Variance Heterogeneity and Non-Normality:

How SAS PROC TTEST® Can Keep Us Honest

Anh P. Kellermann, Aarti P. Bellara, Patricia Rodríguez de Gil, Diep Nguyen, Eun Sook Kim, Yi-Hsin Chen, Jeffrey D. Kromrey

UNIVERSITY OF SOUTH FLORIDA

ABSTRACT

- The independent samples *t*-test is one of the most used tests for detecting true mean differences.
- The SAS system provides the PROC TTEST procedure to conduct a test for the difference between two population means by assuming homogeneity of variance or avoiding it.
- The *t*-test and its alternatives (the Satterthwaite's approximate test and Conditional *t*-test) assume population normality. Past research has provided evidence of the *t*-test's robustness to departures of normality; however questions about the performance of conditional testing when the assumption of normality is not met remain.
- This paper describes previous research on preliminary tests under the normality assumption, extends this research to the evaluation of conditional testing to departures of normality, and provides guidance to researchers on the proper use of this test with non-normal, heteroscedastic population distributions.

PROC TTEST EXAMPLE

Annotated Syntax

The syntax for PROC TTEST is quite simple:

Only requires

a **class** statement

To identify the independent or grouping variable

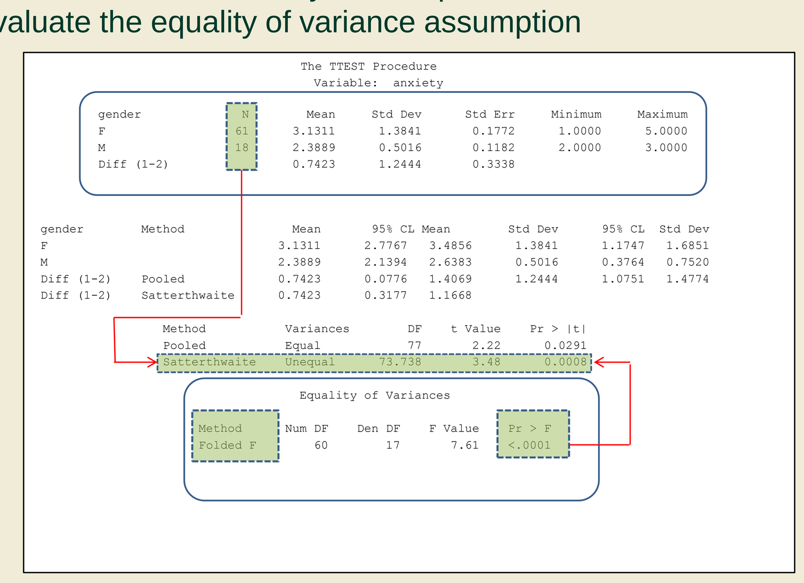
a **var** statement

To identify the dependent or outcome variable

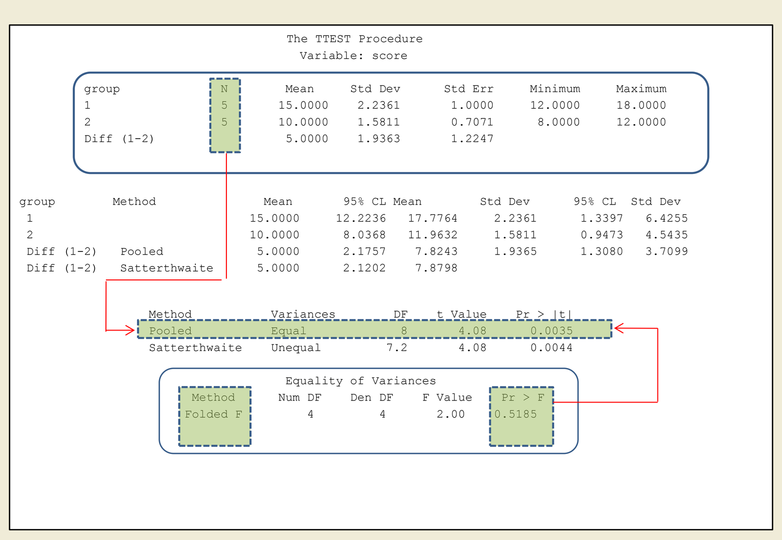
```
PROC TTEST DATA= Survey;
  class Gender;
  var   Anxiety;
run;
```

Annotated Outputs

To determine which *t*-statistic is appropriate, the following outputs show that PROC TTEST by default performs the Folded *F* statistic to evaluate the equality of variance assumption



Output 1. Results of PROC TTEST: Statistically Significant Differences in Variances Observed



Output 2. Results of PROC TTEST: No Statistically significant Differences in Variances Observed

THE SIMULATION STUDY

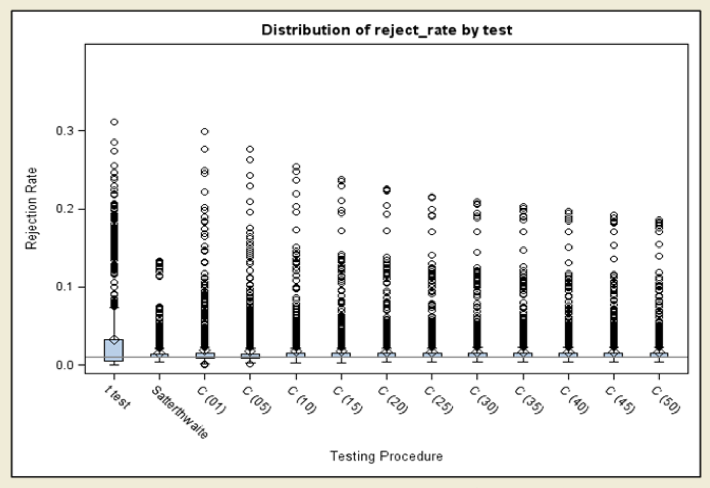
Manipulated Conditions

1) Total sample size (from 10 to 400),
2) Sample size ratio between groups (1:1, 2:3, and 1:4),
3) Variance ratio between populations (1, 2, 4, 8, 12, 16, and 20),

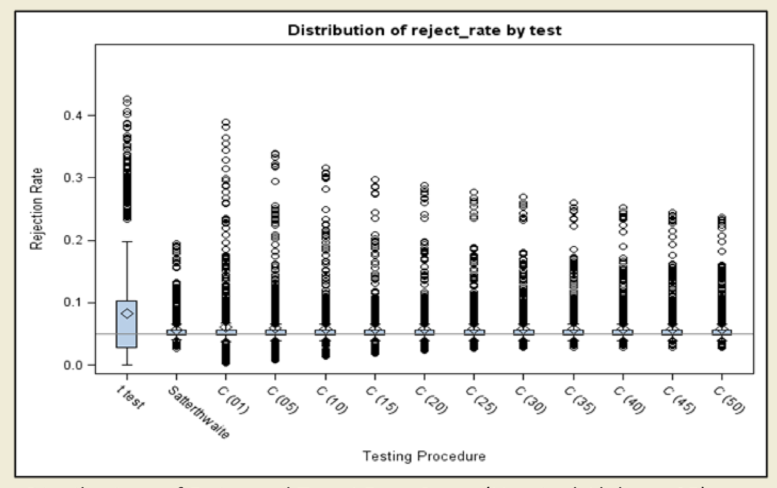
4) Effect size for mean difference between populations (Δ = 0, .2, .5, .8)
5) Alpha set for testing treatment effect (from α = .01 to α = .25),
6) Alpha set for testing homogeneity assumption for the conditional *t*-test (from α = .01 to α = .50),

7) Population distributions with varying skewness and kurtosis values (i.e., γ1 = 1.00 and γ2 = 3.00, γ1 = 1.50 and γ2 =5.00, γ1 = 2.00 and γ2 = 6.00, γ1 = 0.00 and γ2 =25.00, and γ1 = 0.00 and γ2 =00.00).

Distribution of Type I Error Rate Estimates Across All Conditions

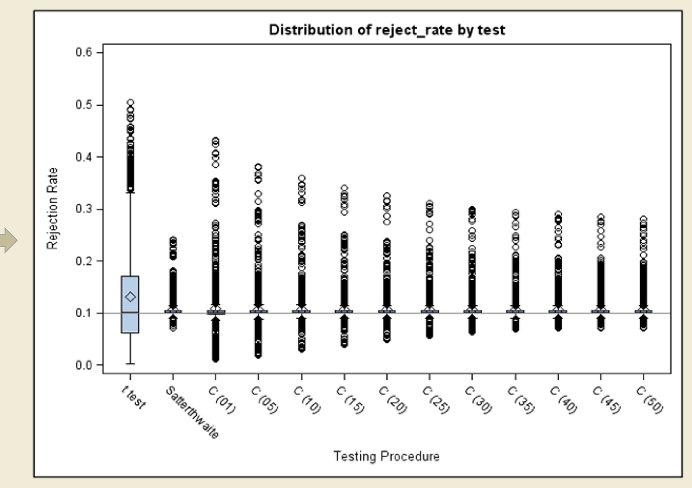


Distributions of Estimated Type I Error Rates (Nominal Alpha = .01)



Distributions of Estimated Type I Error Rates (Nominal Alpha = .05)

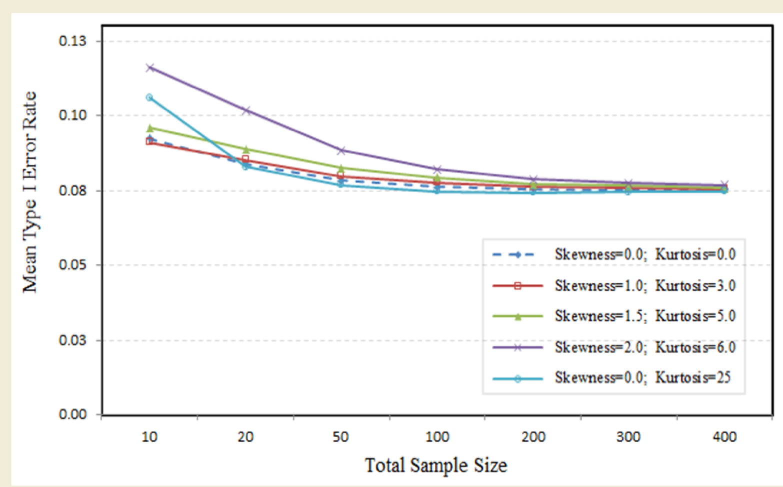
- Great dispersion of Type I error rates for the independent *t*-test
- Satterthwaite's approximate *t*-test provided better Type I error control
- Conditional t-test provided a notable improvement in relative to the independent means *t*-test, and the improvement increases as the alpha level for the Folded F-test is increased.



Distributions of Estimated Type I Error Rates (Nominal Alpha = .1)

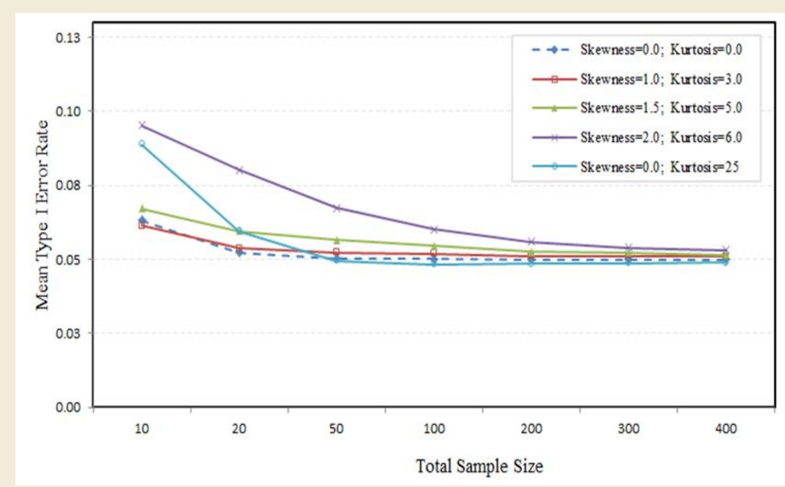
[The plot for C (01) provides the distribution of Type I error rates for the conditional *t*-test when an alpha level of .01 was used with the Folded F-test as the rule to choose between the independent means *t*-test and Satterthwaite's approximate *t*-test.]

Type I Error Control under Different Distribution Shapes



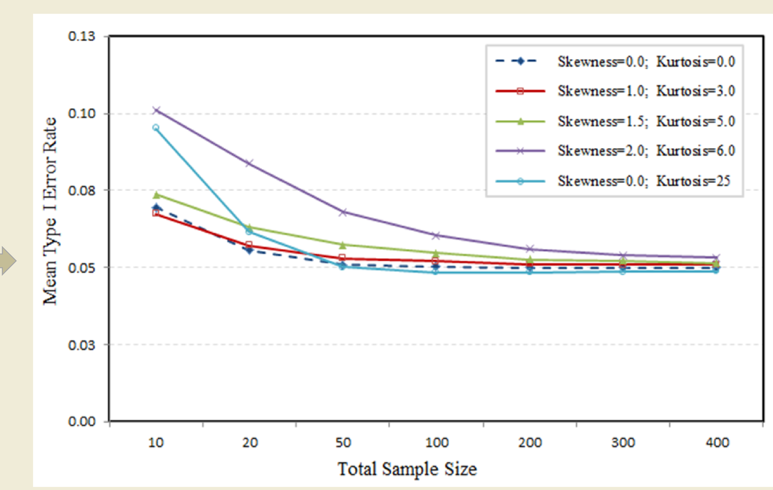
Mean Type I Error Rate for the ***Independent Means*** *(Nominal Alpha = .05).*

Type I error rates of the independent means *t*-test are way above the nominal alpha level regardless of the distribution shapes and total sample sizes



Mean Type I Error Rate for ***Satterthwaite's Approximate T-Test*** *(Nominal Alpha = .05).*

Alternative tests provided much better Type I error control except for extremely small sample size or extremely skewed distribution (i.e., skewness = 2).



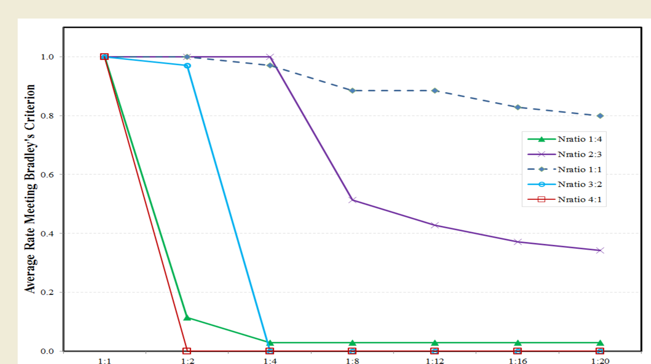
Mean Type I Error Rate for the ***Conditional T-Test*** *(Nominal Alpha = .05)*

Conditions Meeting Bradley's Liberal Criterion

Condition	t-test	Conditional	Satterthwaite	Condition	t-test	Conditional	Satterthwaite
N				Variance ratio			
10	0.451	0.680	0.651	1:1	1.000	0.943	0.920
20	0.486	0.760	0.817	1:2	0.617	0.931	0.949
50	0.451	0.931	0.949	1:4	0.400	0.909	0.926
100	0.434	0.971	0.971	1:8	0.286	0.897	0.914
200	0.417	1.000	1.000	1:12	0.269	0.886	0.891
300	0.406	1.000	1.000	1:16	0.246	0.886	0.891
400	0.406	1.000	1.000	1:20	0.234	0.891	0.897
N ratio				Shape			
1:4	0.180	0.976	0.967	0,0	0.433	0.963	0.967
2:3	0.665	0.976	0.976	1,3	0.437	0.955	0.967
1	0.910	0.967	0.967	1.5,5	0.437	0.931	0.939
3:2	0.282	0.910	0.914	2,6	0.461	0.771	0.780
4:1	0.143	0.702	0.739	0,25	0.412	0.910	0.910

Table 1. The proportions of cases meeting the Bradley's liberal criterion by Tests and Conditions at α = .05

Note. Conditional = the conditional *t*-test at α = .25 for the Folded *F*-test. For shape, two values indicate skewness and kurtosis, respectively.

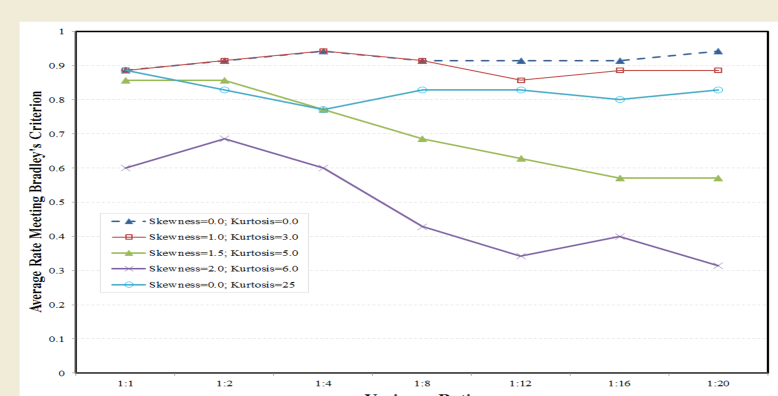


Proportion of Cases Meeting Bradley's Liberal Criterion for the Independent Means T-Test.

- High Bradley rates are warranted for the independent means *t*-tests only under the homogeneity of variance and/or with equal sample size between groups.
- Other than the above conditions, the independent means *t*-tests frequently do not meet Bradley's criterion. For example, when sample sizes are unbalanced between groups, Bradley rates are virtually zero even with moderate heterogeneity of variance.

Bradley's (1978) liberal criterion of robustness was used to examine the Type I error rates across conditions of the study:

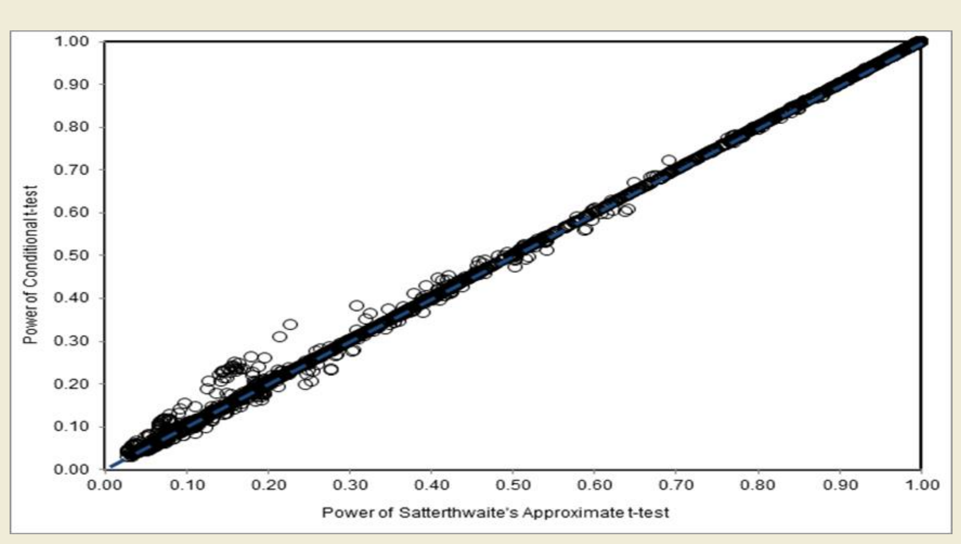
- for independent means *t*-test, variance ratio and sample size ratio are major factors affecting Bradley rates.
- Bradley rates of the conditional *t*-test and Satterthwaite's approximate *t*-test are greatly associated with the shape of data distribution.
- The conditional *t*-test with the Folded F-test alpha set at .25 showed very comparable results to Satterthwaite's tests.



Proportion of Cases Meeting Bradley's Liberal Criterion for Satterthwaite's Approximate T-Test at Alpha=.01.

- when data are not skewed including extremely high kurtosis (i.e., skewness = 0 and kurtosis = 25), Bradley rates are constantly high regardless of variance ratio.
- when data are skewed, Satterthwaite does not work well even under homogeneity of variance.

Statistical Power



Power Estimates for the Conditional T-Test and Satterthwaite's Approximate T-Test

- The independent means *t*-test is the most powerful test for mean differences when the assumptions are met.
- Power comparisons made for conditions in which both Satterthwaite's approximate *t*-test and conditional *t*-test evidenced adequate Type I error control by Bradley's (1978) benchmark showed that there is small power differences between these two alternative tests.
- The differences are such that the conditional testing procedure provides power advantage over the Satterthwaite's approximate *t*-test.

CONCLUSION

- Regardless of the tenability of the normality assumption, the independent means *t*-test performed very well on Type I error control when homogeneity assumption was met.
- Under departures of normality and with heterogeneous variances, both Satterthwaite's approximate *t*-test and conditional *t*-test (using a large alpha level of .25 for the Folded F-test of variances), performed much better than the independent means *t*-test in maintaining adequate Type I error control.
- Extreme skewness (e.g., skewness = 2) contaminated the Type I error control for both alternative testing procedures whereas Kurtosis seemed not to have this kind of impact; and increasing total sample sizes can improve the control of Type I error rates for both alternative tests in case of extreme skewness.
- There is a small power difference between the Satterthwaite's approximate *t*-test and the conditional testing procedure such that the use of the conditional testing procedure may provide a power advantage over the use of Satterthwaite's approximate *t*-test

IMPLICATIONS

To adequately control Type I error and improve power when examining the difference of two independent group means from non-normal populations,

- With existing homogeneity of variance and/or with equal sample size between groups, the independent means *t*-test is the best testing procedure to use regardless of the tenability of the normality assumption.
- With the absence of homogeneity of variance, a Folded F-test with a large alpha level of .25 can provide reasonable guidance in the choice between the independent *t*-test and Satterthwaite's approximate *t*-test:
  - If the F value is NOT statistically significant then use independent means *t*-test.
  - If the F value is statistically significant then use Satterthwaite's approximate *t*-test.

  In addition, if populations are extremely skewed (e.g., skewness = 2), a total sample size of at least 200 is recommended; also, a total sample size of at least 100 is recommended for less skewed populations otherwise the Type I error control procedures may be questionable.