The Sample Size Needed for the Trimmed t Test When One Group Size is Fixed

Author(s): Wei-Ming Luh and Jiin-Huarng Guo

Source: *The Journal of Experimental Education*, 2010, Vol. 78, No. 1 (2010), pp. 14-25
Published by: Taylor & Francis, Ltd.
Stable URL: https://www.jstor.org/stable/27785550

The Journal of Experimental Education, 2010, 78, 14–25

ISSN: 0022-0973 print
DOI: 10.1080/00220970903224578

The Sample Size Needed for the Trimmed *t* Test When One Group Size is Fixed

Wei-Ming Luh
National Cheng Kung University, Taiwan

Jiin-Huarng Guo
National Pingtung University of Education, Taiwan

The sample size determination is an important issue for planning research. However, limitations in size have seldom been discussed in the literature. Thus, how to allocate participants into different treatment groups to achieve the desired power is a practical issue that still needs to be addressed when one group size is fixed. The authors focused on the sample size determination for the second group for the two-sample K. Yuen's (1974) trimmed mean test in the condition of heterogeneous variances. The results show that the sample size needed is less than that of the traditional B.L. Welch's test (1938), especially for nonnormal distributions. Simulation results also demonstrate the accuracy of the proposed formula in terms of Type I error and statistical power.

Keywords: nonnormal distribution, power, robust statistic, Winsorization, Yuen's test

RECENTLY DEVELOPED ROBUST METHODS for data analysis expand the options available for research design. One of these is the trimmed mean, which can be a robust estimate for nonnormal distributions that appear frequently in educational and psychological settings (Micceri, 1989). Many researchers have also recommended the trimmed mean method in the case of heterogeneous variance (Guo & Luh, 2000; Luh & Guo, 2000; Staudte & Sheather, 1990, p. 105; Tukey & McLaughlin, 1963; Wilcox, 1994; Wilcox & Keselman, 2003) which is, again, fairly common for real data (Keselman et al., 1998; Wilcox,

Address correspondence to Wei-Ming Luh, Institute of Education, National Cheng Kung University, Tainan, Taiwan. E-mail: luhwei@mail.ncku.edu.tw

1987). A trimmed mean is based on order statistics and is close to the bulk of scores, representing a compromise between the sample mean and median. The trimmed mean method computes the Winsorized variance by replacing the most extreme observations with less extreme values in both tails of the distribution so that the resulting standard error is much smaller and less affected by nonnormality.

In educational studies, the basic design for a comparative study is a randomized, balanced, two-group parallel design. Although most statistics textbooks show the sample size formulas or introduce some power tables (see Cohen 1988; Kraemer & Thiemann, 1987; Mace, 1974), the determination of the sample size for the trimmed mean method still lacks comprehensive investigation. Although Luh and Guo (2007) and Luh, Olejnik, and Guo (2008) have already developed the sample size determination method for the two-sample trimmed mean, some practical or ethical considerations may constrain the feasibility of the balanced samples. In clinical research, special education, or school studies, one group size may be fixed by circumstances, so the sample size is restricted to another group for treatment comparison (e.g., Heilbrun & McGee, 1985; Hsu, 1993). For example, if we only have $n_2 = 74$ incidence cases available, what is the size required to detect a specified difference in mean for a specified α and $1—\beta$? Unequal-sized groups are not avoidable. Although the literature seldom discusses the limitation in sample size, we think it is practical and important to investigate how to allocate participants into different groups to achieve the desired statistical power.

Heilbrun and McGee (1985) derived a formula for the required number of subjects in a comparison group to test the equality of two means when one sample size is fixed. Considering the testing of the equality of two trimmed means in the context of heterogeneous variances, in the present study, we extended Heilbrun and McGee's work and demonstrated the sample size requirement for Yuen's (1974) two-sample trimmed t statistic when one group size is fixed. In the following three sections, we derive the proposed formula, provide a hypothetical problem to demonstrate the procedure of using the formula, and investigate—through computer simulation—the performance of the proposed formula in terms of Type I error and statistical power. Last, we discuss our findings.

DERIVATION OF THE SAMPLE SIZE FORMULA

Let n denote the sample size for trimming and temporarily consider $X_{(1)} \leq \ldots \leq Xn_t$ as the order statistics of the random sample $X_1 \ldots Xn_t$. Let λ be the proportion of trimming in each tail of the distribution ($0 \leq \lambda < 0.5$); and let $e = [\lambda n_t]$, where $[x]$ is the greatest integer less than or equal to x. Let $k = n_t – 2e$ be the effective sample size. The sample trimmed mean is then

$$\bar{X}_t = \frac{1}{k} \sum_{i=e+1}^{n_t - e} X_{(i)}. \tag{1}$$

Next, we calculate the corresponding sample Winsorized mean:

$$\bar{X}_w = \frac{1}{n_t} \sum_{i=1}^{n_t} Z_i, \tag{2}$$

where

$$Z_i = \begin{cases} X_{(e+1)} & \text{if } X_i \leq X_{(e+1)} \\ X_i & \text{if } X_{(e+1)} < X_i < X_{(n_t - e)} \\ X_{(n_t - e)} & \text{if } X_i \geq X_{(n_t - e)} \end{cases},$$

and let the trimmed sample variance be

$$s_w^2 = \sum_{i=1}^{n_t} (Z_i - \bar{X}_w)^2 / (k - 1). \tag{3}$$

In general, let k_j, $\bar{X}_{tj}$, s_{wj}^2 be the values of k, $\bar{X}_t$, s_w^2 for the jth group. When we compare two trimmed means for the case of heterogeneous variances, the null hypothesis pertains to the equality of population trimmed means H_0: $\mu_{t1} - \mu_{t2} = 0$ versus a one-sided alternative hypothesis H_1: $\mu_{t1} - \mu_{t2} > 0$. We can apply Yuen's two-sample trimmed t statistic, defined as

$$t_w = \frac{\bar{X}_{t1} - \bar{X}_{t2}}{\sqrt{s_{w1}^2/k_1 + s_{w2}^2/k_2}}. \tag{4}$$

The test statistic t_w has a distribution that can be approximated with the student t distribution with degrees of freedom

$$V_w = \frac{(s_{w1}^2/k_1 + s_{w2}^2/k_2)^2}{(s_{w1}^2/k_1)^2/(k_1 - 1) + (s_{w2}^2/k_2)^2/(k_2 - 1)}. \tag{5}$$

Now suppose the sample size for trimming in the first group n_{t1} is fixed; then the resulting effective sample size $k_1 = n_{t1}(1 - 2\lambda)$ is also fixed. Let $\Delta_w = \mu_{t1} - \mu_{t2}$, and we can derive the initial effective sample size for the second group as follows:

$$\frac{\Delta_w}{\sqrt{s_{w1}^2/k_1 + s_{w2}^2/k_2}} = z_{1-\alpha} + z_{1-\beta},$$

$$\frac{s_{w1}^2}{k_1} + \frac{s_{w2}^2}{k_2} = \frac{\Delta_w^2}{(z_{1-\alpha} + z_{1-\beta})^2}. \tag{6}$$

Then, the initial effective sample size of the second group is

$$k_2 = \frac{k_1 s_{w2}^2 (z_{1-\alpha} + z_{1-\beta})^2}{k_1 \Delta_w^2 - s_{w1}^2 (z_{1-\alpha} + z_{1-\beta})^2} \tag{7}$$

if $k_1 \Delta_w^2 - s_{w1}^2 (z_{1-\alpha} + z_{1-\beta})^2 > 0$.

Because the determination of sample size for the t test depends on the degrees of freedom, an iterative procedure is required. With Equation (5), we can calculate the degrees of freedom v_w and obtain values of, $t_{v_w,1-\alpha}$ and $t_{v_w,1-\beta}$ respectively. If $k_1 \Delta_w^2 - s_{w1}^2 (t_{v_w,1-\alpha} + t_{v_w,1-\beta})^2 > 0$, the final effective sample size k_2 is

$$k_2 = \frac{k_1 s_{w2}^2 (t_{v_w,1-\alpha} + t_{v_w,1-\beta})^2}{k_1 \Delta_w^2 - s_{w1}^2 (t_{v_w,1-\alpha} + t_{v_w,1-\beta})^2}. \tag{8}$$

Therefore, the sample size for trimming of the second group is

$$n_{t2} = k_2/(1 - 2\lambda). \tag{9}$$

It should be noted that the minimal effective sample size for the first group has to satisfy

$$k_1 \Delta_w^2 - s_{w1}^2 (z_{1-\alpha} + z_{1-\beta})^2 > 0.$$

That is, $k_1 > s_{w1}^2 (z_{1-\alpha} + z_{1-\beta})^2 / \Delta_w^2$. Usually these values are not integers, so we can round up to the next integer as $k_1 = [s_{w1}^2 (z_{1-\alpha} + z_{1-\beta})^2 / \Delta_w^2] + 1$. In contrast, if we have a large number of subjects—for example, $k_1 \to \infty$, from Equation (6)—we need only the minimum effective sample size of the second group $k_2 = [s_{w2}^2 (z_{1-\alpha} + z_{1-\beta})^2 / \Delta_w^2] + 1$. For a two-sided test, we simply replace α with $\alpha/2$ in the equations.

AN ILLUSTRATIVE EXAMPLE

Suppose the variances of the two comparison groups are heterogeneous, and the sample size for trimming for the first group is fixed at $n_{t1} = 30$. Assume the trimmed sample variances $s_{w1}^2 = 0.689$ and $s_{w2}^2 = 2.756$. Let the trimmed mean difference $|\mu_{t1} - \mu_{t2}| = 1$ and set $\alpha = .05$ for a two-sided test. How many subjects are needed for the second group to achieve the power of 80%?

For $n_{t1} = 30$, assuming the trimming proportion $\lambda = .2$ is appropriate (Rosenberger & Gasko, 1983), we first calculate the effective sample size $k_1 = 30 \times (1-2 \times .2) = 18$. To calculate the size of the second group, with Equation (7), we have the initial effective sample size

$$k_2 = \frac{18 \times 2.756 \times (1.96 + 0.8416)^2}{18 \times 1 - 0.689 \times (1.96 + 0.8416)^2} = 30.9215,$$

and the degrees of freedom

$$v_w = \frac{(s_{w1}^2/k_1 + s_{w2}^2/k_2)^2}{(s_{w1}^2/k_1)^2/(k_1 - 1) + (s_{w2}^2/k_2)^2/(k_2 - 1)} = 46.1568.$$

Then, by using SAS *tinv* function (SAS Institute, 2002), we have

$$t_{v_w,1-\alpha/2} = tinv(1 - \alpha/2, v_w, 0) = tinv(0.975, 46.1568, 0) = 2.01271,$$

and $t_{v_w,1-\beta} = 0.84948$. Therefore, the final effective sample size of the second group is

$$k_2 = \frac{k_1 s_{w2}^2 (t_{v_w,1-\alpha} + t_{v_w,1-\beta})^2}{k_1 \Delta_w^2 - s_{w1}^2 (t_{v_w,1-\alpha/2} + t_{v_w,1-\beta})^2} = 32.8915$$

and its corresponding sample size for trimming is $n_{t2} = 32.8915/0.6 = 54.8192$ (see Table 1, Condition 4).

To evaluate the accuracy of the proposed method, we calculate the resulting power as follows. First, we calculate the noncentral t value (nct_w) by using Yuen's statistic

$$nct_w = \frac{|\mu_{t1} - \mu_{t2}|}{\sqrt{s_{w1}^2/k_1 + s_{w2}^2/k_2}} = 2.86219,$$

and the degrees of freedom $v_w = 48.6417$. Then, the critical value by using SAS function is $cv = tinv(1-\alpha/2, v_w, 0) = tinv(0.975, 48.6417, 0) = 2.00995$.

TABLE 1
The Six Conditions of Sample Size Allocation and the Resulting Power and Increment Ratio for Yuen's Two-Sided Trimmed Mean Test

Condition number	n_{t1}	n_{t2}	$n_{t1} + n_{t2}$	(k_1, k_2)	*Power*	*Increment ratio*[b]%
1	20.2087	71.5466	91.7553	(12.1252, 42.9280)	0.79972	8.424
2	24.2087	61.6172	85.8259	(14.5252, 36.9703)	0.80053	1.417
3[a]	**28.2087**	**56.4174**	**84.6261**	**(16.9252, 33.8504)**	**0.80090**	**0.000**
4	30.0000	54.8192	84.8192	(18.0000, 32.8915)	0.80103	0.228
5	32.2087	53.2246	85.4333	(19.3252, 31.9348)	0.80117	0.954
6	36.2087	51.0600	87.2687	(21.7252, 30.6360)	0.80139	3.123

[a]Condition 3 (in boldface) is the design with the optimal sample size.
[b]The formula for the increment ratio of the total effective sample size is in Equation (A3).

Furthermore, using SAS *probt* function yields

$$\begin{aligned} Power &= 1 - probt(cv, v_w, nct_w) = 1 - probt(2.0095, 48.6417, 2.86219) \\ &= 0.80103. \end{aligned}$$

It shows that the resulting power is very close to the expected power of 0.80.

Note that the previous total effective sample size $k_1 + k_2 = 18 + 32.8915$ is not the optimal sample size. In the optimal condition, the highest discriminating power can be obtained when $k_2/k_1 = s_{w2}/s_{w1}$ (Luh & Guo, 2007; Mace, 1974, p. 82). To determine the optimal sample size for the illustrative example (i.e., $k_2/k_1 = s_{w2}/s_{w1} = 2$), we calculate the initial effective sample sizes (see the Appendix) with $\gamma = 2$. Then, we calculate the degrees of freedom v_w by Equation (5) and obtain values of $t_{v_w,1-\alpha/2}$ and $t_{v_w,1-\beta}$, respectively, to replace $z_{1-\alpha/2}$ and $z_{1-\beta}$ in Equation (A1). Therefore, we obtain the effective sample sizes $k_1 = 16.9252$ and $k_2 = 2 \times 16.9252 = 33.8504$, respectively; and the sample sizes for trimming are $n_{t1} = 16.9252/0.6 = 28.2087$ and $n_{t2} = 33.8504/0.6 = 56.4174$, respectively. These results are reported in Table 1, Condition 3, so that the optimal condition can serve as the baseline to compare with five other conditions of sample size allocation.

In Table 1, Condition 4 is the illustrated example that the first group is fixed at 30, whereas Conditions 1 and 2 are set for the conditions that reduce 8 and 4 subjects for the first group, respectively. Last, Conditions 5 and 6 are set to have additional 4 and 8 subjects for the first group, respectively. It is clear that if the first group size is less than the optimal size ($n_{t1} = 28.2087$), the proposed formula can compensate for it by having more subjects for the second group (see Conditions 1–2) to attain the desired power of 0.8. If the first group size is greater than the optimal size (see Conditions 4–6), the proposed formula will reduce the size of the second group. Note that all the conditions can achieve the desired power, but the amount of increase in the total sample size is different. By comparing to the optimal total size of 84.6261, we find that the fewer the subjects for the first group in the beginning, the more subjects are needed for the second group to compensate for this. Last, the increment ratio (see Equation (A3) in the final column of Table 1 can be an index to show how much the increment of sample size increases relative to the optimal condition. In the optimal condition, the increment ratio should be 0.

SIMULATION DESIGN AND RESULTS

To compare the performance of the test statistics based on the proposed sample size formula and the traditional formula, we conducted a Monte Carlo simulation and followed the designs of Burton, Altman, Royston, and Holder (2006) and Harwell (1992) by using random sample techniques for the conditions of normal–nonnormal distributions and equal–unequal variances to cover a wide range of conditions to mimic the real situations. We used g-and-h distributions (Hoaglin, 1985, p. 503) to manipulate four distribution shapes: normal ($g = h = 0$), symmetrical and heavy-tailed ($g = 0$, $h = .2$), slightly skew with a relatively light tail ($g = 0.3$, $h = 0$), and moderately skew with a relatively light tail ($g = 0.5$, $h = 0$). The corresponding skew and kurtosis were (0, 0), (0, 36.22), (0.95, 4.64), and (1.75, 8.9), respectively, and the corresponding variance (σ^2) was 1, 2.1516, 1.14, and 1.46, respectively (Martinez & Iglewicz, 1984). Furthermore, the corresponding trimmed sample variances s_w^2 for 20% trimming by Equation (3) (by the average of 1,000,000 replications of size 20) are 0.689, 0.820, 0.713, and 0.759, respectively. It is clear that the trimmed sample variances are much smaller than the original variances. To manipulate unequal variances, we set variance ratios as 1 and 4, respectively.

For given $n_1 = n_{t1} = 30$, $\alpha = .05$, $1-\beta = 0.80$ and setting $\mu_1 - \mu_2 = 1$ for Welch's t (1938) and $\mu_{t1} - \mu_{t2} = 1$ for Yuen's t_w, we first computed the resulting sample size n_2 and the sample size for trimming n_{t2} with Equations (7)–(9) with $\lambda = 0$ and $\lambda = .2$, respectively. The results are shown in Columns 4 and 8 of Table 2 for the one-sided test and in Table 3 for the two-sided test, respectively. It is apparent that the second group needs less subjects for the trimmed case than the untrimmed case for the nonnormal distributions of ($g = 0$, $h = 0.2$) and ($g = 0.5$, $h = 0$).

After the sample size needed for untrimming and trimming was specified, we generated the data by setting $\mu_1 - \mu_2 = 0$ and $\mu_{t1} - \mu_{t2} = 0$ to test for Type I error, and setting $\mu_1 - \mu_2 = 1$ and $\mu_{t1} - \mu_{t2} = 1$ to test for the power of test statistics, respectively. We used SAS RANNOR function (SAS Institute, 2002) to generate Z from a standard normal distribution and used the g-and-h distributions

TABLE 2
Given Sample Size $n_{t1} = n_{t2} = 30$, the Resulting Size of the Second Group, and the Percentage of Empirical Type I Error and the Statistical Power of the One-Sided Tests ($\alpha = .05$, $1-\beta = .80$)

		Welch's t test				*Yuen's t_w test*			
(g, h)	*Variance ratio*	σ_1^2	n_2	*Type I error*	*Power*	s_w^2	n_{t2}	*Type I error*	*Power*
(0.0, 0.0)	1	1.00	10	5.22	83.62	0.689	13	5.18	85.68
	4	—	33	5.26	80.87	—	40	5.22	80.51
(0.0, 0.2)	1	2.15	26	4.69	82.06	0.820	15	5.20	83.01
	4	—	99	4.95	80.97	—	50	5.22	81.24
(0.3, 0.0)	1	1.14	11	**6.90**	**79.75**	0.713	13	5.39	81.31
	4	—	39	**5.92**	**80.33**	—	42	**5.45**	80.65
(0.5, 0.0)	1	1.46	15	**6.88**	**80.39**	0.759	14	5.07	80.12
	4	—	54	**6.21**	**79.97**	—	45	5.26	79.64

Note. Boldfaced values indicate that the Type I error rates of the test exceed the criterion $0.05 + 2 \times (0.05 \times 0.95 \times 10^{-4})^{1/2} = 5.44\%$.

to transform Z to reflect the target distribution shapes. The detailed procedure of data generation can be found in Luh and Guo (2005, pp. 83–85). After data were ready, we applied the untrimmed Welch's t and the trimmed mean Yuen's t_w tests to the data ($\alpha = .05$, $1-\beta = 0.80$). We ran 10,000 replications for each experiment and reported the average out of the replications. Tables 2 and 3 present

TABLE 3
Given Sample Size $n_{t1} = n_{t2} = 30$, the Resulting Size of the Second Group, and the Percentage of Empirical Type I Error and the Statistical Power of the Two-Sided Tests ($\alpha = .05$, $1-\beta = .80$)

		Welch's t test				*Yuen's t_w test*			
(g, h)	*Variance ratio*	σ_1^2	n_2	*Type I error*	*Power*	s_w^2	n_{t2}	*Type I error*	*Power*
(0.0, 0.0)	1	1.00	13	4.77	81.95	0.689	17	4.96	83.59
	4	—	45	5.08	80.67	—	55	4.90	80.30
(0.0, 0.2)	1	2.15	42	4.81	81.99	0.820	20	4.79	81.28
	4	—	164	4.91	80.85	—	71	4.93	81.77
(0.3, 0.0)	1	1.14	15	5.11	79.94	0.713	17	5.05	80.37
	4	—	54	4.85	79.28	—	58	4.84	80.29
(0.5, 0.0)	1	1.46	21	4.81	79.92	0.759	19	4.88	80.38
	4	—	77	5.01	79.31	—	63	5.04	79.30

the percentages of empirical Type I error rate and statistical power for the one- and two-sided tests, respectively.

Table 2 shows that Yuen's test can control Type I error well but Welch's test results in liberal Type I error when distributions are skewed for the one-sided test. For the two-sided test (see Table 3), both Welch's test and Yuen's test can control Type I error, but Yuen's test can achieve the power very close to the desired level. In conclusion, considering the size of the sample, Type I errors, and empirical power, we found promising results showing that the proposed formula works not only in the normal condition but also in the nonnormal conditions.

CONCLUSIONS AND DISCUSSION

It is widely recognized that sample size determination is one of the most important tasks researchers face in planning experiments because underestimation will reduce the power to detect an experiment effect (Kazdin & Bass, 1989; Rossi, 1990). Researchers know that power is a function of the significance level, the true alternative hypothesis, the sample size, and the particular test used. When the assumptions of the test are not met, robust statistics such as the trimmed mean method and its corresponding sample size calculation are alternatives that can be used. The present study considers the situation in which the researcher uses the trimmed mean method for variance heterogeneity, has a fixed number of subjects for one group at the time of planning, and derives the sample size determination for another group. The present study shows that the advantage of using the trimmed mean method is obvious with respect to the efficiency of the sample size requirement. Furthermore, the simulation results show a consistent pattern that, by using the proposed method, the Type I error can be controlled and the statistical power can be retained at the desired level.

One more thing to note is that the fixed size for the first group should be large enough to satisfy $k_1\Delta_w^2 - s_{w1}^2(z_{1-\alpha} + z_{1-\beta})^2 > 0$ in a one-sided test, and there should exist n_{t2} large enough to compensate for it by having more subjects. Otherwise, we need to adjust the values of α or β (Heilbrun & McGee, 1985).

Last, it is always advisable that the researcher considers a range of population parameters because the adequacy of the sample size depends on the accuracy of the initial specifications of the assumed parameters in the population. These factors in situations go beyond the coverage of commercial software programs. The existing statistical power analysis software seldom provides enough scope for practical use (Thomas & Krebs, 1997), but the formula developed in the present study fills the gap and provides a good approximation. The optimal allocation plans under group size constraints, presented in the present study, are potentially applicable to a wide variety of research situations. More research on this topic for complex study designs is required, and it is expected to have promising results.

ACKNOWLEDGEMENT

This research was supported by a National Science Council grant (NSC93-2413-H-006-003) to Wei-Ming Luh. The authors wish to thank the editor and reviewers for their insightful suggestions.

AUTHOR NOTES

Wei-Ming Luh is the distinguished professor at the Institute of Education at the National Cheng Kung University in Taiwan. She received her doctorate from the Department of Educational Psychology at the University of Georgia while specializing in educational statistics and data analysis. **Jiin-Huarng Guo** received his doctorate from the Department of Statistics at the University of Georgia and is now a professor in the Department of Applied Mathematics at the National Pingtung University of Education in Taiwan. His research interest is computer simulation.

REFERENCES

Burton, A., Altman, D. G., Royston, P., & Holder, R. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine, 25,* 4279–4292.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Guo, J.-H., & Luh, W.-M. (2000). An invertible transformation two-sample trimmed *t*-statistic under heterogeneity and non-normality. *Statistics & Probability Letters, 49,* 1–7.

Harwell, M. R. (1992). Summarizing Monte Carlo results in methodological research. *Journal of Educational Statistics, 17,* 293–313.

Heilbrun, L. K., & McGee, D. L. (1985). Sample size determination for the comparison of normal means when one sample size is fixed. *Computational Statistics & Data Analysis, 3,* 99–102.

Hoaglin, D. C. (1985). Summarizing shape numerically: The *g*-and-*h* distributions. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), *Exploring data tables, trends, and shapes.* New York: Wiley.

Hsu, L. M. (1993). Using Cohen's tables to determine the maximum power attainable in two-sample tests when one sample is limited in size. *Journal of Applied Psychology, 78,* 303–305.

Kazdin, A. E., & Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of Consulting and Clinical Psychology, 57,* 138–147.

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., et al. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research, 68,* 350–386.

Kraemer, H. C., & Thiemann, S. (1987). *How many subjects? Statistical power analysis in research.* Newbury Park, CA: Sage.
Luh, W.-M., & Guo, J.-H. (2000). Johnson's transformation two-sample trimmed t and its bootstrap method for heterogeneity and non-normality. *Journal of Applied Statistics, 27*, 965–973.
Luh, W.-M., & Guo, J.-H. (2005). Heteroscedastic test statistics for one-way analysis of variance: The trimmed means and Hall's transformation conjunction. *The Journal of Experimental Education, 74*, 75–100.
Luh, W.-M., & Guo, J.-H. (2007). Approximate sample size formulas for the two-sample trimmed mean test with unequal variances. *British Journal of Mathematical and Statistical Psychology, 60*, 137–146.
Luh, W.-M., Olejnik, S., & Guo, J.-H. (2008). Sample size determination for one- and two-sample trimmed mean tests. *The Journal of Experimental Education, 77*, 167–184.
Mace, A. E. (1974). *Sample-size determination.* Huntington, NY: Krieger.
Martinez, J., & Iglewicz, B. (1984). Some properties of the Tukey g and h family of distributions. *Communication in Statistics—Theory and Method, 13*, 353–369.
Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156–166.
Rosenberger, J. L., & Gasko, M. (1983). Comparing location estimators: Trimmed means, medians, and trimean. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), *Understanding robust and exploratory data analysis* (pp. 297–338). New York: Wiley.
Rossi, J. S. (1990). Statistical power of psychological research: What have we gained in 20 years? *Journal of Consulting and Clinical Psychology, 58*, 646–656.
SAS Institute. (2002). *SAS user's guide, version 9.* Cary, NC: Author.
Staudte, R. G., & Sheather, S. J. (1990). *Robust Estimation and Testing.* New York: Wiley.
Thomas, L., & Krebs, C. J. (1997). A review of statistical power analysis software. *Bulletin of the Ecological Society of America, 78*, 126–139.
Tukey, J. W., & McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization I. *Sankhya: The Indian Journal of Statistics: Series A, 25*, 331–352.
Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika, 29*, 350–362.
Wilcox, R. R. (1987). New designs in analysis of variance. *Annual Review of Psychology, 38*, 29–60.
Wilcox, R. R. (1994). Some results on the Tukey-McLaughlin and Yuen methods for trimmed means when distributions are skewed. *Biometrical Journal, 36*, 259–273.
Wilcox, R. R., & Keselman, H. J. (2003). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods, 8*, 254–274.
Yuen, K. (1974). The two-sample trimmed t for unequal population variances. *Biometrika, 61*, 165–170.

APPENDIX

Let $\gamma = n_{t2}/n_{t1} = k_2/k_1$ and $\tau = s_{w1}^2/s_{w2}^2$, we can rewrite Equation (6) as

$$\frac{s_{w1}^2}{k_1}\left(1+\frac{\tau}{\gamma}\right) = \frac{\Delta_w^2}{(z_{1-\alpha}+z_{1-\beta})^2}$$

Then

$$k_1 = \frac{s_{w1}^2 + (z_{1-\alpha} + z_{1-\beta})^2}{\Delta_w^2}\left(1 + \frac{\tau}{\gamma}\right) \quad \text{and} \quad k_1 = \gamma k_1. \tag{A1}$$

The total effective sample size can be expressed as a function of γ:

$$K(\gamma) = k_1 + k_2 = \frac{s_{w1}^2 + (z_{1-\alpha} + z_{1-\beta})^2}{\Delta_w^2}\left(1 + \frac{\tau}{\gamma}\right)(1 + \gamma). \tag{A2}$$

For the optimal condition, $\gamma = \sqrt{\tau}$ (Mace, 1974, p.82), the total effective sample size then is $K(\sqrt{\tau}) = s_{w1}^2(z_{1-\alpha} + z_{1-\beta})^2(1 + \sqrt{\tau})^2/\Delta_w^2$. Therefore, for any value of γ, the increment ratio of the total effective sample size relative to the optimal condition is

$$\frac{K(\gamma) - K(\sqrt{\tau})}{K(\sqrt{\tau})} = \frac{(1 + \tau/\gamma)(1 + \gamma)}{(1 + \sqrt{\tau})^2} - 1. \tag{A3}$$

Take Condition I in Table 1 for example, $\gamma = 71.5466/20.2087 = 3.54$. By Equation (A3), the increment ratio is $(1 + 4/3.54)(1 + 3.54)/9 - 1 = 0.0744$. Note that this value is different from the tabulated value 0.0842. The discrepancy is due to the values of degrees of freedom in the iteration process.