95	0.993745	0.998190	2.5	0	0.95	0.989156	0.996352
1.4	0.2	0.921586	0.992690	0.997798	2.5	0.2	0.921586	0.987554	0.995662
1.4	0.4	0.881232	0.989343	0.996518	2.5	0.4	0.881232	0.98253	0.993443
1.4	0.6	0.826159	0.983155	0.994030	2.5	0.6	0.826159	0.973451	0.989250
1.4	0.8	0.753937	0.973245	0.989780	2.5	0.8	0.753937	0.959334	0.982342
1.4	1	0.662927	0.958455	0.982970	2.5	1	0.662927	0.938958	0.971701
1.6	0	0.95	0.993083	0.997943	3	0	0.95	0.986832	0.995330
1.6	0.2	0.921586	0.991944	0.997508	3	0.2	0.921586	0.984982	0.994490
1.6	0.4	0.881232	0.988334	0.996090	3	0.4	0.881232	0.979206	0.991807
1.6	0.6	0.826159	0.981690	0.993349	3	0.6	0.826159	0.968852	0.986788
1.6	0.8	0.753937	0.971106	0.988699	3	0.8	0.753937	0.952921	0.978626
1.6	1	0.662927	0.955405	0.981300	3	1	0.662927	0.930202	0.966232

confidence bands, will always lead to larger % overlap and to smaller type I error and larger type II error rates, they will not be henceforth considered.

Moreover, Figure 7 shows the three type II errors (β, β', β'_B) at k =1, 1.2, 1.5 and 2. These four figures clearly show the relation that $\beta'_B > \beta' > \beta$. In other words, Bonferroni method will lead to the largest type II error.



Type II error (at K=1)
Pr of type II error
1.2
1
0.8
0.6
0.4
0.2
0
0
0.6
1.2
1.8
2.4
3
3.6
4.2
4.8
d
beta
beta'
beta(bon)
Type II error (at K=1.2)
Pr of typeII error
1.2
1
0.8
0.6
0.4
0.2
0
0
0.4
0.8
1.2
1.6
2
2.4
2.8
3.2
3.6
4
4.4
4.8
d
beta
beta'
beta'(bon)
Type II error (at K=1.5)
Pr of type II error
1.2
1
0.8
0.6
0.4
0.2
0
0
0.4
0.8
1.2
1.6
2
2.4
2.8
3.2
3.6
4
4.4
4.8
d
beta
beta'
beta'(bon)
Type II error (at K=2)
Pr of type II error
1.2
1
0.8
0.6
0.4
0.2
0
0
0.4
0.8
1.2
1.6
2
2.4
2.8
3.2
3.6
4
4.4
4.8
d
beta
beta'
beta'(bon)

Figure 7

5.0 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 (pooled t-test and two-sample t-test) to compare independent normal means when variances are unknown, it is prudent to pretest $H_0: \sigma_x^2 = \sigma_y^2$ at an α–level. Because statistical literature cautions against using the pooled t-test unless there is convincing evidence in favor of $H_0: \sigma_x^2 = \sigma_y^2$, then when testing $H_0: \sigma_x^2 = \sigma_y^2$ just to ascertain to pool or not, the LOS α will be set much higher than 5%.

Consider a random sample of size n_x from the normal universe $N(\mu_x, \sigma_x^2)$. Using the fact that the rv $\frac{(n_x - 1)S_x^2}{\sigma_x^2}$ has a chi-square distribution with $\nu_x = n_x - 1$ degrees of freedom, it follows that the Pr [$\chi^2_{1-\alpha/2,\nu_x} < \frac{(n-1)S_x^2}{\sigma_x^2} < \chi^2_{\alpha/2,\nu_x}$] = 1 − α. Rearranging this last Pr statement results in the (1 − α)100% CI for σ_x^2 →

$$\frac{(n_x - 1)S_x^2}{\chi^2_{\alpha/2,\nu_x}} < \sigma_x^2 < \frac{(n_x - 1)S_x^2}{\chi^2_{1-\alpha/2,\nu_x}} \tag{20a}$$

Hence, the lower CI limit for σ_x^2 is $L(\sigma_x^2) = \frac{\nu_x S_x^2}{\chi^2_{\alpha/2,\nu_x}}$ and the upper CI limit is $U(\sigma_x^2) = \frac{\nu_x S_x^2}{\chi^2_{1-\alpha/2,\nu_x}}$. These CI lower and upper limits result in the confidence interval length

$$\text{CIL}(\sigma_x^2) = \text{U}(\sigma_x^2) - \text{L}(\sigma_x^2) = \frac{\nu_x S_x^2}{\chi^2_{1-\alpha/2,\nu_x}} - \frac{\nu_x S_x^2}{\chi^2_{\alpha/2,\nu_x}}$$

$$= \nu_x S_x^2 \times \left(\frac{1}{\chi^2_{1-\alpha/2,\nu_x}} - \frac{1}{\chi^2_{\alpha/2,\nu_x}}\right) \quad \text{(20b)}$$

The same procedure as above leads to the $(1 - \alpha)100\%$ upper and lower CI limits for σ_y^2

as $\text{L}(\sigma_y^2) = \dfrac{(n_y-1)S_y^2}{\chi^2_{\alpha/2,\nu_y}}$, $\text{U}(\sigma_y^2) = \dfrac{(n_y-1)S_y^2}{\chi^2_{1-\alpha/2,\nu_y}}$ and

$$\text{CIL}(\sigma_y^2) = \nu_y S_y^2 \times \left[\frac{1}{\chi^2_{1-\alpha/2,\nu_y}} - \frac{1}{\chi^2_{\alpha/2,\nu_y}}\right]. \quad \text{(20c)}$$

With the above information, requiring that the two independent CIs must be disjoint in order to reject H_0: $\sigma_x^2 = \sigma_y^2$ at the $\alpha\times100\%$ level is equivalent to either $\text{L}(\sigma_x^2)$ $> \text{U}(\sigma_y^2)$ or $\text{L}(\sigma_y^2) > \text{U}(\sigma_x^2)$, i.e., $\text{L}(\sigma_x^2) > \text{U}(\sigma_y^2) \rightarrow \dfrac{(n_x-1)S_x^2}{\chi^2_{\alpha/2,n_x-1}} > \dfrac{(n_y-1)S_y^2}{\chi^2_{1-\alpha/2,n_y-1}} \rightarrow$

Thus, based on the Overlap procedure, reject H_0 if

$$F_0 = \frac{S_x^2}{S_y^2} > \frac{\nu_y}{\nu_x} \times \frac{\chi^2_{\alpha/2,\nu_x}}{\chi^2_{1-\alpha/2,\nu_y}}, \text{ or } F_0 = \frac{S_x^2}{S_y^2} > \frac{\nu_y}{\nu_x} \times C_{\alpha/2,\nu_x,\nu_y}, \quad \text{(21a)}$$

where $C_{\alpha/2,\nu_x,\nu_y} = \dfrac{\chi^2_{\alpha/2,\nu_x}}{\chi^2_{1-\alpha/2,\nu_y}}$.

Or $\text{L}(\sigma_y^2) > \text{U}(\sigma_x^2) \rightarrow \dfrac{(n_y-1)S_y^2}{\chi^2_{\alpha/2,\nu_y}} > \dfrac{(n_x-1)S_x^2}{\chi^2_{1-\alpha/2,\nu_x}} \rightarrow \dfrac{S_x^2}{S_y^2} < \dfrac{\nu_y}{\nu_x} \times \dfrac{\chi^2_{1-\alpha/2,\nu_x}}{\chi^2_{\alpha/2,\nu_y}}$

$$\rightarrow \text{ Reject } H_0 \text{ if } F_0 = \frac{S_x^2}{S_y^2} < \frac{\nu_y}{\nu_x} \times C_{1-\alpha/2,\nu_x,\nu_y} \quad \text{(21b)}$$

However, the exact (1− α)100% CI for the ratio of the two independent normal variances must be obtained from the Fisher's F distribution as follows:

Consider two samples from N(μ_x, σ_x^2) and N(μ_y, σ_y^2), respectively. Thus,

$$\frac{(n_x-1)S_x^2}{\sigma_x^2} \rightarrow \chi^2_{n_x-1} \;;\quad \frac{(n_y-1)S_y^2}{\sigma_y^2} \rightarrow \chi^2_{n_y-1} \;;\; F_{n_x-1,n_y-1} = \frac{[(n_x-1)S_x^2/\sigma_x^2]/(n_x-1)}{[(n_y-1)S_y^2/\sigma_y^2]/(n_y-1)}$$

$$\rightarrow \Pr(F_{1-\alpha/2,\nu_x,\nu_y} \le \frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} \le F_{\alpha/2,\nu_x,\nu_y}) = 1-\alpha$$

$$\rightarrow \frac{S_x^2}{S_y^2} F_{1-\alpha/2,\nu_y,\nu_x} \le \frac{\sigma_x^2}{\sigma_y^2} \le \frac{S_x^2}{S_y^2} F_{\alpha/2,\nu_y,\nu_x} \quad (22a)$$

$\rightarrow$ CIL = F$_0$($F_{\alpha/2,\nu_y,\nu_x} - F_{1-\alpha/2,\nu_y,\nu_x}$), where $\nu_x = n_x - 1$ and $\nu_y = n_y - 1$. (22b)

Then H$_0$: $\sigma_x^2 = \sigma_y^2$ or H$_0$: $\sigma_x^2/\sigma_y^2 = 1$ must be rejected at the α×100% level of significance if the CI in Eq. (22a) excludes one; otherwise, H$_0$ must not be rejected at the α×100% level. Thus, based on the Standard procedure H$_0$: $\sigma_x^2 = \sigma_y^2$ (or $\frac{\sigma_x^2}{\sigma_y^2} = 1$) must be rejected at the α level iff either F$_0 = \frac{S_x^2}{S_y^2} < F_{1-\alpha/2,\nu_x,\nu_y}$ or F$_0 = \frac{S_x^2}{S_y^2} > F_{\alpha/2,\nu_x,\nu_y}$ (22c)

The Pr of type I error for the exact procedure (i.e., using the Standard method from the null SMD of the ratio S_x^2/S_y^2 which is F_{ν_x,ν_y}) is α. This implies that H$_0$: $\sigma_x^2 = \sigma_y^2$ will be rejected at the α-level iff F$_0 = S_x^2/S_y^2 < F_{1-\alpha/2,\nu_x,\nu_y}$, or F$_0 = S_x^2/S_y^2 > F_{\alpha/2,\nu_x,\nu_y}$. Therefore, the type I error Pr for the two disjoint CIs (α') is given by

α'(two disjoint CIs) = Pr[$U(\sigma_X^2) < L(\sigma_Y^2)$] + Pr[$L(\sigma_X^2) > U(\sigma_Y^2)$]

$$= \Pr[\frac{(n_x-1)S_x^2}{\chi^2_{1-\alpha/2,\nu_x}} < \frac{(n_y-1)S_y^2}{\chi^2_{\alpha/2,\nu_y}}] + \Pr[\frac{(n_x-1)S_x^2}{\chi^2_{\alpha/2,\nu_x}} > \frac{(n_y-1)S_y^2}{\chi^2_{1-\alpha/2,\nu_y}}]$$

$$= \Pr[\frac{S_x^2}{S_y^2} < \frac{\nu_y}{\nu_x} \times \frac{\chi^2_{1-\alpha/2,\nu_x}}{\chi^2_{\alpha/2,\nu_y}}] + \Pr[\frac{S_x^2}{S_y^2} > \frac{\nu_y}{\nu_x} \times \frac{\chi^2_{\alpha/2,\nu_x}}{\chi^2_{1-\alpha/2,\nu_y}}]$$

$$= \Pr(\frac{S_x^2}{S_y^2} < \frac{\nu_y}{\nu_x} \times C_{1-\alpha/2,\nu_x,\nu_y}) + \Pr(\frac{S_x^2}{S_y^2} > \frac{\nu_y}{\nu_x} \times C_{\alpha/2,\nu_x,\nu_y})$$

$$= \Pr(F_{\nu_x,\nu_y} < \frac{\nu_y}{\nu_x} \times C_{1-\alpha/2,\nu_x,\nu_y}) + \Pr(F_{\nu_x,\nu_y} > \frac{\nu_y}{\nu_x} \times C_{\alpha/2,\nu_x,\nu_y}) \quad (23)$$

Table 10 gives the values of α and α' (where α′ represents type I error Pr from the Overlap procedure) for various values of ν_x and ν_y, verifying the same conclusion as before: the Overlap method always leads to a smaller type I error Pr than that of the null sampling distribution of S_x^2 / S_y^2, which is the Fisher's F. Moreover, we have verified that α' value depends on the sizes of ν_x *and* ν_y and not much on their ratio ν_x / ν_y. Eq. (23) can easily verify that at α = 0.01, as ν_x *and* ν_y increase, the Overlap type I error Pr, α', decreases toward 0.000269717, while at α = 0.05 the value of α' decreases (from 0.017800531 at $\nu_x = \nu_y = 1$) toward 0.0055746, similar to the overlapping of CIs for population means. For the special case that $n_x = n_y = n$, the rejection of H_0 from the Overlap method given by Eq. (21a) and Eq. (21b) is reduced to

→ Reject H_0 if either $F_0 = \frac{S_x^2}{S_y^2} < \frac{\chi^2_{1-\alpha/2,n-1}}{\chi^2_{\alpha/2,n-1}} = C_{1-\alpha/2,n-1}$

$$\text{or } F_0 = \frac{S_x^2}{S_y^2} > \frac{\chi^2_{\alpha/2,n-1}}{\chi^2_{1-\alpha/2,n-1}} = C_{\alpha/2,n-1} \quad (24)$$

Table 10. The Values of α and α' for Various Values of ν_x and ν_y

α = 0.05				α = 0.01			
ν_x	ν_y	ν_y/ν_x	α'	ν_x	ν_y	ν_y/ν_x	α'
10	20	2	0.007839	10	20	2	0.000624
10	40	4	0.009969	10	40	4	0.000912
10	60	6	0.011917	10	60	6	0.001182
10	80	8	0.013579	10	80	8	0.001420
30	20	0.666667	0.006640	30	20	0.666667	0.000425
30	40	1.333333	0.006262	30	40	1.333333	0.000368
30	60	2	0.006818	30	60	2	0.000424
30	80	2.666667	0.007546	30	80	2.666667	0.000502
50	20	0.4	0.007611	50	20	0.4	0.000534
50	50	1	0.005954	50	50	1	0.000326
50	80	1.6	0.006228	50	80	1.6	0.000349
50	100	2	0.006611	50	100	2	0.000386
100	60	0.6	0.006233	100	60	0.6	0.000346
100	80	0.8	0.005865	100	80	0.8	0.000308
500	500	1	0.005612	100	100	1	0.000297
1000	1000	1	0.005594	100	120	1.2	0.000300
20	30	1.5	0.006640	20	30	1.5	0.000425
40	60	1.5	0.006230	40	60	1.5	0.000355
80	120	1.5	0.006025	80	120	1.5	0.000322
100	150	1.5	0.005984	100	150	1.5	0.000315
20	40	2	0.007077	20	40	2	0.000473
40	80	2	0.006689	40	80	2	0.000400
80	160	2	0.006493	80	160	2	0.000365
1000	2000	2	0.006313	1000	2000	2	0.000333
40	100	2.5	0.007232	40	100	2.5	0.000456
60	150	2.5	0.007105	60	150	2.5	0.000430
100	250	2.5	0.007003	100	250	2.5	0.000410
1000	2500	2.5	0.006864	1000	2500	2.5	0.000383
30	150	5	0.010087	30	150	5	0.000798
40	200	5	0.009970	40	200	5	0.000765
50	250	5	0.009900	50	250	5	0.000746
100	500	5	0.009759	100	500	5	0.000706
1000	5000	5	0.009630	1000	5000	5	0.000671

From Eqs. (22a & b), the rejection of H_0 using Fisher's F will be simplified as follows:

$F_{n-1,n-1} = \dfrac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2}$, thus, (1 – α)100% CIs for $\dfrac{\sigma_x^2}{\sigma_y^2}$

$$\rightarrow \frac{S_x^2}{S_y^2} F_{1-\alpha/2,n-1,n-1} \le \frac{\sigma_x^2}{\sigma_y^2} \le \frac{S_x^2}{S_y^2} F_{\alpha/2,n-1,n-1} \qquad (25a)$$

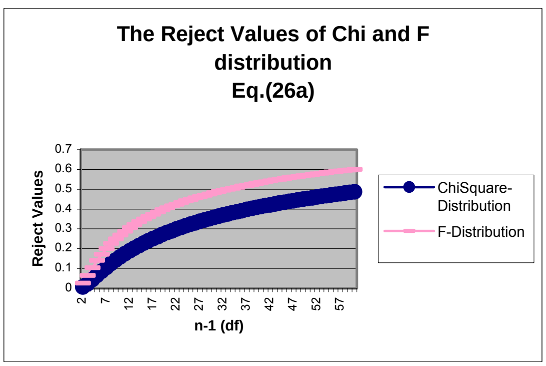
From Eq.(25a), the length of the exact (1-α)% CI is given by $F_0(F_{\alpha/2,n-1,n-1} - F_{1-\alpha,n-1,n-1})$. Thus, H_0 should be rejected if

$$F_0 = \frac{S_x^2}{S_y^2} < F_{1-\alpha/2,n-1,n-1} \quad \text{or } F_0 = \frac{S_x^2}{S_y^2} > F_{\alpha/2,n-1,n-1} . \tag{25b}$$

Comparing Eq.(24) with Eq.(25b), it can be verified (See Figure 8) that at the same

α-level, $\frac{\chi^2_{1-\alpha/2,n-1}}{\chi^2_{\alpha/2,n-1}} = C_{1-\alpha/2,n-1} < F_{1-\alpha/2,n-1,n-1}$ (26a)

and $\frac{\chi^2_{\alpha/2,n-1}}{\chi^2_{1-\alpha/2,n-1}} = C_{\alpha/2,n-1} > F_{\alpha/2,n-1,n-1}$ for all n. (26b)



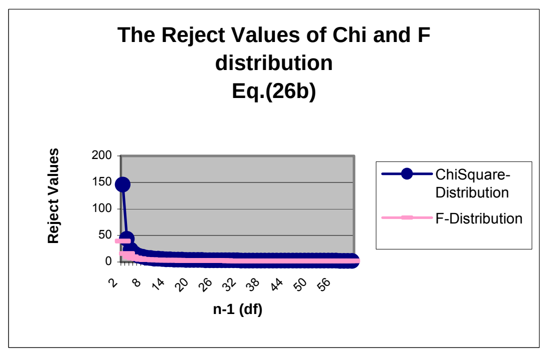


Figure 8

Furthermore, for the balanced $n_x = n_y = n$ case, if the type I error Pr for the Standard Method (Fisher's F distribution) is α, the type I error Pr from the two disjoint CIs (α'), Eq.(23), is reduced to →

$$\alpha' = \Pr[\frac{(n_x-1)S_x^2}{\chi^2_{1-\alpha/2,n_x-1}} < \frac{(n_y-1)S_y^2}{\chi^2_{\alpha/2,n_y-1}}] + \Pr[\frac{(n_x-1)S_x^2}{\chi^2_{\alpha/2,n_x-1}} > \frac{(n_y-1)S_y^2}{\chi^2_{1-\alpha/2,n_y-1}}]$$

$$= \Pr(\frac{S_x^2}{S_y^2} < C_{1-\alpha/2,n-1}) + \Pr(\frac{S_x^2}{S_y^2} > C_{\alpha/2,n-1})$$

$$= \Pr(F_{n-1,n-1} < C_{1-\alpha/2,n-1}) + \Pr(F_{n-1,n-1} > C_{\alpha/2,n-1})$$

$$= 2 \times \Pr(F_{n-1,n-1} > C_{\alpha/2,n-1}), \text{ where } C_{\alpha/2,n-1} = \frac{\chi^2_{\alpha/2,n-1}}{\chi^2_{1-\alpha/2,n-1}} \quad (27)$$

Table 11 shows that α' is much smaller than α for the special case that $n_x = n_y = n$. As in the case of testing H_0: $\mu_x = \mu_y$ at $\alpha = 0.05$, the value of Overlap type I error Pr seems to slowly approach 0.0055746 as $n \to \infty$ and at $\alpha = 0.01$, α' approaches 0.0002697.

Table 11. The Impact of Overlap on Type I Error Pr for the Equal-Sample Size Case When Testing the Ratio σ_x^2/σ_y^2 Against 1

n-1	α	α'	n-1	α	α'
10	0.01	0.000585	10	0.05	0.007468
20	0.01	0.000418	20	0.05	0.006525
50	0.01	0.000326	50	0.05	0.005954
80	0.01	0.000304	80	0.05	0.005812
100	0.01	0.000297	100	0.05	0.005764
130	0.01	0.000291	130	0.05	0.005720
150	0.01	0.000288	150	0.05	0.005701
200	0.01	0.000283	200	0.05	0.005669
250	0.01	0.000281	250	0.05	0.005650
500	0.01	0.000275	500	0.05	0.005612
1000	0.01	0.000272	1000	0.05	0.005594
2000	0.01	0.000271	2000	0.05	0.005584
3000	0.01	0.000271	3000	0.05	0.005581

As before, let O represent the length of overlap between the CIs for σ_x^2 and σ_y^2. Thus, O is larger than 0 only if $U(\sigma_x^2) > U(\sigma_y^2) > L(\sigma_x^2)$ or $U(\sigma_y^2) > U(\sigma_x^2) > L(\sigma_y^2)$. Because both conditions lead to the same result only the case $U(\sigma_x^2) > U(\sigma_y^2) > L(\sigma_x^2)$ is considered, and without loss of generality the X-sample is the one with larger variance so that $S_x^2 / S_y^2 \geq 1$.

→ Because of symmetry, $O = U(\sigma_y^2) - L(\sigma_x^2) = \dfrac{\nu_y S_y^2}{\chi^2_{1-\alpha/2,\nu_y}} - \dfrac{\nu_x S_x^2}{\chi^2_{\alpha/2,\nu_x}}$ (28a)

Let O_r be the maximum value of O at which H_0 is barely rejected at the α level. From Eq. (22c), H_0 must be rejected iff $F_0 = \dfrac{S_x^2}{S_y^2} > F_{\alpha/2,\nu_x,\nu_y}$. Therefore, the borderline value of O will occur when $S_x^2 = S_y^2 \times F_{\alpha/2,\nu_x,\nu_y}$. Inserting this into Eq. (28a) will result in:

$$O_r \le \frac{\nu_y S_y^2}{\chi^2_{1-\alpha/2,\nu_y}} - \frac{\nu_x S_y^2 \times F_{\alpha/2,\nu_x,\nu_y}}{\chi^2_{\alpha/2,\nu_x}} = S_y^2 \times \left(\frac{\nu_y}{\chi^2_{1-\alpha/2,\nu_y}} - \frac{\nu_x F_{\alpha/2,\nu_x,\nu_y}}{\chi^2_{\alpha/2,\nu_x}}\right) \quad (28b)$$

The span of the two individual CIs is $U(\sigma_x^2) - L(\sigma_y^2) = \dfrac{(n_x-1)S_x^2}{\chi^2_{1-\alpha/2,n_x-1}} - \dfrac{(n_y-1)S_y^2}{\chi^2_{\alpha/2,n_y-1}}$

$$= \frac{\nu_x S_y^2 \times F_{\alpha/2,\nu_x,\nu_y}}{\chi^2_{1-\alpha/2,\nu_x}} - \frac{\nu_y S_y^2}{\chi^2_{\alpha/2,\nu_y}} = S_y^2 \times \left(\frac{\nu_x F_{\alpha/2,\nu_x,\nu_y}}{\chi^2_{1-\alpha/2,\nu_x}} - \frac{\nu_y}{\chi^2_{\alpha/2,\nu_y}}\right) \quad (28c)$$

Thus, the percent overlap at the critical limits is

$$\omega_r = \frac{U(\sigma_Y^2) - L(\sigma_X^2)}{U(\sigma_X^2) - L(\sigma_Y^2)} = \left[\frac{\dfrac{\nu_y}{\chi^2_{1-\alpha/2,\nu_y}} - \dfrac{\nu_x F_{\alpha/2,\nu_x,\nu_y}}{\chi^2_{\alpha/2,\nu_x}}}{\dfrac{\nu_x F_{\alpha/2,\nu_x,\nu_y}}{\chi^2_{1-\alpha/2,\nu_x}} - \dfrac{\nu_y}{\chi^2_{\alpha/2,\nu_y}}}\right] \times 100\%$$

$$= \left(\frac{\nu_y \times C_{\alpha/2,\nu_x,\nu_y} \times \chi^2_{\alpha/2,\nu_y} - \nu_x \times F_{\alpha/2,\nu_x,\nu_y} \times \chi^2_{\alpha/2,\nu_y}}{\nu_x \times C_{\alpha/2,\nu_x,\nu_x} \times F_{\alpha/2,\nu_x,\nu_y} \times \chi^2_{\alpha/2,\nu_y} - \nu_y \times \chi^2_{\alpha/2,\nu_x}}\right) \times 100\% \quad (28d)$$

Table 12 shows that as ν_x and ν_y increase, the percentage of the overlap approaches 17.1573% (although not monotonically). Further, once the % overlap exceeds Eq. (28d), then H_0 must not be rejected at the $\alpha \times 100\%$ level. Further, it is the size of ν_x and ν_y that

Table 12. The % Overlap for the Different Combinations of Degree of Freedom at α = 0.05.

ν_x	ν_y	ν_y/ν_x	Overlap (%)	ν_x	ν_y	ν_y/ν_x	Overlap (%)
10	5	0.5	13.92184	10	12	1.2	10.91515
10	10	1	11.54543	20	24	1.2	13.50590
10	15	1.5	10.15131	40	48	1.2	15.04864
10	20	2	9.18956	60	72	1.2	15.62065
10	25	2.5	8.46961	80	96	1.2	15.92389
20	10	0.5	16.24648	100	120	1.2	16.11365
20	20	1	14.11596	150	180	1.2	16.37981
20	30	1.5	12.73993	300	360	1.2	16.67183
20	40	2	11.73402	500	600	1.2	16.80378
20	50	2.5	10.95112	700	840	1.2	16.86607
40	20	0.5	17.20430	800	960	1.2	16.88670
40	40	1	15.57376	900	1080	1.2	16.90327
40	60	1.5	14.35904	1000	1200	1.2	16.91691
40	80	2	13.40830	2000	2400	1.2	16.98483
40	100	2.5	12.63712	3000	3600	1.2	17.01169
60	30	0.5	17.40395	10	15	1.5	10.15131
60	60	1	16.08722	20	30	1.5	12.73993
60	90	1.5	14.98840	40	60	1.5	14.35904
60	120	2	14.08884	60	90	1.5	14.98840
60	150	2.5	13.34073	80	120	1.5	15.33298
80	40	0.5	17.45225	100	150	1.5	15.55403
80	80	1	16.34928	150	225	1.5	15.87382
80	120	1.5	15.33298	300	450	1.5	16.24474
80	160	2	14.47217	500	750	1.5	16.42384
80	200	2.5	13.74343	700	1050	1.5	16.51241
100	50	0.5	17.45418	800	1200	1.5	17.76296
100	100	1	16.50824	900	1350	1.5	16.56697
100	150	1.5	15.55403	1000	1500	1.5	16.58737
100	200	2	14.72328	2000	3000	1.5	16.69266
100	250	2.5	14.01021	3000	4500	1.5	16.73654

determines the % overlap and not the ratio ν_y/ν_x. For the case that $n_x = n_y = n$, the percent overlap in Eq.(28d) reduces to

$$\rightarrow \omega = \frac{U(\sigma_Y^2)-L(\sigma_X^2)}{U(\sigma_X^2)-L(\sigma_Y^2)} = [\frac{(n-1)S_y^2\times[\frac{1}{\chi^2_{1-\alpha/2,n-1}}-\frac{F_{\alpha/2,n-1,n-1}}{\chi^2_{\alpha/2,n-1}}]}{(n-1)S_y^2\times[\frac{F_{\alpha/2,n-1,n-1}}{\chi^2_{1-\alpha/2,n-1}}-\frac{1}{\chi^2_{\alpha/2,n-1}}]}]100\%$$

$$= [\frac{\frac{1}{\chi^2_{1-\alpha/2,n-1}} - \frac{F_{\alpha/2,n-1,n-1}}{\chi^2_{\alpha/2,n-1}}}{\frac{F_{\alpha/2,n-1,n-1}}{\chi^2_{1-\alpha/2,n-1}} - \frac{1}{\chi^2_{\alpha/2,n-1}}}] \times 100\% = [\frac{C_{\alpha/2,n-1} - F_{\alpha/2,n-1,n-1}}{C_{\alpha/2,n-1} \times F_{\alpha/2,n-1,n-1} - 1}] \times 100\% \quad (29)$$

Eq.(29) shows that the rejection percent overlap between the two CIs for the ratio of variances will increase as n increases. Further, ω_r in Eq. (12b) is also a monotonically increasing function of α. For example, at α = 0.05, n = 2, ω_r = 0.1348%; at n = 3, ω_r = 1.8781%; at n = 5, ω_r = 6.0921%; and at n = 20 and α = 0.05, ω_r = 13.9695%, while at n = 20 and α = 0.01, ω_r = 12.0224%. Matlab shows that at ν = n − 1 = 7,819,285 df [note that Matlab 7.6(R2008a) loses accuracy in inverting F at the 7th decimal place beyond 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 (which was 17.157287525%). Further, for very small sample sizes within the interval [2, 4], the Variance-Overlap method is almost an α-level test, like the case of CIs for normal population means when $\sigma_x = \sigma_y$ and K is far away from 1. See the illustration in Table 13.

Table 13. The % Overlap for the Case of α = 0.05 and $n_x = x_y = n$

n-1	Numerator of Eq.(31)	Denominator of Eq.(31)	Overlap (%)	n-1	Numerator of Eq.(31)	Denominator of Eq.(31)	Overlap (%)
10	0.12652	1.09587	11.54543	200	0.00067	0.00397	16.83011
20	0.03214	0.22770	14.11596	400	0.00023	0.00133	16.99303
30	0.01541	0.10223	15.07427	600	0.00012	0.00070	17.04763
40	0.00933	0.05990	15.57376	800	0.00008	0.00045	17.07499
50	0.00637	0.04014	15.88014	1000	0.00005	0.00032	17.09142
60	0.00469	0.02917	16.08722	1200	0.00004	0.00024	17.10239
70	0.00363	0.02237	16.23652	1300	0.00004	0.00021	17.10661
80	0.00291	0.01783	16.34928	1400	0.00003	0.00019	17.11022
90	0.00240	0.01462	16.43743	1500	0.00003	0.00017	17.11336
100	0.00202	0.01227	16.50824	2000	0.00002	0.00011	17.12433

Now, what should the individual confidence level $1 - \gamma$ be so that the two independent CIs lead to the exact α-level test on $H_0: \sigma_x^2 = \sigma_y^2$. The expressions for the two $1-\gamma$ independent CIs are given by $\frac{(n_x-1)S_x^2}{\chi^2_{\gamma/2,\nu_x}} \le \sigma_x^2 \le \frac{(n_x-1)S_x^2}{\chi^2_{1-\gamma/2,\nu_x}}$, (30a)

$$\text{and} \quad \frac{(n_y-1)S_y^2}{\chi^2_{\gamma/2,\nu_y}} \le \sigma_y^2 \le \frac{(n_y-1)S_y^2}{\chi^2_{1-\gamma/2,\nu_y}} \tag{30b}$$

From Eq.(30a) and Eq.(30b), the overlap amount of two individual CIs at confidence level $(1-\gamma)$ is $U'(\sigma_y^2) - L'(\sigma_x^2)$. Therefore, we deduce from (30a &b) that

$$U'(\sigma_y^2) - L'(\sigma_x^2) = \frac{\nu_y S_y^2}{\chi^2_{1-\gamma/2,\nu_y}} - \frac{\nu_x S_x^2}{\chi^2_{\gamma/2,\nu_x}} \tag{30c}$$

Because $H_0: \sigma_x^2 = \sigma_y^2$ must be rejected at the $\alpha\times100\%$-level as soon as Eq.(30c) becomes zero or smaller, we thus impose the rejection criterion $S_x^2 / S_y^2 \ge F_{\alpha/2}$ (where for the sake of convenience $F_{\alpha/2} = F_{\alpha/2,\nu_x,\nu_y}$) into Eq. (30c). In short, we are rejecting $H_0: \sigma_x^2 = \sigma_y^2$ as soon as the two independent CIs in (30a) and (30b) become disjoint. This leads to rejecting $H_0: \sigma_x^2 = \sigma_y^2$ iff $U'(\sigma_y^2) - L'(\sigma_x^2) = \frac{\nu_y S_y^2}{\chi^2_{1-\gamma/2,\nu_y}} - \frac{\nu_x F_{\alpha/2} S_y^2}{\chi^2_{\gamma/2,\nu_x}} \le 0.$ (31a)

At the borderline value, we set the overlap amount at LOS γ in inequality (31a) to 0 in order to solve for $\gamma \rightarrow \frac{\nu_y S_y^2}{\chi^2_{1-\gamma/2,\nu_y}} - \frac{\nu_x F_{\alpha/2} S_y^2}{\chi^2_{\gamma/2,\nu_x}} = 0 \rightarrow \frac{\nu_y}{\chi^2_{1-\gamma/2,\nu_y}} - \frac{\nu_x F_{\alpha/2}}{\chi^2_{\gamma/2,\nu_x}} = 0$

$$\rightarrow \frac{\nu_x F_{\alpha/2}}{\chi^2_{\gamma/2,\nu_x}} = \frac{\nu_y}{\chi^2_{1-\gamma/2,\nu_y}} \rightarrow \frac{\nu_x F_{\alpha/2}}{\nu_y} = \frac{\chi^2_{\gamma/2,\nu_x}}{\chi^2_{1-\gamma/2,\nu_y}} \rightarrow \frac{\nu_x F_{\alpha/2}}{\nu_y} = C_{\gamma/2,\nu_x,\nu_y} \tag{31b}$$

where $F_{\alpha/2} = F_{\alpha/2,\nu_x,\nu_y}$ and $C_{\gamma/2,\nu_x,\nu_y} = \chi^2_{\gamma/2,\nu_x} / \chi^2_{1-\gamma/2,\nu_y}$. Eq. (31b) clearly shows that the value of γ depends 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. (31b) reduces to $2F_{0.025,20,10} = C_{\gamma/2,20,10} = \chi^2_{\gamma/2,20} / \chi^2_{1-\gamma/2,10} \rightarrow 2\times3.4185 = \chi^2_{\gamma/2,20} / \chi^2_{1-\gamma/2,10} \rightarrow 6.8371 = \chi^2_{\gamma/2,20} / \chi^2_{1-\gamma/2,10}$. Through trial & error the solution to this last inequality is $\gamma/2 = 0.0712$ ($\gamma = 0.1424$). In turns out that as long as $\nu_x = 2\nu_y$, the required confidence level for the two independent CI on σ_x^2 and σ_y^2 must be set approximately equal to $1-2\times0.0712 = 85.76\%$. Further, if $\nu_y = 2\nu_x$ the required confidence level for the two independent CI on σ_x^2 and σ_y^2 must be set approximately equal to $1-2\times0.083 = 83.40\%$. In the case of balanced design (i.e., when $\nu_x = \nu_y$) Eq. (31b) reduces to

$$F_{\alpha/2,n-1,n-1} = C_{\gamma/2,n-1} \tag{31c}$$

It can be verified that the approximate solution to Eq. (31c) when $\alpha = 0.05$, n-1=10, through trial & error, is $\gamma/2 = 0.079$. Therefore, the individual CIs have to be set at 84.20%. For the moderate sample $10 \le n \le 30$, the approximate solution is 0.08. We used Matlab to determine that un the limit (as $n \rightarrow 7{,}819{,}286$ at 7 decimal accuracy), $\gamma/2 \rightarrow 0.08288800$. because MS Excel 2003 cannot invert χ^2_ν for df beyond $\nu = 1119$. Table14 shows the value of γ to make the two sides of Eq.(31b) equal for different ν_x and ν_y combinations. Table 15 shows the cases when ν_y is kept fixed at 20 but the ratio ν_x / ν_y changes from 0.5 to 50 causing γ to become smaller and smaller.

Table 14. The Overlap Significance Level, γ, that Yields the Same 5%-Level Test or 1%-Level Test by the Standard Method

$\alpha = 0.05$						$\alpha = 0.01$					
ν_x	ν_y	$\frac{\nu_x}{\nu_y}$	$\frac{\nu_x F_{\alpha/2}}{\nu_y}$	γ	$C_{\gamma/2,\nu_x,\nu_y}$	ν_x	ν_y	$\frac{\nu_x}{\nu_y}$	$\frac{\nu_x F_{\alpha/2}}{\nu_y}$	γ	$C_{\gamma/2,\nu_x,\nu_y}$
10	20	0.5	1.38684	0.16658	1.38684	10	20	0.5	1.92350	0.06875	1.92350
20	40	0.5	1.03386	0.16560	1.03386	20	40	0.5	1.29921	0.06878	1.29921
30	60	0.5	0.90760	0.16489	0.90760	30	60	0.5	1.09372	0.06847	1.09372
40	80	0.5	0.83952	0.16440	0.83952	40	80	0.5	0.98697	0.06819	0.98697
50	100	0.5	0.79585	0.16402	0.79585	50	100	0.5	0.92002	0.06796	0.92002
60	120	0.5	0.76497	0.16373	0.76497	60	120	0.5	0.87343	0.06776	0.87343
80	160	0.5	0.72348	0.16329	0.72348	80	160	0.5	0.81179	0.06745	0.81179
100	200	0.5	0.69635	0.16297	0.69635	100	200	0.5	0.77209	0.06722	0.77209
200	400	0.5	0.63291	0.16211	0.63291	200	400	0.5	0.68121	0.06658	0.68121
500	1000	0.5	0.58092	0.16128	0.58092	500	1000	0.5	0.60881	0.06592	0.60881
1000	2000	0.5	0.55615	0.16117	0.55611	1000	2000	0.5	0.57498	0.06552	0.57499
2000	4000	0.5	0.53918	0.16075	0.53915	2000	4000	0.5	0.55207	0.06551	0.55203
10	10	1	3.71679	0.15810	3.71679	10	10	1	5.84668	0.05981	5.84668
20	20	1	2.46448	0.16189	2.46448	20	20	1	3.31779	0.06400	3.31779
30	30	1	2.07394	0.16317	2.07394	30	30	1	2.62778	0.06548	2.62778
40	40	1	1.87520	0.16382	1.87520	40	40	1	2.29584	0.06623	2.29584
50	50	1	1.75195	0.16421	1.75195	50	50	1	2.09671	0.06669	2.09671
60	60	1	1.66679	0.16447	1.66679	60	60	1	1.96217	0.06699	1.96217
80	80	1	1.55488	0.16480	1.55488	80	80	1	1.78924	0.06738	1.78924
100	100	1	1.48325	0.16499	1.48325	100	100	1	1.68089	0.06761	1.68089
200	200	1	1.32045	0.16538	1.32045	200	200	1	1.44159	0.06808	1.44159
500	500	1	1.19185	0.16562	1.19185	500	500	1	1.25956	0.06836	1.25956
1000	1000	1	1.13205	0.16570	1.13205	1000	1000	1	1.17708	0.06846	1.17708
2000	2000	1	1.09164	0.16568	1.09164	2000	2000	1	1.12214	0.06854	1.12213
10	5	2	13.23831	0.13278	13.23831	10	5	2	27.23636	0.04228	27.23640
20	10	2	6.83709	0.14239	6.83709	20	10	2	10.54803	0.04970	10.54800
30	15	2	5.28747	0.14629	5.28747	30	15	2	7.37349	0.05296	7.37349
40	20	2	4.57464	0.14848	4.57464	40	20	2	6.04306	0.05483	6.04306
50	25	2	4.15744	0.14990	4.15744	50	25	2	5.30448	0.05607	5.30448
60	30	2	3.88002	0.15092	3.88002	60	30	2	4.83030	0.05696	4.83030
80	40	2	3.52875	0.15230	3.52875	80	40	2	4.24979	0.05817	4.24979
100	50	2	3.31170	0.15320	3.31170	100	50	2	3.90249	0.05896	3.90249
200	100	2	2.84057	0.15532	2.84057	200	100	2	3.17944	0.06083	3.17944
500	250	2	2.48968	0.15705	2.48968	500	250	2	2.66931	0.06234	2.66931
1000	500	2	2.33277	0.15786	2.33277	1000	500	2	2.44932	0.06305	2.44932
2000	1000	2	2.22893	0.15825	2.22900	2000	1000	2	2.30657	0.06352	2.30658

Table 15. The Overlap Significance Level, γ, That Yields the Same 5%-Level Test or 1%-Level Test by the Standard Method at Fixed ν_y and Changing ν_x

$\alpha = 0.05$						$\alpha = 0.01$					
ν_x	ν_y	$\frac{\nu_x}{\nu_y}$	$\frac{\nu_x F_{\alpha/2}}{\nu_y}$	γ	$C_{\gamma/2,\nu_x,\nu_y}$	ν_x	ν_y	$\frac{\nu_x}{\nu_y}$	$\frac{\nu_x F_{\alpha/2}}{\nu_y}$	γ	$C_{\gamma/2,\nu_x,\nu_y}$
10	20	0.5	1.38684	0.16658	1.38684	10	20	0.5	1.92350	0.06875	1.92350
12	20	0.6	1.60550	0.16628	1.60550	12	20	0.6	2.20674	0.06803	2.20674
16	20	0.8	2.03723	0.16445	2.03723	16	20	0.8	2.76540	0.06611	2.76540
20	20	1	2.46448	0.16189	2.46448	20	20	1	3.31779	0.06400	3.31779
24	20	1.2	2.88907	0.15910	2.88907	24	20	1.2	3.86643	0.06192	3.86643
28	20	1.4	3.31194	0.15629	3.31194	28	20	1.4	4.41266	0.05995	4.41266
32	20	1.6	3.73362	0.15356	3.73362	32	20	1.6	4.95722	0.05811	4.95722
36	20	1.8	4.15445	0.15095	4.15445	36	20	1.8	5.50059	0.05641	5.50059
40	20	2	4.57464	0.14848	4.57464	40	20	2	6.04306	0.05483	6.04306
50	20	2.5	5.62323	0.14287	5.62323	50	20	2.5	7.39659	0.05139	7.39659
60	20	3	6.67008	0.13802	6.67008	60	20	3	8.74765	0.04853	8.74765
70	20	3.5	7.71585	0.13380	7.71585	70	20	3.5	10.09722	0.04612	10.09722
80	20	4	8.76092	0.13010	8.76092	80	20	4	11.44579	0.04407	11.44579
90	20	4.5	9.80550	0.12683	9.80550	90	20	4.5	12.79368	0.04229	12.79368
100	20	5	10.84972	0.12392	10.84972	100	20	5	14.14107	0.04074	14.14107
110	20	5.5	11.89369	0.12130	11.89369	110	20	5.5	15.48809	0.03936	15.48809
120	20	6	12.93745	0.11894	12.93745	120	20	6	16.83483	0.03814	16.83482
130	20	6.5	13.98105	0.11679	13.98105	130	20	6.5	18.18134	0.03705	18.18134
140	20	7	15.02453	0.11482	15.02453	140	20	7	19.52768	0.03606	19.52768
1000	20	50	104.70358	0.07610	104.70360	1000	20	50	135.22825	0.01898	135.22158

Next, the type Π error Pr for both the F-distribution and separate CIs cases are discussed. Comparing equations (21 a & b) with Eq. (22c), because $(\nu_y/\nu_x)\times C_{\alpha/2,\nu_x,\nu_y} > F_{\alpha/2,\nu_x,\nu_y}$, and $(\nu_y/\nu_x)\times C_{1-\alpha/2,\nu_x,\nu_y} < F_{1-\alpha/2,\nu_x,\nu_y}$ (see the illustration in the Figure 7), it follows that the disjoint CIs provide more stringent requirement for rejecting H_0. Thus, the rejecting rule from two disjoint CIs will always lead to a larger Type Π error Pr (or much less statistical power) as illustrated below. By definition, β =Pr(Type Π error) = Pr(not rejecting $H_0|H_0$ is false). Since H_0 is assumed false, it follows that $\sigma_x^2 \neq \sigma_y^2$. Let λ $= \sigma_x/\sigma_y \rightarrow \sigma_x^2 = \lambda^2\sigma_y^2$. Thus, $\beta(\lambda) = \Pr\,(F_{1-\alpha/2,\nu_x,\nu_y} \le F_0 \le F_{\alpha/2,\nu_x,\nu_y} \mid \lambda = \sigma_x/\sigma_y)$

$$\rightarrow \beta(\lambda) = cdfF_{\nu_x,\nu_y}(F_{\alpha/2,\nu_x,\nu_y}/\lambda^2) - cdfF_{\nu_x,\nu_y}(F_{1-\alpha/2,\nu_x,\nu_y}/\lambda^2) \quad (32a)$$

And the Type Π error Pr of the two independent CIs is given by

$$\beta'(\lambda) = \Pr\{[L(\sigma_x^2) \le U(\sigma_y^2)] \cap [L(\sigma_y^2) \le U(\sigma_x^2)] | \lambda = \frac{\sigma_x}{\sigma_y}\}$$

$$= \Pr\{[\frac{\nu_x S_x^2}{\chi^2_{\alpha/2,\nu_x}} \le \frac{\nu_y S_y^2}{\chi^2_{1-\alpha/2,\nu_y}}] \cap [\frac{\nu_y S_y^2}{\chi^2_{\alpha/2,\nu_y}} \le \frac{\nu_x S_x^2}{\chi^2_{1-\alpha/2,\nu_x}}] | \lambda = \frac{\sigma_x}{\sigma_y}\}$$

$$= \Pr\{[\frac{\nu_y}{\nu_x} \times \frac{\chi^2_{1-\alpha/2,\nu_x}}{\chi^2_{\alpha/2,\nu_y}} \le \frac{S_x^2}{S_y^2} \le \frac{\nu_y}{\nu_x} \times \frac{\chi^2_{\alpha/2,\nu_x}}{\chi^2_{1-\alpha/2,\nu_y}}] | \lambda = \frac{\sigma_x}{\sigma_y}\}$$

$$= \Pr[\frac{\nu_y}{\nu_x} \times C_{1-\alpha/2,\nu_x,\nu_y} \le \frac{S_x^2}{S_y^2} \le \frac{\nu_y}{\nu_x} \times C_{\alpha/2,\nu_x,\nu_y} | \lambda = \frac{\sigma_x}{\sigma_y}]$$

$$= \Pr[\frac{\sigma_y^2}{\sigma_x^2} \times \frac{\nu_y}{\nu_x} \times C_{1-\alpha/2,\nu_x,\nu_y} \le F_{\nu_x,\nu_y} \le \frac{\sigma_y^2}{\sigma_x^2} \times \frac{\nu_y}{\nu_x} \times C_{\alpha/2,\nu_x,\nu_y} | \lambda = \frac{\sigma_x}{\sigma_y}]$$

$$= \Pr[\frac{1}{\lambda^2} \times \frac{\nu_y}{\nu_x} \times C_{1-\alpha/2,\nu_x,\nu_y} \le F_{\nu_x,\nu_y} \le \frac{1}{\lambda^2} \times \frac{\nu_y}{\nu_x} \times C_{\alpha/2,\nu_x,\nu_y}]$$

$$= cdfF_{\nu_x,\nu_y}(\frac{1}{\lambda^2} \times \frac{\nu_y}{\nu_x} \times C_{\alpha/2,\nu_x,\nu_y}) - cdfF_{\nu_x,\nu_y}(\frac{1}{\lambda^2} \times \frac{\nu_y}{\nu_x} \times C_{1-\alpha/2,\nu_x,\nu_y}) \quad (32b)$$

Table 16 illustrates that the Type Π Error Pr from the two overlapping CIs (Eq.(32b)) is larger than the corresponding exact value from the F distribution (Eq.(32a)). For the case $n_x = n_y = n$, the type II error Pr, β(λ) in Eq.(32a), is reduced to

$$\rightarrow \beta(\lambda) = cdfF(F_{\alpha/2,n-1,n-1}/\lambda^2) - cdfF(F_{1-\alpha/2,n-1,n-1}/\lambda^2) \quad (33a)$$

As n increases, the second term on the RHS of Eq.(33a), $cdfF(F_{1-\alpha/2,n-1,n-1}/\lambda^2)$, becomes smaller. For example at λ =1.6, Table 17 shows that when n ≥ 10, the 2nd term is less

Table 16. The Relative Power of the Overlap to the Standard Method for Different df Combinations at λ=1.2

ν_x	ν_y	β	β'	$(\frac{\beta'-\beta}{1-\beta})100\%$	ν_x	ν_y	β	β'	$(\frac{\beta'-\beta}{1-\beta})100\%$
10	10	0.91766	0.98481	81.55134	100	10	0.91207	0.96043	54.99654
10	20	0.89163	0.98041	81.92535	100	40	0.74692	0.91585	66.74959
10	30	0.87733	0.97509	79.69653	100	70	0.63397	0.87179	64.97326
10	40	0.86842	0.96990	77.12143	100	100	0.55858	0.83125	61.77049
10	50	0.86236	0.96505	74.61099	100	150	0.47954	0.75978	53.84462
20	10	0.91548	0.97991	76.22917	120	20	0.84992	0.94123	60.84118
20	20	0.87759	0.97518	79.72614	120	50	0.69250	0.89059	64.41798
20	30	0.85247	0.96921	79.12977	120	80	0.58356	0.84178	62.00594
20	40	0.83499	0.96306	77.61297	120	110	0.50857	0.79724	58.74004
20	50	0.82223	0.95707	75.85031	120	150	0.44078	0.74501	54.40270
40	20	0.86430	0.96531	74.43748	150	30	0.78694	0.91608	60.61311
40	30	0.82554	0.95716	75.44521	150	70	0.59415	0.83886	60.29563
40	40	0.79525	0.94876	74.97338	150	100	0.49836	0.78518	57.17595
40	50	0.77124	0.94041	73.95230	150	130	0.43027	0.73690	53.81970
40	60	0.75187	0.93229	72.71069	150	160	0.38045	0.69411	50.62713
70	20	0.85581	0.95410	68.16592	200	50	0.66561	0.85840	57.65517
70	40	0.76416	0.93059	70.56976	200	80	0.52932	0.79054	55.49954
70	60	0.69858	0.90724	69.22532	200	120	0.40562	0.70859	50.97174
70	80	0.65095	0.88509	67.07999	200	150	0.34220	0.65484	47.52817
70	100	0.61528	0.86454	64.78992	200	180	0.29514	0.60752	44.31852

Table 17. Type II Error for Different Degrees of Freedom (The Case of $n_x = n_y = n$)

ν	λ	Eq.(33a) 1st term	Eq.(33a) 2nd term	β(λ)	ν	λ	Eq.(33a) 1st term	Eq.(33a) 2nd term	β(λ)
5	1	0.975	0.025	0.95	21	1.6	0.4451093	5.2063E-05	0.4450573
5	1.2	0.9482959	0.0115	0.9367963	22	1.6	0.4243906	4.2048E-05	0.4243486
5	1.4	0.9090074	0.005788	0.9032193	23	1.6	0.4043861	3.4049E-05	0.4043521
5	1.6	0.8578487	0.003139	0.8547093	24	1.6	0.3850938	2.7639E-05	0.3850662
5	1.8	0.7971565	0.001811	0.7953453	25	1.6	0.3665092	2.2487E-05	0.3664867
5	2	0.7301745	0.0011	0.7290745	30	1.6	0.2838905	8.2591E-06	0.2838822
5	2.1	0.6954029	0.000872	0.6945313	35	1.6	0.2172298	3.1597E-06	0.2172266
10	1.2	0.9246265	0.006969	0.9176572	40	1.6	0.1644551	1.2481E-06	0.1644538
10	1.4	0.8361705	0.002117	0.8340535	45	1.6	0.123329	5.0597E-07	0.1233284
10	1.6	0.7168104	0.000705	0.716105	50	1.6	0.0917083	2.0959E-07	0.0917081
15	1.2	0.9024968	0.004697	0.8977997	55	1.6	0.067677	8.8419E-08	0.0676769
15	1.4	0.7639183	0.000929	0.7629895	60	1.6	0.0495987	3.7894E-08	0.0495986
15	1.6	0.5840969	0.000201	0.5838961	65	1.6	0.0361208	1.6465E-08	0.0361207
20	1.1	0.9400273	0.009206	0.9308214	70	1.6	0.0261534	7.2419E-09	0.0261534
20	1.2	0.880935	0.003341	0.877594	75	1.6	0.0188356	3.2198E-09	0.0188356
20	1.3	0.7969205	0.001215	0.7957056	80	1.6	0.0134984	1.4455E-09	0.0134984

than 0.001 so that the 1st term on the RHS of (33a) gives the approximate value of β(λ=1.6) to 3 decimal accuracy once n ≥ 10 and λ ≥1.6, where ν = n − 1 = degrees of freedom. Thus, for the $n_x = n_y = n$ case, if the acceptance criterion is based on overlapping of the two CIs, then Eq.(32b) will be changed to

$$\beta'(\lambda) = \Pr[\frac{\sigma_y^2}{\sigma_x^2} \times \frac{\chi^2_{1-\alpha/2,n-1}}{\chi^2_{\alpha/2,n-1}} \le F_{n_{x-1},n_{y-1}} \le \frac{\sigma_y^2}{\sigma_x^2} \times \frac{\chi^2_{\alpha/2,n-1}}{\chi^2_{1-\alpha/2,n-1}} \mid \lambda = \frac{\sigma_x}{\sigma_y}]$$

$$= \Pr[\frac{1}{\lambda^2} \frac{\chi^2_{1-\alpha/2,n-1}}{\chi^2_{\alpha/2,n-1}} \le F_{n_{x-1},n_{y-1}} \le \frac{1}{\lambda^2} \frac{\chi^2_{\alpha/2,n-1}}{\chi^2_{1-\alpha/2,n-1}}]$$

$$= cdfF(C_{\alpha/2,n-1}/\lambda^2) - cdfF(C_{1-\alpha/2,n-1}/\lambda^2), \text{ where } C_{\alpha/2,n-1} = \frac{\chi^2_{\alpha/2,n-1}}{\chi^2_{1-\alpha/2,n-1}}. \qquad (33b)$$

As the degrees of freedom, ν (= n − 1) or λ increases, the second term of the Eq.(33b), $cdfF(1/\lambda^2 C_{\alpha/2,n-1})$, becomes smaller. For example, if ν is fixed at 10, the cumulative probability of the second term on the RHS of Eq.(33b) will be less than 0.001 when λ = 1.2. Conversely, if λ is kept at 1.2, the 2nd term is less than 0.001 if $\nu \ge 9$ (see the illustration in Table 18). Based on the above discussion, Eqs.(33) can be approximated as

$$\beta(\lambda) = cdfF(F_{\alpha/2,n-1,n-1}/\lambda^2) \quad \text{and} \quad \beta'(\lambda) = cdfF(C_{\alpha/2,n-1}/\lambda^2) \qquad (33c)$$

In Eqs.(26), $\frac{\chi^2_{\alpha/2,n-1}}{\chi^2_{1-\alpha/2,n-1}} = C_{\alpha/2,n-1} > F_{\alpha/2,n-1,n-1}$ for all n and $C_{\alpha/2,n-1}/\lambda^2 > F_{\alpha/2,n-1,n-1}/\lambda^2$

and as a result $\beta'(\lambda) = cdfF(C_{\alpha/2,n-1}/\lambda^2) > cdfF(F_{\alpha/2,n-1,n-1}/\lambda^2) = \beta(\lambda)$, i.e., β' is larger than β for all n. This conclusion is the same as that of testing the difference in population means. Thus, the disjoint confidence intervals always lead to less statistical power ($1-\beta' < 1-\beta$) than the Standard method as illustrated in Table 19.

Table 18. Type II Error Pr for Overlap Method at Different ν and λ Combinations

ν	λ	cdfF[$C_{\alpha/2,n-1}/\lambda^2$]	cdfF[$1/(\lambda^2 C_{\alpha/2,n-1})$]	$\beta'(\lambda)$
5	1	0.995360773	0.004639227	0.9907215
5	1.1	0.992857478	0.002991257	0.9898662
5	1.2	0.989488051	0.001992928	0.9874951
5	1.3	0.985109718	0.001366314	0.9837434
5	1.4	0.979590495	0.000960523	0.97863
5	1.5	0.972814945	0.000690369	0.9721246
5	1.6	0.964688941	0.000506045	0.9641829
10	1	0.996265764	0.003734236	0.9925315
10	1.1	0.992320511	0.001754061	0.9905664
10	1.2	0.985665823	0.000856962	0.9848089
10	1.3	0.975349647	0.000434812	0.9749148
10	1.4	0.960474199	0.000228614	0.9602456
10	1.5	0.940325727	0.000124248	0.9402015
5	1.2	0.989488051	0.001992928	0.9874951
6	1.2	0.988854079	0.00162283	0.9872312
7	1.2	0.988136592	0.001352309	0.9867843
8	1.2	0.98735811	0.001146785	0.9862113
9	1.2	0.986531801	0.000985905	0.9855459
10	1.2	0.985665823	0.000856962	0.9848089

To evaluate the approximate RELEFF of Overlap relative to the Standard method, we make use of Eq. (33c) and determine n_x & n_y by equating $\beta'(\lambda)$ to $\beta(\lambda)$, i.e.,

$$cdf\mathrm{F}\left(\frac{1}{\lambda^2} \times \frac{\nu_y'}{\nu_x'} \times C_{0.025,\nu_x',\nu_y'}\right) \cong cdfF_{\nu_x,\nu_y}\left(F_{0.025,\nu_x,\nu_y} / \lambda^2\right) \qquad (33d)$$

The approximate solution from (33d) is quite accurate for n_x & $n_y \geq 10$ and moderately large values of $\lambda \geq 1.40$. It is impossible to find a general closed-form solution from (33d) for n_x' and n_y', the values of which depend on n_x, n_y and λ. Accordingly, we used MS Excel 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 RELEFF is equal to 45.60%.

Table 19. Comparison of Exact Type II Error Pr with That of the Overlap Method for Different df and λ Combinations

ν	λ	β	β'	$(\frac{\beta'-\beta}{1-\beta})100\%$	ν	λ	β	β'	$(\frac{\beta'-\beta}{1-\beta})100\%$
10	1	0.95000	0.99253	85.06306	10	1.5	0.77819	0.94020	73.04082
20	1	0.95000	0.99348	86.95072	20	1.5	0.57950	0.84388	62.87376
40	1	0.95000	0.99395	87.90145	40	1.5	0.28357	0.59492	43.45754
60	1	0.95000	0.99411	88.21829	60	1.5	0.12404	0.36520	27.53091
80	1	0.95000	0.99419	88.37660	80	1.5	0.05021	0.20201	15.98204
10	1.1	0.94135	0.99057	83.91660	10	1.6	0.71610	0.91441	69.85200
20	1.1	0.93082	0.98916	84.32517	20	1.6	0.46648	0.76677	56.28565
40	1.1	0.90933	0.98446	82.86491	40	1.6	0.16445	0.43432	32.29809
60	1.1	0.88746	0.97886	81.21252	60	1.6	0.04960	0.19999	15.82399
80	1.1	0.86531	0.97259	79.64573	80	1.6	0.01350	0.07962	6.70295
10	1.2	0.91766	0.98481	81.55134	10	1.7	0.65039	0.88283	66.48510
20	1.2	0.87759	0.97518	79.72614	20	1.7	0.36257	0.67786	49.46347
40	1.2	0.79525	0.94876	74.97338	40	1.7	0.08785	0.29250	22.43585
60	1.2	0.71313	0.91510	70.40623	60	1.7	0.01745	0.09560	7.95426
80	1.2	0.63369	0.87553	66.01992	80	1.7	0.00305	0.02598	2.29994
10	1.3	0.88125	0.97491	78.87485	10	1.8	0.58355	0.84576	62.96360
20	1.3	0.79571	0.94822	74.65377	20	1.8	0.27316	0.58303	42.63294
40	1.3	0.62785	0.87213	65.64046	40	1.8	0.04381	0.18296	14.55280
60	1.3	0.47869	0.77595	57.02151	60	1.8	0.00553	0.04061	3.52738
80	1.3	0.35514	0.67031	48.87405	80	1.8	0.00060	0.00723	0.66354
10	1.4	0.83405	0.96025	76.04383	10	1.9	0.51777	0.80383	59.32096
20	1.4	0.69286	0.90490	69.03678	20	1.9	0.20032	0.48840	36.02383
40	1.4	0.44469	0.74960	54.90758	40	1.9	0.02065	0.10721	8.83884
60	1.4	0.26610	0.57438	42.00565	60	1.9	0.00162	0.01561	1.40221
80	1.4	0.15124	0.41297	30.83713	80	1.9	0.00011	0.00177	0.16613

- The asymptotic RELEFF is 100% as n_x & $n_y \to \infty$. The larger λ is, the more rapidly the ARE (asymptotic RELEFF) approaches 100%. For example at $n_x = n_y = 100$ and $\lambda = 1.20$, the RELEFF is 48.10% while the corresponding RELEFF at $\lambda = 1.6$ is equal to 73% and at $\lambda = 2$ is equal to 81%.

Unfortunately, our Overlap results in this section is not applicable to non-normal underlying distributions as $n \to \infty$ because $(n - 1)S^2/\sigma^2$ is a quadratic form unlike $\bar{x}$.

6.0 The Impact of Overlap on Type I Error Probability of H_0: $\mu_x = \mu_y$ for Unknown Normal Process Variances and Small Sample Sizes

Since the population variances σ_x^2 and σ_y^2 are unknown, then their point unbiased estimators S_x^2 and S_y^2, respectively, must be used in order to make statistical inferences regarding σ_x^2/σ_y^2 and $\mu_x - \mu_y$. Thus, $\dfrac{(\bar{x}-\mu_x)}{S_x/\sqrt{n_x}}$ is not normally distributed but its sampling distribution (SMD) follows that of the Student's t [or simply "Student's"] with $(n_x - 1)$ degrees of freedom. As a result

$$\rightarrow \ \Pr(\bar{x} - t_{\alpha/2, n_x-1} S_x/\sqrt{n_x} \le \mu_x \le \bar{x} + t_{\alpha/2, n_x-1} S_x/\sqrt{n_x}) = 1-\alpha \qquad (34a)$$

Hence, the lower $(1-\alpha)$% CI for μ_x is $L(\mu_x) = \bar{x} - t_{\alpha/2, n_x-1} S_x/\sqrt{n_x}$, the corresponding upper limit is $U(\mu_x) = \bar{x} + t_{\alpha/2, n_x-1} S_x/\sqrt{n_x}$, and

$$\rightarrow \ \mathrm{CIL}(\mu_x) = 2\times t_{\alpha/2, n_x-1} S_x/\sqrt{n_x} \qquad (34b)$$

Similarly, $L(\mu_y) = \bar{y} - t_{\alpha/2, n_y-1} S_y/\sqrt{n_y}$, $U(\mu_y) = \bar{y} + t_{\alpha/2, n_y-1} S_y/\sqrt{n_y}$ and

$$\rightarrow \ \mathrm{CIL}(\mu_y) = 2\times t_{\alpha/2, n_y-1} S_y/\sqrt{n_y} . \qquad (34c)$$

6.1 The Case of H$_0$: $\sigma_x = \sigma_y = \sigma$ Not Rejected Leading to the Pooled t-Test

Assuming that X ~ N(μ_x, σ^2) and Y~N(μ_y, σ^2), then $\overline{X} - \overline{Y}$ has the N($\mu_x - \mu_y$, $\sigma^2 / n_x + \sigma^2 / n_y$) sampling distribution, where it is assumed that σ^2 is the common value of unknown $\sigma_x^2 = \sigma_y^2 = \sigma^2$. With the above assumptions, $\overline{x} - \overline{y}$ is an unbiased estimator of $\mu_x - \mu_y$ with Var($\overline{x} - \overline{y}$) $= \sigma^2 (1/n_x + 1/n_y)$. 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 sample t-test. If the assumption $\sigma_x = \sigma_y = \sigma$ is tenable and because statistical theory dictates that the total resources to be allocated according to n_x = $\sigma_x N / (\sigma_x + \sigma_y)$ = N/2 = n_y, then the most common application of the pooled t-test occurs under equal sample sizes. This assertion is consistent with Montgomery and Runger (1994, p. 411) and J. L. Devore (2004, p. 377). Although the LOS of a statistical test is rarely set beyond 10%, to be on the conservative side, we will use the pooled t-test iff the *P-value* of the pretest on $H_0 : \sigma_x = \sigma_y = \sigma$ exceeds 20%. Further, J. L Devore (2008, p. 340) states that " *Unfortunately, the usual " F test" of equal variances (Section 9.5) 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 10, because of Devore's quoted statement, pooling should be avoided unless the *P-value* of testing $H_0 : \sigma_x = \sigma_y = \sigma$ exceeds 40%. When $\sigma_x = \sigma_y = \sigma$ is tenable, the unbiased estimators S_x^2 and S_y^2 should be pooled to obtain one unbiased estimator of σ^2 , which is given by their weighted average based on their degrees of freedom, i.e.,

$$S_p^2 = \frac{\nu_x S_x^2 + \nu_y S_y^2}{\nu_x + \nu_y} = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2} \quad (35)$$

Note that E(S_p^2) $= \sigma^2$. Therefore, the sample $se(\bar{x} - \bar{y}) = S_p\sqrt{1/n_x + 1/n_y}$ and as a result the rv $[(\bar{x} - \bar{y})$-$(\mu_x - \mu_y)]/(S_p\sqrt{1/n_x + 1/n_y})$ has a Student's t sampling distribution with $\nu = \nu_x + \nu_y = n_x + n_y - 2$. Accordingly, the exact two sided $(1-\alpha)$100% CI for $\mu_x - \mu_y$ by the Standard method is given by

$$\bar{x} - \bar{y} - t_{\alpha/2,\nu} \times S_p\sqrt{1/n_x + 1/n_y} \le \mu_x - \mu_y \le \bar{x} - \bar{y} + t_{\alpha/2,\nu} \times S_p\sqrt{1/n_x + 1/n_y} \quad (36a)$$

resulting in the CIL of

$$2\,t_{\alpha/2,\nu} \times S_p\sqrt{1/n_x + 1/n_y}\text{, where } \nu = n_x + n_y - 2. \quad (36b)$$

Thus, H_0: $\mu_x - \mu_y = 0$ must be rejected at LOS $= \alpha$ iff

$$|\bar{x} - \bar{y}| > t_{\alpha/2,\nu} \times S_p\sqrt{1/n_x + 1/n_y} \quad (36c)$$

But, for the individual two t-CIs, the rejection condition is either L(μ_x) > U(μ_y) or L(μ_y) > U(μ_x). Using the definition of type I error Pr, bearing in mind that $t_\nu^2 = F_{1,\nu}$, leads to

α' = Pr(reject H_0| $\mu_x - \mu_y = 0$) = Pr[L(μ_x) >U(μ_y)] + Pr[L(μ_y) >U(μ_x)]

$$= \Pr[\bar{x} - t_{\alpha/2,\nu_x}\frac{S_x}{\sqrt{n_x}} > \bar{y} + t_{\alpha/2,\nu_y}\frac{S_y}{\sqrt{n_y}}] + \Pr[\bar{y} - t_{\alpha/2,\nu_y}\frac{S_y}{\sqrt{n_y}} > \bar{x} + t_{\alpha/2,\nu_x}\frac{S_x}{\sqrt{n_x}}]$$

$$= \Pr[\bar{x} - \bar{y} > t_{\alpha/2,\nu_x}\frac{S_x}{\sqrt{n_x}} + t_{\alpha/2,\nu_y}\frac{S_y}{\sqrt{n_y}}] + \Pr[\bar{x} - \bar{y} < -t_{\alpha/2,\nu_x}\frac{S_x}{\sqrt{n_x}} - t_{\alpha/2,\nu_y}\frac{S_y}{\sqrt{n_y}}]$$

$$= \Pr[|\bar{x} - \bar{y}| > t_{\alpha/2,\nu_x} S_x/\sqrt{n_x} + t_{\alpha/2,\nu_y} S_y/\sqrt{n_y}] \quad (37a)$$

$$= \Pr[|t_\nu| > \frac{t_{\alpha/2,\nu_x} S_x / \sqrt{n_x} + t_{\alpha/2,\nu_y} S_y / \sqrt{n_y}}{S_p \sqrt{1/n_x + 1/n_y}}]$$

$$= \Pr[F_{1,\nu} > \frac{(t_{\alpha/2,\nu_x} S_x / \sqrt{n_x} + t_{\alpha/2,\nu_y} S_y / \sqrt{n_y})^2}{S_p^2 (1/n_x + 1/n_y)}] \quad (37b)$$

Without loss of generality, we name the sample with the larger variance as X and let F_0 $= S_x^2/S_y^2 \geq 1$. Multiplying the argument on the RHS of Eq. (37b) by $n_x n_y$ for both numerator and denominator and substituting $F_0 = S_x^2/S_y^2 \geq 1$ into (37b) results in →

$$\alpha' = \Pr[F_{1,\nu} > \frac{\nu(t_{\alpha/2,\nu_x} \sqrt{F_0 n_y} + t_{\alpha/2,\nu_y} \sqrt{n_x})^2}{(\nu_y + F_0 \nu_x)(n_x + n_y)}]$$

$$\rightarrow \alpha' = \Pr[F_{1,\nu} > \frac{\nu(t_{\alpha/2,\nu_x} \sqrt{F_0 R_n} + t_{\alpha/2,\nu_y})^2}{(\nu_y + F_0 \nu_x)(1 + R_n)}] \quad (37c)$$

$$= \Pr[F_{1,\nu} > \frac{\nu(k \times t_{\alpha/2,\nu_x} + t_{\alpha/2,\nu_y})^2}{(\nu_y + F_0 \nu_x)(1 + R_n)}] \quad (37d)$$

where $\nu = n_x + n_y - 2$, $R_n = n_y/n_x$ and $k = \sqrt{R_n F_0} = (S_x \sqrt{n_y})/(S_y \sqrt{n_x})$ is the sample *se* ratio. Eq.(37c) clearly shows that, besides α, the value α' depends only on n_x, n_y and F_0 $= S_x^2/S_y^2$ and not on the specific values of S_x and S_y. For the pooled t-test, in the most common case of balanced design (i.e., $n_x = n_y = n$), Eq. (37c) reduces to

$$\alpha' = \Pr[F_{1,\nu} > \frac{F_{\alpha,1,n-1}(1 + \sqrt{F_0})^2}{1 + F_0}] \quad (37e)$$

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 (37e) attains its maximum at $F_0 = 1$ and its minimum at $F_{0.10,n-1,n-1}$ or at $F_{0.90,n-1,n-1}$. 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 (37e) is a monotonically increasing function of n.

Further, Matlab has verified that the limiting value of α' in Eq. (37e), as n → 7,819,286 lies in the interval [0.005574595835 (at $F_0 = F_{0.10}$), 0.005574597084 (at $F_0 = 1$)], both of which are very close to the known-Variance case of testing H_0: $\mu_x = \mu_y$. Eq. (37e) 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 sample t-statistic, discussed in the next section, while (37e) pertains to the pooled t-test. Further, it will be shown that the denominator df of the F statistic for their general case will not equal n −1 as stated by Payton *et al.* (2000).

The next objective is to show that $\alpha' < \alpha$ for all n_x, n_y, S_x and S_y for which

$$F_{0.90,n_x-1,n_y-1} < F_0 = S_x^2 / S_y^2 < F_{0.10,n_x-1,n_y-1}.$$

First, comparing inequality (36c) with Eq.(37a), it follows that if

$$t_{\alpha/2,v_x}\frac{S_x}{\sqrt{n_x}} + t_{\alpha/2,v_y}\frac{S_y}{\sqrt{n_y}} > t_{\alpha/2,v} \times S_p\sqrt{\frac{1}{n_x}+\frac{1}{n_y}} \tag{38a}$$

then the same conclusion as the case of known and equal variances will be reached, i.e., $\alpha' < \alpha$. For the case of balanced design where $n_x = n_y = n$, inequality (38a) reduces to

$$\rightarrow \; t_{\alpha/2,n-1}\frac{S_x}{\sqrt{n}} + t_{\alpha/2,n-1}\frac{S_y}{\sqrt{n}} > t_{\alpha/2,2(n-1)} \times \sqrt{(S_x^2 + S_y^2)/2} \times \sqrt{2/n}$$

$$\rightarrow \; \text{Is } t_{\alpha/2,n-1}(S_x + S_y) > t_{\alpha/2,2(n-1)} \times \sqrt{S_x^2 + S_y^2}\ ? \tag{38b}$$

Substituting $F_0 = S_x^2 / S_y^2$ into Eq.(38b) results in

$$t_{\alpha/2,n-1}(1+\sqrt{F_0}) > t_{\alpha/2,2(n-1)}\sqrt{1+F_0}\ ? \tag{38c}$$

It is clear that the inequality in (38c) easily holds because $t_{\alpha/2,n-1} > t_{\alpha/2,2(n-1)}$ for all finite n and $(1+\sqrt{F_0}) > \sqrt{1+F_0}$ for all values of F_0 because F_0 is never negative. Therefore, the inequality (38a) is true for the case of equal sample sizes but it is not always so for the unequal sample sizes case. In the unbalanced case, the difficulty in inequality (38a) occurs when the larger sample size (which will be denoted by n_x) also has much larger variance for which inequality (38a) will not be true. For example, if n_y = 20, S_y^2 = 0.30, n_x = 60 and S_x^2 = 1.8, the LHS of inequality (38a) becomes 0.6029 and its RHS becomes 0.6157 so that the inequality is violated. However, in such a case the value of F-statistic is $F_0 = S_x^2 / S_y^2 = 6$ whose *P-value* for pre-testing $H_0 : \sigma_x = \sigma_y = \sigma$ will equal to 0.00007736, i.e., this last hypothesis is easily rejected so that pooling is disallowed. Again to be on the conservative side, we allow pooling iff the *P-value* of the variance-pretest exceeds 20%. Otherwise, the two-independent sample t-statistic will be used for testing $H_0 : \mu_x = \mu_y$. This is consistent with Devore's (2004, p. 377) assertion of "using the two-sample t procedure unless there is compelling evidence for doing otherwise, particularly when the two sample sizes are different". Further, unlike the case of balanced design, when $n_x > n_y$ the value of α' is an increasing function (but not monotonically) of F_0 but when $n_x < n_y$, the value of α' is almost always a decreasing function of F_0. Thus, for fixed $n_x > n_y$, the maximum occurs at $F_{0.10,n_x-1,n_y-1}$, and when $n_x < n_y$ the maximum occurs at $F_{0.90,n_x-1,n_y-1}$. As n_x and n_y both increase at the same F_0, so does the value of α' in (37c). When one sample size is twice the other, the limiting value of α' (in terms of n_x and n_y) is 0.006286690. When one sample size is three times the other, the limiting value of α' is 0.00733793. When one sample size is four times the other, the limiting value of α' is 0.008390775. When one sample size is 5 times the other, the limiting value of α' is 0.0093831123. When one sample size is 10 times the other limiting value of α' is 0.01336332. Finally, as limit of $R_n = n_y/n_x \to \infty$ or 0, the Overlap type I Pr approaches that of an exact α-level test. Table 20 gives the exact α' from Eq. (37c) for different n_x and n_y combinations. The values of F_0 in Table 20 are restricted such that *P-value* of the pretest $H_0: \sigma_x = \sigma_y$ exceeds 20%.

6.2 The Case of H_0: $\sigma_x = \sigma_y$ Rejected Leading to the Two-Independent Sample t-Test

Assuming that $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$, than $\overline{X} - \overline{Y}$ is also $N(\mu_x - \mu_y$, $\sigma_x^2 / n_x + \sigma_y^2 / n_y)$, but now the null hypothesis of H_0: $\sigma_x = \sigma_y$ is rejected at the 20% level leading to the assumption that the F-statistic $F_0 = S_x^2 / S_y^2 > 2$ for all sample sizes $16 \le n_x$ & n_y. Note that for larger sample sizes such as n_x & $n_y = 41$, F_0 can be as small as 1.510 and H_0: $\sigma_x = \sigma_y$ can still be rejected at the 20% level because $F_{0.10,40,40} = 1.5056$, while for n_x & $n_y = 11$, an F_0 as large as 2.323 is needed because $F_{0.10,10,10} = 2.3226$. Note that an $F_0 = 2$ is significant at the level 20% once n_x & $n_y \ge 16$ because $F_{0.10,15,15} = 1.9722$. It has been shown in statistical theory that if the assumption $\sigma_x = \sigma_y$ is not

Table 20. The Pooled α' Values for Different n_x, n_y and F_0 Combinations at $\alpha = 0.05$

n_x	n_y	F_0	α'	n_x	n_y	F_0	α'
20	40	0.8	0.0071355	20	40	1.4	0.0039812
20	60	0.8	0.0087831	20	60	1.4	0.0035984
20	80	0.8	0.0103096	20	80	1.4	0.0035130
30	10	0.8	0.0034279	30	10	1.4	0.0085406
30	20	0.8	0.0047292	30	20	1.4	0.0068050
30	30	0.8	0.0054425	30	30	1.4	0.0055284
40	20	0.8	0.0044541	40	20	1.4	0.0080709
40	40	0.8	0.0054943	40	40	1.4	0.0055823
40	80	0.8	0.0075641	40	80	1.4	0.0042383
40	100	1	0.0063762	40	100	1.4	0.0040584
20	40	1	0.0056135	20	40	1.5	0.0037263
20	60	1	0.0062056	20	60	1.5	0.0032137
20	80	1	0.0068563	20	80	1.5	0.0030428
30	10	1	0.0049815	30	10	1.5	0.0094908
30	20	1	0.0053938	30	20	1.5	0.0071647
30	30	1	0.0053753	30	30	1.5	0.0055980
40	20	1	0.0056135	40	10	1.5	0.0111374
40	40	1	0.0054254	40	20	1.5	0.0086996
40	80	1	0.0059601	40	30	1.5	0.0067838
40	100	1	0.0063762	40	40	1.5	0.0056536
20	40	1.2	0.0046424	1000	2000	$F_{0.90,vx,vy}$	0.0067706
20	60	1.2	0.0046283	100000	200000	$F_{0.90,vx,vy}$	0.0063436
20	80	1.2	0.0048067	10000000	20000000	$F_{0.90,vx,vy}$	0.0063019
30	10	1.2	0.0067025	1000000000	2000000000	$F_{0.90,vx,vy}$	0.0062871
30	30	1.2	0.0054202	1000	3000	$F_{0.90,vx,vy}$	0.0081935
40	20	1.2	0.0068266	100000	300000	$F_{0.90,vx,vy}$	0.0074960
40	40	1.2	0.0054714	10000000	30000000	$F_{0.90,vx,vy}$	0.0074280
40	80	1.2	0.0049359	1000000000	3000000000	$F_{0.90,vx,vy}$	0.0073386
20	40	1.3	0.0042824	1000	4000	$F_{0.90,vx,vy}$	0.0095652
20	60	1.3	0.0040623	100000	400000	$F_{0.90,vx,vy}$	0.0086485
20	80	1.3	0.0040901	10000000	40000000	$F_{0.90,vx,vy}$	0.0085592
30	10	1.3	0.0076096	1000000000	4000000000	$F_{0.90,vx,vy}$	0.0083917
30	30	1.3	0.0054683	1000	5000	$F_{0.90,vx,vy}$	0.0108398
40	20	1.3	0.0074460	100000	500000	$F_{0.90,vx,vy}$	0.0097352
40	40	1.3	0.0055207	10000000	50000000	$F_{0.90,vx,vy}$	0.0096277
40	80	1.3	0.0045561	1000000000	5000000000	$F_{0.90,vx,vy}$	0.0093842

tenable**,** the statistic $[(\bar{x}-\bar{y})-(\mu_x-\mu_y)]/\sqrt{(S_x^2/n_x)+(S_y^2/n_y)}$ has the approximate

Student's t-distribution with degrees of freedom

$$\nu = \frac{(S_x^2/n_x + S_y^2/n_y)^2}{\frac{(S_x^2/n_x)^2}{n_x - 1} + \frac{(S_y^2/n_y)^2}{n_y - 1}} = \frac{[V(\bar{x}) + V(\bar{y})]^2}{\frac{(V(\bar{x}))^2}{\nu_x} + \frac{(V(\bar{y}))^2}{\nu_y}} = \frac{\nu_x \nu_y [V(\bar{x}) + V(\bar{y})]^2}{\nu_y (V(\bar{x}))^2 + \nu_x (V(\bar{y}))^2} \quad (39a)$$

$$\rightarrow \nu = \frac{\nu_x \nu_y (F_0 R_n + 1)^2}{\nu_y (F_0 R_n)^2 + \nu_x} = \frac{\nu_x \nu_y (k^2 + 1)^2}{\nu_y k^4 + \nu_x} \quad (39b)$$

where $V(\bar{x}) = S_x^2 / n_x$, $R_n = n_y/n_x$, $k = (S_x/\sqrt{n_x})/(S_y/\sqrt{n_y})$, and $F_0 = S_x^2 / S_y^2$. Eq. (39b) shows that ν depends only on n_x, n_y, and the *se* ratio $\sqrt{F_0 R_n}$. The formulas for degrees of freedom in (39a &b) rarely lead to an integer and ν is generally rounded down to make the test of H_0: $\mu_x - \mu_y = 0$ conservative, i.e., rounding down ν increases the *P-value* of this last test. However, programs like Matlab and Minitab will provide the cdf and percentage points of the t-distribution for non-integer values of ν in Eqs. (39). It has been verified using a spreadsheet that $\text{Min}(\nu_x, \nu_y) < \nu < \nu_x + \nu_y$ is a certainty, and hence this t-test is less powerful than the pooled t-test. In fact, it is easy to algebraically prove that for the case of $n_x = n_y = n$, the value of ν always exceeds (n − 1) and is always less than 2(n − 1). It can also be verified that the maximum of ν in Eqs. (39) occurs when the larger sample also has much larger variance, but yet its value can never exceed the df, $n_x + n_y - 2$, of the pooled t-test, as illustrated in Table 21.

When H_0: $\sigma_x = \sigma_y$ is rejected at the 20% level (i.e., *P-value* of the test < 0.20), the approximate (1−α)×100% CI for $\mu_x - \mu_y$ is given by

$$\bar{x} - \bar{y} - t_{\alpha/2,\nu} \times \sqrt{S_x^2/n_x + S_y^2/n_y} \le \mu_x - \mu_y \le \bar{x} - \bar{y} + t_{\alpha/2,\nu} \times \sqrt{S_x^2/n_x + S_y^2/n_y} \quad (40a)$$

resulting in the approximate CIL is $2\, t_{\alpha/2,\nu} \times \sqrt{S_x^2/n_x + S_y^2/n_y}$, and H_0: $\mu_x - \mu_y = 0$ can

Table 21. Verifying the Inequality that min(ν_x, ν_y) < ν < ν_x+ ν_y for Different ν_x and ν_y Combinations

ν_x	ν_y	$Var(x)$	$Var(y)$ (at $F_0 = F_{0.=90,\nu_x,\nu_y}$)	ν	ν_x	ν_y	$Var(x)$	$Var(y)$ (at $F_0 = F_{0.=90,\nu_x,\nu_y}$)	ν
1	11	20	6.201	1.106	11	1	20	0.331	11.992
6	16	20	9.181	8.371	16	6	20	6.987	18.725
11	21	20	10.550	17.483	21	11	20	9.449	30.068
16	26	20	11.451	27.422	26	16	20	10.779	40.883
26	31	20	12.364	49.013	31	26	20	12.174	56.686
36	41	20	13.220	69.455	41	36	20	13.110	76.486
46	51	20	13.830	89.823	51	46	20	13.758	96.317
66	71	20	14.664	130.370	71	66	20	14.626	136.052
86	91	20	15.222	170.752	91	86	20	15.198	175.855
106	111	20	15.631	211.034	111	106	20	15.615	215.702
126	131	20	15.947	251.252	131	126	20	15.935	255.579
176	181	20	16.505	351.633	181	176	20	16.498	355.351
226	231	20	16.878	451.882	231	226	20	16.874	455.192
326	331	20	17.361	652.198	331	326	20	17.359	654.979
426	431	20	17.669	852.395	431	426	20	17.668	854.840
526	531	20	17.889	1052.532	531	526	20	17.888	1054.739
1200	3000	20	18.811	2151.446	3000	1200	20	18.788	2275.082
1500	3500	20	18.921	2768.370	3500	1500	20	18.903	2912.635
2000	4000	20	19.037	3913.944	4000	2000	20	19.026	4090.453
2500	4500	20	19.120	5067.119	4500	2500	20	19.112	5263.906
2800	5000	20	19.166	5693.779	5000	2800	20	19.159	5901.618

be rejected at LOS $=\alpha$ if $|\bar{x}-\bar{y}|> t_{\alpha/2,\nu} \times \sqrt{S_x^2/n_x + S_y^2/n_y}$, i.e.,

$$\alpha \cong \Pr(|\bar{x}-\bar{y}|> t_{\alpha/2,\nu} \times \sqrt{S_x^2/n_x + S_y^2/n_y} \mid \mu_x - \mu_y = \delta = 0)$$

$$\cong \Pr(F_{1,\nu} > t_{\alpha/2,\nu}^2 \mid \delta = 0) = \Pr(F_{1,\nu} > F_{\alpha,1,\nu} \mid \delta = 0) \qquad (40b)$$

As in the case of pooled t-test, for the individual two t-CIs, the rejection requirement is either L(μ_x) > U(μ_y) or L(μ_y) > U(μ_x) leading to the same condition as before in Eq. (37a). That is,

$$\alpha' = \Pr(\text{reject } H_0 \mid \delta = 0) = \Pr[\,|\bar{x}-\bar{y}| > t_{\alpha/2,\nu_x} S_x/\sqrt{n_x} + t_{\alpha/2,\nu_y} S_y/\sqrt{n_y}\,] \qquad (37a)$$

It is impossible to studentize the argument of Eq. (37a) because when $\sigma_x \neq \sigma_y$, the expression for $t_\nu = Z/\sqrt{\chi^2_\nu/\nu}$ will show that $|\bar{x}-\bar{y}|/\sqrt{S_x^2/n_x + S_y^2/n_y}$ is not central t distributed with $n_x+n_y - 2$ df. In other words, there does not exist a central χ^2_ν rv that reduces $t_\nu = Z/\sqrt{\chi^2_\nu/\nu}$ to the form $|\bar{x}-\bar{y}|/\sqrt{S_x^2/n_x + S_y^2/n_y}$ iff $\sigma_x \neq \sigma_y$. However, $|\bar{x}-\bar{y}|/\sqrt{S_x^2/n_x + S_y^2/n_y}$ is approximately t distributed with df ν given in Eqs.(39).

Therefore, Eq. (37a) can approximately be written as

$$\alpha' \approx \Pr\{|t_\nu| > (t_{\alpha/2,\nu_x} S_x/\sqrt{n_x} + t_{\alpha/2,\nu_y} S_y/\sqrt{n_y})/\sqrt{S_x^2/n_x + S_y^2/n_y}\}$$

$$= \Pr\{F_{1,\nu} > (t_{\alpha/2,\nu_x} S_x/\sqrt{n_x} + t_{\alpha/2,\nu_y} S_y/\sqrt{n_y})^2/(S_x^2/n_x + S_y^2/n_y)\}$$

Or $$\alpha' \approx \Pr\{F_{1,\nu} > (t_{\alpha/2,\nu_x} S_x\sqrt{n_y} + t_{\alpha/2,\nu_y} S_y\sqrt{n_x})^2/(n_y S_x^2 + n_x S_y^2)\}$$

$$= \Pr[F_{1,\nu} > (k \times t_{\alpha/2,\nu_x} + t_{\alpha/2,\nu_y})^2/(1+k^2)] \quad (41a)$$

Let $R_n = n_y/n_x$ (or $n_y = R_n n_x$) and $F_0 = S_x^2/S_y^2$. Substituting $R_n = n_y/n_x$ and F_0 into Eq. (41a) results in

$$\alpha' \approx \Pr[F_{1,\nu} > (t_{\alpha/2,\nu_x}\sqrt{R_n F_0} + t_{\alpha/2,\nu_y})^2/(R_n F_0 + 1)] \quad (41b)$$

α' can be also represents as $\alpha' \approx \Pr[F_{1,\nu} > (k \times t_{\alpha/2,\nu_x} + t_{\alpha/2,\nu_y})^2/(k^2+1)]$, where $k = (S_x/\sqrt{n_x})/(S_y/\sqrt{n_y})$. When $n_x = n_y = n$, $R_n = 1$, and the above formula for α' reduces to

$$\alpha' \approx \Pr[F_{1,\nu} > F_{\alpha,1,n-1}(\sqrt{F_0} + 1)^2/(F_0 + 1)] \quad (41c)$$

which is similar to (37e) but ν is given by Eqs. (39) instead of $n_x+n_y - 2$ in the case of the pooled t-test, or instead of n − 1 as stated by Payton *et al.* (2000). For equal sample sizes

Eq. (39b) simplify to $\nu = (\mathrm{n}-1)(1+\mathrm{F}_0)^2/(1+\mathrm{F}_0^2)$ and not (n −1) as reported by Payton *et al.* (2000). 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 $F_0 = 1$ is in perfect agreement with H_0: $\sigma_x^2 = \sigma_y^2$. Further, Eq. (41b) shows that α' does not depend on the specific values of S_x^2 *and* S_y^2 but only on their ratio $k = \sqrt{\mathrm{F}_0 R_n}$. For Payton *et al.*'s (2000) reported example of $n_1 = n_2 = 10$, $S_1 = 0.80$ and $S_2 = 1.60$, Eq. (41c) shows that at n = 10 and $F_0 = 0.25$, ν = 13.2353 resulting in the value of $\alpha' \approx 0.00940573$, which is different from 0.0149 reported by Payton *et al.* (2000, p. 549). The df used by them was 9 which caused the % relative error in their reported $\alpha' = 0.0149$ to be equal 54.414%.

Payton *et al.* (2000, p. 549) also make the following statement about 1/3 of the way from the top of their p. 549: *"If the samples are collected from the same normal population, the quantity* $\dfrac{n(\overline{Y}_1 + \overline{Y}_2)^2}{S_1^2 + S_2^2}$ *is F-distributed with 1 and n−1 degrees of freedom."* The statement should go as follows: If the samples are collected from the two normal populations with identical means and variances, [our Eq. (37e) shows that] the statistic $\dfrac{n(\overline{Y}_1 - \overline{Y}_2)^2}{S_1^2 + S_2^2}$ is F-distributed with 1 and 2(n−1) degrees of freedom (not n −1 as stated).

Payton *et al.* (2000, p. 550) also make the following statement in the second paragraph leading to their Eq. [1]: *" If the researcher is willing to assume that S_1 and S_2 are estimating the same parameter value (i.e., homogenous variances), then the above equation simplifies to 0.95 =* $Pr[F_{1,9} < 2F_{\alpha,1,9}]$ *[1]"*

Their above quote should be stated as follows: In the unlikely event that $F_0 = S_1^2 / S_2^2$ is realized to equal 1, then the above equation simplifies to 0.95 = Pr($F_{1,13.2353} < 2F_{\gamma,1,9}$). Note that they are using (1 − α) also as the Overlap confidence level, and secondly just because two independent population variances are equal, it does not imply that the corresponding point estimates S_1^2 *and* S_2^2 will be the same. Further, Payton *et al.* limit their sample means, $\overline{Y}_1$ and $\overline{Y}_2$, originating from the same normal population on p. 548. Our work herein is not limited to the same normal population but to any two distinct normal populations.

We now proceed to obtain the LUB (least upper bound) and the GLB (greatest lower bound) for α' in Eq. (41b). The LUB occurs when the argument on the RHS of the Pr in Eq. (41b) is smallest. To this end, let ν_2 = Max(ν_x, ν_y) and thus

$$\alpha' \geq \Pr\{F_{1,\nu} > [t_{\alpha/2,\nu_2}\sqrt{R_nF_0} + t_{\alpha/2,\nu_2}]^2/(R_nF_0 + 1)\}$$

$$\rightarrow \text{LUB}(\alpha') = \Pr\{F_{1,\nu} > F_{\alpha,1,\nu_2}(\sqrt{R_nF_0} + 1)^2/(R_nF_0 + 1)\}$$

Conversely, the greatest lower bound occurs when the argument on the RHS of (41b) is largest. Letting ν_1 = Min(ν_x, ν_y) in (41b) results in

$$\alpha' \leq \Pr\{F_{1,\nu} > [t_{\alpha/2,\nu_1}\sqrt{R_nF_0} + t_{\alpha/2,\nu_1}]^2/(R_nF_0 + 1)\}$$

$$\rightarrow \text{GLB}(\alpha') = \Pr\{F_{1,\nu} > F_{\alpha,1,\nu_1}(\sqrt{R_nF_0} + 1)^2/(R_nF_0 + 1)\}\text{ , or}$$

$$\Pr\{F_{1,\nu} > F_{\alpha,1,\nu_1}(\sqrt{R_nF_0} + 1)^2/(R_nF_0 + 1)\} < \alpha' < \Pr\{F_{1,\nu} > F_{\alpha,1,\nu_2}(\sqrt{R_nF_0} + 1)^2/(R_nF_0 + 1)\},$$

while the expression for exact type I Pr from (40b) is $\alpha \cong \Pr(F_{1,\nu} > F_{\alpha,1,\nu} \mid \mu_x - \mu_y = 0)$. The function $(\sqrt{R_nF_0} + 1)^2/(R_nF_0 + 1)$ clearly always exceeds 1 because $R_nF_0 =$

$(n_y / n_x) \times (S_x^2 / S_y^2) = V(\bar{x}) / V(\bar{y})$, which is also a *se* ratio, can never equal to zero and the function is bounded by $1 < (\sqrt{R_n F_0} + 1)^2 / (R_n F_0 + 1) \le 2$, the maximum occurring when $R_n F_0 = 1$. Because we are seeking to establish that α' in (41b) is always smaller that $\alpha \cong \Pr(F_{1,\nu} > F_{\alpha,1,\nu} \mid \mu_x - \mu_y = 0)$, we consider the very worst-case scenario where the smallest sample has the largest variance. For example, at $\nu_x = 1$, $\nu_y = 11$ (so that $R_n = n_y/n_x = 12/2 = 6$), $S_x^2 = 8.00$, $S_y^2 = 2.4805$, $F_0 = F_{0.10,1,11} = 3.2252$, $t_{0.025,1} = 12.7062$, and $t_{0.025,11} = 2.2010$ the value of $\nu = 1.105755$. Substituting these into (41b) results in $\alpha' = \Pr\{F_{1,1.105755} > 165.842616) = 0.0386 < 0.05$. It has also been verified that S_y^2 can be as small as $S_x^2 / F_{0.0001,\nu x,\nu y}$ and still $\alpha' < \alpha$. Note that if $F_{0.90,n_x-1,n_y-1} < F_0 = S_x^2 / S_y^2 < F_{0.10,n_x-1,n_y-1}$, then we are recommending using the pooled t-test so that the value of F_0 must lie outside the interval $(F_{0.90,n_x-1,n_y-1}, F_{0.10,n_x-1,n_y-1})$ in order to apply the two-independent sample t-test. This is consistent with statistical literature (see J. L. Devore, p.377) that suggests not to use the pooled t-test unless there is compelling evidence in favor of H_0: $\sigma_x = \sigma_y$.

Keeping $F_0 \ge F_{0.10,n_x-1,n_y-1}$ fixed, α' in Eq. (41a) attains its minimum at $R_n = n_y/n_x = 1$, and the limit of α'as $R_n \to 0$ or ∞ is α; similarly, if $F_0 \le F_{0.90,n_x-1,n_y-1}$ is kept fixed, α' is minimum at $R_n = 1$ and its limit approaches α as $R_n \to 0$ or ∞. As $F_0 \to \infty$, α' approaches the value of α, i.e., the Overlap converges to an α-level test; however, the farther R_n is above 1, the faster is the limiting approach of α' to α as $F_0 \to \infty$. As $F_0 \to 0$,

α' also approaches the value of α, and the farther R_n is below 1, the faster is the limiting approach of α' to α as $F_0 \to 0$.

For example, if $n_x = n_y = 50$ (i.e., $R_n = 1$), $F_0 = 10^6$ then $\alpha' = 0.04978024$ (nearly 5%). If $n_x = 50$, $n_y = 100$, $R_n = 2$, $F_0 = 10^6$, then $\alpha' = 0.04984645882$. However, If $n_x = 50$, $n_y = 25$, $R_n = 0.5$, $F_0 = 10^6$, $\alpha' = 0.0496811387$ but if $F_0 = 10^{-6}$, then $\alpha' = 0.0498546652$.

Further, if $R_n = 1$, then the limiting value of α' as $n_x \to \infty$ is equal to 0.0055751 as long as $F_{0.10} < F_0 < \infty$. The limiting value of α' as $F_0 \to \infty$ is equal to α. If $R_n = 2$, the limiting value of α' as $n_x \to \infty$ is equal to 0.00632067 and as $F_0 \to \infty$, $\alpha' \to \alpha$; if If $R_n = 3$, the limit is equal to 0.0074346; if If $R_n = 4$, then $\alpha' \to 0.008555$; if $R_n = 5$, the limiting value is 0.00962217; if $R_n = 10$, the value of the limit is 0.01391355, etc. As $R_n \to \infty$ (or 0), $\alpha' \to \alpha$, where the rate of approach to α increases with increasing $F_0 > F_{0.10}$ (or decreasing $F_0 < F_{0.90}$), respectively. This is in total agreement with Table 4 on p. 23 where $\alpha' \to \alpha$ as the SE ratio [$= (\sigma_x / \sqrt{n_x})/(\sigma_y / \sqrt{n_y})$] became larger and larger. Note that as n_x and $n_y \to \infty$, $\sqrt{F_0} \to$ the SE ratio.

In conclusion, it is clear that the Overlap does reduce type I error Pr substantially and only the limiting value of α' as $R_n \to \infty$ (or 0) is close but always less than α .

6.3 Comparing the Paired t-CI with Two Independent t-CIs

Unlike, the independent t-CI for $\mu_x - \mu_y$ for a completely randomized design (CRD), the paired t-CI must be formed for a randomized complete block design (RCBD), where the rvs X and Y are paired observations or a bivariate random vector $\begin{bmatrix} x \\ y \end{bmatrix}$ from a bivariate normal population. The most common example is when a measurement X is made on a subject (such his/her weight) and a treatment (such as a diet plan) is applied and 3 months later the same subject's weight Y is measured and the difference D = X − Y is formed for that subject to ascertain the effectiveness of the diet plan. The paired t-test is sometimes misused when X and Y are independent random variables, i.e., they do not belong to the same block. Assuming that the rv D = X − Y is N($\mu_x - \mu_y$, σ_D^2), then $(\bar{d} - \mu_d)/se(\bar{d}) = [\bar{d} - (\mu_x - \mu_y)]\sqrt{n}/S_d$ has the exact Student's t-distribution with ν = n − 1 degrees of freedom. Thus under the null hypothesis H_0: $\mu_x - \mu_y = 0$, the statistic $\bar{d}\sqrt{n}/S_d$ can be used to make a decision about the validity of H_0 at the α- level, i.e., we reject H_0: $\mu_x - \mu_y = 0$ at the α×100% level iff $t_0 = \bar{d}\sqrt{n}/S_d$ exceeds $t_{\alpha/2,n-1}$. Because X and Y are correlated, then the

$$V(D) = V(X - Y) = V(X) + V(Y) - 2\text{Covariance}(X, Y) \tag{42}$$

where the Covariance(X, Y) = COV(X, Y) = $\sigma_{xy} = E[(X - \mu_x)(Y - \mu_y)]$. Because the population correlation coefficient $\rho_{xy} = \sigma_{xy}/(\sigma_x\sigma_y)$ always lies within the closed interval [−1, 1], then it follows that $\sigma_x\sigma_y \geq |\sigma_{xy}|$. A point unbiased estimate of the V(D) is given by $\hat{\sigma}_d^2 = S_x^2 + S_y^2 - 2\hat{\sigma}_{xy}$, where $\hat{\sigma}_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})/(n-1)$ and $S_d = \sqrt{S_x^2 + S_y^2 - 2\hat{\sigma}_{xy}}$. The sample correlation coefficient r = $\hat{\sigma}_{xy}/(S_xS_y)$ is also constrained to the closed interval [−1, 1], implying that for certain that both $\hat{\sigma}_{xy}$ and

$|\hat{\sigma}_{xy}| \le S_x S_y$. Although, comparing the overlapping of two independent CIs (generally from a completely randomized design) with a CI where observations are related is not appropriate and clearly misused, we do such comparison here just to ascertain the impact of Overlap on type I probability when the two samples are correlated, as also done in Overlap literature.

For paired observations, the (1−α)×100% CI for the expected difference in means is given by: $\bar{x} - \bar{y} - t_{\alpha/2,n-1} \times S_d / \sqrt{n} \le \mu_x - \mu_y \le \bar{x} - \bar{y} + t_{\alpha/2,n-1} \times S_d / \sqrt{n}$ (43)

so that H_0 is rejected at the α×100% level if $|\bar{d}| = |\bar{x} - \bar{y}| > t_{\alpha/2,n-1} \times S_d / \sqrt{n}$, i.e., α = $\Pr(|\bar{x} - \bar{y}| \sqrt{n} / S_d > t_{\alpha/2,n-1}) = \Pr(|t_{n-1}| \ge t_{\alpha/2,n-1})$. For the two separate CIs, the rejection requirement is either $L(\mu_x) > U(\mu_y)$ or $L(\mu_y) > U(\mu_x)$ leading to the same condition as in Eq. (37a).

$$\alpha' = \Pr[|\bar{x} - \bar{y}| > t_{\alpha/2,n-1} S_x / \sqrt{n} + t_{\alpha/2,n-1} S_y / \sqrt{n}]$$

$$= \Pr[|\bar{d}| \sqrt{n} > t_{\alpha/2,n-1}(S_x + S_y)] = \Pr[|\bar{d}| \sqrt{n} / S_d > t_{\alpha/2,n-1}(S_x + S_y) / S_d] \quad (44a)$$

Because the null SMD of $\bar{d} \sqrt{n} / S_d$ is the Student's t with (n − 1) df, then (44a) can be written as $\alpha' = \Pr[|t_{n-1}| > t_{\alpha/2,n-1}(S_x + S_y) / S_d] = \Pr[F_{1,n-1} > F_{\alpha,1,n-1}(S_x + S_y)^2 / S_d^2]$

$$= \Pr[F_{1,n-1} > F_{\alpha,1,n-1}(\sqrt{F_0} + 1)^2 / (F_0 + 1 - 2r\sqrt{F_0})] \quad (44b)$$

We now proceed to show that $(S_x+S_y)^2 \ge S_d^2 = S_x^2 + S_y^2 - 2\hat{\sigma}_{xy}$ for all values S_x and S_y. There are two possibilities: (1) $r > 0 \rightarrow \hat{\sigma}_{xy} > 0$, in which case it is obvious that $(S_x+S_y)^2 > S_d^2$. (2) $r < 0 \rightarrow \hat{\sigma}_{xy} < 0$ and S_d^2 attains its maximum when $r = -1 \rightarrow \mathrm{Max}(S_d^2)$

$= S_x^2 + S_y^2 + 2|\hat{\sigma}_{xy}|$. In this worst-scenario case, it is clear that $(S_x+S_y)^2 =$

$S_x^2 + S_y^2 + 2S_xS_y \geq S_x^2 + S_y^2 + 2|\hat{\sigma}_{xy}|$ because $|\hat{\sigma}_{xy}| \leq S_xS_y$. Thus, as before, $\alpha' < \alpha$.

How much smaller α' is than α depends both on the sign and magnitude of the sample correlation coefficient r . The glb occurs when $(S_x+S_y)^2$ is largest relative to S_d^2, i.e., when S_d^2 attains its minimum value. This minimum occurs when X and Y are highly positively correlated and in the limit $\hat{\sigma}_{xy} \rightarrow S_xS_y$. From Eq. (44b) we obtain

$$\alpha' = \Pr[F_{1,n-1} > F_{\alpha,1,n-1}(S_x+S_y)^2/S_d^2] \geq \Pr[F_{1,n-1} > F_{\alpha,1,n-1}(S_x+S_y)^2/(S_x^2+S_y^2-2S_xS_y)]$$

$$\rightarrow \alpha' > \Pr[F_{1,n-1} > F_{\alpha,1,n-1}(S_x+S_y)^2/(S_x-S_y)^2]$$

$$\rightarrow \text{GLB}(\alpha') = \Pr[F_{1,n-1} > F_{\alpha,1,n-1}(S_x+S_y)^2/(S_x-S_y)^2]$$

$$\rightarrow \text{GLB}(\alpha') = \Pr[F_{1,n-1} > F_{\alpha,1,n-1}(\sqrt{F_0}+1)^2/(\sqrt{F_0}-1)^2] \qquad (44c)$$

There seems to exist a problem in the inequality (44c), i.e., when $S_x \rightarrow S_y$, the expression on the RHS of the Pr is not defined. However, this occurs iff the correction coefficient r = 1 which occurs only if the values of the rv Y is precisely a linear function of X, i.e., Y must equal to ax + b +ε and the constant a > 0 and b can be any real number. Because a > 0, the variance of Y cannot equal to that of X.

Secondly, the largest value of α' occurs when $F_{\alpha,1,n-1}(S_x+S_y)^2/S_d^2$ attains its minimum value which in turn occurs when S_d attains its maximum value of $S_x^2 + S_y^2 +$ $2|\hat{\sigma}_{xy}|$. Thus,

$$\alpha' = \Pr[F_{1,n-1} > F_{\alpha,1,n-1}(S_x+S_y)^2/S_d^2] \leq \Pr[F_{1,n-1} > F_{\alpha,1,n-1}(S_x+S_y)^2/(S_x^2+S_y^2+2\left|\hat{\sigma}_{xy}\right|)]$$

The quantity $S_x^2 + S_y^2 + 2|\hat{\sigma}_{xy}|$ attains its maximum when the sample correlation coefficient r = –1, in which case $|\hat{\sigma}_{xy}|$ = S_xS_y so the maximum value of α' reduces to

$\alpha' \leq \Pr[F_{1,n-1} > F_{\alpha,1,n-1}(S_x + S_y)^2 / (S_x^2 + S_y^2 + 2S_xS_y)] = \Pr[F_{1,n-1} > F_{\alpha,1,n-1}) = \alpha.$

Hence, we have the result

$\Pr[F_{1,n-1} > F_{\alpha,1,n-1}(S_x + S_y)^2 / (S_x - S_y)^2] \leq \alpha' \leq \Pr[|T_{n-1}| > t_{\alpha/2,n-1}] = \alpha$

For any r, the value of α' can never exceed α. For example, for α = 0.05, F_0 = 1, r = – 0.25, the value of α' ranges in the interval 0.01369406 (at n = 101) ≤ α' ≤ 0.0321416 (at n = 3). For α = 0.05, F_0 = 1.5, r = – 0.25, it ranges in the interval 0.013975172 (at n = 101) ≤ α' ≤ 0.032328946 (at n = 3); at F_0 = 2.0, 0.014512 (at n = 101) ≤ α' ≤ 0.032681748 (at n =3). As $F_0 \to \infty$, $\alpha' \to \alpha$.

For α = 0.05, F_0 =1, r = – 0.5, the value of α' ranges in the interval 0.02406825415 (at n = 101) ≤ α' ≤ 0.03820582 (at n = 3). For α = 0.05, F_0 =1.5, r = –0.5, it ranges in the interval 0.02430291 (at n = 101) ≤ α' ≤ 0.03832841 (at n = 3); at F_0 = 2.0, 0.0247473226 (at n = 101) ≤ α' ≤ 0.038559305 (at n =3). As $F_0 \to \infty$, $\alpha' \to \alpha$.

For α = 0.05, F_0 =1, r = – 0.75, the value of α' ranges in the interval 0.0363984432 (at n = 101) ≤ α' ≤ 0.0441575 (at n = 3). For α = 0.05, F_0 =1.5, r = –0.75, it ranges in the interval 0.03653187 (at n = 101) ≤ α' ≤ 0.04421765 (at n = 3); at F_0 = 2.0, 0.036783673 (at n = 101) ≤ α' ≤ 0.04433101 (at n =3). As $F_0 \to \infty$, $\alpha' \to \alpha$.

Moreover, the effect of negative correlation is to increase α' toward α as n –1 goes toward 1. For example, when n – 1 =1, F_0 = 2 and r = – 0.90, then α' = 0.0487767; at n – 1 =1, F_0 = 2 and r = – 0.95, then α' = 0.0493921346 while at r =

$-$ 0.99 and $F_0 = 2$, $\alpha' = 0.04987903$. For fixed F_0 and $-1 \le r < 0$, the limiting behavior of α' as $n \to \infty$ is difficult to investigate because as $n \to \infty$, then per force $F_0 = S_x^2 / S_y^2 \to \sigma_x^2 / \sigma_y^2$ (an unknown parameter), and $r \to \rho_{xy}$ (the population correlation coefficient which is another unknown parameter). Most importantly, Matlab loses accuracy in inverting $F_{1,n-1}$ once $n-1$ far exceeds 1,000,000. For example, Matlab gave α'(at n = 1,000,000, F_0 =1, r = –0.50) = 0.0236254, α'(at n = 10,000,000, F_0 =1, r = –0.50) = 0.0237475, but α'(at n = 100,000,000, F_0 =1, r = -0.50) = 0.0938949. We are fairly certain that α'(as $n \to \infty$, F_0 = 1, r = –0.50) = 0.0236254 is the correct answer and the last one is inaccurate because Matlab gave finv(0.95,1,100000000) = 2.10472249984741 instead of the correct value of 3.841458914. For negative correlation we have verified that as F_0 and $n \to \infty, \alpha' \to \alpha$.

For $\alpha = 0.05$, $F_0 = 1.00$, and r = 0.25, the value of α' ranges in the interval 0.0016243942 (at n =101) $\le \alpha' \le$ 0.01966083 (at n = 3). For $\alpha = 0.05$, $F_0 = 1.5$, and r = 0.25, the value of α' ranges in the interval 0.0017703 (at n = 101) $\le \alpha' \le$ 0.0199853 (at n = 3). While, For $\alpha = 0.05$, $F_0 = 2.00$, and r = 0.25, the value of α' ranges in the interval 0.00206727 (at n = 101) $\le \alpha' \le$ 0.02059583 (at n = 3).

For $\alpha = 0.05$, F_0 =1, r = 0.50, the value of α' lies in the interval 0.000136523 (at n = 101) $\le \alpha' \le$ 0.0132366264 (at n = 3); at F_0 =1.5, α' lies in the interval 0.000169103 (at n = 101) $\le \alpha' \le$ 0.013633621 (at n = 3); while at F_0 = 2.00, α' lies in the interval 0.0002456622 (at n = 101) $\le \alpha' \le$ 0.01438047707496 (at n = 3).

For $\alpha = 0.05$, F_0 =1, r = 0.75, the value of α' lies in the interval 0.0000001795166(at n = 101) $\le \alpha' \le$ 0.00668445233 (at n = 3);at F_0 =1.5, α' lies in the interval 0.00000041140283 (at n = 101) $\leq \alpha' \leq$ 0.00715685 (at n = 3); while at F_0 = 2.00, α' lies in the interval 0.000001550553 (at n = 101) $\leq \alpha' \leq$ 0.0080453 (at n = 3).

The impact of positive correlation is to reduce α' toward zero as $n \rightarrow \infty$. For example, at α = 0.05, F_0 = 1.5, r = 0.75 and n = 500, α' reduces to 0.00000012177 from its value of 0.00000041140283 at n = 101, and at n = 101, r = 0.90, α' reduces to $1.25244259407 \times 10^{-12}$.

Because, we have coded Matlab functions (see Appendix B) to compute the α' values for all three above cases (pooled t-test, two-independent t-test, and the paired t-test), no extra tables are provided.

7.0 The Percent Overlap that Leads to the Rejection of $H_0:\mu_x = \mu_y$

7.1 The case of Unknown $\sigma_x = \sigma_y = \sigma$

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.

As before, let O represent the amount of overlap length between the two individual CIs on process means. Then O will be 0 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 < α. Thus, the overlap amount O is larger than 0 when $U(\mu_x) > U(\mu_y) > L(\mu_x)$ or $U(\mu_y) > U(\mu_x) > L(\mu_y)$. In these two cases, both $U(\mu_x) > U(\mu_y) > L(\mu_x)$ and $U(\mu_y) > U(\mu_x) > L(\mu_y)$ will lead to the same result. Therefore, only $U(\mu_x) > U(\mu_y) > L(\mu_x)$ is discussed here so that we are making the assumption that $\bar{x} - \bar{y} \geq 0 \rightarrow$

$$O = U(\mu_y) - L(\mu_x) = (\,\bar{y} + t_{\alpha/2,\nu_y} S_y / \sqrt{n_y}\,) - (\,\bar{x} - t_{\alpha/2,\nu_x} S_x / \sqrt{n_x}\,)$$

$$= (\,t_{\alpha/2,\nu_x} S_x / \sqrt{n_x} + t_{\alpha/2,\nu_y} S_y / \sqrt{n_y}\,) - (\,\bar{x} - \bar{y}\,) \qquad (45a)$$

Further, the span of the two individual CIs is

$$U(\mu_x) - L(\mu_y) = (\,\bar{x} + t_{\alpha/2,\nu_x} \times S_x / \sqrt{n_x}\,) - (\,\bar{y} - t_{\alpha/2,\nu_y} \times S_y / \sqrt{n_y}\,)$$

$$= (\,t_{\alpha/2,\nu_x} S_x / \sqrt{n_x} + t_{\alpha/2,\nu_y} S_y / \sqrt{n_y}\,) + (\,\bar{x} - \bar{y}\,) \qquad (45b)$$

From equations (45a &b) the % overlap is given by

$$\omega = \frac{(t_{\alpha/2,\nu_x} S_x / \sqrt{n_x} + t_{\alpha/2,\nu_y} S_y / \sqrt{n_y}) - (\bar{x} - \bar{y})}{(t_{\alpha/2,\nu_x} S_x / \sqrt{n_x} + t_{\alpha/2,\nu_y} S_y / \sqrt{n_y}) + (\bar{x} - \bar{y})} \times 100\% \qquad (45c)$$

As $|\bar{x} - \bar{y}|$ increases, the *P-value* of the test decreases (i.e., H_0: $\mu_x = \mu_y$ must be rejected more strongly) and ω in Eq. (45c) decreases. Because H_0: $\mu_x = \mu_y$ must be rejected at the α- level if $|\bar{x} - \bar{y}| \geq t_{\alpha/2,\nu} \times S_p \sqrt{1/n_x + 1/n_y}$, where $\nu = n_x + n_y - 2$, then from (45c) H_0 must be barely rejected at α×100% level or less iff

$$\omega \leq \frac{(t_{\alpha/2,\nu_x} S_x / \sqrt{n_x} + t_{\alpha/2,\nu_y} S_y / \sqrt{n_y}) - (t_{\alpha/2,\nu} \times S_p \sqrt{1/n_x + 1/n_y})}{(t_{\alpha/2,\nu_x} S_x / \sqrt{n_x} + t_{\alpha/2,\nu_y} S_y / \sqrt{n_y}) + (t_{\alpha/2,\nu} \times S_p \sqrt{1/n_x + 1/n_y})} \times 100\% \qquad (46a)$$

Putting $R_n = n_y / n_x$ into Eq. (46a) leads to the result →

$$\omega \leq \frac{(t_{\alpha/2,\nu_x} S_x + t_{\alpha/2,\nu_y} S_y / \sqrt{R_n}) - (t_{\alpha/2,\nu} \times S_p \sqrt{1 + 1/R_n})}{(t_{\alpha/2,\nu_x} S_x + t_{\alpha/2,\nu_y} S_y / \sqrt{R_n}) + (t_{\alpha/2,\nu} \times S_p \sqrt{1 + 1/R_n})} \times 100\%$$

$$\rightarrow \omega \leq \frac{(t_{\alpha/2,\nu_x} S_x \sqrt{R_n} + t_{\alpha/2,\nu_y} S_y) - (t_{\alpha/2,\nu} \times S_p \sqrt{R_n + 1})}{(t_{\alpha/2,\nu_x} S_x \sqrt{R_n} + t_{\alpha/2,\nu_y} S_y) + (t_{\alpha/2,\nu} \times S_p \sqrt{R_n + 1})} \times 100\% \qquad (46b)$$

As defined before, letting $F_0 = S_x^2 / S_y^2$ into Eq. (46b) and recalling $\nu = n_x + n_y - 2$ results in

$$\omega \leq \frac{\sqrt{F_0 R_n} \times t_{\alpha/2,\nu_x} + t_{\alpha/2,\nu_y} - t_{\alpha/2,\nu} \sqrt{(1+R_n)(\nu_x F_0 + \nu_y)/\nu}}{\sqrt{F_0 R_n} \times t_{\alpha/2,\nu_x} + t_{\alpha/2,\nu_y} + t_{\alpha/2,\nu} \sqrt{(1+R_n)(\nu_x F_0 + \nu_y)/\nu}} \times 100\% \qquad (46c)$$

where $F_0 \times R_n = (S_x^2 / S_y^2) \times \frac{n_y}{n_x} = \frac{S_x^2 / n_x}{S_y^2 / n_y} = \frac{V(\bar{x})}{V(\bar{y})}$ = (*se ratio*)^2. Thus, the percent overlap at which H_0 should be rejected exactly at the α-level is given by

$$\omega_r = \frac{\sqrt{F_0 R_n} \times t_{\alpha/2,\nu_x} + t_{\alpha/2,\nu_y} - t_{\alpha/2,\nu}\sqrt{(1+R_n)(\nu_x F_0 + \nu_y)/\nu}}{\sqrt{F_0 R_n} \times t_{\alpha/2,\nu_x} + t_{\alpha/2,\nu_y} + t_{\alpha/2,\nu}\sqrt{(1+R_n)(\nu_x F_0 + \nu_y)/\nu}} \times 100\% \quad (46d)$$

$$= \frac{k \times t_{\alpha/2,\nu_x} + t_{\alpha/2,\nu_y} - t_{\alpha/2,\nu}\sqrt{(1+R_n)(\nu_x F_0 + \nu_y)/\nu}}{k \times t_{\alpha/2,\nu_x} + t_{\alpha/2,\nu_y} + t_{\alpha/2,\nu}\sqrt{(1+R_n)(\nu_x F_0 + \nu_y)/\nu}} \times 100\% \quad (46e)$$

Eq. (46d) shows that the % overlap at which H_0 must be rejected at the α-level depends only on α, n_x, n_y and F_0 and not on the specific values of S_x and S_y. For larger values of n_x and $n_y > 30$, the dependency on α is negligible because $t_{\alpha/2,\nu_x}$, $t_{\alpha/2,\nu_y}$ and $t_{\alpha/2,\nu}$ are close in values and are almost equal once n_x and $n_y > 60$.

For the case of balanced completely randomized design (i.e., $n = n_x = n_y \rightarrow R_n=1$), the inequality in (46d) reduces to

$$\omega_r = \frac{t_{\alpha/2,n-1}(1+\sqrt{F_0}) - t_{\alpha/2,2(n-1)}\sqrt{1+F_0}}{t_{\alpha/2,n-1}(1+\sqrt{F_0}) + t_{\alpha/2,2(n-1)}\sqrt{1+F_0}} \times 100\% \quad (46e)$$

We first discuss the limiting property of the Eq. (46e). Because this is the case of pooled t-test, $F_0 = (S_x/S_y)^2$ (the ratio of the two sample variances) must lie within the acceptance interval ($F_{0.90,n-1,n-1}, F_{0.10,n-1,n-1}$); otherwise H_0: $\sigma_x = \sigma_y$ must be rejected at the 20% level. Further, the first derivative of ω_r vanishes at $F_0 = 1$ so that the maximum of ω_r occurs at $F_0 = 1$ and is equal to 0.613625686 at n = 2 and its maximum approaches $(2-\sqrt{2})/(2+\sqrt{2}) = 0.171573$ as $n \rightarrow \infty$. Note that as $n \rightarrow \infty$, the value of F_0 that must lie within ($F_{0.90,n-1,n-1} \le F_0 \le F_{0.10,n-1,n-1}$) must per force also go towards 1 because the limiting value of both $F_{0.90,n-1,n-1}$ and $F_{0.10,n-1,n-1}$ is nearly 1. That is, for all F_0 values where H_0: $\sigma_x = \sigma_y$ cannot be rejected, the limiting value of ω_r in terms of n cannot be less than 0.171573. At n = 61, ω_r = 0.17603 if F_0 = 1.2 so that the approach of ω_r toward 0.171573 occurs fairly rapidly in terms of n as long as $F_{0.90,n-1,n-1} \le F_0 \le F_{0.10,n-1,n-1}$. At n = 31 and F_0 =1.30, the value of ω_r = 0.1806 so that H_0: $\mu_x = \mu_y$ must be rejected at 5% or less if the % overlap is less than or equal to 18.06%.

In the unbalanced case if R_n = 0.50 or 2, the limiting value of ω_r at F_0 =1 is equal to $(\sqrt{2}+1-\sqrt{3})/(\sqrt{2}+1+\sqrt{3})$ = 0.164525. Further, as $R_n \neq 1$ deviates farther from 1, the limiting value of ω_r decreases for a fixed F_0 as long as $F_{0.90,n-1,n-1} \le F_0 \le F_{0.10,n-1,n-1}$. For example, at R_n = 3 (or 1/3), the limiting value of ω_r is 0.15470; at R_n = 4 (or 0.25), its limiting value is 0.14590; at R_n = 5 (or 0.20), its limiting value is 0.13835; at R_n = 0.10 (or 10), the limiting value of ω_r is 0.11307, while at Rn = 20 (or 0.05) the limiting value of ω_r is equal to 0.088472. Clearly, as R_n deviates farther from 1, the limiting value of ω_r decreases, implying that the Overlap approaches an α-level test. See the illustration in Table 22. Finally, it must be noted that as n_x and n_y become very large, the limit of Eq. (46d) becomes identical to $\omega_r = \dfrac{(1+k-\sqrt{1+k^2})}{(1+k+\sqrt{1+k^2})}100\%$ given in Eq. (12e).

Next, what should each individual confidence level $1-\gamma$ be so that the two independent CIs lead to the exact $\alpha\times$100%-level test on H_0: $\mu_x = \mu_y$. The expressions for the two $1-\gamma$ independent CIs are given by

$$\bar{x} - t_{\gamma/2,\nu_x} \times S_x / \sqrt{n_x} \le \mu_x \le \bar{x} + t_{\gamma/2,\nu_x} \times S_x / \sqrt{n_x} \qquad (47a)$$

$$\bar{y} - t_{\gamma/2,\nu_y} \times S_y / \sqrt{n_y} \le \mu_y \le \bar{y} + t_{\gamma/2,\nu_y} \times S_y / \sqrt{n_y} \qquad (47b)$$

It is clear that H_0: $\mu_x = \mu_y$ must be rejected at the α-level iff the amount of overlap

Table 22. The Value of ω_r for Different F_0 and R_n Combinations

F_0	ν_x	ν_y	R_n	ω_r	F_0	ν_x	ν_y	R_n	ω_r
0.5	1	1	1	60.90835	0.8	11	5	0.5	24.33944
0.5	11	11	1	19.33138	0.8	21	10	0.5	20.81321
0.5	21	21	1	17.90985	0.8	51	25	0.5	18.92389
0.5	31	31	1	17.42757	0.8	61	30	0.5	18.72412
0.5	41	41	1	17.18504	0.8	1001	500	0.5	17.81175
0.5	51	51	1	17.03911	0.8	11	23	2.0	17.54347
0.5	100	100	1	16.74930	0.8	31	63	2.0	15.87565
0.5	150	150	1	16.64982	0.8	51	103	2.0	15.53555
0.5	200	200	1	16.60028	0.8	1001	2003	2.0	15.04753
0.8	1	1	1	61.31421	1	11	5	0.5	22.76531
0.8	11	11	1	19.95417	1	21	10	0.5	19.37801
0.8	21	21	1	18.53612	1	41	20	0.5	17.85949
0.8	31	31	1	18.05497	1	81	40	0.5	17.14235
0.8	41	41	1	17.81299	1	151	75	0.5	16.81705
0.8	51	51	1	17.66738	1	501	250	0.5	16.56104
0.8	100	100	1	17.37823	1	1001	500	0.5	16.50667
0.8	500	500	1	17.14089	1	10001	5000	0.5	16.45788
0.8	1000	1000	1	17.11144	1	500	1001	2.0	16.50667
1	1	1	1	61.36257	1	2	5	2.0	35.76007
1	10	10	1	20.33789	1	10	21	2.0	19.37801
1	50	50	1	17.75437	1	20	41	2.0	17.85949
1	100	100	1	17.45340	1	50	101	2.0	17.00221
1	10000	10000	1	17.16022	1	100	201	2.0	16.72519
1	500000	500000	1	17.15735	1	10000	20001	2.0	16.45517
1	10000000	10000000	1	17.15729	1	10000000	20000001	2.0	16.45247
1.2	1	1	1	61.33025	1	100	302	3.0	15.77155
1.2	10	10	1	20.28821	1	10000	30002	3.0	15.47304
1.2	20	20	1	18.63655	1	1000000	3000002	3.0	15.47008
1.2	60	60	1	17.60330	1	100	403	4.0	14.91845
1.3	10	10	1	20.23527	1	10000	40003	4.0	14.59306
1.3	20	20	1	18.58326	1	1000000	4000003	4.0	14.58984
1.3	30	30	1	18.05974	1	10000	50004	5.0	13.83815
1.6	5	5	1	23.67919	1	1000000	5000004	5.0	13.83471
1.6	10	10	1	20.01202	1	10000	100009	10.0	11.31132
1.6	13	13	1	19.23368	1	10000000	100000009	10.0	11.30718

between (47a) and (47b) barely becomes zero or less. Without loss of generality, the x-sample will be denoted such that $\bar{x} - \bar{y} \geq 0$. Therefore, we deduce from (47a & b) that

$$U'(\mu_y) - L'(\mu_x) = (\,\bar{y} + t_{\gamma/2,\nu_y} \times S_y / \sqrt{n_y}\,) - (\,\bar{x} - t_{\gamma/2,\nu_x} \times S_x / \sqrt{n_x}\,)$$

$$= t_{\gamma/2,\nu_y} \times S_y / \sqrt{n_y} \; + t_{\gamma/2,\nu_x} \times S_x / \sqrt{n_x} - (\,\bar{x} - \bar{y}\,) \qquad (48a)$$

Because H_0: $\mu_x = \mu_y$ must be rejected at the α-level as soon as the RHS of Eq. (48a) becomes 0 or smaller, we impose the borderline rejection criterion $|\bar{x} - \bar{y}| = t_{\alpha/2,\nu} \times S_p\sqrt{1/n_x + 1/n_y}$ into Eq. (48a). In short, we are rejecting H_0:$\mu_x = \mu_y$ as soon as the two independent CIs in (47a) and (47b) become disjoint. This leads to rejecting H_0:$\mu_x = \mu_y$ iff

$$t_{\gamma/2,\nu_y} \times S_y / \sqrt{n_y} \; + \; t_{\gamma/2,\nu_x} \times S_x / \sqrt{n_x} - (\,\bar{x} - \bar{y}\,) \le 0. \qquad (48b)$$

At the borderline value, we set $\bar{x} - \bar{y} = t_{\alpha/2,\nu} \times S_p\sqrt{1/n_x + 1/n_y}$ and set the LHS of inequality (48b) to 0 in order to solve for γ.

$$\rightarrow t_{\gamma/2,\nu_y} \times S_y / \sqrt{n_y} \; + \; t_{\gamma/2,\nu_x} \times S_x / \sqrt{n_x} - t_{\alpha/2,\nu} \times S_p\sqrt{1/n_x + 1/n_y} = 0$$

$$t_{\gamma/2,\nu_y} \times S_y \; + \; t_{\gamma/2,\nu_x} \times S_x \sqrt{R_n} - t_{\alpha/2,\nu} \times S_p\sqrt{R_n + 1} = 0.$$

$$t_{\gamma/2,\nu_y} \; + \; t_{\gamma/2,\nu_x} \times \sqrt{F_0 R_n} - t_{\alpha/2,\nu} \times \sqrt{(R_n+1)(\nu_x F_0 + \nu_y)/\nu} = 0.$$

$$\rightarrow t_{\gamma/2,\nu_y} + t_{\gamma/2,\nu_x} \times \sqrt{F_0 R_n} = t_{\alpha/2,\nu} \times \sqrt{(R_n+1)(\nu_x F_0 + \nu_y)/\nu}$$

$$\text{or } t_{\gamma/2,\nu_y} + k \times t_{\gamma/2,\nu_x} = t_{\alpha/2,\nu} \times \sqrt{(R_n+1)(\nu_x F_0 + \nu_y)/\nu} \qquad (49a)$$

where $F_0 = S_x^2 / S_y^2$, $R_n = n_y/n_x$, $\nu = n_x + n_y - 2$ and $k = \sqrt{F_0 R_n}$ = se ratio of samples. Eq. (49a) clearly shows that the value of γ depends on the LOS α of testing H_0: $\mu_x = \mu_y$, also on $F_0 = S_x^2 / S_y^2$, and the sample sizes n_x, n_y. For the case of balanced design ($n_x = n_y = n$), (49a) reduces to

$$t_{\gamma/2,n-1} \times S_y \; + \; t_{\gamma/2,n-1} \times S_x - t_{\alpha/2,2(n-1)} \times S_p\sqrt{2} = 0$$

$$\rightarrow \; t_{\gamma/2,n-1}(S_x + S_y) \; - \; t_{\alpha/2,2(n-1)} \times \sqrt{S_x^2 + S_y^2} \; = 0$$

$$\rightarrow \; t_{\gamma/2,n-1} = t_{\alpha/2,2(n-1)} \times \sqrt{S_x^2 + S_y^2} \; / \, (S_x + S_y)$$

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

$$\rightarrow \; F_{\gamma,1,n-1} = F_{\alpha,1,2(n-1)} \times (1 + F_0)/(1 + \sqrt{F_0}\,)^2 \qquad (49b)$$

For example, when $\alpha = 0.05$, n_x & $n_y = 21$, Eq. (49b) gives $\gamma = 0.16807$ so that the two independent CIs have to be set at the confidence level $1-\gamma = 0.83193$ in order for the Overlap to provide an exact 5% level test. The values of $1-\gamma$ range from 0.2020062 at $n-1 = 1$ down to 0.16596 at $n-1 = 100$. In order to obtain the limiting value of γ, we let $n \rightarrow \infty$ in (49b) resulting in Lim $t_{\gamma/2,n-1}$($n \rightarrow \infty$) = $1.96 \times \sqrt{1+1} / (\sqrt{1} + 1) = 1.96/\sqrt{2}$ = 1.38593 → Limit γ (as $n \rightarrow \infty$) = $\Pr(|Z| \geq 1.38593) = 0.16578$, which is identical to the know-&-equal-variances case from Eq. (13) at K =1.

7.2 The Case of H_0: $\sigma_x = \sigma_y$ Rejected Leading to the Two-Independent Sample t-Test

Assuming that X~N(μ_x, σ_x^2) and Y~N(μ_y, σ_y^2) and $\overline{X} - \overline{Y}$ is N($\mu_x - \mu_y$, $\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}$), where now the null hypothesis of H_0: $\sigma_x = \sigma_y$ is rejected at the 20% level (i.e., the *P-value* of the pre-test is less than 20%) leading to the assumption that the F-statistic $F_0 = S_x^2 / S_y^2$ is outside the interval ($F_{0.90,n_x-1,n_y-1}$, $F_{0.10,n_x-1,n_y-1}$), where without loss of generality the sample with the larger mean will be called X. It has been shown in statistical theory that if the assumption $\sigma_x = \sigma_y$ is not tenable, the statistic

$[(\bar{x}-\bar{y})-(\mu_x-\mu_y)]/\sqrt{(S_x^2/n_x)+(S_y^2/n_y)}$ has approximately the Student's t-distribution

with degrees of freedom given by Eq. (39).

$$\nu = \frac{(S_x^2/n_x + S_y^2/n_y)^2}{\frac{(S_x^2/n_x)^2}{n_x-1}+\frac{(S_y^2/n_y)^2}{n_y-1}} = \frac{[V(\bar{x})+V(\bar{y})]^2}{\frac{(V(\bar{x}))^2}{\nu_x}+\frac{(V(\bar{y}))^2}{\nu_y}} = \frac{\nu_x\nu_y[V(\bar{x})+V(\bar{y})]^2}{\nu_y(V(\bar{x}))^2+\nu_x(V(\bar{y}))^2} \tag{39a}$$

$$\rightarrow \nu = \frac{\nu_x\nu_y(F_0R_n+1)^2}{\nu_y(F_0R_n)^2+\nu_x} = \frac{\nu_x\nu_y(k^2+1)^2}{\nu_y k^4+\nu_x} \tag{39b}$$

As before, the amount of overlap between the two individual CIs is given by

$$O = U(\mu_y) - L(\mu_x) = (\bar{y}+t_{\alpha/2,\nu_y}S_y/\sqrt{n_y}) - (\bar{x}-t_{\alpha/2,\nu_x}S_x/\sqrt{n_x})$$

$$= (t_{\alpha/2,\nu_x}S_x/\sqrt{n_x}+t_{\alpha/2,\nu_y}S_y/\sqrt{n_y}) - (\bar{x}-\bar{y}) \tag{50a}$$

Further, the span of the two individual CIs is

$$U(\mu_x) - L(\mu_y) = (\bar{x}+t_{\alpha/2,\nu_x}\times S_x/\sqrt{n_x}) - (\bar{y}-t_{\alpha/2,\nu_y}\times S_y/\sqrt{n_y})$$

$$= (t_{\alpha/2,\nu_x}S_x/\sqrt{n_x}+t_{\alpha/2,\nu_y}S_y/\sqrt{n_y}) + (\bar{x}-\bar{y}) \tag{50b}$$

From equations (50 a &b) the % overlap is given by

$$\omega = \frac{(t_{\alpha/2,\nu_x}S_x/\sqrt{n_x}+t_{\alpha/2,\nu_y}S_y/\sqrt{n_y})-(\bar{x}-\bar{y})}{(t_{\alpha/2,\nu_x}S_x/\sqrt{n_x}+t_{\alpha/2,\nu_y}S_y/\sqrt{n_y})+(\bar{x}-\bar{y})}\times 100\% \tag{50c}$$

As $|\bar{x}-\bar{y}|$ increases, the *P-value* of the test decreases (i.e., H_0: $\mu_x = \mu_y$ must rejected more strongly) and ω in Eq. (50c) decreases. Because H_0: $\mu_x = \mu_y$ must be barely rejected at the α- level if $|\bar{x}-\bar{y}| = t_{\alpha/2,\nu}\times\sqrt{S_x^2/n_x+S_y^2/n_y}$, where ν is given in Eq. (39), then from (50c)

$$\omega_r = \frac{(t_{\alpha/2,\nu_x} S_x / \sqrt{n_x} + t_{\alpha/2,\nu_y} S_y / \sqrt{n_y}) - t_{\alpha/2,\nu} \times \sqrt{S_x^2 / n_x + S_y^2 / n_y}}{(t_{\alpha/2,\nu_x} S_x / \sqrt{n_x} + t_{\alpha/2,\nu_y} S_y / \sqrt{n_y}) + t_{\alpha/2,\nu} \times \sqrt{S_x^2 / n_x + S_y^2 / n_y}} \times 100\% \qquad (51a)$$

In order to simplify (51a), we multiply throughout by $\sqrt{n_y}$, divide throughout by S_y and replace n_y/n_x by R_n and S_x^2 / S_y^2 by F_0 resulting in

$$\omega_r = \frac{(t_{\alpha/2,\nu_x} \sqrt{R_n F_0} + t_{\alpha/2,\nu_y}) - t_{\alpha/2,\nu} \times \sqrt{R_n F_0 + 1}}{(t_{\alpha/2,\nu_x} \sqrt{R_n F_0} + t_{\alpha/2,\nu_y}) + t_{\alpha/2,\nu} \times \sqrt{R_n F_0 + 1}} \times 100\%$$

$$= \frac{(k\, t_{\alpha/2,\nu_x} + t_{\alpha/2,\nu_y}) - t_{\alpha/2,\nu} \times \sqrt{k^2 + 1}}{(k\, t_{\alpha/2,\nu_x} + t_{\alpha/2,\nu_y}) + t_{\alpha/2,\nu} \times \sqrt{k^2 + 1}} \times 100\% \qquad (51b)$$

where F_0 lies outside the 20% acceptance interval ($F_{0.90,n_x-1,n_y-1}$, $F_{0.10,n_x-1,n_y-1}$). The % overlap in Eq. (51b) changes very little as α changes, increasing a bit as α decreases while other parameters n_x, n_y and F_0 are kept fixed. As F_0 increases, the value of ω_r decreases such that as $F_0 \to \infty$, $\omega_r \to 0$ so that the overlap becomes an exact α-level test. The limiting (in terms of n_x and n_y) values of ω_r at R_n = 2, 3, 4, 5, 10 and 20 are independent of α (because for large n_x and n_y all 3 t inverse functions in (51b) are almost equal) and are almost identical to those of the pooled t-test, namely 0.164509, 0.154679, 0.1458744, 0.138322, 0.11305, and 0.08845, respectively.

When the design is balanced ($n_x = n_y = n$), the % overlap in Eq.(51b) that still leads to the rejection of H_0: $\mu_x = \mu_y$ at the α-level reduces to

$$\omega_r = \frac{t_{\alpha/2,n-1}(\sqrt{F_0} + 1) - t_{\alpha/2,\nu} \times \sqrt{F_0 + 1}}{t_{\alpha/2,n-1}(\sqrt{F_0} + 1) + t_{\alpha/2,\nu} \times \sqrt{F_0 + 1}} \times 100\% \qquad (51c)$$

where in the balanced case $\nu = \frac{(n-1)(S_x^2+S_y^2)^2}{S_x^4+S_y^4} = \frac{(n-1)(F_0+1)^2}{F_0^2+1}$. If the % overlap exceeds Eq.(51c), then H_0: $\mu_x = \mu_y$ can no longer be rejected at the $\alpha\times100\%$ level of significance. For values of F_0 outside the range ($F_{0.90,n_x-1,n_y-1}$, $F_{0.10,n_x-1,n_y-1}$), the limiting value of ω_r (at any α) as $F_0 \rightarrow 1$ from Eq. (51c) is, as before, equal to $(2-\sqrt{2})/(2+\sqrt{2})$ = 0.171573. Again, as $F_0 \rightarrow \infty$, $\omega_r \rightarrow 0$, which is consistent with the results in Chapter 3 with known but unequal sample case of the *SE* ratio $k \rightarrow \infty$.

Now, what should each individual confidence level $1-\gamma$ be so that the two independent CIs lead to the exact $\alpha\times100\%$-level test on H_0: $\mu_x = \mu_y$. As before, the expressions for the two $1-\gamma$ independent CIs are given by

$$\bar{x} - t_{\gamma/2,\nu_x} \times S_x / \sqrt{n_x} \le \mu_x \le \bar{x} + t_{\gamma/2,\nu_x} \times S_x / \sqrt{n_x} \quad (52a)$$

$$\bar{y} - t_{\gamma/2,\nu_y} \times S_y / \sqrt{n_y} \le \mu_y \le \bar{y} + t_{\gamma/2,\nu_y} \times S_y / \sqrt{n_y} \quad (52b)$$

It is clear that H_0: $\mu_x = \mu_y$ must be rejected at the α-level iff the amount of overlap between (52a) and (52b) barely becomes zero or less. Without loss of generality, the x-sample will be denoted such that $\bar{x} - \bar{y} \ge 0$. Therefore, we deduce from (52a &b) that

$$U'(\mu_y) - L'(\mu_x) = (\bar{y} + t_{\gamma/2,\nu_y} \times S_y / \sqrt{n_y}) - (\bar{x} - t_{\gamma/2,\nu_x} \times S_x / \sqrt{n_x})$$

$$= t_{\gamma/2,\nu_y} \times S_y / \sqrt{n_y} + t_{\gamma/2,\nu_x} \times S_x / \sqrt{n_x} - (\bar{x} - \bar{y}) \quad (53a)$$

Because H_0: $\mu_x = \mu_y$ must be rejected at the α-level as soon as the RHS of (53a) becomes 0 or smaller, we impose the critical limit of rejection $|\bar{x} - \bar{y}| = t_{\alpha/2,\nu} \times \sqrt{S_x^2/n_x + S_y^2/n_y}$ into Eq. (53a), where ν is given in Eq. (39). In short, we are rejecting

$H_0{:}\mu_x = \mu_y$ as soon as the two independent CIs in Eq.(52a) and Eq.(52b) become disjoint. This leads to rejecting $H_0{:}\mu_x = \mu_y$ iff

$$t_{\gamma/2,\nu_y} \times S_y / \sqrt{n_y} + t_{\gamma/2,\nu_x} \times S_x / \sqrt{n_x} - (\bar{x} - \bar{y}) \le 0. \quad (53b)$$

At the borderline value, we set $\bar{x} - \bar{y} = t_{\alpha/2,\nu} \times \sqrt{S_x^2 / n_x + S_y^2 / n_y}$ and set the LHS of inequality (53b) to 0 in order to solve for γ.

$$\rightarrow t_{\gamma/2,\nu_y} \times S_y / \sqrt{n_y} + t_{\gamma/2,\nu_x} \times S_x / \sqrt{n_x} - t_{\alpha/2,\nu} \times \sqrt{S_x^2 / n_x + S_y^2 / n_y} = 0.$$

$$\rightarrow t_{\gamma/2,\nu_y} \times S_y + t_{\gamma/2,\nu_x} \times S_x \sqrt{R_n} - t_{\alpha/2,\nu} \times \sqrt{S_x^2 R_n + S_y^2} = 0.$$

$$\rightarrow t_{\gamma/2,\nu_y} + t_{\gamma/2,\nu_x} \times \sqrt{F_0 R_n} - t_{\alpha/2,\nu} \times \sqrt{F_0 R_n + 1} = 0.$$

Or: $t_{\gamma/2,\nu_y} + t_{\gamma/2,\nu_x} \times \sqrt{F_0 R_n} = t_{\alpha/2,\nu} \times \sqrt{F_0 R_n + 1}$

$$\rightarrow t_{\gamma/2,\nu_y} + k \times t_{\gamma/2,\nu_x} = t_{\alpha/2,\nu} \times \sqrt{k^2 + 1} \quad (54a)$$

where $F_0 = S_x^2 / S_y^2$, $R_n = n_y/n_x$ and ν is given in Eq. (39). Eq. (54a) clearly shows that the value of γ depends on the LOS α of testing $H_0{:}\mu_x = \mu_y$, F_0, and the sample sizes n_x and n_y. For the case of balanced design ($n_x = n_y = n$), (54a) reduces to

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

$$\text{or} \quad F_{\gamma,1,n-1} = F_{\alpha,1,\nu} \times (1 + F_0)/(1 + \sqrt{F_0})^2 \quad (54b)$$

where $\nu = \dfrac{(n-1)(F_0+1)^2}{F_0^2 + 1}$. The limiting value of γ in Eq. (54b), as $n \rightarrow \infty$, can easily be obtained from $Z_{\gamma/2} = Z_{\alpha/2} \times \sqrt{F_0 + 1} / (1 + \sqrt{F_0})$. The results will be the same for Eq. (54a). For example, using (54b) at $\alpha = 0.05$, $n = 10$, $F_0 = 4.0$, $\nu = 13.2353$ resulting in $\gamma =$

0.1424. Payton *et al.* (2000) report this value as 0.1262 because the denominator df of $F_{\gamma,1,\nu}$ in the formula atop their page 550 is inaccurate. For n_x & n_y > 100, as $F_0 \to \infty$, $\gamma \to \alpha$ so that the Overlap approaches an α-level test. See the illustration in Table 23.

7.3 Comparing the Paired t-CI with Two Independent t-CIs

As before, let O represent the amount of overlap length between the two individual CIs. Then, O will be 0 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 < α. Thus, O is larger than 0 when $U(\mu_x) > U(\mu_y) > L(\mu_x)$ or $U(\mu_y) > U(\mu_x) > L(\mu_y)$. In these two cases, both $U(\mu_x) > U(\mu_y) > L(\mu_x)$ and $U(\mu_y) > U(\mu_x) > L(\mu_y)$ will lead to the same result. Therefore, only $U(\mu_x) > U(\mu_y) > L(\mu_x)$ is discussed here so that we are making the assumption that $\bar{x} - \bar{y} \ge 0$.

$$. \quad O = U(\mu_y) - L(\mu_x) = (\bar{y} + t_{\alpha/2,n-1} S_y / \sqrt{n}) - (\bar{x} - t_{\alpha/2,n-1} S_x / \sqrt{n})$$

$$= (t_{\alpha/2,n-1} S_x / \sqrt{n} + t_{\alpha/2,n-1} S_y / \sqrt{n}) - (\bar{x} - \bar{y}) \qquad (55a)$$

Further, the span of the two individual CIs is

$$U(\mu_x) - L(\mu_y) = (\bar{x} + t_{\alpha/2,n-1} \times S_x / \sqrt{n}) - (\bar{y} - t_{\alpha/2,n-1} \times S_y / \sqrt{n})$$

$$= (t_{\alpha/2,n-1} S_x / \sqrt{n} + t_{\alpha/2,n-1} S_y / \sqrt{n}) + (\bar{x} - \bar{y}) \qquad (55b)$$

From equations (55a &b) the % overlap is given by

$$\omega = \frac{(t_{\alpha/2,n-1} S_x / \sqrt{n} + t_{\alpha/2,n-1} S_y / \sqrt{n}) - (\bar{x} - \bar{y})}{(t_{\alpha/2,n-1} S_x / \sqrt{n} + t_{\alpha/2,n-1} S_y / \sqrt{n}) + (\bar{x} - \bar{y})} \times 100\% \qquad (55c)$$

As $|\bar{x} - \bar{y}|$ increases, the *P-value* of the test decreases (i.e., H_0: $\mu_x = \mu_y$ must rejected more strongly) and ω in Eq. (55c) decreases. Because H_0: $\mu_x = \mu_y$ must be rejected at the

Table 23. The γ Value for Different Combinations of n_x, n_y and R_n at Either $F_0 = F_{0.90,\nu_x,\nu_y}$ or $F_0 = F_{0.05,\nu_x,\nu_y}$

n_x	n_y	R_n	$F_0 = F_{0.90,\nu_x,\nu_y}$	γ	$F_0 = F_{0.05,\nu_x,\nu_y}$	γ
5	5	1	0.24347	0.13970	4.10725	0.08485
20	20	1	0.54873	0.16301	1.82240	0.13999
60	60	1	0.71470	0.16510	1.39918	0.15333
500	500	1	0.89152	0.16571	1.12168	0.16206
1000	1000	1	0.92208	0.16574	1.08451	0.16320
100000	100000	1	0.99193	0.16762	1.00814	0.16553
5	6	1.2	0.24688	0.15190	3.52020	0.08565
20	24	1.2	0.55765	0.16613	1.75251	0.13711
60	72	1.2	0.72272	0.16635	1.37397	0.15122
500	600	1.2	0.89554	0.16580	1.11574	0.16098
1000	1200	1.2	0.92509	0.16568	1.08054	0.16231
100000	120000	1.2	0.99227	0.16537	1.00779	0.16505
10	15	1.5	0.42534	0.16860	2.12195	0.11484
20	30	1.5	0.56715	0.16794	1.68491	0.13282
80	120	1.5	0.76388	0.16603	1.29555	0.15011
500	750	1.5	0.89977	0.16466	1.10959	0.15858
1000	1500	1.5	0.92825	0.16437	1.07642	0.16011
100000	150000	1.5	0.99262	0.16371	1.00742	0.16329
5	10	2	0.25409	0.17390	2.69268	0.08157
20	40	2	0.57733	0.16722	1.61932	0.12632
80	160	2	0.77239	0.16349	1.27469	0.14452
500	1000	2	0.90426	0.16127	1.10318	0.15392
1000	2000	2	0.93160	0.16082	1.07212	0.15564
100000	200000	2	0.99300	0.15980	1.00704	0.15928
10	50	5	0.45063	0.15367	1.76252	0.08712
50	250	5	0.73697	0.14463	1.30352	0.11609
500	2500	5	0.91313	0.14027	1.09087	0.13122
1000	5000	5	0.93819	0.13961	1.06381	0.13320
100000	500000	5	0.99373	0.13810	1.00629	0.13746
10	100	10	0.45673	0.13019	1.69556	0.07319
50	500	10	0.74397	0.12427	1.28494	0.09796
500	5000	10	0.91637	0.12052	1.08649	0.11195
1000	10000	10	0.94059	0.11992	1.06084	0.11383
100000	1000000	10	0.99400	0.11851	1.00602	0.11790
100000	2000000	20	0.99413	0.10086	1.00588	0.10032
100000	3000000	30	0.99418	0.09215	1.00583	0.09166
100000	5000000	50	0.99422	0.08297	1.00579	0.08254
100000	10000000	100	0.99424	0.07341	1.00576	0.07305
100000	20000000	200	0.99426	0.06654	1.00575	0.06622
100000	50000000	500	0.99427	0.06042	1.00574	0.06015

α- level if $|\bar{x}-\bar{y}| \geq t_{\alpha/2,n-1} \times S_d / \sqrt{n}$, then from (55c) H_0 must be barely rejected at α×100% or less iff

$$\omega \leq \frac{(t_{\alpha/2,n-1}S_x/\sqrt{n}+t_{\alpha/2,n-1}S_y/\sqrt{n})-t_{\alpha/2,n-1}\times S_d/\sqrt{n}}{(t_{\alpha/2,n-1}S_x/\sqrt{n}+t_{\alpha/2,n-1}S_y/\sqrt{n})+t_{\alpha/2,n-1}\times S_d/\sqrt{n}}\times 100\%$$

$$\rightarrow \ \omega \leq \frac{S_x+S_y-S_d}{S_x+S_y+S_d}\times 100\% \ \rightarrow \ \omega \leq \frac{S_x+S_y-\sqrt{S_x^2+S_y^2-2rS_xS_y}}{S_x+S_y+\sqrt{S_x^2+S_y^2-2rS_xS_y}}\times 100\%$$

$$\rightarrow \ \omega \leq \frac{\sqrt{F_0}+1-\sqrt{F_0+1-2r\sqrt{F_0}}}{\sqrt{F_0}+1+\sqrt{F_0+1-2r\sqrt{F_0}}}\times 100\% \qquad (56a)$$

$$\text{Or} \quad \omega_r = \frac{\sqrt{F_0}+1-\sqrt{F_0+1-2r\sqrt{F_0}}}{\sqrt{F_0}+1+\sqrt{F_0+1-2r\sqrt{F_0}}}\times 100\% \qquad (56b)$$

where $F_0 = S_x^2/S_y^2$, $S_d^2 = S_x^2+S_y^2-2\hat{\sigma}_{xy}$ and $\hat{\sigma}_{xy} = \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})/(n-1)$. Just like the case of known variances, the % overlap ω_r in (56b) depends only on the correlation coefficient r and the ratio of the two sample variances, 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., the two samples are independent) and $F_0 = 1$, then Eq. (56b) reduces to $\frac{2-\sqrt{2}}{2+\sqrt{2}}\times 100\%$ = 17.1573%, which was the % overlap for the case of independent samples and equal variances given in Eq. (3f). When r = 1 and $F_0 \geq 1$, ω_r in Eq. (56b) reduces to $1/\sqrt{F_0}$, while if r = 1 and $F_0 \leq 1$, ω_r in Eq. (51) reduces to $\sqrt{F_0}$. On the other hand, when r = −1, as expected ω_r in Eq. (56b) reduces to zero regardless of values of F_0 so that the Overlap becomes an exact α-level test.

Finally, what should the individual confidence levels $1-\gamma$ be so that the two independent CIs lead to the exact α-level test on H_0: $\mu_x = \mu_y$. As before, the expressions for the two $1-\gamma$ independent CIs are given by

$$\bar{x} - t_{\gamma/2,n-1} \times S_x / \sqrt{n} \le \mu_x \le \bar{x} + t_{\gamma/2,n-1} \times S_x / \sqrt{n} \tag{52c}$$

$$\bar{y} - t_{\gamma/2,n-1} \times S_y / \sqrt{n} \le \mu_y \le \bar{y} + t_{\gamma/2,n-1} \times S_y / \sqrt{n} \tag{52d}$$

It is clear that H_0: $\mu_x = \mu_y$ must be rejected at the α-level iff the amount of overlap between (52c) and (52d) barely becomes zero or less. Without loss of generality, the x-sample will be denoted such that $\bar{x} - \bar{y} \ge 0$. Therefore, we deduce from (52c & d) that

$$U'(\mu_y) - L'(\mu_x) = (\bar{y} + t_{\gamma/2,n-1} \times S_y / \sqrt{n}) - (\bar{x} - t_{\gamma/2,n-1} \times S_x / \sqrt{n})$$

$$= t_{\gamma/2,n-1} \times S_y / \sqrt{n} + t_{\gamma/2,n-1} \times S_x / \sqrt{n} - (\bar{x} - \bar{y}) \tag{53c}$$

Because H_0: $\mu_x = \mu_y$ must be rejected at the α-level as soon as the RHS of (53d) becomes 0 or smaller, we impose the rejection criterion $|\bar{x} - \bar{y}| \ge t_{\alpha/2,n-1} \times S_d / \sqrt{n}$ into Eq. (53c). This leads to rejecting H_0:$\mu_x = \mu_y$ iff

$$t_{\gamma/2,n-1} \times S_y / \sqrt{n} + t_{\gamma/2,n-1} \times S_x / \sqrt{n} - t_{\alpha/2,n-1} \times S_d / \sqrt{n} \le 0. \tag{53d}$$

At the borderline value, we set the LHS of inequality (53d) equal to 0 in order to solve for γ.

→ $t_{\gamma/2,n-1} \times S_y + t_{\gamma/2,n-1} \times S_x - t_{\alpha/2,n-1} \times S_d = 0.$

→ $t_{\gamma/2,n-1} \times (S_y + S_x) = t_{\alpha/2,n-1} \times S_d$

→ $t_{\gamma/2,n-1} = t_{\alpha/2,n-1} \times \dfrac{\sqrt{S_x^2 + S_y^2 - 2rS_xS_y}}{S_x + S_y}$

$$\rightarrow \quad t_{\gamma/2,n-1} = t_{\alpha/2,n-1} \times \frac{\sqrt{F_0 + 1 - 2r\sqrt{F_0}}}{\sqrt{F_0} + 1} \tag{54c}$$

where $F_0 = S_x^2 / S_y^2$. Eq. (54c) clearly shows that the value of γ depends on the LOS α of testing H_0: $:\mu_x = \mu_y$, F_0, and the sample size n. When r = −1, Eq. (54c) shows that $\gamma \equiv \alpha$ so that the Overlap becomes an exact α-level test; while if r = 1 the RHS attains its minimum value leading to maximum value for γ. When r = 0 (i.e., uncorrelated X & Y), Eq. (54c) shows that for very large or very small values of F_0, the Overlap in the limit becomes an α-level test.

8.0 The Impact of Overlap on Type II Error Probability for the Case of Unknown Process Variances σ_x^2, σ_y^2 and Small to Moderate Sample Sizes

Since the population variances σ_x^2 and σ_y^2 are unknown, then their point unbiased estimators S_x^2 and S_y^2, respectively, must be used for the purpose of statistical inference. As mentioned in Chapter 6 the rv $\dfrac{(\bar{x}-\mu_x)}{S_x/\sqrt{n_x}}$ is not normally distributed but its sampling follows that of W. S. Gosset's t-distribution with (n_x-1) degrees of freedom. As a result, the acceptance interval of the test statistic $\dfrac{(\bar{x}-\mu_{x0})}{S_x/\sqrt{n_x}}$ at the LOS α is $(-t_{\alpha/2,n_{x-1}},\ t_{\alpha/2,n_x-1})$, where $t_{\alpha/2,\nu} > 0$ for all 0 < α < 0.50, and it also follows that

$$\Pr(\bar{x}-t_{\alpha/2,n_x-1}S_x/\sqrt{n_x} \le \mu_x \le \bar{x}+t_{\alpha/2,n_x-1}S_x/\sqrt{n_x}) = 1-\alpha \tag{57a}$$

Hence, the lower $(1-\alpha)$% CI for μ_x is L(μ_x) $=\bar{x}-t_{\alpha/2,n_x-1}S_x/\sqrt{n_x}$, the corresponding upper limit is U(μ_x) $=\bar{x}+t_{\alpha/2,n_x-1}S_x/\sqrt{n_x}$, and the

$$\text{CIL}(\mu_x) = 2\times t_{\alpha/2,n_x-1}S_x/\sqrt{n_x} \tag{57b}$$

Similarly, L(μ_y) = $\bar{y}-t_{\alpha/2,n_y-1}S_y/\sqrt{n_y}$, U($\mu_y$) $=\bar{y}+t_{\alpha/2,n_y-1}S_y/\sqrt{n_y}$ and

$$\text{CIL}(\mu_y) = 2\times t_{\alpha/2,n_y-1}S_y/\sqrt{n_y}\,. \tag{57c}$$

8.1 The Case of H$_0$: $\sigma_x = \sigma_y = \sigma$ Not Rejected Leading to the Pooled t-Test

Assuming that X ~ N(μ_x, σ^2) and Y~N(μ_y, σ^2), then $\overline{X} - \overline{Y}$ has the N($\mu_x - \mu_y$, $\sigma^2 / n_x + \sigma^2 / n_y$) distribution, where it is assumed that σ^2 is the common value of the unknown $\sigma_x^2 = \sigma_y^2 = \sigma^2$. With the above assumptions, $\bar{x} - \bar{y}$ is an unbiased estimator of $\mu_x - \mu_y$ with Var($\bar{x} - \bar{y}$) $= \sigma^2 (1/n_x + 1/n_y)$. In practice a pretest on H$_0$: $\sigma_x^2 = \sigma_y^2 = \sigma^2$ is required before deciding to use either the pooled t-test or the two-independent-sample t-test. If the assumption $\sigma_x = \sigma_y = \sigma$ is tenable and because statistical theory dictates that the total resources be allocated according to $n_x = \sigma_x N / (\sigma_x + \sigma_y) = N/2 = n_y$, then the most common application of the pooled t-test occurs under equal sample sizes. Henceforth, the pooled t-test will be used iff the *P-value* of the pretest $H_0 : \sigma_x = \sigma_y = \sigma$ exceeds 20%, and for very small sample sizes n_x & $n_y < 10$, a *P-value* of at least 40% for the pretest is recommended. Since the common value of the process variances σ^2 is unknown, its unbiased estimators S_x^2 and S_y^2 should be pooled to obtain one unbiased estimator of σ^2, which as before is given by their weighted average based on their degrees of freedom, i.e.,

$$S_p^2 = \frac{\nu_x S_x^2 + \nu_y S_y^2}{\nu_x + \nu_y} = \frac{(n_x - 1) S_x^2 + (n_y - 1) S_y^2}{n_x + n_y - 2} \qquad (35)$$

Note that E(S_p^2) $= \sigma^2$. Therefore, the $se(\bar{x} - \bar{y}) = S_p \sqrt{1/n_x + 1/n_y}$ and as a result the rv $[(\bar{x} - \bar{y}) - (\mu_x - \mu_y)]/(S_P \sqrt{1/n_x + 1/n_y})$ has a central "Student's" sampling distribution with $\nu = n_x + n_y - 2$. Accordingly, the AI (acceptance interval) for a 5%-level test of H$_0$: $\mu_x - \mu_y = 0$ is given by

$$-t_{0.025,\nu} \times S_p\sqrt{1/n_x + 1/n_y} \le \bar{x} - \bar{y} \le t_{0.025,\nu} \times S_p\sqrt{1/n_x + 1/n_y} \qquad (58a)$$

where $\nu = \nu_x + \nu_y = n_x + n_y - 2$. Henceforth, in this section we let $t_{0.025}$ represent $t_{0.025,n_x+n_y-2}$ only for notational convenience. Thus the AI in (58a) reduces to

$$-t_{0.025} \times S_p\sqrt{1/n_x + 1/n_y} \le \bar{x} - \bar{y} \le t_{0.025} \times S_p\sqrt{1/n_x + 1/n_y} \qquad (58b)$$

Under the null hypothesis H_0: $\mu_x - \mu_y = 0$, the SMD of $t_0 = (\bar{x} - \bar{y})/S_p\sqrt{1/n_x + 1/n_y}$ is that of the central t with $\nu = \nu_x + \nu_y = n_x + n_y - 2$. Put differently, the null distribution of $(\bar{x} - \bar{y})/S_p\sqrt{1/n_x + 1/n_y}$ is $T_{n_x+n_y-2}$. Thus, the AI for in (58b) reduces to

$$\text{AI: } -t_{0.025} \le t_0 \le t_{0.025} \qquad (58c)$$

where $t_0 = (\bar{x} - \bar{y})/S_p\sqrt{1/n_x + 1/n_y}$ and $t_{0.025} = t_{0.025,n_x+n_y-2}$. However, if H_0: $\mu_x - \mu_y =$ 0 is false (so that a type II error can occur), the SMD of $(\bar{x} - \bar{y})/S_p\sqrt{1/n_x + 1/n_y}$ is no longer the central $T_{n_x+n_y-2}$. Thus, we next derive the SMD of the test statistic $t_0 =$ $(\bar{x} - \bar{y})/S_p\sqrt{1/n_x + 1/n_y}$ under the alternative H_1: $\mu_x - \mu_y = \delta \ne 0$.

From statistical theory, the SMD of the rv $U/\sqrt{\chi^2_\nu/\nu}$ is that of the central Student's t with df equal to that of χ^2_ν, where $U \sim N(0,1)$, i.e., U is a unit normal rv, and χ^2_ν is chi-squared distributed rv with ν df and independent of U. However, if $E(U) \ne 0$, then $U/\sqrt{\chi^2_\nu/\nu}$ is no longer central t distributed, but the rv $(Z+\xi)/\sqrt{\chi^2_\nu/\nu}$, where $Z \sim N(0,1)$, has the noncentral t distribution with ν df and noncentrality parameter ξ and the distribution is almost universally denoted by $t'_\nu(\xi)$, i.e., $(Z+\xi)/\sqrt{\chi^2_\nu/\nu} \sim t'_\nu(\xi)$. We will now illustrate how the above noncentral t distribution is used to compute type II error Pr when testing equality of two normal means with unknown but equal process variances. (This result has already been known in statistical literature for over 35 years.) By definition

$$\beta = \Pr(\text{Accepting } H_0\text{: } \mu_x - \mu_y = 0 \text{ if } H_0 \text{ is false})$$

$$= \Pr(-t_{0.025} \leq t_0 \leq t_{0.025} \mid \mu_x - \mu_y = \delta)$$

$$= \Pr(-t_{0.025} \leq t_0 = (\bar{x} - \bar{y})/S_p\sqrt{1/n_x + 1/n_y} \leq t_{0.025} \mid \mu_x - \mu_y = \delta) \qquad (59)$$

If H_0 is assumed false so that $E(\bar{x} - \bar{y}) = \mu_x - \mu_y \neq 0$, the SMD of $(\bar{x} - \bar{y})/S_p\sqrt{1/n_x + 1/n_y}$ is no longer the $t'_\nu(\xi = 0)$, which is the central t with $\nu = \nu_x + \nu_y = n_x + n_y - 2$ degrees of freedom. Thus, we first standardize $\bar{x} - \bar{y}$ in Eq.(59) as shown below, assuming $\sigma_x = \sigma_y = \sigma$.

$$\beta = \Pr\left(-t_{0.025} \leq \frac{[(\bar{x} - \bar{y}) - (\mu_x - \mu_y) + (\mu_x - \mu_y)]/\sqrt{\sigma^2/n_x + \sigma^2/n_y}}{S_p\sqrt{1/n_x + 1/n_y}\,/\sqrt{\sigma^2/n_x + \sigma^2/n_y}} \leq t_{0.025} \mid \mu_x - \mu_y = \delta\right)$$

$$= \Pr\left(-t_{0.025} \leq \frac{Z + [(\mu_x - \mu_y)/\sqrt{\sigma^2/n_x + \sigma^2/n_y}\,]}{\sqrt{S_p^2/\sigma^2}} \leq t_{0.025} \mid \mu_x - \mu_y = \delta\right)$$

$$= \Pr\left(-t_{0.025} \leq \frac{Z + [(\mu_x - \mu_y)/\sqrt{\sigma^2/n_x + \sigma^2/n_y}\,]}{\sqrt{[(n_x + n_y - 2)S_p^2/\sigma^2]/(n_x + n_y - 2)}} \leq t_{0.025} \mid \mu_x - \mu_y = \delta\right)$$

$$= \Pr\left(-t_{0.025} \leq \frac{Z + [\delta/\sqrt{\sigma^2/n_x + \sigma^2/n_y}\,]}{\sqrt{\chi^2_{n_x + n_y - 2}/(n_x + n_y - 2)}} \leq t_{0.025}\right)$$

$$= \Pr\left(-t_{0.025} \le \frac{Z+\xi}{\sqrt{\chi_\nu^2/\nu}} \le t_{0.025}\right) \qquad (60a)$$

where $\xi = (\mu_x - \mu_y)/\sqrt{\sigma^2/n_x + \sigma^2/n_y} = \delta/\sqrt{\sigma^2/n_x + \sigma^2/n_y}$ and $\nu = n_x+n_y-2$.

However, as stated above, the SMD of $\frac{Z+\xi}{\sqrt{\chi_\nu^2/\nu}}$ is the noncentral t with $\nu = n_x+n_y-2$ and noncentrality parameter $\xi = \delta/\sqrt{\sigma^2/n_x + \sigma^2/n_y} = \delta/SE(\bar{x}-\bar{y})$, i.e.,

$$\frac{Z+\xi}{\sqrt{\chi_\nu^2/\nu}} \sim t'_{n_x+n_y-2}\left(\frac{\delta}{\sigma\sqrt{1/n_x+1/n_y}}\right) = t'_{n_x+n_y-2}\left(\frac{\delta}{\sigma}\sqrt{\frac{n_x n_y}{n_x+n_y}}\right).$$

Thus, $\beta = \Pr\left(-t_{0.025} \le \frac{Z+\xi}{\sqrt{\chi_\nu^2/\nu}} \le t_{0.025}\right)$

$$= \Pr\left(-t_{0.025} \le t'_{n_x+n_y-2}\left(\frac{\delta}{\sigma}\sqrt{\frac{n_x n_y}{n_x+n_y}}\right) \le t_{0.025}\right) \qquad (60b)$$

Note that when $\delta = 0$, the argument in (60b) becomes the central t and β becomes equal to $1-\alpha$.

When the design is balanced, the SMD of the test statistic t_0 under H_1 reduces to $t'_{2(n-1)}\left(\frac{\delta}{\sigma}\sqrt{\frac{n}{2}}\right)$, i.e., when $n_x = n_y = n$ →

$$\beta = \Pr\left(-t_{0.025} \le \frac{Z+\xi}{\sqrt{\chi_\nu^2/\nu}} \le t_{0.025}\right) = \Pr\left(-t_{0.025} \le t'_{2(n-1)}\left(\frac{\delta}{\sigma}\sqrt{\frac{n}{2}}\right) \le t_{0.025}\right) \qquad (60c)$$

As an example, suppose we draw a random sample $n_x = 7$ from a $N(\mu_x, \sigma^2)$ and one of size $n_y = 11$ 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 $\alpha = 5\%$ versus the 2-sided alternative H_1: $\mu_x - \mu_y \ne 0$.

We wish to answer the question as to what is the Pr of accepting H_0 if the true mean difference $\mu_x - \mu_y$ were not zero but were equal to 0.50σ, i.e., we wish to compute the type II error Pr at $\delta = 0.50\sigma$. Then the corresponding value of the noncentrality parameter is equal to $\xi = (\delta / \sigma)\sqrt{n_x n_y / (n_x + n_y)} = (0.50\sigma / \sigma)\sqrt{77/18} = 1.03414$ and type II error Pr from Eq. (60b) is equal to →

$$\beta = \Pr(-t_{0.025,16} \le t'_{16}(1.03414) \le t_{0.025,16}) = \Pr(-2.119905 \le t'_{16}(1.03414) \le 2.119905)$$

$$= cdf[\text{of } t'_{16}(1.03414) \text{ at } 2.119905] - cdf[\text{of } t'_{16}(1.03414) \text{ at } (-2.119905)].$$

Fortunately, both Minitab and Matlab provide the cdf of the noncentral t distribution. Using Minitab, we obtain $cdf[\text{of } t'_{16}(1.03414) \text{ at } 2.119905] = 0.838156$ and cdf[of $t'_{16}(1.03414)$ at –2.119905] = 0.0016652; thus, β = 0.838156–0.0016652 = 0.836491 so that the power of the test at δ = 0.50σ is equal to 1–β = 1–0.836491 = 0.163509. Clearly as $\delta = \mu_x - \mu_y$ departs further from zero, the power of the test must increase, which is illustrated next. Suppose now δ = 0.80σ; then $\xi = (0.80\sigma / \sigma)\sqrt{77/18} = 1.65462315$ and $\beta(\text{at } \delta = 0.80\sigma,\ n_x = 7, n_y = 11)$

$$= \Pr(-2.119905 \le t'_{16}(1.65462315) \le 2.119905)$$

$$= cdf[\text{of } t'_{16}(1.65462315) \text{ at } 2.119905] - cdf[\text{of } t'_{16}(1.65462315) \text{ at } -2.119905]$$

$$= 0.656987 - 0.0002142 = 0.656733,$$

and hence the power of the test increases from 0.163509 to 1–0.656773 = 0.343227. It is interesting to note that if the design is balanced, then the power of the test always increases for the same parameter values. For example, if $n_x = n_y = 9$ so that ν = 16 stays in tact, then at δ = 0.80σ, the noncentrality parameter $\xi = (0.80\sigma / \sigma)\sqrt{81/18} =$

1.6970563 and

$$\beta\,(\text{at } \delta = 0.80\sigma,\ n_x = 9, n_y = 9) = \Pr(-2.119905 \le t'_{16}(1.6970563) \le 2.119905)$$

$$= cdf[\text{of } t'_{16}(1.6970563) \text{ at } 2.119905] - cdf[\text{of } t'_{16}(1.6970563) \text{ at } -2.119905]$$

$$= 0.642235 - 0.0001839 = 0.6420511,$$

so that Power(at $\delta = 0.80\sigma$) = 0.357949, which exceeds the value of 0.343227 for the unbalanced case. The syntax for Matlab noncentral t *cdf* is *nctcdf*(t, ν, ξ).

As in the case of known variances, the type II error Pr from the Overlap is computed similar to Eqs. (7) shown below.

$$\beta' = \Pr(\text{Overlap} | \delta > 0) = \Pr\{[L(\mu_x) \le U(\mu_y)] \cap [L(\mu_y) \le U(\mu_x)] | \mu_x - \mu_y = \delta\}$$

$$= \Pr\{[\bar{x} - t_{\alpha/2,\nu_x} S_x/\sqrt{n_x} \le \bar{y} + t_{\alpha/2,\nu_y} S_y/\sqrt{n_y}] \cap$$

$$[\bar{y} - t_{\alpha/2,\nu_y} S_y/\sqrt{n_y} \le \bar{x} + t_{\alpha/2,\nu_x} S_x/\sqrt{n_x}] | \delta\}$$

$$= \Pr\{[\bar{x} - \bar{y} \le t_{\alpha/2,\nu_x} S_x/\sqrt{n_x} + t_{\alpha/2,\nu_y} S_y/\sqrt{n_y}] \cap$$

$$[-t_{\alpha/2,\nu_y} S_y/\sqrt{n_y} - t_{\alpha/2,\nu_x} S_x/\sqrt{n_x} \le \bar{x} - \bar{y}] | \delta\}$$

$$= \Pr\{[-t_{\alpha/2,\nu_y} S_y/\sqrt{n_y} - t_{\alpha/2,\nu_x} S_x/\sqrt{n_x} \le \bar{x} - \bar{y} \le t_{\alpha/2,\nu_x} S_x/\sqrt{n_x} + t_{\alpha/2,\nu_y} S_y/\sqrt{n_y}] | \delta\}$$

$$= \Pr\{[-A \le \bar{x} - \bar{y} \le +A] | \delta\} \qquad (61)$$

where $A = t_{\alpha/2,\nu_x} S_x/\sqrt{n_x} + t_{\alpha/2,\nu_y} S_y/\sqrt{n_y}$.

In order to apply the noncentral t-distribution to compute the Pr in Eq. (61), not available in statistical literature, we must first inside brackets divide throughout by $S_p\sqrt{1/n_x + 1/n_y}$ and then standardize $\bar{x} - \bar{y}$ as illustrated below.

$$\beta' = \Pr\{[-A/S_p\sqrt{1/n_x + 1/n_y} \le (\bar{x} - \bar{y})/S_p\sqrt{1/n_x + 1/n_y} \le A/S_p\sqrt{1/n_x + 1/n_y}] | \delta\}$$

$$= \Pr[-A_p \le t'_{n_x+n_y-2}(\frac{\delta}{\sigma}\sqrt{\frac{n_x n_y}{n_x+n_y}}) \le A_p] \quad (62a)$$

where $A_p = \dfrac{t_{\alpha/2,\nu_x} S_x/\sqrt{n_x} + t_{\alpha/2,\nu_y} S_y/\sqrt{n_y}}{S_p\sqrt{1/n_x+1/n_y}}$. Note that if δ = 0, Eq. (62a) reduces to $1-\alpha'$

as was shown in Eq. (37a). In the case of balanced design, Eq. (62a) reduces to

$$\beta' = \Pr[-\frac{t_{\alpha/2,n-1}(S_x+S_y)}{\sqrt{S_x^2+S_y^2}} \le t'_{2(n-1)}(\frac{\delta}{\sigma}\sqrt{\frac{n}{2}}) \le \frac{t_{\alpha/2,n-1}(S_x+S_y)}{\sqrt{S_x^2+S_y^2}}]$$

$$= \Pr[-\frac{t_{\alpha/2,n-1}(\sqrt{F_0}+1)}{\sqrt{F_0+1}} \le t'_{2(n-1)}(\frac{\delta}{\sigma}\sqrt{\frac{n}{2}}) \le \frac{t_{\alpha/2,n-1}(\sqrt{F_0}+1)}{\sqrt{F_0+1}}] \quad (62b)$$

As an example, suppose samples of sizes $n_x = n_y = 9$ are drawn from two independent normal universes with unknown but equal 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 S_x= 0.65 and $S_y = 0.54$. Note that it is sufficient to provide the ratio $F_0 = S_x^2/S_y^2$ instead of the specific values of S_x and S_y. Because $t_{\alpha/2,n-1} = t_{0.025,8} = 2.306004$, ξ = $(0.80\sigma/\sigma)\sqrt{81/18}$ = 1.6970563, Eq. (62b) yields β'(at $\delta = \mu_x - \mu_y = 0.80\sigma$, $n_x = 9$, $n_y = 9$) = $\Pr[-3.247338 \le t'_{16}(1.6970563) \le 3.247338]$ = 0.904239–0.0000078 = 0.904231.

The above value of β' is much larger than $\beta = 0.6420511$ using the Standard method. It can easily be verified that the random function $\dfrac{S_x+S_y}{\sqrt{S_x^2+S_y^2}}$ lies within the interval

$1 < \dfrac{S_x+S_y}{\sqrt{S_x^2+S_y^2}} = \dfrac{\sqrt{F_0}+1}{\sqrt{F_0+1}} \le \sqrt{2}$. However, we are using the pooled t-test only if

$F_{0.90,n-1,n-1} \le F_0 = S_x^2 / S_y^2 \le F_{0.10,n-1,n-1}$ and hence $\frac{\sqrt{F_{0.90,n-1,n-1}}+1}{\sqrt{F_{0.90,n-1,n-1}+1}} = \frac{\sqrt{F_{0.10,n-1,n-1}}+1}{\sqrt{F_{0.10,n-1,n-1}+1}}$

$\le \frac{S_x + S_y}{\sqrt{S_x^2 + S_y^2}} = \frac{\sqrt{F_0}+1}{\sqrt{F_0+1}} \le \sqrt{2}$. Note that the equality on the most LHS of this last equation

follows from the fact that $F_{0.10,n-1,n-1} = 1/ F_{0.90,n-1,n-1}$ for all n. Therefore, for a balanced

design the GLB of β' for a 5%-level test is given by

$$\text{GLB}(\beta') = \Pr[-t_{0.025,n-1}\frac{\sqrt{F_{0.10}}+1}{\sqrt{F_{0.10}+1}} \le t'_{2(n-1)}(\frac{\delta}{\sigma}\sqrt{\frac{n}{2}}) \le t_{0.025,n-1}\frac{\sqrt{F_{0.10}}+1}{\sqrt{F_{0.10}+1}}] \quad (63a)$$

where $F_{0.10} = F_{0.10,n-1,n-1}$, and the LUB is given by

$$\text{LUB}(\beta') = \Pr[-t_{0.025,n-1}\sqrt{2} \le t'_{2(n-1)}(\frac{\delta}{\sigma}\sqrt{\frac{n}{2}}) \le t_{0.025,n-1}\sqrt{2}\,] \quad (63b)$$

i.e.,

$$\Pr[-t_{0.025,n-1}\frac{\sqrt{F_{0.10}}+1}{\sqrt{F_{0.10}+1}} \le t'_{2(n-1)}(\frac{\delta}{\sigma}\sqrt{\frac{n}{2}}) \le \frac{\sqrt{F_{0.10}}+1}{\sqrt{F_{0.10}+1}}\, t_{0.025,n-1}] \le \beta' \le$$

$$\Pr[-t_{0.025,n-1}\sqrt{2} \le t'_{2(n-1)}(\frac{\delta}{\sigma}\sqrt{\frac{n}{2}}) \le t_{0.025,n-1}\sqrt{2}\,] \quad (63c)$$

Thus, for the example with $n_x = n_y = 9$, the GLB that the overlap type II error Pr can

attain is given by Eq. (63a) and is computed below.

GLB (β' at $\delta = \mu_x - \mu_y = 0.80\sigma$)

$$= \Pr[-2.306004\frac{\sqrt{2.589349}+1}{\sqrt{2.589349+1}} \le t'_{16}(1.6970563) \le 3.175781]$$

$$= \Pr[-3.175781 \le t'_{16}(1.6970563) \le 3.175781]$$

$$= 0.89451811 - 0.00000954 = 0.89450857.$$

Thus, the smallest % relative error for the power of the test from Overlap is

$$[(0.357949-0.105492)/0.357949]\times 100\% = 70.53\%.$$

Furthermore, the LUB that the overlap type II error Pr can become is given by Eq.(63b) and is calculated as following:

LUB (β'at $\delta = \mu_x - \mu_y = 0.80\sigma$)

$$= \Pr[-2.306004\times\sqrt{2} \le t'_{16}(1.6970563) \le 3.26118232]$$

$$= \Pr[-3.26118232 \le t'_{16}(1.6970563) \le 3.26118232]$$

$$= 0.90602841 - 0.00000756 = 0.90602085$$

Therefore, the worst % relative error for the power of the test from Overlap is

$$[(0.357949-0.093979)/0.357949]\times 100\% = 73.75\%.$$

8.2 The Case of $H_0: \sigma_x = \sigma_y$ Rejected Leading to the Two-Independent Sample t-Test (or the t-Prime Test)

Assuming that X~N(μ_x, σ_x^2) and Y~N(μ_y, σ_y^2), then $\bar{X} - \bar{Y}$ is N($\mu_x - \mu_y$, $\sigma_x^2/n_x + \sigma_y^2/n_y$), but now the null hypothesis of $H_0: \sigma_x = \sigma_y$ is rejected at the 20% level leading to the assumption that the F-statistic $F_0 = S_x^2/S_y^2 > 2$ for all sample sizes 16 $\le n_x$ & n_y.

It has been shown in statistical theory that if the assumption $\sigma_x = \sigma_y$ is not tenable, the statistic $[(\bar{x}-\bar{y})-(\mu_x-\mu_y)]/\sqrt{(S_x^2/n_x)+(S_y^2/n_y)}$ has the approximate central Student's t-distribution with degrees of freedom

$$\nu = \frac{[v(\bar{x})+v(\bar{y})]^2}{\frac{(v(\bar{x}))^2}{\nu_x}+\frac{(v(\bar{y}))^2}{\nu_y}} = \frac{\nu_x\nu_y[v(\bar{x})+v(\bar{y})]^2}{\nu_y(v(\bar{x}))^2+\nu_x(v(\bar{y}))^2} = \frac{\nu_x\nu_y(F_0R_n+1)^2}{\nu_y(F_0R_n)^2+\nu_x} \qquad (39)$$

where $v(\bar{x}) = S_x^2/n_x$ and $F_0 = S_x^2/S_y^2$. The formula for degrees of freedom in (39) rarely leads to an integer and is generally rounded down to make the test of $H_0: \mu_x - \mu_y = 0$ conservative, i.e., the rounding down ν increases the *P-value* of the test. However, programs like Matlab and Minitab will provide probabilities of the t-distribution for non-integer values of ν in Eq. (39). It has been verified by the authors that ν in Eq. (39) attains its maximum when the larger sample also has much larger variance than the sample whose size is much smaller. Even then, it is for certain that $\text{Min}(\nu_x, \nu_y) < \nu < \nu_x + \nu_y$, and hence the two-sample t-test is less powerful than the pooled t-test. When $H_0: \sigma_x = \sigma_y$ is rejected at the 20% level (i.e., *P-value* < 0.20), the type II error Pr of a 5%-level test is given by β = Pr (Accepting H_0: $\mu_x - \mu_y = 0$ if H_0 is false) →

$$\beta \doteq \Pr(-t_{0.025,\nu} \le t_0 \le t_{0.025,\nu} \mid \mu_x - \mu_y = \delta) \qquad (64)$$

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, ν, given in Eq. (39). Henceforth in this section we let $t_{0.025}$ represent $t_{0.025,\nu}$ only for notational convenience. 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 $\sigma_x = \sigma_y$, is intractable using central χ^2. As far as we know, the exact power of the t-Prime (or the two-sample independent t-test) test has not yet been obtained in statistical literature. That is, the SMD of t_0 is not the noncentral t with some noncentrality parameter ξ. The development that follows, the results already existing in statistical literature, is only an approximation because there does not exist an exact solution for type II error Pr of testing H$_0$: $\mu_x - \mu_y = 0$ when the variances are unknown and unequal. We first approximately studentize the expression for β in Eq.(64).

$$\beta \doteq \Pr(-t_{0.025,\nu} \le t_0 \le t_{0.025,\nu} \mid \mu_x - \mu_y = \delta)$$

$$\doteq \Pr(-t_{0.025,\nu} \le (\bar{x} - \bar{y}) / \sqrt{(S_x^2 / n_x) + (S_y^2 / n_y)} \le t_{0.025,\nu} \mid \mu_x - \mu_y = \delta)$$

$$\doteq \Pr\{-t_{0.025} - \delta / \sqrt{(S_x^2 / n_x) + (S_y^2 / n_y)} \le [(\bar{x} - \bar{y}) - \delta] / \sqrt{(S_x^2 / n_x) + (S_y^2 / n_y)}$$

$$\le t_{0.025} - \delta / \sqrt{(S_x^2 / n_x) + (S_y^2 / n_y)}\}$$

$$\doteq \Pr\{-t_{0.025} - \delta / \sqrt{(S_x^2 / n_x) + (S_y^2 / n_y)} \le t_\nu \le t_{0.025} - \delta / \sqrt{(S_x^2 / n_x) + (S_y^2 / n_y)}\}$$

$$\doteq \Pr\{-t_{0.025} - \Delta \le t_\nu \le t_{0.025} - \Delta\} \qquad (65)$$

where the studentized mean difference $\Delta = (\mu_x - \mu_y) / \sqrt{(S_x^2 / n_x) + (S_y^2 / n_y)}$

$= \delta / \sqrt{(S_x^2 / n_x) + (S_y^2 / n_y)}$. Unfortunately, the approximate expression for β in Eq.(65) 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 Eq.(65) can be carried out iff δ is specified in units of the $se(\bar{x} - \bar{y})$, 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 $n_x = n_y = 9$ are drawn from two independent normal populations with unknown but unequal variances. We wish to compute the Pr of accepting H$_0$: $\mu_x - \mu_y = 0$ at $\alpha = 0.05$ if $\mu_x - \mu_y = \delta = 0.4$ and the sample statistics are S_x= 0.65 and S_y = 0.54. Eq.(39) gives $\nu = \dfrac{\nu_x\nu_y[V(\bar{x})+V(\bar{y})]^2}{\nu_y(V(\bar{x}))^2+\nu_x(V(\bar{y}))^2}$ = 15.48, $t_{0.025} = t_{0.025,15.48}$ = 2.1257, $\Delta = (\mu_x-\mu_y)/\sqrt{(S_x^2/n_x)+(S_y^2/n_y)}$ = 0.40/0.281681 = 1.420044 so that $-t_{0.025}-\Delta$ = – 3.545744, $t_{0.025}-\Delta$ = 2.1257–1.420044 = 0.7057, and β(at δ = 0.40) $\doteq$ Pr$(-3.5457 \le t_{15.48} \le 0.7056)$ = 0.75454016–0.00140620 = 0.75313396. If $\mu_x-\mu_y$ = 0.60, similar calculations will show that β(at $\mu_x-\mu_y$ = 0.6) $\doteq$ Pr$(-4.25577 \le t_{15.48} \le -0.00437)$ = 0.4982845– 0.0003233 = 0.4979612. Note that the above approximate type II error Prs would be in exact agreement with what UCLA's Statistics Department Power Calculator lists on their website (www.stat.ucla.edu). If n_x = 7, S_x = 0.65, n_y =11 and S_y = 0.54 the type II error Pr increases a bit from 0.7532 to β(at $\mu_x-\mu_y$ = 0.4) = 0.79083377 − 0.0022143 = 0.7886.

Again, the type II error Pr from the Overlap is computed similar to Eq. (7), just like the case of the pooled t-test, as shown below.

$$\beta' = \Pr(\text{Overlap} \mid \delta > 0) = \Pr\{[L(\mu_x) \le U(\mu_y)] \cap [L(\mu_y) \le U(\mu_x)] \mid \mu_x - \mu_y = \delta\}$$

Note that the event $[L(\mu_x) \le U(\mu_y)] \cap [L(\mu_y) \le U(\mu_x)]$ is equivalent to either $L(\mu_x) \le U(\mu_y) \le U(\mu_x)$ or $L(\mu_y) \le U(\mu_x) \le U(\mu_y)$. Thus,

$$\beta' = \Pr\{[\bar{x} - t_{\alpha/2,\nu_x} S_x/\sqrt{n_x} \le \bar{y} + t_{\alpha/2,\nu_y} S_y/\sqrt{n_y}] \cap$$

$$[\bar{y} - t_{\alpha/2,\nu_y} S_y/\sqrt{n_y} \le \bar{x} + t_{\alpha/2,\nu_x} S_x/\sqrt{n_x}] \mid \delta\}$$

$$= \Pr\{[\bar{x} - \bar{y} \le t_{\alpha/2,\nu_x} S_x/\sqrt{n_x} + t_{\alpha/2,\nu_y} S_y/\sqrt{n_y}] \cap$$

$$[-t_{\alpha/2,\nu_y} S_y/\sqrt{n_y} - t_{\alpha/2,\nu_x} S_x/\sqrt{n_x} \le \bar{x} - \bar{y}] \mid \delta\}$$

$$= \Pr\{[-t_{\alpha/2,\nu_y}\ S_y/\sqrt{n_y} - t_{\alpha/2,\nu_x}\ S_x/\sqrt{n_x} \le \bar{x} - \bar{y} \le t_{\alpha/2,\nu_x}\ S_x/\sqrt{n_x} + t_{\alpha/2,\nu_y}\ S_y/\sqrt{n_y}\ ]|\ \delta\}$$

$$= \Pr\{[-A \le \bar{x} - \bar{y} \le +A]|\ \delta\} \quad (66)$$

where A = $t_{\alpha/2,\nu_x}\ S_x/\sqrt{n_x} + t_{\alpha/2,\nu_y}\ S_y/\sqrt{n_y}$. Studentizing inside the brackets of Eq.(66) results in:

$$\beta' = \Pr\{[-A-(\mu_x-\mu_y) \le (\bar{x} - \bar{y}) - (\mu_x-\mu_y) \le +A-(\mu_x-\mu_y)]|\ \delta\}$$

$$= \Pr\{[-A-(\mu_x-\mu_y)]/\ se(\bar{x}-\bar{y}) \le [(\bar{x} - \bar{y}) - (\mu_x-\mu_y)]\ /se(\bar{x}-\bar{y}) \le (A-\delta)/\ se(\bar{x}-\bar{y})\}$$

where $se(\bar{x}-\bar{y}) = \sqrt{(S_x^2/n_x)+(S_y^2/n_y)}$. Thus,

$$\beta' \doteq \Pr\{(-A-\delta)/\ se(\bar{x}-\bar{y}) \le t_\nu \le (A-\delta)/\ se(\bar{x}-\bar{y})\} \quad (67)$$

For the example, if $n_x = n_y = 9$, $S_x = 0.65$ and $S_y = 0.54$, A = 0.914715, ν = $\dfrac{\nu_x \nu_y [V(\bar{x})+V(\bar{y})]^2}{\nu_y (V(\bar{x}))^2 + \nu_x (V(\bar{y}))^2}$ = 15.48 as before, Eq. (67) now gives $\beta'(\delta = 0.40) \doteq$ $\Pr[-4.66738 \le t_{15.48} \le 1.827295]$ = 0.95650-0.00014 = 0.9564 as compared to exact value of β(at δ =0.40) = 0.7532 and a % relative error in power ($[(\beta'-\beta)/(1-\beta)] \times 100\%/(1-\beta)$) equal to 82.33%.

8.3 The Impact of Overlap on Type II Error Probability for the Paired t-Test (i.e., the Randomized Block Design) when Process Variances are Unknown

Consider the 5%-level test of H_0: $\mu_x - \mu_y = \mu_d = 0$ versus the 2-sided alternative H_1: $\mu_d \ne 0$, 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. The appropriate test statistic for testing H_0: $\mu_d = 0$ is $t_0 = \bar{d}\sqrt{n}/S_d$, where S_d $= \sqrt{S_x^2 + S_y^2 - 2\hat{\sigma}_{xy}}$ and $\hat{\sigma}_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})/(n-1)$. The decision rule is to reject H_0 at the 0.05-level iff $|\bar{d}\sqrt{n}/S_d| > t_{0.025,n-1}$. Thus, for a 5%-level test by definition

$$\beta = \Pr(\text{Accepting } H_0\text{: } \mu_x - \mu_y = 0 \text{ if } H_0 \text{ is false})$$

$$= \Pr(-t_{0.025,n-1} \le t_0 \le t_{0.025,n-1} \mid \mu_d = \delta)$$

Fortunately, just like in the case of the pooled t-test, the exact SMD of t_0 under the alternative H_1: $\mu_d \neq 0$ has been known for well over 35 years. That is to say, an exact expression for the OC curve of the paired t-test already exists in statistical literature as illustrated below.

$$\beta = \Pr(-t_{0.025,n-1} \le \bar{d}\sqrt{n}/S_d \le t_{0.025,n-1} \mid \mu_d = \delta) \tag{68}$$

where for notational convenience we will let $t_{0.025} = t_{0.025,n-1}$ in this section.

Standardizing $\bar{d}\sqrt{n}/S_d$ in Eq. (68) under the alternative H_1: $\mu_d \neq 0$ leads to

$$\beta = \Pr(-t_{0.025,n-1} \le \bar{d}\sqrt{n}/S_d \le t_{0.025,n-1} \mid \mu_d = \delta)$$

$$= \Pr\left(-t_{0.025,n-1} \le \frac{[(\bar{d} - \mu_d) + \mu_d]\sqrt{n}/\sigma_d}{S_d/\sigma_d} \le t_{0.025,n-1} \mid \mu_d = \delta\right)$$

$$= \Pr\left(-t_{0.025,n-1} \le \frac{Z + \mu_d\sqrt{n}/\sigma_d}{\sqrt{S_d^2/\sigma_d^2}} \le t_{0.025,n-1} \mid \mu_d = \delta\right)$$

$$= \Pr\left(-t_{0.025,n-1} \le \frac{Z + \xi}{\sqrt{\frac{(n-1)S_d^2/\sigma_d^2}{(n-1)}}} \le t_{0.025,n-1} \mid \mu_d = \delta\right)$$

$$= \Pr(-t_{0.025,n-1} \le \frac{Z+\xi}{\sqrt{\chi^2_{n-1}/(n-1)}} \le t_{0.025,n-1} \mid \mu_d = \delta)$$

$$= \Pr(-t_{0.025,n-1} \le t'_{n-1}(\xi) \le t_{0.025,n-1}) \qquad (69)$$

Eq. (69) shows that the exact SMD of $t_0 = \bar{d}\sqrt{n}/S_d$ under the alternative H_1: $\mu_d \ne 0$ is the noncentral t with noncentrality parameter $\xi = \mu_d\sqrt{n}/\sigma_d$ and $\nu = n-1$ df, 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: $\mu_x - \mu_y = \mu_d = 0$ at the 5% level with a random sample of size n = 10 blocks from a bivariate normal distribution versus the alternative H_1: $\mu_d = 0.50\sigma_d$. Thus from Eq. (69), $\beta(\text{at } \mu_d = 0.50\sigma_d) = \Pr(-t_{0.025,9} \le t'_{n-1}(\xi) \le t_{0.025,9})$, where $\xi = 0.50\sigma_d\sqrt{10}/\sigma_d = 1.581139$. Consequently using Matlab we obtain

$$\beta(\text{at } \mu_d = 0.50\sigma_d, n=10) = \Pr(-2.262157 \le t'_9(1.581139) \le 2.262157)$$

$$= nctcdf(2.262157, 9, 1.581139) - nctcdf(-2.262157, 9, 1.581139)$$

$$= 0.7071714 - 0.00034704 = 0.70682435,$$

so that the power of the test is given by PWF(at $0.50\sigma_d$, n =10) = 0.29317565. It is common knowledge in the field of Statistics that the power of a test should increase with increasing sample size; a statistical test for which the limit of its PWF does not approach 1 as $n \to \infty$, is said to be inconsistent. It is also estimated that in order to double the power of a test, roughly more than twice the sample size is needed. For this reason, consider this last example where the PWF(at $0.50\sigma_d$) was equal to 0.29317565 with n = 10 but now we set the value of n at 20. Then, at n = 20, $\beta(\text{at } \mu_d = 0.50\sigma_d, n = 20) = \Pr(-t_{0.025,19} \le t'_{n-1}(\xi) \le t_{0.025,19})$, where $\xi = 0.50\sigma_d\sqrt{20}/\sigma_d = 2.236068$ and

$t_{0.025,19} = 2.093024$

→ β(at $\mu_d = 0.50\sigma_d$, n = 20)

$= nctcdf(2.0930240,19,\ 2.236068) - \text{nctcdf}(-2.0930240,19,\ 2.236068)$

$= 0.43551707475811 - 2.152176224301527\times 10^{-5}$

$= 0.43549555299587$

→ PWF(at $0.50\sigma_d$, n =20)

$= 0.564504447 < 2\times$ PWF(at $0.50\sigma_d$, n =10).

On the other hand, in order to have the same value of PWF at $\mu_d = (1/2)\times 0.50\sigma_d$, roughly four times the sample size is needed.

As in sections 8.1 and 8.2, the type II error Pr using the Overlap is given by

$$\beta' = \Pr\{[-t_{\alpha/2,\nu_y}\ S_y/\sqrt{n_y} - t_{\alpha/2,\nu_x}\ S_x/\sqrt{n_x} \le \bar{x} - \bar{y} \le$$

$$t_{\alpha/2,\nu_x}\ S_x/\sqrt{n_x} + t_{\alpha/2,\nu_y}\ S_y/\sqrt{n_y}\ ]|\ \delta \ne 0\}$$

$$= \Pr\{[-A \le \bar{d} \le +A]|\ \delta \ne 0\} \tag{70a}$$

where as before $A = t_{\alpha/2,\nu_x}\ S_x/\sqrt{n_x} + t_{\alpha/2,\nu_y}\ S_y/\sqrt{n_y}$ and $\bar{d} = \bar{x} - \bar{y}$. However, because this is a block design, then per force $n_x = n_y = n$, and as a result $A = t_{\alpha/2,n-1}(S_x+S_y)/\sqrt{n}$.

Following the exact same development that leads to Eq. (69), we obtain

$$\beta' = \Pr[-t_{\alpha/2,n-1}(S_x+S_y)/\sqrt{n} \le \bar{d} \le t_{\alpha/2,n-1}(S_x+S_y)/\sqrt{n}\ ]$$

$$= \Pr[-t_{\alpha/2,n-1}(S_x+S_y)/S_d \le \bar{d}\ \sqrt{n}/S_d \le t_{\alpha/2,n-1}(S_x+S_y)/S_d]$$

$$= \Pr[-t_{\alpha/2,n-1}(S_x+S_y)/S_d \le \frac{(\bar{d} - \mu_d + \mu_d)\sqrt{n}/\sigma_d}{S_d/\sigma_d} \le t_{\alpha/2,n-1}(S_x+S_y)/S_d]$$

$$= \Pr[-t_{\alpha/2,n-1}(S_x+S_y)/S_d \le \frac{(\bar{d}-\mu_d+\mu_d)\sqrt{n}/\sigma_d}{\sqrt{\frac{(n-1)S_d^2/\sigma_d^2}{n-1}}} \le t_{\alpha/2,n-1}(S_x+S_y)/S_d]$$

$$= \Pr[-t_{\alpha/2,n-1}(S_x+S_y)/S_d \le t'_{n-1}(\xi) \le t_{\alpha/2,n-1}(S_x+S_y)/S_d] \quad (70b)$$

where $\xi = \mu_d\sqrt{n}/\sigma_d$. Because $S_d = \sqrt{S_x^2+S_y^2-2rS_xS_y}$, it follows that $S_d \le S_x+S_y$, and equality occurring iff the sample correlation coefficient $r = -1$. On comparing the expression for β in Eq. (69) with that of β′ in (70b), it is clear that $\beta \le \beta'$ because $t_{\alpha/2,n-1}(S_x+S_y)/S_d \ge t_{\alpha/2,n-1}$. Further, dividing the numerator and denominator of $(S_x+S_y)/S_d$ by S_y, we obtain $(S_x+S_y)/S_d = (\sqrt{F_0}+1)/\sqrt{F_0+1-2r\sqrt{F_0}}$. Substituting this last into (70b) results in

$$\beta' = \Pr[-t_{\alpha/2,n-1}(\frac{\sqrt{F_0}+1}{\sqrt{F_0+1-2r\sqrt{F_0}}}) \le t'_{n-1}(\xi) \le t_{\alpha/2,n-1}(\frac{\sqrt{F_0}+1}{\sqrt{F_0+1-2r\sqrt{F_0}}})] \quad (70c)$$

The final expression for Overlap type II error Pr in Eq. (70c) clearly shows that the value of β′ depends only on the sample size n, noncentrality parameter $\xi = \mu_d\sqrt{n}/\sigma_d$, the sample correlation coefficient r, and the sample variance ratio $F_0 = S_x^2/S_y^2$ but does not depend on the specific values of S_x^2 and S_y^2. Only when $r = -1$, the value of β′ equals to β; otherwise $\beta' > \beta$. Further, the limit of β′ as $F_0 \to 0$ or as $F_0 \to \infty$ is also equal to β. It can easily be shown using calculus that the function $(\sqrt{F_0}+1)/\sqrt{F_0+1-2r\sqrt{F_0}}$ in the argument of β′ in Eq. (70c) attains its maximum at $F_0 = 1$ with its value equal to $\sqrt{2/(1-r)}$. Thus for a 5%-level test, the least upper bound for β′ is given by

$$\text{LUB}(\beta') = \Pr[-t_{0.025,n-1}\sqrt{2/(1-\text{r})} \leq t'_{n-1}(\xi) \leq t_{0.025,n-1}\sqrt{2/(1-\text{r})}\,] \qquad (71)$$

As r→ −1, the LUB(β′)→ β, but as r→ +1, LUB(β′) → +1. Thus the impact of negative correlation is to reduce the Overlap type II error Pr while the impact of positive correlation is to increase β′. As an example, for a random bivariate pair of size n = 10, a 5%-level test, and r = – 0.50, Eq. (71) at $\mu_d = 0.50\sigma_d$ yields

$$\text{LUB}(\beta') = \Pr[-t_{0.025,9}\sqrt{2/1.5} \leq t'_9(1.581139) \leq 2.262157\sqrt{2/1.5}\,]$$

$$= \Pr[-2.612114 \leq t'_9(1.581139) \leq 2.612114]$$

$$= 0.793446 - 0.0001628 = 0.79328344$$

as compared to β = 0.70682435 from the Standard method. However, if r were equal to +0.50, then

$$\text{LUB}(\beta') = \Pr[-t_{0.025,9}\sqrt{2/0.5} \leq t'_9(1.581139) \leq 2.262157\times 2]$$

$$= \Pr[-4.524314 \leq t'_9(1.581139) \leq 4.524314] = 0.976860.$$

9.0 Conclusions and Future Research

Chapter 3 used the normal theory with known variances to prove results that already existed in Overlap literature, some of which were obtained through simulation. It was proved that for a nominal significance level $\alpha = 0.05$, the corresponding 95% overlapping CIs provide a much smaller LOS $\alpha' = 0.0055746$, which fully agrees with the value computed from Eq. (7) on p. 184 of Schenker *et al.* (2001). Schenker *et al.* provide their results without any proof. Further, Chapter 3 proved that for a LOS of 0.01, the corresponding Overlap LOS was $\alpha' = 0.0002697$, while the literature provides results only for the nominal LOS of 5%. Further the smaller the LOS of the Standard method becomes, the larger is the % relative error of the Overlap LOS. Although, the Overlap literature has never considered the one-sided alternative, Chapter 3 showed that the Overlap LOS is ½ of the corresponding two-sided alternative (i.e., the Overlap procedure becomes even more conservative for a one-sided alternative).

Second, a concept that has not been discussed in Overlap literature is the maximum % overlap that the two independent CIs can have and H_0: $\mu_x = \mu_y$ cannot still be rejected at a pre-assigned LOS α. It was proven that this maximum % overlap depends only on the *SE* ratio [$k = (\sigma_y / \sqrt{n_y}) / (\sigma_x / \sqrt{n_x})$ or $k = (\sigma_x / \sqrt{n_x}) / (\sigma_y / \sqrt{n_y})$] and is equal to 17.1573% at k = 1 and diminishes to zero as k $\to\infty$ or zero. At k = 10, it was shown that the maximum % overlap reduces to 4.5137%

so that the Overlap procedure converges to an exact α-level test for limiting values of k.

Third, the chapter showed that the two independent CIs must each have a confidence level of $1-\gamma = 1-2\Phi(-Z_{\alpha/2}/\sqrt{2})$ in order to provide an exact α-level test. This last formula gives a confidence levels of 0.931452 for both independent intervals at $\alpha = 0.01$, and $1-\gamma = 0.83422373$ at $\alpha = 0.05$. This latter value is in perfect agreement with Overlap literature while the former value of $1-\gamma = 0.931452$ has not been reported.

Finally, the Overlap procedure leads to less statistical power compared to the Standard method and its RELEFF for small sample sizes is poor and heavily depends on δ/σ, but its asymptotic RELEFF is 100% as $n \to \infty$. For the simplest case of $\sigma_x = \sigma_y$ and $n_x = n_y$ an exact formula (15e) was obtained for the RELEFF of Overlap to the Standard method.

Chapter 4 investigated the Bonferroni Overlap CIs against the Standard procedure and determined that the Bonferroni concept makes the Overlap even more conservative and loses more statistical power.

Chapter 5 examined the overlapping CIs for two process variances against the Standard method that uses the Fisher's F distribution; the Overlap literature has not investigated the Overlap procedure for variance ratios. As in the case of process means, the Overlap reduces the LOS of the test and the limiting value of α' at $\alpha = 0.05$ and k = 1 is roughly 0.0055746, while as $k \to \infty$ or zero, the Overlap approaches an exact α-level test.

Second, the limiting value of maximum % overlap that does not reject H_0: $\sigma_x = \sigma_y$ is exactly 17.15726% as was in the case of two process means.

Third, the individual confidence levels have to be set at γ obtained from Eq. (31b)

$\chi^2_{\gamma/2,\nu_x} / \chi^2_{1-\gamma/2,\nu_y} = \nu_x F_{\alpha/2,\nu_x,\nu_y} / \nu_y$, where the limiting value of γ is 0.165766 at k =1 just like the case of means; further, as k → ∞ or zero, γ → 0.

Last, the power of Overlap procedure is always less than Standard but approaches that of the Standard as k → ∞ or zero. The asymptotic RELEFF of Overlap to the Standard method is 100% as n_x & n_y →∞.

Chapter 6 examined the impact of Overlap on type I error Pr, in the normal case with unknown variances but samples sizes ≤ 50, using the pooled t and two-independent sample t statistics, and also the effect of positive and negative correlations on the Overlap procedure. Specific formulas for $\alpha'$of the pooled t-test (37c), the two-independent sample t-test (41b), and the paired t-test (44a) were derived and documented. The Overlap literature has not considered the pooled t-test.

Chapter 7 used the pooled t-statistic to derive an expression for the % overlap, ω_r, below which H_0: $\mu_x = \mu_y$ cannot be rejected at the α level. Unlike the simple case of known variances where ω_r depends only on the SE ratio k, when the process variances are unknown and sample sizes are not large, ω_r depends on n_x, n_y, $F_0 = S_x^2 / S_y^2$, and α. For the case of two-independent t-statistic, ω_r depends on n_x, n_y, k, and α, while for the paired t-test, it depends only on the correlation coefficient between X and Y and $F_0 = S_x^2 / S_y^2$. For all 3 cases, Chapter 7 also derived expressions for the individual confidence levels, 1−γ, that provide an α-level test by the Overlap method. In the case of pooled t-test, γ depends only on n_x, n_y, and F_0. For the two-independent sample t-test γ depends on n_x, n_y and F_0. While for the paired t-test, it depends only on n, r and F_0.

Chapter 8 used the noncentral t-distribution to derive formulas for the OC curves (and also power functions) for the case of underlying normal distribution with unknown variances and moderate to small sample sizes $n \le 50$, the results of which have been available in statistical literature for more than 35 years. However, the chapter also derived formulas for type II error Pr of Overlap (β') using the noncentral t. The exact results obtained for this latter case have not been available in statistical literature.

As further research, one could consider the Overlap problem for other normal parameters, such as the coefficient of variation σ/μ [see Vangel (1996) and Payton (1996)] and quantiles $\mu + Z_\alpha\sigma$, $0 < \alpha < 1$. Further, we suspect that the SMD of S (= the standard deviation for a sample) from a non-normal population approaches normality, toward $N[\sigma, \sigma^2/(2n)]$, but very agonizingly slowly ($n > 100$). The exact SMD of S from a $N(\mu, \sigma^2)$ has been documented in statistical literature more than 50 years ago. For an underlying normal population, it is also widely known that an $n > 75$ is needed in order for S to be roughly normally distributed according to $N[c_4\sigma, (1 - c_4^2)\sigma^2]$, where $c_4 = \Gamma(n/2)\sqrt{2}$ / $\{\Gamma[(n-1)/2] \times \sqrt{n-1}\} < 1$ is a well-known QC constant. Note that the approximate V(S) generally reported in statistical literature is $\sigma^2/(2n)$, but we know for fact that $\sigma^2/[2(n - 0.745)]$ is a better approximation to the exact variance of S from a $N(\mu, \sigma^2)$, which is given by $V(S) = (1 - c_4^2)\sigma^2$. Unfortunately, the farther the skewness and kurtosis of an underlying non-normal distribution are from zero, the much larger sample size is needed for the SMD of S to exhibit normality. Thus, if the underlying distribution is non-normal, only the limiting comparison of Standard CI to Overlap may be accomplished based on CIs of σ_x and σ_y. Also, we have not yet seen the impact of overlapping CIs on parameters of other underlying distributions such as Uniform, Weibull, and Beta.

10.0 References

[1] Brownlee, K. A. (1965), *Statistical Theory and Methodology in Science and Engineering*, Wiley, NY.

[2] Cole, S. R. and Blair, R.C. (1999), Overlapping Confidence Intervals. *Journal of the American Academy of Dermatology,* 41(6), pp.1051-1052.

[3] Devore, J. L. (2008), *Probability and Statistics*, Thomson Brooks/Cole, Canada.

[4] Djordjevic, M. V., Stellman, S. D. and Zang, E.(2000), Doses of Nicotine and Lung Carcinogens Delivered to Cigarette Smokers. *Journal of the National Cancer Institute*, 92(2), pp.106-111.

[5] Goldstein H. Healy MJR. (1995), The Graphical Presentation of a Collection of Means. *Journal of the Royal Statistical Society A*,158: pp.175-177.

[6] Hool, J. N. and Maghsoodloo, S. (1980), Normal Approximation to Linear Combinations of Independently Distributed Random Variables. *AIIE Transactions*, 12, pp.140-144.

[7] Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995), *Continuous Univariate Distributions*, 2nd edition, John Wiley & Sons, Inc.

[8] Kelton, W. D., Sadowski, R. P. and Sturrock, D. T. (2004) *Simulation with Arena*, pp. 265-268, McGraw-Hill Companies, Inc., NY.

[9] Kendall, M. G. and Stuart, A. (1963) The Advanced Theory of Statistics, Charles Griffin & Company Limited, London.

[9] Maghsoodloo S., and Hool, J. N. (1981), On Normal Approximation of Simple Linear Combinations. *The Journal of the ALABAMA ACADEMY of SCIENCES*, 52, October 1981, No. 4, pp.207-219.

[10] Mancuso, C. A., Peterson, M.G. E. and Charlson, M. E. (2001) Comparing Discriminative Validity Between a Disease-Specific and a General Health Scale in Patients with Moderate Asthma. *Journal of Clinical Epidemiology*, 54, pp.263-274.

[11] Montgomery D. C. and Runger G. C. (1994), *Applied Statistics and Probability for Engineers,* John Wiley & sons, Inc., p. 411.

[12] Payton, M. E., (1996), " Confidence intervals for the coefficient of variation," Proc. Kansas State Univ. Conf. Appl. Statistical Agriculture, 8, pp. 82-87.

[13] Payton, M. E., Miller, A. E. and Raun, W. R. (2000) Testing Statistical Hypotheses Using Standard Error Bars and Confidence Intervals. *Communication in Soil Science and Plant Analysis*, 31, pp.547-552.

[14] Payton, M. E., Greenstone, M. H. and 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, pp.34-39

[15] Schenker, N. and Gentleman, J. E. (2001) On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals. *The American Statistician*, 55, pp.182-186.

[16] Sont, W. N., Zielinski, J. M., Ashmore, J. P., Jiang, H., Krewski, D., Fair, M. E., Band, P. R. and Letourneau, E. G. (2001) First Analysis of Cancer Incidence and Occupation Radiation Exposure Based on the National Dose Registry of Canada. *American Journal of Epidemiology*, 153(4), pp.309-318.

[17] Tersmette, A. C., Petersen, G. M., Offerhaus, G.J.A., Falatko, F. C., Brune, K. A., Goggins, M., Rozenblum, E., Wilentz, R. E., Yeo, C.J., Cameron, J. L., Kern, S. E. and Hruban, H. (2001) Increased Risk of Incident Pancreatic Cancer Among First-degree Relatives of Patients with Familial Pancreatic Cancer. *Clinical Cancer Research,* 7, pp.738-744.

[18] Vangel, M.G. (1996), " Confidence intervals for a normal coefficient of variation," *The American Statistician*, 50, pp. 21-26.

APPENDICES

Appendix A: The Kurtosis of the sum of n independent Uniform, U(0, 1), distributions.

Appendix B: Matlab functions

Appendix A: The Kurtosis of the sum of n independent uniform, U(0, 1), distributions

Suppose $x_1, x_2, \ldots, x_n$ are independently and each uniformly distributed over the real interval [0, 1]. It is well known that the 1st four moments of each x_i is given by $\mu_i = 1/2$, $\mu_2 = V(X_i) = 1/12 = \sigma^2$, $\mu_3 = 0$ (by symmetry), and $\mu_4 = 1/80$ so that $\alpha_4 = \mu_4/\sigma^4 = 144/80 = 1.80$ and the kurtosis of each x_i is equal to $\beta_4 = \alpha_4 - 3 = -1.20$.

Now consider the sum $Y_n = \sum_{i=1}^{n} x_i$; our objective is to compute the 1st four moments of Y_n from the known moment of each x_i, i =1,2,…, n. Clearly, the mean of Y_n is given by $E(Y_n) = n/2$, the variance is given by $V(Y_n) = nV(X_i) = n/12$, $\mu_3(Y_n) = 0$ by symmetry, and $\mu_4(Y_n)$ is computed below.

$$\mu_4(Y_n) = E[\sum_{i=1}^{n} x_i - (n/2)]^4 = E[\sum_{i=1}^{n}(x_i - 1/2)]^4$$

$$= E[\sum_{i=1}^{n}(x_i - 1/2)^4 + {}_4C_2 \times \sum_{i=1}^{n-1}\sum_{j>1}^{n}(x_i - 1/2)^2(x_j - 1/2)^2]$$

$$= n\mu_4(x_i) + 6 \times {}_nC_2 \times V(X_i) \times V(X_j)$$

Note that in the binomial expansion of $[\sum_{i=1}^{n}(x_i - 1/2)]^4$ the expectation of odd products such as $E[(x_1 - 1/2)(x_2 - 1/2)^3]$ vanish due to mutual independence of x_i and x_j for all $i \neq j$.

Hence, $\mu_4(Y_n) = n/80 + 3n(n-1)\sigma^4 = n/80 + 3n(n-1)/144 = n/80 + n(n-1)/48$

Thus, $\alpha_4(Y_n) = \dfrac{\mu_4(Y_n)}{V(Y_n)} = \dfrac{n/80 + n(n-1)/48}{(n/12)^2} = \dfrac{144/80 + 144(n-1)/48}{n}$

$$= \frac{1.80 + 3(n-1)}{n} \rightarrow \alpha_4(Y_n) = 3 - 1.20/n$$

$\rightarrow \beta_4(Y_n) = \alpha_4(Y_n) - 3 = -1.20/n$

Thus for a 2-fold convolution of U(0, 1), the kurtosis is $-1.20/n = -0.60$ while for a 6-fold convolution the kurtosis of $\sum_{i=1}^{6} x_i$ is equal to $-1.20/n = -0.20$.

Appendix B: Matlab functions

(a) The following Three Matlab functions compute the Overlap significance level, α', for a pooled t-test, two-independent t-test, and the paired t-test, respectively, at a given significance level α = a, sample sizes n_x & n_y and sample variance ratio $F_0 = S_x^2 / S_y^2$.

1.
```
function y = aprP(a,nx,ny,F0);
tx=tinv(1-a/2,nx-1);ty=tinv(1-a/2,ny-1);nu = nx+ny-2;
RHS=nu*(tx*sqrt(F0*ny)+ty*sqrt(nx))^2/((ny-1+F0*(nx-1))*(nx+ny));
 y=1-fcdf(RHS,1,nu);
```

2.
```
function y = apr(a,nx,ny,F0);
tx=tinv(1/a/2,nx-1); ty=tinv(1-a/2,ny-1);Rn=ny/nx; nu=
(nx-1)*(ny-1)*(1+F0*Rn)^2/(nx-1+(ny-1)*(F0*Rn)^2);
RHS=(tx*sqrt(F0*Rn)+ty)^2/(1+F0*Rn);
y=1-fcdf(RHS,1,nu);
```

3.
```
function y = aprc(a,n,F0,r);
F1=finv(1-a,1,n-1);
RHS=F1*(sqrt(F0)+1)^2/(1+F0-2*r*sqrt(F0));
y=1-fcdf(RHS,1,n-1);
```

(b) The following Matlab functions compute the overlap proportion for a pooled t-test, two-independent sample t-test, and the paired test, respectively at a given significance level α = a, sample sizes n_x & n_y and sample variance ratio $F_0 = S_x^2 / S_y^2$.

1. function y = OmegaP(a,nx,ny,F0);

```
Rn = ny/nx;nu=nx+ny-2;n1=nx-1; n2=ny-1;
NUM = tinv(1-a/2,n1)*sqrt(F0*Rn)+tinv(1-a/2,n2)-
tinv(1- a/2,nu)*sqrt((1+Rn)*(n1*F0+n2)/nu);
DEN = tinv(1-a/2,n1)*sqrt(F0*Rn)+tinv(1-a/2,n2)+tinv(1-
a/2,nu)*sqrt((1+Rn)*(n1*F0+n2)/nu);
y= NUM./DEN;
```

2. function y = Omega(a,nx,ny,F0);

```
Rn= ny/nx;n1= nx-1; n2=ny-1;
nu=(n1*n2*(F0*Rn+1)^2)/(n2*(Rn*F0)^2+n1);
NUM=tinv(1-a/2,n1)*sqrt(Rn*F0)+tinv(1-a/2,n2)-tinv(1-a/2,nu)*sqrt(F0*Rn +1);
DEN = tinv(1-a/2,n1)*sqrt(Rn*F0)+tinv(1-a/2,n2)+tinv(1-a/2,nu)*sqrt(F0*Rn +1);
y= NUM./DEN;
```

3. function y = OmegaC(F0, r);

```
NUM = sqrt(F0)+1-sqrt(1+F0-2*r*sqrt(F0));
DEN = sqrt(F0)+1+sqrt(1+F0-2*r*sqrt(F0));
y= NUM./DEN;
```

(c) The following Matlab codes compute the value of γ that provides an α-level test for the two-independent t-test.

```
a=0.05;
nx=4;
ny=8;
F0=1.5;
Rn= ny/nx;n1=nx-1;
n2=ny-1;
nu=(n1*n2*(F0*Rn+1)^2)/(n2*(Rn*F0)^2+n1);
RHS =tinv(1-a/2,nu)*sqrt(Rn*F0+1);
```

```
c(1)=a;
for i= 2:25
c(i)= c(i-1)+0.005
LHS(i) = tinv(1-c(i)/2,n2)+tinv(1-c(i)/2,n1)*sqrt(Rn*F0);
end
for i = 2: 25
if RHS-0.005<=LHS(i) & LHS(i) <= RHS + 0.005
break
g=c(i)
    end
end
```