Approximate sample size formulas for the two-sample trimmed mean test with unequal variances

Wei-Ming Luh¹* and Jiin-Huarng Guo²
¹Institute of Education, National Cheng Kung University, Taiwan
²Department of Mathematical Sciences, Chinese Air Force Academy, Taiwan

Yuen's two-sample trimmed mean test statistic is one of the most robust methods to apply when variances are heterogeneous. The present study develops formulas for the sample size required for the test. The formulas are applicable for the cases of unequal variances, non-normality and unequal sample sizes. Given the specified \( \alpha \) and the power \((1 - \beta)\), the minimum sample size needed by the proposed formulas under various conditions is less than is given by the conventional formulas. Moreover, given a specified size of sample calculated by the proposed formulas, simulation results show that Yuen's test can achieve statistical power which is generally superior to that of the approximate \( t \) test. A numerical example is provided.

I. Introduction

Testing the equality of two means from independent samples is very common in scientific inquiry. Moreover, deciding on the sample size needed for the study is an important and serious business for planning research. Selecting an insufficient sample size will yield a study with inadequate sensitivity, whereas an excessive sample size will waste resources (Adcock, 1997; Levin, 1997; Muller, LaVange, Ramey, & Ramey, 1992; Sedlmeier & Gigerenzer, 1989). Several textbooks on sample size determination and power analysis are available (Cohen, 1988; Desu & Raghavarao, 1990; Kraemer & Thiemann, 1987).

Given the specified \( \alpha \) and the power \((1 - \beta)\), the sample size formula for a two-sided test is

\[
n_1 = n_2 = 2\sigma^2 \left( Z_{1-\alpha/2} + Z_{1-\beta} \right)^2 / d^2,
\]

(1)

*Correspondence should be addressed to Wei-Ming Luh, Institute of Education, National Cheng Kung University, Tainan City, 701, Taiwan, Republic of China (e-mail: luhwei@mail.ncku.edu.tw).
where $\sigma^2$ is the common variance, $Z$ refers to quantiles of the standard normal distribution, and $d = |\mu_1 - \mu_0|$ is the width of the tolerable region in which the discriminating power of the test is less than $1 - \beta$ (Mace, 1974, p. 78). It should be noted that for the case of unequal group sizes, the power of the test will be less than $1 - \beta$. The formula above is popular because of its stress on the magnitude of the effect of interest rather than statistical significance (Kupper & Hafner, 1989).

However, the standard formulas are usually for cases in which certain assumptions are met. Moreover, their use is limited to situations involving the well-known parametric statistics. Little attention has been given to sample size determination for other test statistics. If the variances are heterogeneous (i.e. $\sigma_1^2 \neq \sigma_2^2$), Mace (1974, p. 81) used Welch’s $t$ and developed formulas for the sample size needed for each group. He noted that the test of the highest discriminating power for a given total sample size would be obtained when $n_2/n_1 = \sigma_2/\sigma_1$ (Mace, 1974, p. 82). For a two-sided test,

$$n_1 = \sigma_1(\sigma_1 + \sigma_2)(Z_{1-\alpha/2} + Z_{1-\beta})^2/d^2,$$

and

$$n_2 = \sigma_2(\sigma_1 + \sigma_2)(Z_{1-\alpha/2} + Z_{1-\beta})^2/d^2.$$  \(\text{(2)}\)

Moreover, Schouten (1999) developed generalized sample size formulas for a two-sided test,

$$n_1 = \sigma_1(\sigma_1 + \sigma_2)(Z_{1-\alpha/2} + Z_{1-\beta})^2/d^2 + Z_{1-\alpha/2}^2/(2(1 + \sqrt{\tau})),$$

and

$$n_2 = \sigma_2(\sigma_1 + \sigma_2)(Z_{1-\alpha/2} + Z_{1-\beta})^2/d^2 + Z_{1-\alpha/2}^2\sqrt{\tau}/(2(1 + \sqrt{\tau})),$$

where $\tau = \sigma_2^2/\sigma_1^2$. Note the adjustment in the formulas by using the variance ratio (i.e. $\tau$) can improve the determination of sample size.

Yuen (1974) proposed a two-sample trimmed $t$ test for unequal variances for the case of symmetric, heavy-tailed distributions. Her simulation results showed that her test performed better than Welch’s (1938) approximate test. Luh and Guo (2000) and Guo and Luh (2000) extended Yuen’s method in conjunction with Johnson (1978) or Hall’s (1992) transformation to deal with heterogeneity and non-normality. Although trimmed mean methods are becoming popular, and statistical packages such as SAS have included them in their subroutines, the gaps in the availability of sample size determination procedures are most notable in robust statistics (Adcock, 1997). Wilcox (2003, pp. 259–261) used the percentile bootstrap method with 20% trimming for comparing two independent groups to estimate power, but not sample size. Luh, Olejnik, and Guo (2006) have already developed sample size formulas for one- and two-sample trimmed mean tests with homogeneous variances. Since trimmed mean methods have been recommended for the heterogeneous variance or non-normal conditions (Oosterhoff, 1994; Staudte & Sheather, 1990; Tukey & McLaughlin, 1963; Wilcox, 1994b; Yuen, 1974), their corresponding sample size determination procedures need to be developed. Therefore, analogous to Schouten’s
(1999) formulas, sample size formulas for Yuen’s two-sample trimmed mean test are developed in the present study.

The organization of the paper is as follows. Section 2 demonstrates the derivation of the proposed formulas for one- and two-sided tests. Section 3 presents an illustrative example to show the minimum sample size needed under various conditions for the proposed formulas and the conventional formulas. As to Yuen’s two-sample trimmed \( t \) test, Section 4 presents the computer simulation design and results. That is, given a specified size of sample calculated by the proposed formulas, the resulting Type I error and statistical power for Yuen’s test and the approximate \( t \) test are compared. A brief discussion concludes the paper.

2. Derivation of the proposed sample size formulas

Temporarily let \( x_i \) be the \( i \)th observation \((i = 1, \ldots, n)\), and \( X_{(1)} \leq \ldots \leq X_{(n)} \) be the order statistics. To calculate the trimmed mean, first let \( \gamma \) be the amount of trimming, \( 0 \leq \gamma < 0.5 \). Then \( e = [\gamma n] \), where \([x]\) is the greatest integer less than or equal to \( x \). The value of \( \gamma \) will be specified later. Let \( f = n - 2e \), the effective sample size. Then the sample trimmed mean is computed by removing the \( e \) largest and \( e \) smallest observations and averaging the remaining values:

\[
\bar{X}_t = \frac{1}{f} \sum_{i=e+1}^{n-e} X_{(i)}.
\]

The corresponding sample Winsorized mean is

\[
\bar{X}_w = \frac{1}{n} \sum Z_i,
\]

where

\[
Z_i = \begin{cases} 
X_{(e+1)} & \text{if } X_i \leq X_{(e+1)}, \\
X_i & \text{if } X_{(e+1)} < X_i < X_{(n-e)}, \\
X_{(n-e)} & \text{if } X_i \geq X_{(n-e)].}
\end{cases}
\]

The Winsorized sample variance is then

\[
S^2_w = \frac{1}{n} \sum (Z_i - \bar{X}_w)^2 / (n - 1),
\]

and let the trimmed sample variance be

\[
\hat{\sigma}^2_w = \frac{1}{f} \sum (Z_i - \bar{X}_w)^2 / (f - 1).
\]

For two independent groups, let \( X_{ij} \) be the \( i \)th random sample observation from the \( j \)th group, \( i = 1, \ldots, n_j, j = 1, 2 \). Let \( n_j \) be the sample size of the \( j \)th group. When trimmed means are being compared, the null hypothesis pertains to the equality of two population trimmed means:

\[
H_0 : \mu_{1t} = \mu_{2t}.
\]
In general, let \( f_j, \bar{X}_{ij}, \hat{\sigma}^2_{ij} \) be the values of \( f, \bar{X}, \hat{\sigma}^2 \) for the \( j \)th group. Then Yuen’s two-sample trimmed \( t \) statistic (Yuen, 1974) is defined as

\[
t_w = \frac{\bar{X}_{1t} - \bar{X}_{2t}}{\sqrt{\hat{\sigma}^2_{w1}/f_1 + \hat{\sigma}^2_{w2}/f_2}}.
\]  

The test statistic \( t_w \) is approximately Student \( t \) distributed with degrees of freedom

\[
\nu = \left( \hat{\sigma}^2_{w1}/f_1 + \hat{\sigma}^2_{w2}/f_2 \right)^2 / \left[ (\hat{\sigma}^2_{w1}/f_1)^2 / (f_1 - 1) + (\hat{\sigma}^2_{w2}/f_2)^2 / (f_2 - 1) \right].
\]

The test of the highest discriminating power for a given total sample size \( f_1 + f_2 \) will be obtained when \( f_2/f_1 = \hat{\sigma}_{w2}/\hat{\sigma}_{w1} \). Therefore, analogous to Schouten’s formulas, the effective sample sizes for a two-sided test are

\[
f_1 = \hat{\sigma}_{w1}(\hat{\sigma}_{w1} + \hat{\sigma}_{w2})(Z_{1-\alpha/2} + Z_{1-\beta})^2 / d^2 + Z_{1-\alpha/2}^2 / (2(1 + \sqrt{\tau_w})�)
\]  

and

\[
f_2 = \hat{\sigma}_{w2}(\hat{\sigma}_{w1} + \hat{\sigma}_{w2})(Z_{1-\alpha/2} + Z_{1-\beta})^2 / d^2 + Z_{1-\alpha/2}^2 \sqrt{\tau_w} / (2(1 + \sqrt{\tau_w})�)
\]

where \( d = |\mu_1 - \mu_2| \) and \( \tau_w = \hat{\sigma}_{w2}^2/\hat{\sigma}_{w1}^2 \). For the given effective sample sizes \( f_1 \) and \( f_2 \), the trimmed sample sizes \( n_{t1} \) and \( n_{t2} \) are

\[
n_{t1} = \left[ f_1 / (1 - 2\gamma) \right] + 1
\]

and

\[
n_{t2} = \left[ f_2 / (1 - 2\gamma) \right] + 1
\]

respectively, where \( [x] \) is the greatest integer less than or equal to \( x \).

3. Illustrative examples

This section shows the minimum sample size needed under various distribution shapes and heterogeneous variances for the proposed and conventional formulas. The following numerical examples all use simulated data. First, we considered five \( g \)-and-\( b \) distributions, \( (g = 0, b = 0), (g = 0, b = 0.1), (g = 0, b = 0.2), (g = 0.5, b = 0), (g = 0.5, b = 0.2) \), representing various skewed and/or heavy-tailed scenarios (Hoaglin, 1985, p. 503). The larger the \( b \) value, the heavier the distribution tail. The skewness and kurtosis for these distributions are \( (0, 0), (0, 5.5), (0, 36.22), (1.75, 8.9) \) and \( (13.16, 42.895.9) \), respectively (Wilcox, 1995b). The \( g \)-and-\( b \) family of distributions was generated by a single transformation of the standard normal to allow for a wide spectrum of distribution shapes (Hoaglin, 1985). Let

\[
\varepsilon = g^{-1}(\exp(gZ) - 1) \exp(bZ^2/2) \text{ when } g = 1 \varepsilon = Z \exp(bZ^2/2).
\]

Then, the simulated observation was

\[
X = \mu + \sqrt{\tau} \times \sigma \times \varepsilon.
\]

For the normal case, we considered three variance patterns: \( (1, 1), (1, 4) \) and \( (1, 9) \). Therefore, \( \tau \) was set to \( 1, 4 \) and \( 9 \) to give homogeneous and heterogeneous variance cases as well. For the non-normal cases, we needed to calculate the variances by using
Equations (2.5) and (2.6) in Martinez and Iglewicz (1984). The resulting variances are listed in column 2 of Table 1.

Table 1. Calculated sample size \( n \) and trimmed sample size \( n_t \) for the one-sided case

<table>
<thead>
<tr>
<th>((g, h))</th>
<th>((\sigma_1^2, \sigma_2^2))</th>
<th>((\hat{\sigma}<em>{\omega 1}^2, \hat{\sigma}</em>{\omega 2}^2))</th>
<th>(n_{1}, n_{2})</th>
<th>(n_{1t}, n_{2t})</th>
</tr>
</thead>
<tbody>
<tr>
<td>((0, 0))</td>
<td>((1, 1))</td>
<td>18, 18</td>
<td>(0.69, 0.69)</td>
<td>21, 21</td>
</tr>
<tr>
<td>((1.4))</td>
<td>27, 53</td>
<td>(0.69, 2.76)</td>
<td>31, 61</td>
<td></td>
</tr>
<tr>
<td>((1.9))</td>
<td>35, 104</td>
<td>(0.69, 6.21)</td>
<td>40, 120</td>
<td></td>
</tr>
<tr>
<td>((0, 0.1))</td>
<td>((1.4, 1.4))</td>
<td>25, 25</td>
<td>(0.75, 0.75)</td>
<td>23, 23</td>
</tr>
<tr>
<td>((1.4, 5.6))</td>
<td>37, 73</td>
<td>(0.75, 3.0)</td>
<td>33, 66</td>
<td></td>
</tr>
<tr>
<td>((1.4, 12.6))</td>
<td>49, 145</td>
<td>(0.75, 6.75)</td>
<td>44, 131</td>
<td></td>
</tr>
<tr>
<td>((0, 0.2))</td>
<td>((2.15, 2.15))</td>
<td>38, 38</td>
<td>(0.81, 0.81)</td>
<td>25, 25</td>
</tr>
<tr>
<td>((2.15, 8.61))</td>
<td>56, 112</td>
<td>(0.81, 3.24)</td>
<td>36, 71</td>
<td></td>
</tr>
<tr>
<td>((2.15, 19.36))</td>
<td>75, 223</td>
<td>(0.81, 7.29)</td>
<td>47, 141</td>
<td></td>
</tr>
<tr>
<td>((0.5, 0))</td>
<td>((1.46, 1.46))</td>
<td>26, 26</td>
<td>(0.75, 0.75)</td>
<td>23, 23</td>
</tr>
<tr>
<td>((1.46, 5.84))</td>
<td>38, 76</td>
<td>(0.75, 3.01)</td>
<td>33, 66</td>
<td></td>
</tr>
<tr>
<td>((1.46, 13.13))</td>
<td>51, 151</td>
<td>(0.75, 6.77)</td>
<td>44, 131</td>
<td></td>
</tr>
<tr>
<td>((0.5, 0.2))</td>
<td>((4.18, 4.18))</td>
<td>73, 73</td>
<td>(0.89, 0.89)</td>
<td>27, 27</td>
</tr>
<tr>
<td>((4.18, 16.73))</td>
<td>108, 216</td>
<td>(0.89, 3.56)</td>
<td>39, 78</td>
<td></td>
</tr>
<tr>
<td>((4.18, 37.65))</td>
<td>144, 432</td>
<td>(0.89, 8.01)</td>
<td>52, 155</td>
<td></td>
</tr>
</tbody>
</table>

Finally, we calculated the corresponding trimmed variances \((\hat{\sigma}_{\omega 1}^2, \hat{\sigma}_{\omega 2}^2)\) for 20% trimming by using (9). The resulting variances are listed in column 4 of Table 1. The reason why we focused on 20% symmetric trimming was that it was recommended by Rosenberg and Gasko (1983) and Wilcox (1997).

Suppose we set \(\alpha = .05\), \(1 - \beta = .90\), \(d = d_t = 1\) and \(\tau = \tau_\omega = 2\). With \((g = 0, h = 0)\) and \((\hat{\sigma}_{\omega 1}^2, \hat{\sigma}_{\omega 2}^2) = (1, 4)\), from Equations (4) and (5) (for the one-sided case, replacing \(\alpha/2 \) by \(\alpha\)), we have

\[
n_1 = 1(1 + 2)(1.645 + 1.282)^2/1 + 1.645^2/(2(1 + 2)) = 27,
\]

\[
n_2 = 2(1 + 2)(1.645 + 1.282)^2/1 + 1.645^2 \times 2/(2(1 + 2)) = 53.
\]

From Equations (11) and (12), \((\hat{\sigma}_{\omega 1}^2, \hat{\sigma}_{\omega 2}^2) = (0.69, 2.76)\),

\[
f_1 = \sqrt{0.69} (\sqrt{0.69} + \sqrt{2.76})(1.645 + 1.282)^2/1 + 1.645^2/(2(1 + 2)) = 18.19,
\]

\[
f_2 = \sqrt{2.76} (\sqrt{0.69} + \sqrt{2.76})(1.645 + 1.282)^2/1 + 1.645^2 \times 2/(2(1 + 2)) = 36.37.
\]

Then, from Equations (13) and (14), we have

\[
n_{1t} = [18.19/(1 - 2(0.2))] + 1 = 31.
\]

and

\[
n_{2t} = [36.37/(1 - 2(0.2))] + 1 = 61.
\]

We calculated the sample sizes needed for the one-sided case under various conditions in the same manner and the results are presented in Table 1. In untrimmed cases, \((n_1, n_2)\) are calculated from Equations (4) and (5) (replacing \(\alpha/2 \) by \(\alpha\)), respectively. Trimmed sample sizes \((n_{1t}, n_{2t})\) are calculated from Equations (11), (12), (13) and (14), respectively. When the variances are equal and the distributions are...
normal \((g = 0, b = 0)\), it is known that the trimmed mean method requires larger sample sizes than the conventional \(t\). It is also true that when the variance increases, \(n_t\) is still larger than \(n\). However, when the distribution is non-normal, \(n_t\) is smaller than \(n\). The more extreme the distribution shape, the smaller the \(n_t\) needed compared to \(n\). The dramatic increase in efficiency gained by applying the proposed formulas can greatly reduce the cost of sampling.

It should also be noted that the larger the variance, the larger the sample size needed. Moreover, when the variances are equal, the sample sizes needed for each group are also equal. Table 2 shows the sample sizes for the two-sided cases; the pattern of the results is consistent with Table 1.

**Table 2.** Calculated sample size \(n\) and trimmed sample size \(n_t\) for the two-sided case

<table>
<thead>
<tr>
<th>((g, h))</th>
<th>((\sigma_1^2, \sigma_2^2))</th>
<th>(n_1, n_2)</th>
<th>((\hat{\sigma}<em>{w1}^2, \hat{\sigma}</em>{w2}^2))</th>
<th>(n_{t1}, n_{t2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>((0, 0))</td>
<td>((1, 1))</td>
<td>22, 22</td>
<td>(0.69, 0.69)</td>
<td>26, 26</td>
</tr>
<tr>
<td></td>
<td>((1.4))</td>
<td>33, 65</td>
<td>(0.69, 2.76)</td>
<td>38, 75</td>
</tr>
<tr>
<td></td>
<td>((1.9))</td>
<td>43, 128</td>
<td>(0.69, 6.21)</td>
<td>50, 148</td>
</tr>
<tr>
<td>((0, 0.1))</td>
<td>((1.4, 1.4))</td>
<td>31, 31</td>
<td>(0.75, 0.75)</td>
<td>28, 28</td>
</tr>
<tr>
<td></td>
<td>((1.4, 5.6))</td>
<td>45, 90</td>
<td>(0.75, 3.0)</td>
<td>41, 81</td>
</tr>
<tr>
<td></td>
<td>((1.4, 12.6))</td>
<td>60, 178</td>
<td>(0.75, 6.75)</td>
<td>54, 161</td>
</tr>
<tr>
<td>((0, 0.2))</td>
<td>((2.15, 2.15))</td>
<td>47, 47</td>
<td>(0.81, 0.81)</td>
<td>30, 30</td>
</tr>
<tr>
<td></td>
<td>((2.15, 8.61))</td>
<td>69, 137</td>
<td>(0.81, 3.24)</td>
<td>44, 88</td>
</tr>
<tr>
<td></td>
<td>((2.15, 19.36))</td>
<td>91, 273</td>
<td>(0.81, 7.29)</td>
<td>58, 173</td>
</tr>
<tr>
<td>((0.5, 0))</td>
<td>((1.46, 1.46))</td>
<td>32, 32</td>
<td>(0.75, 0.75)</td>
<td>28, 28</td>
</tr>
<tr>
<td></td>
<td>((1.46, 5.84))</td>
<td>47, 94</td>
<td>(0.75, 3.01)</td>
<td>41, 82</td>
</tr>
<tr>
<td></td>
<td>((1.46, 13.13))</td>
<td>62, 186</td>
<td>(0.75, 6.77)</td>
<td>54, 161</td>
</tr>
<tr>
<td>((0.5, 0.2))</td>
<td>((4.18, 4.18))</td>
<td>89, 89</td>
<td>(0.89, 0.89)</td>
<td>33, 33</td>
</tr>
<tr>
<td></td>
<td>((4.18, 16.73))</td>
<td>133, 266</td>
<td>(0.89, 3.56)</td>
<td>48, 96</td>
</tr>
<tr>
<td></td>
<td>((4.18, 37.65))</td>
<td>177, 530</td>
<td>(0.89, 8.01)</td>
<td>64, 190</td>
</tr>
</tbody>
</table>

### 4. Simulation design and results

To evaluate the validity of the proposed formulas, we performed extensive computer simulations for a wide range of conditions (keeping constants \(\alpha = .05, 1 - \beta = .90, \gamma = 0.2\)), and we ran 10,000 replications for each experiment to ensure that stability had been reached (Robey & Barcikowski, 1992; Serlin, 2000). The SAS RANNOR function (SAS Institute, 1999) was used to produce the simulated observations. The same five distribution shapes were selected as before. We generated data by using (15) and (16), and set \(\tau = \tau_w = 1, 4, \) or 9 respectively. In total, there were five distribution shapes and three variance patterns, resulting in 15 configurations.

It was noted that when comparing means, the \(g\)-and-\(b\) distribution has a mean of zero when \(g = 0\), so multiplying each \(\epsilon_y\) by \(\sigma_j\) to obtain unequal variances did not affect the null hypothesis. For \(g > 0\), the mean of the \(g\)-and-\(b\) distribution is

\[
\mu_{gb} = \left(\exp\left\{g^2/(2(1 - b))\right\} - 1\right)/\left\{g(1 - b)^{1/2}\right\}
\]

(Hoaglin, 1985, p. 503; Martinez & Iglewicz, 1984; Wilcox, 1994a). In this case, \(\mu_{gb}\) should be subtracted from \(\epsilon_y\) before multiplying by \(\sigma_j\). When dealing with trimmed means, the data generation can be referred to Luh and Guo (2000). Based on the size of
the trimmed sample needed under the specific condition, the generated data were first
tested to investigate the empirical Type I error (setting \( d = d_t = 0 \)) of the
corresponding Welch’s \( t \) and Yuen’s \( t_w \), respectively. Every empirical Type I error was
recorded for each experiment and then the average of 10,000 replications was reported.
Second, by setting \( d = d_t = 1 \), the generated data were applied to Welch’s \( t \) and Yuen’s
\( t_w \), respectively, to obtain the empirical power value. The average value from 10,000
replications was then reported.

Table 3 shows the resulting Type I error and power of the test statistics \( t \) and \( t_w \) when
the sample size \((n_{11}, n_{12})\) is drawn for the one-sided case. Note that the \( t \) is liberal in
terms of Type I error when the distribution is \((g = 0.5, b = 0)\) or \((g = 0.5, b = 0.2)\). On
the other hand, the \( t_w \) is more consistent in retaining the Type I error at around the .05
level. Moreover, the resulting power of \( t_w \) can achieve the nominal value of .9 but the \( t \)
cannot, because of undersampling. The more extreme the distribution shape is, the
greater the loss of the empirical power for \( t \). Finally, for the two-sided case (see Table 4),
note that the empirical power of the \( t \) does not achieve the predetermined power
whereas the \( t_w \) does.

### Table 3. For trimmed sample size \((n_{11}, n_{12})\), the resulting Type I error (\(\hat{\alpha}\)) and power of Welch’s \( t \) and Yuen’s \( t_w \) for the one-sided case

<table>
<thead>
<tr>
<th>((g, h))</th>
<th>((\sigma_1^2, \sigma_2^2))</th>
<th>((n_{11}, n_{12}))</th>
<th>(\hat{\alpha})</th>
<th>(\text{Power})</th>
</tr>
</thead>
<tbody>
<tr>
<td>((0, 0))</td>
<td>(1, 1)</td>
<td>(21, 21)</td>
<td>4.85</td>
<td>4.99</td>
</tr>
<tr>
<td></td>
<td>(1, 4)</td>
<td>(31, 61)</td>
<td>4.74</td>
<td>4.70</td>
</tr>
<tr>
<td></td>
<td>(1, 9)</td>
<td>(40, 120)</td>
<td>4.75</td>
<td>4.73</td>
</tr>
<tr>
<td>((0, 0.1))</td>
<td>(1.4, 1.4)</td>
<td>(23, 23)</td>
<td>4.90</td>
<td>5.09</td>
</tr>
<tr>
<td></td>
<td>(1.4, 5.6)</td>
<td>(33, 66)</td>
<td>5.10</td>
<td>4.99</td>
</tr>
<tr>
<td></td>
<td>(1.4, 12.6)</td>
<td>(44, 131)</td>
<td>5.17</td>
<td>5.02</td>
</tr>
<tr>
<td>((0, 0.2))</td>
<td>(2.15, 2.15)</td>
<td>(25, 25)</td>
<td>4.82</td>
<td>4.66</td>
</tr>
<tr>
<td></td>
<td>(2.15, 8.61)</td>
<td>(36, 71)</td>
<td>4.74</td>
<td>4.75</td>
</tr>
<tr>
<td></td>
<td>(2.15, 19.36)</td>
<td>(47, 141)</td>
<td>4.64</td>
<td>4.81</td>
</tr>
<tr>
<td>((0.5, 0))</td>
<td>(1.46, 1.46)</td>
<td>(23, 23)</td>
<td>5.07</td>
<td>4.94</td>
</tr>
<tr>
<td></td>
<td>(1.46, 5.84)</td>
<td>(33, 66)</td>
<td>5.62</td>
<td>5.35</td>
</tr>
<tr>
<td></td>
<td>(1.46, 13.13)</td>
<td>(44, 131)</td>
<td>5.82</td>
<td>5.57</td>
</tr>
<tr>
<td>((0.5, 0.2))</td>
<td>(4.18, 4.18)</td>
<td>(27, 27)</td>
<td>4.76</td>
<td>4.66</td>
</tr>
<tr>
<td></td>
<td>(4.18, 16.73)</td>
<td>(39, 78)</td>
<td>6.53</td>
<td>5.38</td>
</tr>
<tr>
<td></td>
<td>(4.18, 37.65)</td>
<td>(52, 155)</td>
<td>7.19</td>
<td>5.71</td>
</tr>
</tbody>
</table>

### 5. Conclusion and discussion

Today, there are many robust methods for achieving more accurate confidence intervals,
control over the probability of a Type I error, and for revealing important differences
that are missed by conventional methods based on means (Wilcox, 1995a). However,
although the trimmed mean methods are well known and easy to apply, the
determination of sample size for trimmed means with Winsorized variance has not
yet been fully studied. Therefore, the present study develops sample size formulas for
the heterogeneous two-sample Yuen’s trimmed \( t_w \) test to fill the gap.
The proposed sample size formulas are applicable for the cases of unequal variances, non-normality and unequal sample sizes. Based on the formulas, if the distribution is normal, applying Yuen’s method will result in a slightly larger sample size than the conventional method. However, if the distribution shape is heavy-tailed, the advantages of applying Yuen’s method are (i) the correction of assumption violations, and (ii) the total sample size needed with trimming is actually much smaller than the conventional $t$ for the specified $\alpha$ and power ($1-\beta$). Yuen’s trimmed mean method and the proposed sample size formulas are easily applied and this paper recommends them both.

As to the future, sample size formulas for one- and two-way analysis of variance are currently being investigated. Sample size determinations on confidence intervals for one- and two-sample problems are also being considered.

**Acknowledgements**

This research was supported by a National Sciences Council grant (NSC-93-2413-H-006-03) Taiwan, ROC, to the first author. The authors would like to express their gratitude to the editor and referees for their helpful comments.

**References**


---

**Table 4.** For trimmed sample size $(n_{t1}, n_{t2})$, the resulting Type I error ($\hat{\alpha}$) and power of Welch’s $t$ and Yuen’s $t_w$ for the two-sided case

<table>
<thead>
<tr>
<th>$(g, h)$</th>
<th>$(\sigma^2_1, \sigma^2_2)$</th>
<th>$(n_{t1}, n_{t2})$</th>
<th>$\hat{\alpha}$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$t$</td>
<td>$t_w$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$t$</td>
<td>$t_w$</td>
</tr>
<tr>
<td>(0, 0)</td>
<td>(1, 1)</td>
<td>(26, 26)</td>
<td>5.16</td>
<td>5.53</td>
</tr>
<tr>
<td></td>
<td>(1, 4)</td>
<td>(38, 75)</td>
<td>4.95</td>
<td>5.39</td>
</tr>
<tr>
<td></td>
<td>(1, 9)</td>
<td>(50, 148)</td>
<td>4.63</td>
<td>4.67</td>
</tr>
<tr>
<td>(0, 0.1)</td>
<td>(1.4, 1.4)</td>
<td>(28, 28)</td>
<td>4.99</td>
<td>5.24</td>
</tr>
<tr>
<td></td>
<td>(1.4, 5.6)</td>
<td>(41, 81)</td>
<td>4.77</td>
<td>4.76</td>
</tr>
<tr>
<td></td>
<td>(1.4, 12.6)</td>
<td>(54, 161)</td>
<td>4.94</td>
<td>4.99</td>
</tr>
<tr>
<td>(0, 0.2)</td>
<td>(2.15, 2.15)</td>
<td>(30, 30)</td>
<td>5.22</td>
<td>4.74</td>
</tr>
<tr>
<td></td>
<td>(2.15, 8.61)</td>
<td>(44, 88)</td>
<td>5.12</td>
<td>5.06</td>
</tr>
<tr>
<td></td>
<td>(2.15, 19.36)</td>
<td>(58, 173)</td>
<td>5.03</td>
<td>4.89</td>
</tr>
<tr>
<td>(0.5, 0)</td>
<td>(1.46, 1.46)</td>
<td>(28, 28)</td>
<td>4.74</td>
<td>4.42</td>
</tr>
<tr>
<td></td>
<td>(1.46, 5.84)</td>
<td>(41, 82)</td>
<td>5.06</td>
<td>5.22</td>
</tr>
<tr>
<td></td>
<td>(1.46, 13.13)</td>
<td>(54, 161)</td>
<td>4.80</td>
<td>5.24</td>
</tr>
<tr>
<td>(0.5, 0.2)</td>
<td>(4.18, 4.18)</td>
<td>(33, 33)</td>
<td>4.37</td>
<td>4.40</td>
</tr>
<tr>
<td></td>
<td>(4.18, 16.73)</td>
<td>(48, 96)</td>
<td>4.89</td>
<td>4.71</td>
</tr>
<tr>
<td></td>
<td>(4.18, 37.65)</td>
<td>(64, 190)</td>
<td>4.87</td>
<td>4.61</td>
</tr>
</tbody>
</table>


Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika, 29*, 350–362.


Received 19 April 2005; revised version received 16 January 2006