Paper 1660-2014

Trimmed_t: A SAS® Macro for the Trimmed *t* Test

Patricia Rodríguez de Gil, Anh P. Kellermann, Diep T. Nguyen,
Eun Sook Kim, Jeffrey D. Kromrey University of South Florida

ABSTRACT

The independent means *t*-test is commonly used for testing the equality of two population means. However, this test is very sensitive to violations of the population normality and homogeneity of variance assumptions. In such situations, Yuen's (1974) trimmed *t*-test is recommended as a robust alternative. The purpose of this paper is to provide a SAS macro that allows easy computation of Yuen's symmetric trimmed *t*-test. The macro output includes a table with trimmed means for each of two groups, Winsorized variance estimates, degrees of freedom, and obtained value of *t* (with two-tailed *p*-value). In addition, the results of a simulation study are presented and provide empirical comparisons of the Type I error rates and statistical power of the independent samples *t*-test, Satterthwaite's approximate *t*-test and the trimmed *t*-test when the assumptions of normality and homogeneity of variance are violated.

Keywords: heteroscedasticity, non-normality, Independent means *t*-test, Satterthwaite approximate *t*-test, trimmed means, simulation

INTRODUCTION

The independent means *t*-test and Satterthwaite's approximate *t*-test are provided in SAS with PROC TTEST. However, in some conditions, neither of these tests provides adequate Type I error control. The purpose of the present paper is to provide a SAS macro for computing a more robust inferential procedure: Yuen's (1974) *t*-test based on trimmed means. The macro programming language and an output example (including the trimmed values of two groups' means, Winsorized variances, the degrees of freedom and *p*-value corresponding with the trimmed *t*) are provided. In addition, the results of a simulation study comparing the Type I error control and power of the trimmed *t* test against the independent sample *t*-test and Satterthwaite's approximate test are presented.

PROBLEMS WITH THE *T*-TEST

Non-normality and Heterogeneity of Variance. Testing for the equality of means across independent groups is "a common inferential problem" (Keselman, Wilcox, Lix, Algina, & Fradette, 2007; p. 267). The independent means *t*-test relies on a strong assumption of equal variances (homoscedasticity) as the test statistic is a ratio of the difference in sample means to an estimate of the standard error of the difference, using a pooled variance estimate. Alternative approaches (e.g., Satterthwaite's approximate test) relax this assumption, approximating the *t* distribution and the corresponding degrees of freedom. Although the *t*-test may be one of the most basic and widely used statistical procedures to compare two group means (Hayes & Cai, 2007), statisticians to date are still evaluating the various conditions and factors for which this test is robust under the violation of the equality of variances assumption. Research on preliminary tests suggests that the choice between the *t*-test and the Satterthwaite's test, conditioning on a preliminary test of the assumption of homogeneity of variance, is not effective (Grissom, 2000; Hayes & Cai, 2007; Moser, Stevens, & Watts, 1989; .Rasch, Kubinger, & Moder, 2011; Zimmerman, 2004, 2010).

Keselman, Wilcox, Othman, and Fradette (2002) have suggested the use of trimmed means to achieve robustness in the presence of non-normality and variance heterogeneity. Lix and Keselman (1998) studied the performance of the Welch test (a close relative of Satterthwaite's test) in addition to the performance of Alexander and Govern, James, and Brown and Forsythe tests for testing mean equality in the presence of unequal variances. These tests can generally control Type I error rate when group variances are heterogeneous and data are normally distributed. However, these tests become liberal when the assumptions of normality and homogeneity of variances are violated, and they become even more liberal with unbalanced groups. For all the investigated distributions in the Lix and Keselmann study, a symmetric trimming was applied by removing 20% of the observations from each tail of the groups' set of scores. Their results showed that the studied methods generally exhibited a very good Type I error control rate when computed with trimmed means and Winsorized variances. Using a one-way completely randomized experiment, Keselmann, Wilcox, Algina, Fradette, and Othman (2004) compared seven methods known to be robust to the effects of non-normality and variance heterogeneity. For six methods (WJ or Welch-James-type heteroscedastic tests) known to provide good Type I error control and power (Algina & Keselman, 1998), using either symmetric or asymmetric trimming, Winsorized means and variances, were applied. The power of these tests was compared to the power of the one-step-M-Estimator trimmed means (MOMT; Wilcox & Keselman; 2003), test for the detection of treatment effects. Preliminary power results showed minor differences between the WJ tests due to data transformation or sample size. However, there were power differences favoring the WJ tests over the MOMT (.13).

Robust methods such as modified F_t and modified S_1 have been also recommended to overcome the sensitivity of the *t*-test to variance heterogeneity. Yusof, Abdullah, Yahaya, and Othman (2012) proposed the use of the trimmed mean, as a central tendency measure in the F_t test, and the median as the central tendency measure in S_1 when comparing the equality of two groups. These methods were compared in terms of Type I error under conditions of normality and non-normality represented by skewed-*g* and *h*-distributions.

Nguyen et al. (2012) conducted a simulation study to investigate the performance of the *t*-test, Satterthwaite's approximate *t*-test, and the conditional *t*-test in terms of Type I error control and statistical power. Factors manipulated in the study included total sample size (10, 20, 50, 100, 200, 300, 400), sample size ratio between groups (1:1, 1:3, 1:4), variance ratio between groups (1:1, 1:2, 1:4, 1:8, 1:12, 1:16, 1:20), population effect size (0, .2, .5, .8), alpha for testing treatment effect (.01, .05, .10, .15, .20, and .25), and alpha for testing the homogeneity of variance (01, .05, .10, .15, .20, .25, .30, .40, .45, and .50). For each condition, 100,000 replications were simulated, which provided a maximum standard error of an observed proportion (e.g., Type I error rate estimates) of .0015, and a 95% CI no wider than ± .003 (Robey & Barcikowski, 1992). Overall, the Satterthwaite's approximate *t*-test performed best in control of Type I error rates under all conditions. Results indicated that to maintain adequate Type I error control, the independent means *t*-test required that the homogeneity of variance assumption was met in addition to equal sample size between groups, regardless of the tenability of the normality assumption. The alpha level used for the Folded *F* test had an impact on Type I error control for the conditional *t*-test. The more conservative the alpha level, the larger Type I error. Because of lower statistical power of the Folded *F* test, the study recommended the conditional *t*-test using relatively large alpha levels for the test of variances. The results also showed that an increase in total sample size did not improve the control of Type I error rate for the independent means *t*-test, but larger samples provided better Type I error control for the conditional *t*-test.

Kellermann et al. (2013) extended the Nguyen et al. (2012) study to investigate the performance of the *t*-test, Satterthwaite's approximate *t*-test, and conditional *t*-test under heteroscedastic populations. In addition to the normal population, four non-normal populations where studied, with varying values of skewness and kurtosis (γ1 = 1.00, γ2 = 3.00; γ1 = 1:50, γ2 =5:00; γ1 = 2.00, γ2 = 6.00; γ1 = 0.00, γ2 =25.00) respectively. Findings were similar to the Nguyen et al. (2012) results with normal populations. Both the Satterthwaite's and conditional *t*-tests provided tremendous improvements in Type I error control compared to the independent means *t*-test when group variances were unequal. However, extreme skewness contaminated the Type I error control for these tests. On the other hand, kurtosis did not seem to have the same effect. Increasing sample size ($n \geq 200$) helped improve the Type I error control for the Satterthwaite's and conditional tests, but not for the independent *t*-test.

THE TRIMMED T TEST

Yuen (1974) proposed the *Trimmed t* test for the independent two-sample case, under unequal population variances. In each sample, the trimmed mean is computed by removing *g* observations from each tail of the distribution:

$$\bar{X}_t = \frac{1}{n-2g}\left(x_{g+1} + x_{g+2} + \ldots + x_{n-g}\right),$$

where
$x_1,\ldots,x_n$ are the ordered values in a sample
g = observations trimmed from each tail of the sample distribution
$n - 2g$ = the number of observations in the trimmed sample.

In addition to the trimmed mean, the Winsorized mean is required to compute the appropriate variance estimate. Instead of "trimming," this method replaces the most extreme *g* observations by the next-most-extreme value.

$$\bar{X}_w = \frac{1}{n}\left(\left[g+1\right]x_{g+1} + x_{g+2} + \ldots + \left[g+1\right]x_{n-g}\right).$$

Given the Winsorized mean, the Winsorized sum-of-squared deviations is computed as:

$$SSD_w = \left[g+1\right]\left[x_{g+1} - \bar{X}_w\right]^2 + \left[x_{g+2} - \bar{X}_w\right]^2 + \ldots + \left[g+1\right]\left[x_{n-g} - \bar{X}_w\right]^2.$$

Note that this is just the regular sum-of-squares approach using the replaced values and the Winsorized mean. From the Winsorized sum-of-squared deviations, the Winsorized variance is obtained as:

$$S_w^2 = \frac{SSD_w}{n - 2g - 1}.$$

Note that we have returned to the trimmed sample size here.

Finally, the obtained value of the trimmed *t* is obtained by dividing the difference between the trimmed means by the estimated standard error of the difference:

$$t = \frac{\bar{X}_{t1} - \bar{X}_{t2}}{\sqrt{\frac{S_{w1}^2}{n_1 - 2g} + \frac{S_{w2}^2}{n_2 - 2g}}}$$

The degrees of freedom are obtained from $\frac{1}{df} = \frac{c^2}{n_1 - 2g - 1} + \frac{(1-c)^2}{n_2 - 2g - 1}$

where

$$c = \frac{S_{w1}^2 \big/ (n_1 - 2g - 1)}{\left[S_{w1}^2 \big/ (n_1 - 2g - 1)\right] + \left[S_{w2}^2 \big/ (n_2 - 2g - 1)\right]}$$

Generally, the Welch and Satterthwaite approximate *t*-tests become conservative with leptokurtic distributions (Yuen, 1974). However, there is caveat for its use with small samples. The trimmed *t* (Yuen, 1974) is recommended instead because of its advantages (e.g., easy to compute and critical values from the standard *t* table can be used). Yuen (1974) conducted a study to determine whether the use of trimmed means and Winsorized variances resulted in robust tests for mean equality. Variables manipulated included sample sizes (10 or 20), standard deviation ratios (0.25, 0.5, 2.0 and 4.0), trimming rate (*g*) (from 1 observation to .25n_j observations), and a variety of distribution shapes. For unequal sample sizes, the amount of trimming was in fixed proportions. Ten thousand replications per condition were generated. Results showed a Type I error control for the trimmed means closer to the nominal alpha level than those obtained with the Welch's test, although some still deviated quite a bit from the nominal level. Yuen suggested an adaptive trimming approach; that is, the number of observations trimmed (*g*) should be chosen depending on the degree of leptokurtosis.

SAS MACRO

```
%macro trimmed_t (data = _LAST_, IV = X, DV = Y, trimpct = 0, trim = 0);
*remove observations with missing values if present;
data &data;
set &data;
if MISSING(&IV) + MISSING(&DV) > 0 then delete;

* Sort the data by DV value and number the observations in each group;
proc sort data = &data;
      by &IV &DV;
data sorted;
      set &data;
      by &IV;
      if first.&IV then count = 0;
      count + 1;
```

```
* Count number of observations per group, compute full-sample means and
variances;
proc means noprint data = sorted;
      by &IV;
      var &DV;
      output out = counts N = group_size mean = sample_mean var = sample_var;
data _null_;
      set counts;
      by &IV;
      if _N_ = 1 then call symput('Group1',&IV);
      if _N_ = 2 then call symput('Group2',&IV);
run;

*check if N - 2g - 1 > 0;
data check_1;
      set counts;
      if _N_ = 1 then do;
            if &trimpct NE 0 then ncheck_1 = group_size - 2*(ROUND(group_size
* &trimpct/100)) -1 ;
            if &trim NE 0 then ncheck_1 = group_size - 2*&trim -1 ;
            call symput('ncheck1',ncheck_1);
      end;
      if _N_ = 2 then do;
            if &trimpct NE 0 then ncheck_2 = group_size - 2*(ROUND(group_size
* &trimpct/100)) -1 ;
            if &trim NE 0 then ncheck_2 = group_size - 2*&trim -1 ;
            call symput('ncheck2',ncheck_2);
      end;
run;
%if (&ncheck1 le 0)|(&ncheck2 le 0) %then %do;
            data _null_;
            file print notitles;
                  put @1 'Trim too much' /
                  @1 'Please check your data' /
                  @1 'Macro has exited';
            return;
            run;
%end;
%if (&ncheck1 le 0)|(&ncheck2 le 0) %then %return;

* Check to be sure two groups (and only two groups) are present;
proc means noprint data = counts;
      var group_size;
      output out = check_2 N = n_groups;
run;
data check_2;
      set check_2;
      call symput('N_check',n_groups);
run;
      %if &n_check NE 2 %then %do;
            data _null_;
            file print notitles;
                  put @1 'Analysis must be based on two groups only' /
                  @1 'Please check your data' /
                  @1 'Macro has exited';
            return;
            run;
```

```
      %end;
      %if &n_check NE 2 %then %return;

* Determine n of cases to trim from each tail (trim) and delete those cases;
data trimmed;
      merge sorted counts;
      by &IV;
      if &trimpct = 0 then do; * if trim is specified, this executes;
            trim = &trim;
            if count LE &trim then delete;
            if count GT group_size - &trim then delete;
      end;
      if &trimpct NE 0 then do; * if trimpct is specified, this executes;
            trim = ROUND(group_size * &trimpct/100);
            if count LE trim then delete;
            if count GT group_size - trim then delete;
      end;

* Compute trimmed means and both largest and smallest remaining value in each
group (to be used in winsorization);
proc means noprint data = trimmed;
      var &DV;
      by &IV;
      output out = trimmed_mn mean = trimmed_mean min = smallest max =
largest;

* Replace trimmed observations with largest or smallest remaining value;
data winsorized;
      merge sorted trimmed_mn counts;
      by &IV;
      if &trimpct = 0 then do; * if trim is specified, this executes;
            trim = &trim;
            if count LE &trim then &DV = smallest;
            if count GT group_size - &trim then &DV = largest;
      end;
      if &trimpct NE 0 then do; * if trimpct is specified, this executes;
            trim = ROUND(group_size * &trimpct/100);
            if count LE trim then &DV = smallest;
            if count GT group_size - trim then &DV = largest;
      end;

* Compute winsorized mean and winsorized SS;
proc means noprint data = winsorized n mean css;
      var &DV;
      by &IV;
      output out = winsor2 n = group_n mean = winsorized_mean css =
winsorized_ss;

* Extract summary statistics needed for trimmed t calculation and final
output;
data group_stats;
      merge trimmed_mn winsor2 counts;
      by &IV;
      if &trimpct = 0 then trim = &trim;
      if &trimpct NE 0 then trim = ROUND(group_n * &trimpct/100);
      if _N_ = 1 then do;
            trim_mn1 = trimmed_mean;
```

```
            n1 = group_n;
            SS1 = winsorized_ss;
            trim1 = trim;
            sample_mean1 = sample_mean;
            sample_var1 = sample_var;
      end;
      if _N_ = 2 then do;
            trim_mn2 = trimmed_mean;
            n2 = group_n;
            SS2 = winsorized_ss;
            trim2 = trim;
            sample_mean2 = sample_mean;
            sample_var2 = sample_var;
      end;

* Place summary statistics into a single observation in SAS;
proc means noprint data = group_stats;
      var trim_mn1 n1 SS1 trim1 sample_mean1 sample_var1 trim_mn2 n2 SS2
trim2 sample_mean2 sample_var2;
      output out = one_row mean = ;

* Compute trimmed t, df, and p-value;
data ttest;
      set one_row;
      win_var1 = SS1/(n1 - 2*trim1 - 1);
      win_var2 = SS2/(n2 - 2*trim2 - 1);
      trimmed_t = (trim_mn1 - trim_mn2) / SQRT(win_var1/(n1 - 2*trim1) +
win_var2/(n2 - 2*trim2));

      c = (win_var1/(n1 - 2*trim1)) / (win_var1/(n1 - 2*trim1) + win_var2/(n2
- 2*trim2));
      df = (c**2/(n1 - 2*trim1 - 1) + (1 - c)**2/(n2 - 2*trim2 - 1))**-1;
      pvalue = 2*(1 - PROBT(abs(trimmed_t),df));

data _null_;
      set ttest;
      trim_n1 = n1 - 2*trim1;
      trim_n2 = n2 - 2*trim2;
      totalTrim1 = 2*trim1;
      totalTrim2 = 2*trim2;
      file print notitles;
      put @1 'Trimmed t-test Macro Output' //
      @1 'Before trimming:' /
      @10 'Group' @31 'N' @39 ' Mean' @47 ' Variance' /
      @10 '---------------' @28 '-------' @ 37 '---------' @47 '----------' /
      @10 "&Group1" @28 n1 5. @37 sample_mean1 BEST8. @48 sample_var1 BEST8.
/     @10 "&Group2" @28 n2 5. @37 sample_mean2 BEST8. @48 sample_var2 BEST8.
//    @1 'Trimmed Cases:'
      @30 'Total N' /
      @10 'Group' @30 'Trimmed'  /
      @10 '---------------' @30 '-------'  /
      @10 "&Group1" @30 totalTrim1 5.  /
      @10 "&Group2" @30 totalTrim2 5.  //
      @1 'After trimming:' /
      @37 ' Trimmed' @47 'Winsorized' /
      @10 'Group' @31 'N' @39 ' Mean' @47 ' Variance' /
      @10 '---------------' @28 '-------' @ 37 '---------' @47 '----------' /
```

```
      @10 "&Group1" @28 trim_n1 5. @37 trim_mn1 BEST8. @48 win_var1 BEST8. /
      @10 "&Group2" @28 trim_n2 5. @37 trim_mn2 BEST8. @48 win_var2 BEST8. //
      @1 'Trimmed t-test:' /
            @10 'Obtained t:' @30 trimmed_t BEST8. /
            @10 'df: '              @30 df BEST8. /
            @10 'p-value:'          @30 pvalue BEST8.;
return;
%mend trimmed_t;
```

MACRO EXECUTION

In order to use the trimmed_t macro, users first need to create a SAS dataset that inputs their own data. Then the macro is called using as arguments:

data = name of user's dataset

IV = independent variable

DV = dependent variable

trimpct = percentage of trimmed observations

trim = number of trimmed observations

In this execution, 26 observations from two groups are used to illustrate the macro. The observations are read into the SAS dataset EXAMPLE.

```
data EXAMPLE;
input group score;
cards;
 1 12
 1 14
 1 18
 1 25
 1 32
 1 44
 2 17
 2 22
 2 14
 2 12
 2 30
 2 29
 2 19
 1 12
 1 14
 1 18
 1 25
 1 32
 1 44
 2 17
 2 22
 2 14
 2 12
 2 30
 2 29
 2 19
 ;
```

The following code is used to call the macro after the date are read. This code identifies the dataset (named "EXAMPLE" in this example) to be employed for analysis, the names of the independent variable ("group" in this

example) and dependent variable ("score" in this example), the percentage of trimmed observations (specified as 5% in the example).

```
%trimmed_t (data=EXAMPLE,IV = group, DV = score, trimpct = 5);
```

If the users want to trim the data according to number of observations rather than percentage of observations, they can indicate the number of observations that need to be trimmed in the "trim" argument. For example, if there are 3 observations selected to be trimmed, the macro call would be:

```
%trimmed_t (data=EXAMPLE,IV = group, DV = score, trim = 3);
```

OUTPUT EXAMPLE OF TRIMMED_T MACRO

Sample output of the trimmed_t macro is shown in Table 1. In this table, there is descriptive information of the two groups including the number of observations, the mean and variance of each group before trimming, and the trimmed mean and Winsorized variance for each group after trimming. Then the numbers of observations trimmed for each group are presented. Finally, the obtained *t* value, the degrees of freedom and *p*-value of the trimmed *t*-test are also included.

```
Before trimming:
         Group                N         Mean     Variance
         ---------------    -------  --------- ----------
                     1         12    24.16667   135.4242
                     2         14    20.42857   45.18681

Trimmed Cases:               Total N
         Group               Trimmed
         ---------------     -------
                     1           2
                     2           2

After trimming:
                                        Trimmed  Winsorized
         Group                N         Mean     Variance
         ---------------    -------  --------- ----------
                     1         10        23.4    165.5185
                     2         12     20.33333    53.4026

Trimmed t-test:
         Obtained t:         0.669169
         df:                   13.681
         p-value:            0.514522
```

Table 1. Sample output for the Trimmed_t Macro

SIMULATION

We extended Kellermann et al. (2013) study including the trimmed *t* test and compared the performance of the independent means *t* test, Satterthwaite's test, and the trimmed *t* test when testing the mean difference of two groups. The comparison also included two types of conditional tests: conditional *t* test and conditional trimmed *t* test. In the conditional *t* test, either the independent means *t* test or Satterthwaite's test is conducted depending on the statistical decision about the homogeneity of variance. That is, if the Folded *F* test rejects the null hypothesis of homogeneous variance, Satterthwaite's test is conducted. Otherwise, the independent means *t* test is selected. Similarly, in the conditional trimmed *t* test, if the homogeneity of variance is not met based on the Folded *F* test, the trimmed *t* test is conducted. Otherwise, the independent means *t* test is used to test the two-group mean difference. We investigated the behaviors of the aforementioned tests with nonnormal data as well as with normal data under either homogeneity or heterogeneity of variance.

The simulation design factors included total sample size (10, 20, 50, 100, 200, 300, 400), sample size ratio between groups (1:1, 1:3, 1:4), variance ratio between groups (1:1, 1:2, 1:4, 1:8, 1:12, 1:16, 1:20), population effect size (0, .2, .5, .8), alpha for testing treatment effect (.01, .05, .10, .15, .20, and .25), alpha for testing the homogeneity of variance for Folded *F* test (01, .05, .10, .15, .20, .25, .30, .40, .45, and .50) and percent trimming (10, 20, 30) for trimmed *t* test. For each data generation condition, 100,000 replications were generated. We majorly examined Type I error and

statistical power as the simulation outcomes. For Type I error, we further investigated the robustness of Type I error control using the Bradley's liberal criterion (for example, .between 025 and .075 when the significance level is .05).

It should be noted that we only report the results of the alpha at .05 because different levels of alpha for testing the treatment effect did not yield saliently different results. Figure 1 displays the distribution of Type I error rates of the independent means t test, Satterthwaite's test, and the trimmed *t* test at alpha = .05. Satterthwaite's test controlled Type I error around the predetermined alpha level. On the other hand, the independent means *t* test showed a considerable variability in Type I error rates. The Type I error rates of the trimmed *t* test depends on the degree of trimming. As more observations were trimmed (from 5% to 20%), the overall behaviors of trimmed *t* test deteriorated showing larger variability in Type I error control. For the trimmed *t* test, the majority of simulation conditions including large sample and nonnormal conditions yielded Type I error over the nominal alpha level.

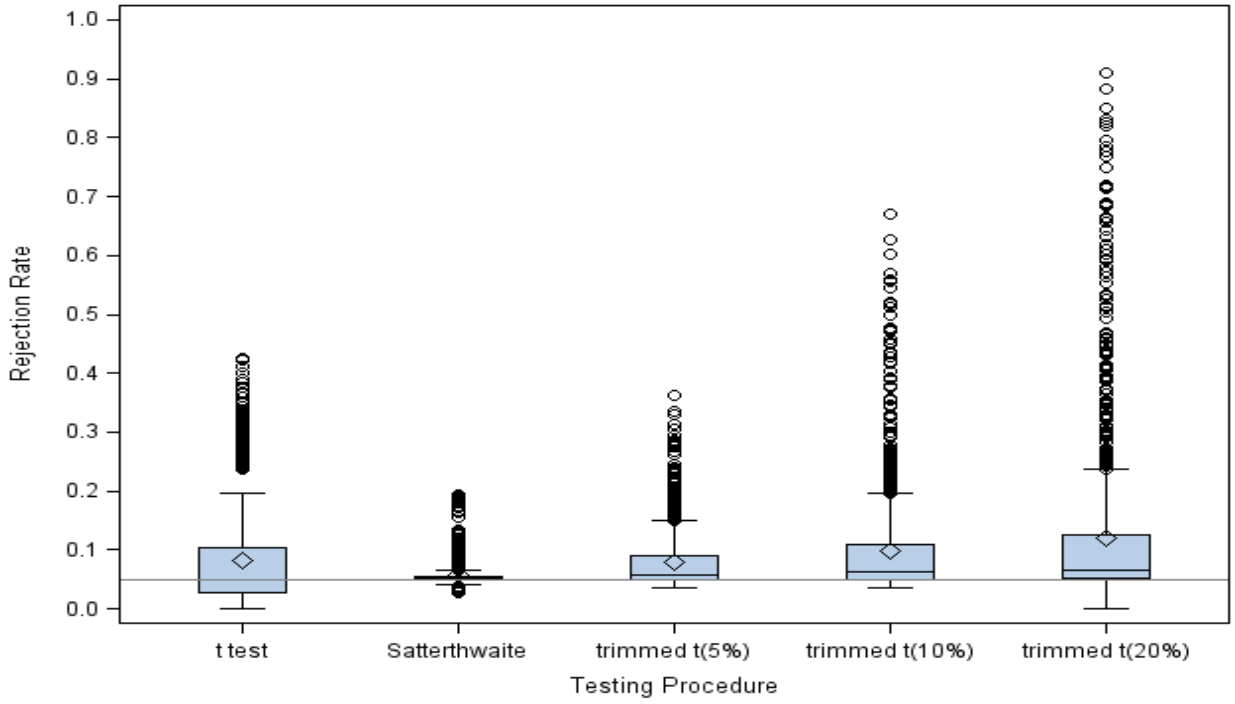

Figure 1. The Distribution of Estimated Type I Error rates of Independent Means *t* Test, Satterthwaite's Test, and Trimmed *t* Test at alpha = .05

The performance of the two conditional tests was investigated in comparison to independent means *t* test and Satterthwaite's test. As shown in Figure 2, the performance of conditional *t* test in which the rejection of homogeneous variance leads to Satterthwaite's test instead of the independent means *t* test became comparable to that of Satterthwaite's test as the alpha for testing the homogeneity of variance increased over .20. On the contrary, the significance level set for the homogeneity of variance test did not make an impact on the performance of the conditional trimmed *t* test. In other words, conditional decisions between the independent means *t* test and trimmed *t* test yielded inflated Type I error rates regardless of the power of Folded *F* test as presented in Figure 3 when 5% of the observations were trimmed. Similar patterns emerged with apparently increased variability in Type I error rates for 10% and 20% trimming.

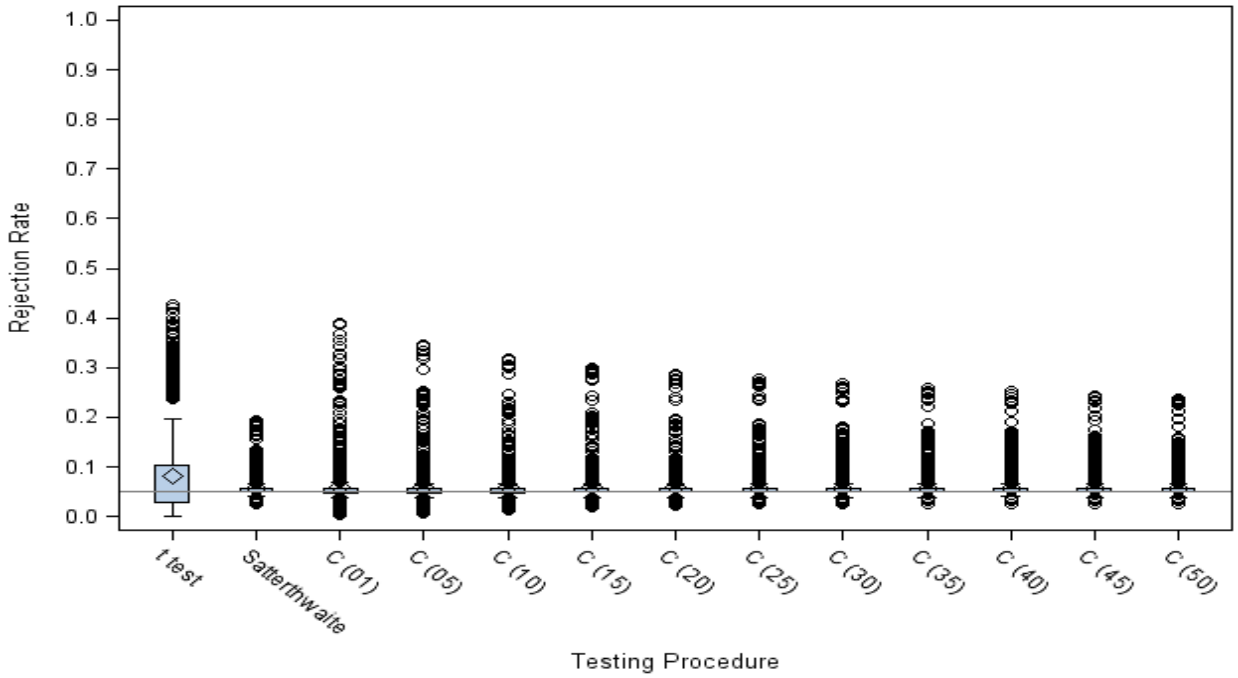

Figure 2. The Distribution of Estimated Type I Error rates of Independent Means *t* Test, Satterthwaite's Test, and Conditional *t* test at alpha = .05. C (01) means conditional *t* test when the alpha for testing the homogeneity of variance equals .01.

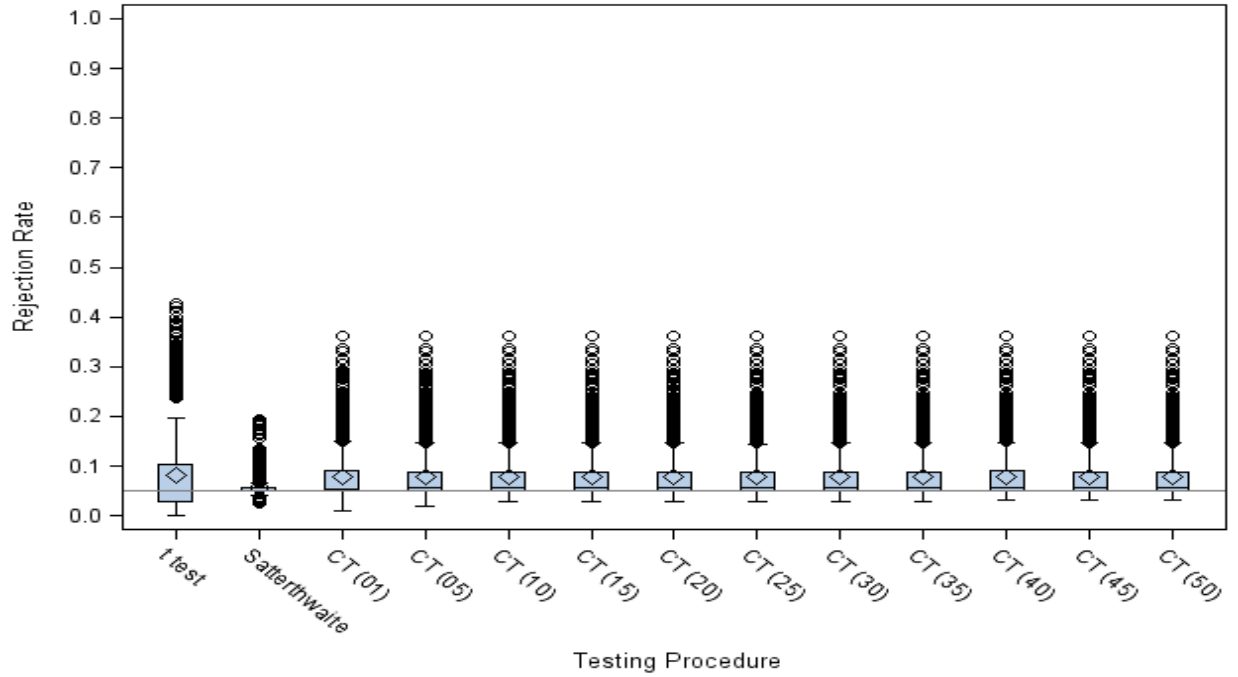

Figure 3. The Distribution of Estimated Type I Error rates of Independent Means *t* Test, Satterthwaite's Test, and Conditional Trimmed *t* test at alpha = .05. CT (01) means conditional trimmed *t* test with 5% trimming when the alpha for testing the homogeneity of variance equals .01.

Table 2 presents the proportion of conditions meeting the Bradley's liberal criterion for Type I error control. Satterthwaite's test always met the liberal criterion when the total sample size exceeded 200. Unless the total sample size was very small such as 10, Satterthwaite's test showed reasonable performance in controlling Type I error within the Bradley's criterion. The trimmed *t* test in general outperformed the independent means *t* test across the total sample size conditions. Interestingly, the trimmed *t* test performed better with the total sample size about 50 and 100 than with larger samples. Even when the total sample size was very small (i.e., 10), the proportion of conditions meeting the Bradley's criterion was over 60% except the conditions of large degree of trimming (20%). Similar patterns were observed with the conditional trimmed *t* test regardless of the alpha set for testing the homogeneity of variance.

Total sample size	*t* test	Satterthwaite	Trimmed *t* (5%)	Trimmed *t* (10%)	Trimmed *t* (20%)
10	.47	.65	.62	.62	.56
20	.49	.81	.59	.54	.45
50	.45	.94	.77	.67	.53
100	.44	.97	.74	.64	.59
200	.42	1.00	.69	.59	.55
300	.41	1.00	.63	.56	.52
400	.41	1.00	.59	.56	.52

Table 2. The Proportions of Conditions Meeting the Bradley's Liberal Criterion of Type I Error Control

We conducted power analysis only with the conditions meeting the Bradley's liberal criterion because high power associated with high Type I error is not compelling. Overall, Satterthwaite's test on average produced higher power followed by the trimmed *t* test and the independent means *t* test. For the trimmed *t* test, more trimming with the loss of observations possibly leads to lower power as demonstrated in Figure 4.

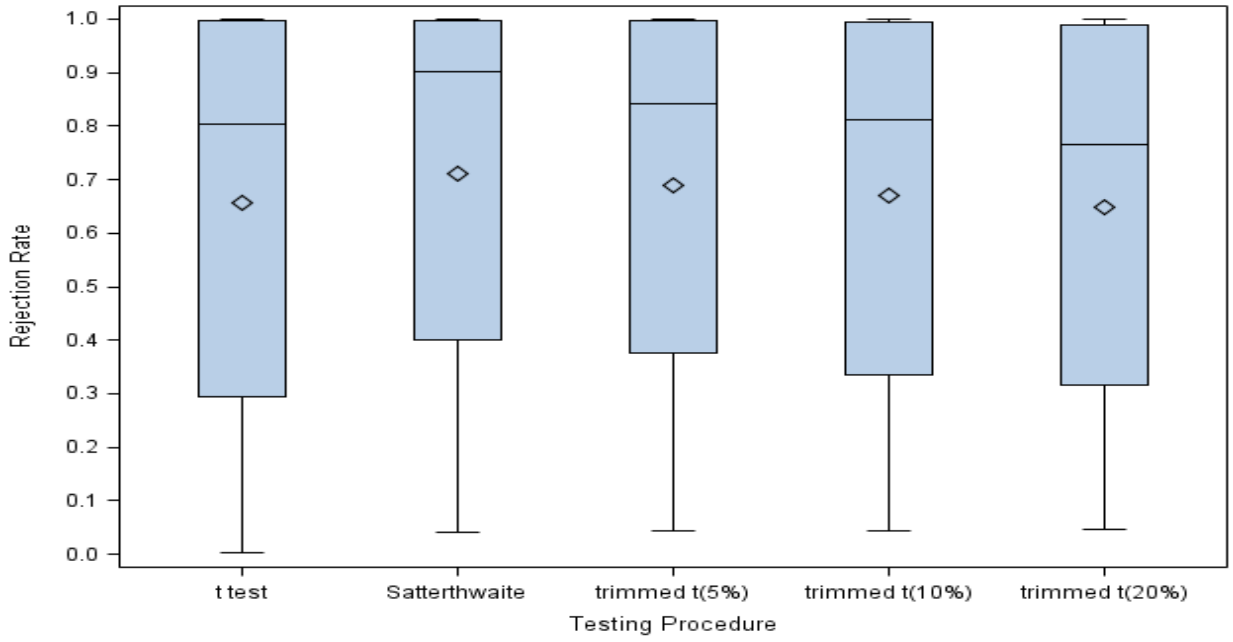

Figure 4. The Distribution of Power of Independent Means *t* Test, Satterthwaite's Test, and Trimmed *t* Test at alpha = .05

CONCLUSION

The trimmed *t* test has been proposed as a robust alternative to the independent means *t*-test. The macro trimmed_t presented in this paper offers a simple vehicle for computing the trimmed *t* and provides flexibility in the specification of either the percentage of cases to be trimmed from each tail of the sample distributions or the number of cases to be trimmed. The macro uses only SAS/BASE programming to provide utility without the need for more advanced components of the SAS system.

In its current form, the macro provides symmetric trimming (the same number of observations are trimmed from each tail of each sample). However, this can be readily modified to trim samples asymmetrically. Additional enhancements, such as adding graphic presentations of the samples before and after trimming can also be easily incorporated into the present macro code.

REFERENCES

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31,* 144-152.

Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology*, *68*, 155-165.

Hayes, A. F. & Cai, L. (2007). Further evaluating the conditional decision rule for comparing independent means. *British Journal of Mathematical and Statistical Psychology*, *60*, 217-244.

Kellermann, A., Bellara, A. P., De Gil, P. R., Nguyen, D., Kim, E. S., Chen, Y-H, & Kromey, J. D. (2013). Variance heterogeneity and Non-Normality: How SAS PROC TTEST® can keep us honest. *Proceedings of the Annual SAS Global Forum Conference, Cary, NC: SAS Institute Inc*.

Keselman, H. J., Othman, A. R., Wilcox, R. R., and Fradette, K. (2004). The new and improved two-sample t-test. *Psychological Science, 15*(1), 47-51.

Keselman, H. J., Wilcox, R. R., Algina, J., Fradette, K., & Othman, A. R. (2004). A power comparison of robust test statistics based on adaptive estimators. *Journal of Modern Applied Statistical Methods,* 3(1), 27-38.

Keselman, H. J., Wilcox. R. R., Algina, J., Othman, A. R., and Fradette, K. (2008). A comparative study of robust tests for spread: Asymmetric trimming strategies. *British Journal of Mathematical and Statistical Psychology, 61,* 235-253.

Keselman, H. J., Wilcox, R. R., and Lix, L. M. (2003). A generally robust approach to hypothesis testing in independent and correlated groups designs. *Psychophysiology, 40,* 586-596.

Keselman, H. J., Wilcox, R. R., Lix, L. M., Algina, J., and Fradette, K. (2007). Adaptive robust estimation and testing. *British Journal of Mathematical and Statistical Psychology, 60,* 267-293.

Lix, L. M., and Keselman, H. J. (1998). To trim or not to trim: Test of location equality under heteroscedasticity and nonnormality. *Educational and Psychological Measurement, 54*(3) 409-429.

Moser, B. K., Stevens, G. R., & Watts, C. L. (1989). The two-sample t-test versus Satterthwaite's approximate F test. *Communications in Statistics: Theory and Methods*, *18*, 3963-3975.

Nguyen, D., De Gil, P. R., Kim, E. S., Bellara, A. P., Kellermann, A., Chen, Y-H., & Kromrey, J. D. (2012). PROC TTest® (Old Friend), What are you trying to tell us?. *Proceedings of the South East SAS Group Users*, Cary, NC.

Rasch, D., Kubinger, K. D., & Moder, K. (2011). The two-sample t-test: pre-testing its assumptions does not pay off. *Statistical Papers*, *52*, 219-231.

Robey, R. R., & Barcikowski, R. S. (1992). Type I error and the number of iterations in Monte Carlo studies of robustness. British Journal of Mathematical and Statistical Psychology, 45, 283-288.

SAS Institute, Inc. (2002-2003). SAS version 9.2

Yuen, K. K. (1974). The two-sample trimmed *t* for unequal population variances. *Biometrika, 61*(1), 165-170.

Yusof, Z., Abdullah, S., Yahaya, S. S. S., Othman, A. R. (2012). A robust alternative to the t-test. *Modern Applied Science, 6*(5), 27-33.

Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology*, *57*, 173-181.

Zimmerman, D. W. (2010). A simple and effective decision rule for choosing a significance test to protect against non-normality. *British Journal of Mathematical and Statistical Psychology, 64,* 388-409.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Patricia Rodríguez de Gil
University of South Florida
4202 East Fowler Avenue, EDU 105
Tampa, FL 33620
prodrig6@usf.edu