

December 2017

# JMASM 47: ANOVA\_HOV: A SAS Macro for Testing Homogeneity of Variance in One-Factor ANOVA Models (SAS)

Isaac Li

*University of South Florida, Tampa, FL, liy1@mail.usf.edu*

Yi-Hsin Chen

*University of South Florida, Tampa, FL*

Yan Wang

*University of South Florida, Tampa, FL*

Patricia Rodríguez de Gil

*University of South Florida, Tampa, FL*

Thanh Pham

*University of South Florida, Tampa, FL*

*See next page for additional authors*

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

## Recommended Citation

Li, I., Chen, Y.-H., Wang, Y., de Gil, P. R., Pham, T., Nguyen, D., Kim, E. S., & Kromrey, J. D. (2017). JMASM 47: ANOVA\_HOV: A SAS Macro for Testing Homogeneity of Variance in One-Factor ANOVA Models (SAS). *Journal of Modern Applied Statistical Methods*, 16(2), 506-539. doi: 10.22237/jmasm/1509496080

---

# JMASM 47: ANOVA\_HOV: A SAS Macro for Testing Homogeneity of Variance in One-Factor ANOVA Models (SAS)

## **Authors**

Isaac Li, Yi-Hsin Chen, Yan Wang, Patricia Rodríguez de Gil, Thanh Pham, Diep Nguyen, Eun Sook Kim, and Jeffrey D. Kromrey

## JMASM 47: ANOVA\_HOV: A SAS Macro for Testing Homogeneity of Variance in One-Factor ANOVA Models (SAS)

**Isaac Li**

University of South Florida  
Tampa, FL

**Yi-Hsin Chen**

University of South Florida  
Tampa, FL

**Yan Wang**

University of South Florida  
Tampa, FL

**Patricia Rodríguez de Gil**

University of South Florida  
Tampa, FL

**Thanh Pham**

University of South Florida  
Tampa, FL

**Diep Nguyen**

University of South Florida  
Tampa, FL

**Eun Sook Kim**

University of South Florida  
Tampa, FL

**Jeffrey D. Kromrey**

University of South Florida  
Tampa, FL

---

Variance homogeneity (HOV) is a critical assumption for ANOVA whose violation may lead to perturbations in Type I error rates. Minimal consensus exists on selecting an appropriate test. This SAS macro implements 14 different HOV approaches in one-way ANOVA. Examples are given and practical issues discussed.

*Keywords:* ANOVA, homogeneity, SAS

---

### Introduction

The analysis of variance (ANOVA) is regarded as one of the most powerful and flexible methods for testing hypotheses about population means. When the underlying assumptions are satisfied, ANOVA provides powerful statistical tests, but if they are not satisfied, Type I error rates are inflated and statistical power reduced. A critical assumption of ANOVA is homogeneity of variance (HOV), that is, the compared populations have equal variances, which can be expressed as

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2. \quad (1)$$

---

*Isaac Li is a graduate student of Educational Measurement and Research. Email him at liy1@mail.usf.edu.*

Although it was found that moderate deviations from the HOV assumption might not seriously affect the results in ANOVA (Glass, Peckham, & Sanders, 1972), there remains the concern about large violations of HOV that can lead to invalid inferences. In fact, careful attention has been paid to this issue, and numerous procedures have been developed for the purpose of evaluating the HOV assumption although each of these procedures has limitations. For example, the classic  $F$  test has been commonly applied, but was found to be sensitive to departures from normality. Similarly, Snedecor and Cochran (1989) discovered that Bartlett's test (Bartlett, 1937) is sensitive to violations of the normality assumption. Therefore, it is advisable to turn to alternative tests that are less dependent on the assumption of normality.

Cases of these alternative approaches which are relatively insensitive to departures from normality include the Levene (Levene, 1960), Brown-Forsythe (BF; Brown & Forsythe, 1974), and O'Brien (OB; O'Brien, 1979, 1981) tests. Simulation studies (Algina, Olejnick, & Ocanto, 1989; Conover, Johnson, & Johnson, 1981; Lee, Katz, & Restori, 2010; Olejnick & Algina, 1987) have revealed differences in performance among these tests. For example, the OB test provided Type I error rates near the nominal alpha in unbalanced samples, but with platykurtic distributions which exhibit lighter tails than a normal distribution, it was more sensitive to variance differences than the BF test. When sample sizes were equal, OB had a power advantage used with platykurtic distributions and had a slight power advantage when used with mesokurtic distributions (i.e., exhibiting same tail weight as a normal distribution) regardless of whether the distributions were skewed or not. On the other hand, the BF test had a power advantage used with leptokurtic distributions (i.e., having heavier tails than a normal distribution) regardless of the skewness. When sample sizes were unequal, results were different because the relative power of these tests depended on the direction of the relationship between the population variances and sample sizes. For example, the power of the OB test was improved when used with skewed-platykurtic and symmetric platykurtic distributions. The power of BF was also enhanced when the relationship between population variance and sample size was direct (i.e. larger samples come from populations with larger variances) and with leptokurtic or mesokurtic distributions. With other distributions, the tests had similar power.

The appropriate test should be selected in the context of their research questions, sample data, and analysis plan. It would be difficult to reach a consensus in answering the question what method should be applied for testing the HOV assumption, because every study is different and no procedure is

superior to the rest in an absolute sense. Furthermore, some testing procedures are less known, or how to implement them, which makes it harder to be selected as a suitable HOV method. The goal of the current study is two-fold: 1) introduce fourteen different approaches that can be used to evaluate the HOV assumption and 2) provide a SAS macro designed to implement these statistical tests.

### Statistical Methods for Testing the HOV Assumption

**Bartlett Test.** Bartlett (1937) proposed a special use of the chi-square test for testing the HOV assumption. The computation of the Bartlett's chi-square is as follows:

$$\chi^2 = \frac{(N-k) \log \left[ \frac{\sum_{j=1}^k (n_j - 1) S_j^2}{(N-k)} \right] - \sum_{j=1}^k (n_j - 1) \log(S_j^2)}{1 + \frac{\left( \sum_{j=1}^k \frac{1}{n_j - 1} \right) - \frac{1}{(N-k)}}{3(k-1)}}, \quad (2)$$

where  $N$  is total sample size,  $n_j$  is the group  $j$  sample size,  $k$  refers to number of groups, and  $S_j^2$  denotes group  $j$  variance. The null hypothesis that the subpopulation variances are equal will be rejected if Bartlett's  $\chi^2$  statistic is greater than the critical  $\chi^2$  value with degrees of freedom  $k - 1$ .

**Levene's Test (Absolute and Squared).** To overcome departures from normality, Levene (1960) proposed to use the absolute values or squared residuals, which transforms a test of variances into a test of means that is relatively robust to the violation of the normality assumption (Sayago & Asuero, 2004). The statistics of the absolute values and squared residuals used for the Levene's test are defined as follows (Katz, Restori, & Lee, 2009):

$$Z_{ij} = |Y_{ij} - \bar{Y}_{.j}| \text{ and } Z_{ij}^2 = (Y_{ij} - \bar{Y}_{.j})^2, \quad (3)$$

where  $Y_{ij}$  is the raw score and  $\bar{Y}_{.j}$  denotes the mean of the  $j^{\text{th}}$  group.

The test statistic,  $W$ , which is approximately distributed like  $F$ , is defined as:

$$W = \frac{(N - k) \sum_{j=1}^k n_j (\bar{Z}_{.j} - \bar{Z}_{..})^2}{(k - 1) \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Z}_{ij} - \bar{Z}_{.j})^2}, \quad (4)$$

where  $\bar{Z}_{.j}$  represents the group mean of  $Z_{ij}$  and  $\bar{Z}_{..}$  is the grand mean. The obtained  $W$  value is compared to the  $F$  critical value ( $F_{crit}$ ) with degrees of freedom  $(k - 1)$  and  $(N - k)$  for numerator and denominator, respectively. The null hypothesis that the group variances are equal is rejected if  $W > F_{crit}$ .

**Brown-Forsythe Test.** This test follows the idea of Levene's test but uses the group median instead of the group mean in calculation of absolute deviation. This test first transforms the dependent variable values into absolute values of deviations from the group median:

$$z_{ij} = |Y_{ij} - \tilde{Y}_j| \quad (5)$$

where  $\tilde{Y}_j$  is the median of group  $j$  and  $z_{ij}$  refers to the transformed value of  $Y_{ij}$ .

The obtained statistic  $W$  is computed using the same formula as that in the Levene's Test presented above. The critical values of  $W$  are obtained from the  $F$  table with  $(k - 1)$  and  $(N - k)$  degrees of freedom. If the  $p$ -value  $< \alpha$ ,  $H_0$  is rejected and we conclude that the variances are not equal. The Brown-Forsythe test is more robust than Levene's test when the population distribution is skewed (Brown & Forsythe, 1974).

**O'Brien's Test.** O'Brien (1979) proposed a method that transforms original scores so they would represent sample variances. The transformed scores are used for ANOVA or the Welch test as a dependent variable. The transformation he proposed is the weighted average of a modified Levene's squared difference that is computed as

$$r_{ij}(w) = \frac{(w + n_j - 2)n_j(Y_{ij} - \bar{Y}_{.j})^2 - wS_j^2(n_j - 1)}{(n_j - 1)(n_j - 2)}, \quad (6)$$

where  $S_j^2$  represents the within-group unbiased estimate of variance for sample  $j$  and  $w$  ( $0 \leq w \leq 1$ ) denotes the weighting factor.

## ANOVA\_HOV: A SAS MACRO

The mean of the transformed values per group,  $\bar{r}_j = \frac{\sum r_{ij}}{n_j} = S_j^2$ , will equal the variance computed for that group. The weighted average,  $r_{ij}(w)$  is a modification of Levene's squared difference from the group mean ( $w = 0$ ), and a jackknife pseudo value of  $S_j^2(w = 1)$ . O'Brien (1981) suggested setting  $w = .5$  as default. Then the  $F$  statistic computed on  $r_{ij}$  in the standard ANOVA procedure would be the test statistic for O'Brien's test.

**Bootstrap Brown-Forsythe Test.** Boos and Brownie (2004) recommended a bootstrap approach for comparing variances where power outweighs computational simplicity. The test they investigated in earlier work (Boos & Brownie, 1989) focused on drawing bootstrap samples from

$$\bar{S} = \{Y_{ij} - \bar{Y}_j\}, \quad (7)$$

which will simply be the residuals from the original sample. Instead of drawing a separate bootstrap sample from each of the groups, bootstrapping is performed by pooling the residuals across groups. In each bootstrap sample, a test statistic for variances is computed and the  $p$ -value for the bootstrap test is obtained as the proportion of bootstrap samples with a statistic's value that is greater than that observed in the original data. The authors recommended using the median version of Levene's test statistic (that is, the Brown-Forsythe statistic), then obtaining the  $p$ -value via the bootstrap rather than the  $F$  distribution. This recommendation was also made by Lim and Loh (1996): their simulation study found that the bootstrapped version of Levene's test provided more power than the  $F$  distribution version.

**Ramsey Conditional Test: Brown-Forsythe or O'Brien.** Ramsey (1994) proposed a conditional procedure based on the Brown-Forsythe method and the O'Brien method. He suggested the appropriate test between the two methods should be selected conditional on a test of kurtosis. To decide which procedure to use, the kurtosis for each of  $k$  groups is computed by using Pearson's traditional measure:

$$b_2 = \frac{m_4}{m_2^2}, \quad (8)$$

where  $m_r = \Sigma(Y_{ij} - \bar{Y}_{.j})^r / n_j$  and  $m_2$  is the second moment about the mean or the biased sample variance. Thus, the formula for kurtosis of group  $j$ ,  $b_{2j}$ , is

$$b_{2j} = \frac{\sum(Y_{ij} - \bar{Y}_{.j})^4}{n_j} \cdot \left[ \frac{\sum(Y_{ij} - \bar{Y}_{.j})^2}{n_j} \right]^{-2}. \quad (9)$$

The kurtosis value for each group is then compared to critical values obtained from a table provided by Ramsey and Ramsey (1993). A score of  $-1$ ,  $0$ , or  $1$  is recorded depending on the test being significantly platykurtic, nonsignificant, or significantly leptokurtic, respectively. These scores of kurtosis were added across groups as a total score,  $S$ , which was used to identify the population as platykurtic if  $S \leq -1$ , mesokurtic if  $S = 0$ , or leptokurtic if  $S \geq 1$ . The O'Brien method will be applied if the data are platykurtic and the Brown-Forsythe method will be used if the data are mesokurtic or leptokurtic.

**Cochran's Test (Arithmetic and Harmonic).** Cochran's  $C$  (Cochran, 1941) is essentially an outlier test that defines a test statistic as the ratio of the largest group variance to the sum of the sample variances:

$$C = \frac{S_{\max}^2}{\sum S_j^2} \quad (10)$$

Critical values of Cochran's  $C$  can be obtained from the central  $F$  distribution:

$$\text{Critical } C = \frac{1}{1 + \frac{k-1}{F_{\frac{\alpha}{k}, n-1, (k-1)(n-1)}}}, \quad (11)$$

where  $n$  is the number of observations in each group (for the balanced design) and  $F$  refers to the critical value of  $F$  at  $\alpha/k$  with  $df = n - 1, (k - 1)(n - 1)$ . For an unbalanced design one could either use an arithmetic average of  $n$  or the harmonic

mean of  $n$  to calculate degrees of freedom in the numerator. If the obtained value of  $C$  exceeds the critical value, the null hypothesis of variance homogeneity is rejected.

***G Test.*** As an alternative to Cochran's  $C$ , 't Lam (2010) suggested using the  $G$  test which is a ratio of the product of the largest variance and its degrees of freedom to the sum of the products of each variance and its degrees of freedom:

$$G = \frac{v_{\max} S_{\max}^2}{\sum v_j S_j^2} \quad (12)$$

Critical values of the  $G$  test can be obtained from the central  $F$  distribution:

$$\text{Critical } G = \frac{1}{1 + \frac{v_{\text{pool}} / (v_{\max} - 1)}{F_{\frac{\alpha}{k}, \bar{n}_j - 1, (k-1)(\bar{n}_j - 1)}}}, \quad (13)$$

where  $v_{\text{pool}}$  is the pooled degrees of freedom;  $v_{\max}$  denotes the degrees of freedom for the group with the largest variance;  $\bar{n}_j$  is the mean number of observations in each group;  $F_{\alpha/k}$  denotes the critical value of  $F$  at  $\alpha/k$  with  $df = \bar{n}_j - 1, (k-1)(\bar{n}_j - 1)$ .

***The Maximum F-ratio Test ( $F_{\max}$ ) (Arithmetic and Harmonic).*** Hartley (1950) developed the  $F_{\max}$  test (a.k.a. Hartley's test) for comparing three or more population variances. Hartley's test requires independent random samples of the same size from normally distributed populations (Ott & Longnecker, 2010). The  $F_{\max}$  test is computed by taking the ratio of the largest group variance ( $S_{\max}^2$ ) to the smallest group variance ( $S_{\min}^2$ ) and its formula is presented as follows:

$$F_{\max} = \frac{S_{\max}^2}{S_{\min}^2}. \quad (14)$$

The value of  $F_{max}$  is compared to a critical value from the table containing the sampling distribution of Hartley's test. A  $F_{max}$  value less than the critical value provides evidence that the groups have similar or equal variances. Similar to the Cochran's  $C$  test for an unbalanced design, an arithmetic average of the sample size or a harmonic mean can be used to calculate degrees of freedom in the numerator.

**Z-Variance Test.** Overall and Woodward (1974) proposed the Z-variance test for the HOV assumption. This test is an approximate test based on Fisher and Yates (1963), which transforms the chi-square statistics with large degrees of freedom ( $df$ ) into standard normal deviate  $z$ -scores. Fisher and Yates' (1963) formula provided a  $z$ -score transformation for chi-square statistics based on large  $df$ :

$$z = \sqrt{2\chi^2} - \sqrt{2(df) - 1}. \quad (15)$$

Because sample variance estimates have a chi-square distribution,

$$\chi_{n-1}^2 = \frac{(n-1)S^2}{\sigma^2}, \quad (16)$$

where  $S^2$  is the sample variance estimate;  $\sigma^2$  is the true population variance. A  $z$ -score transformation of within-cell variance estimates that appears adequate for samples as small as  $n = 10$  or larger is then computed as

$$Z_j = \sqrt{\frac{c(n_j - 1)S_j^2}{MS_w}} - \sqrt{c(n_j - 1) - 1}, \quad (17)$$

where  $c = 2 + 1/n_j$ ,  $n_j$  is the sample size for the  $j^{\text{th}}$  group;  $S_j^2$  denotes unbiased variance estimate for the  $j^{\text{th}}$  group, and  $MS_w$  denotes the pooled within-group mean square (or  $MS_E$  as the pooled within-group error variance).

The  $F$ -ratio with  $k - 1$  and  $\infty$  degrees of freedom for testing the homogeneity of variance of  $k$  samples variance is given as

$$F = \frac{\sum_{j=1}^k Z_j^2}{k-1}. \quad (18)$$

Note the division of the sum of squared  $z$ -scores by  $k - 1$ , as the estimate of  $\sigma^2$ . The right-hand side of the equation can be conceived as a between groups sum of squares divided by  $df = k - 1$ . Overall and Woodward (1974) found that the  $Z$ -variance test performed very well with normally distributed data but produced too many Type I errors when samples were from leptokurtic or skewed distributions.

**Modified Z-variance Test.** To improve the performance of the  $Z$ -variance test when sample distributions are leptokurtic or skewed, Overall and Woodward (1976) proposed the modified  $Z$ -variance approach to testing the HOV. Based on their examination of the robustness of this modification to the  $z$ -variance against four other tests, in which the authors investigated a  $c$  value based on sample size, skewness and kurtosis, they determined  $c$  to be a scaling coefficient that affects the variability of the  $Z_j$  values. The formula for  $c$  is,

$$c = 2.0 \left( \frac{2.9 + .2 / n_j}{K} \right)^{1.6(n_j - 1.8K + 14.7)/n_j}, \quad (19)$$

where  $n_j$  is the sample size of the  $j^{\text{th}}$  group and  $K$  refers to the mean of the kurtosis indices from all groups. Overall and Woodward's kurtosis index is the 4<sup>th</sup> power of the  $Z$ -scores within each group divided by  $n_j - 2$  degrees of freedom,

$$K = \frac{\sum Z_{ij}^4}{n_j - 2}, \quad (20)$$

where

$$Z_{ij} = \frac{Y_{ij} - \bar{Y}_{.j}}{\sqrt{\frac{n_j - 1}{n_j} S_j^2}}. \quad (21)$$

## Methodology

### ANOVA\_HOV: A SAS Macro

ANOVA\_HOV is a macro is written to implement the 14 procedures to test the homogeneity of variance assumption in one-way ANOVA models. Before the macro is defined, two datasets of critical values are created for the  $F_{max}$  test and Ramsey's conditional test (see [Appendix A](#) for SAS code). Critical values at alpha levels of .05 and .01 for the  $F_{max}$  test are included. To use the ANOVA\_HOV macro researchers need to input their data set into SAS, identify variables, and then call the macro. The macro will run each of the fourteen HOV tests using the data set and input variables.

Macro input parameters. Use of the macro requires specification of four arguments for the macro call, which are briefly described below.

- data.* Specifies the name of the input data set.
- iv.* Specifies the name of the independent variable.
- dv.* Specifies the name of the dependent variable.
- nboots.* Specifies the number of bootstrap samples used in the Bootstrap Brown-Forsythe Test. A large number of replications are conducive to consistency in estimation.

## Results

To illustrate the use of the macro, two examples are provided. Data in the examples were downloaded from the database of the Integrated Postsecondary Education Data System (IPEDS), a system of interrelated surveys conducted annually by the U.S. Department of Education's National Center for Education Statistics (NCES) (NCES, 2014). IPEDS gathers information from every college, university, and technical and vocational institution that participates in the federal student financial aid programs. The latest wave of data available was collected in 2014 and only excerpts of the data were involved in the examples below.

For Example 1, the dependent variable was enrollment of Hispanic students and the HOV was tested on the independent variable, institutions of different categories. The four categories of institutions ("cat") were 1) degree-granting with primarily baccalaureate degrees or above; 2) degree-granting but less than 50% above bachelor's level; 3) degree-granting at levels of Associate's and

## ANOVA\_HOV: A SAS MACRO

certificates; and 4) non-degree-granting. The number of enrolled Hispanic students per institution (“sumef”) ranged from 0 to several thousands. To evaluate the macro in a situation of equal group sizes, a random sample of 10 institutions were drawn from each category to create a sample of 40 out of a total of over 7,000 institutions in the data set. The syntax below shows how the data were read into SAS to be analyzed.

```
data enrfrn;
  input cat sumef @@;
  datalines;
2 94 2 3 2 278 2 19 2 2 2 3 2 18 2 10 2 1 2 50
3 8 3 2 3 12 3 19 3 79 3 8 3 20 3 23 3 15 3 1
4 2582 4 983 4 743 4 2 4 4 4 19 4 13 4 339 4 197 4 4
5 10 5 20 5 1 5 153 5 1 5 60 5 1 5 3 5 21 5 4
;;;
```

The statement below invokes the macro by calling the variables.

```
%hov (data = enrfrn, iv = cat, dv = sumef, nboots = 5000);
```

Independent Variable:	cat	
N of Groups:	4	
Dependent Variable:	sumef	
Total N of Observations:	40	
	Obtained	
Test	Value	p
Brown and Forsythe	3.5791	.0231
Levene's (absolute values)	9.3372	.0001
Levene's (squared values)	1.9722	.1356
O'Brien	1.747	.1748
Bartlett	102.025	0
Bootstrap Brown-Forsythe	3.5791	.3402
Ramsey's Conditional Test	3.5791	.0231
Z-variance	20.0296	0
MZ-variance test	3.5384	.014
Cochran's C test (with arithmetic mean)	0.9848	p < .10
Cochran's C test (with harmonic mean)	0.9848	p < .10
G test	0.9848	p < .10
F-max test (with arithmetic mean)	1312.1533	p < .01
F-max test (with harmonic mean)	1312.1533	p < .01

**Figure 1.** Output for Example 1.

The output from the macro run is shown in Figure 1. The top paragraph gives a summary of the data while the rest of output includes the name of the

HOV tests, the obtained values for each test, and their associated  $p$ -values. Note that only comparisons with critical values are reported for the bottom five tests because their  $p$ -values are not calculated.

In the second example, the ANOVA\_HOV macro was applied to a large data set from IPEDS with unequal group sizes. The independent variable was a classification of how the institution is operated (“control”): 1) public institution, 2) private not-for-private institution, and 3) private for-profit institution. Both of the first two categories included more than 1,000 institutions whereas there were fewer than 200 in the third group. The dependent variable in the analysis was total salaries and wages expenditures as part of the institution’s library budget (“lsalwag”), which has a mean of \$1,143,651. The data file in analysis contains nearly 3,000 observations and is named “libsal”. The macro was invoked using the statement below and the number of bootstraps was set at 5,000 once again.

```
%hov (data = libsal, iv = control, dv = lsalwag, nboots = 5000);
```

The output of this analysis is provided in Figure 2.

Independent Variable:	control	
N of Groups:	3	
Dependent Variable:	lsalwag	
Total N of Observations:	2964	
	Obtained	
Test	Value	p
Brown and Forsythe	12.0649	0
Levene's (absolute values)	20.569	0
Levene's (squared values)	1.0383	.3542
O'Brien	1.0376	.3544
Bartlett	713.5098	0
Bootstrap Brown-Forsythe	12.0649	.1590
Ramsey's Conditional Test	12.0649	0
Z-variance	163.616	0
MZ-variance test	24.4983	0
Cochran's C test (with arithmetic mean)	0.5688	p < .10
Cochran's C test (with harmonic mean)	0.5688	p < .10
G test	0.5317	p < .10
F-max test (with arithmetic mean)	158.3482	p < .01
F-max test (with harmonic mean)	158.3482	p < .01

Figure 2. Output for Example 2.

## A Simulation Study

To investigate the accuracy and precision of the fourteen approaches for testing the homogeneity of variance assumption in one-way ANOVA models, a simulation study (Wang, Rodríguez de Gil, Chen, Kromrey, Kim, Pham, Nguyen,

& Romano, 2017) was conducted employing six design factors: number of groups, average number of observations per group, sample size pattern, variance pattern, maximum group variance ratio, and population distributions measured by skewness and kurtosis. The performance of the fourteen methods was examined at six nominal alpha levels (.01, .05, .10, .15, .20, and .25). For each condition, 5,000 samples were generated and Type I error and statistical power examined. Bradley's (1978) liberal criterion for robustness was set at  $\pm 0.5\alpha$  around the nominal alpha and eta-square effect sizes were estimated to explore the significant impacts of the research design factors on the variability in the Type I error estimates. Lastly, statistical power was estimated only for conditions in which Type I error was adequately controlled.

Of the approaches, Levene's test with the squared residuals, Brown-Forsythe, O'Brien, Ramsey's conditional procedure, and Bootstrap Brown-Forsythe tests were the five tests that maintained adequate Type I error control and were superior to other tests. Meanwhile, the power of these five tests was found to be acceptable and the power differences between them were subtle. O'Brien's test had slightly less power than the other four tests and the Bootstrap Brown-Forsythe test tended to have slightly greater power.

Population shape and average group size were two design factors that affected Type I error rates. Specifically, Ramsey and O'Brien were the only two tests that maintained adequate Type I error control across all the population shapes investigated. Among the five best tests, Brown-Forsythe and Bootstrap Brown-Forsythe tests had conservative Type I error rates if the distribution shape was extremely leptokurtic. In contrast, Levene's test with the squared residuals tended to have liberal Type I error rates when the shape is skewed. With normal populations, 12 out of 14 approaches maintained adequate Type I error control. Bartlett, Cochran with the harmonic mean, and Z-variance tests maintained Type I error control extremely well. Levene's test with squared values,  $F_{max}$  with the harmonic mean, and Z-variance had larger power. Z-variance test seemed to be the best choice for a normal distribution. Average group size had significant impact on Type I error control for four of the five best performers including Ramsey, Brown-Forsythe, Bootstrap Brown-Forsythe, and Levene's test with squared values. Increasing the sample size of a group improved Type I error control; on the other hand, when the average per-group sample size falls to five, only O'Brien's test maintained adequate Type I error control, which improves for all tests as the nominal alpha level grew.

## Conclusion

When examining the differences between two or more group means, ANOVA is among the most commonly used procedures. The assumptions of variance homogeneity and normality continue to be critical; in particular, violations of HOV may impact Type I error rates. This macro provides the researcher with the ability to apply fourteen different evaluation measures mainly for two purposes: first, for testing substantive hypotheses about population variances (e.g., do educational enrichment programs increase heterogeneity of student achievement scores?), and secondly, for assessing the tenability of homogeneity of variance in consideration of a subsequent test of mean differences. Different nominal alpha levels may be indicated for these uses.

HOV tests must be selected with care because it is difficult to assess population distribution shape based on samples, more so with small samples, which might result in severe consequences of non-normality for most of these tests. It is recommended that when average group is fewer than 10, the O'Brien test should be used because it gives a better Type I error control; for average group sizes between 10 and 20, the Ramsey test provides the most power; as for large groups with more than 20 observations, the Bootstrap Brown-Forsythe test will supply the most power.

## References

- Algina, J., Olejnik, S., & Ocanto, R. (1989). Type I error rates and power estimates for selected two-sample tests of scale. *Journal of Educational and Behavioral Statistics, 14*(4), 373-384. doi: 10.3102/10769986014004373.
- Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 160*(901), 268-282. doi: 10.1098/rspa.1937.0109
- Boos, D. D., & Brownie, C. (1989). Bootstraps methods for testing homogeneity of variances. *Technometrics, 31*(1), 69-82. doi: 10.1080/00401706.1989.10488477.
- Boos, D. D., & Brownie, C. (2004). Comparing variances and other measures of dispersion. *Statistical Science, 19*(4), 571-578. doi: 10.1214/088342304000000503
- Bradley, J. V. (1978). Robustness?. *British Journal of Mathematical and Statistical Psychology, 31*(2), 144-152. doi: 10.1111/j.2044-8317.1978.tb00581.x.

## ANOVA\_HOV: A SAS MACRO

Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, *69*(346), 364-367. doi: 10.1080/01621459.1974.10482955.

Cochran, W. G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. *Annals of Human Genetics*, *11*(1), 47-52. doi: 10.1111/j.1469-1809.1941.tb02271.x.

Conover, W. J., Johnson, M. E., & Johnson, M. M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, *23*(4), 351-361. doi: 10.1080/00401706.1981.10487680.

Fisher, R. A., & Yates, F. (1963). *Statistical tables for biological agricultural and medical research*. New York: Hafner.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, *42*(3), 237-288. doi: 10.3102/00346543042003237

Hartley, H. O. (1950). The maximum f-ratio as a short-cut test for heterogeneity of variance. *Biometrika*, *37*(3/4), 308-312. doi: 10.2307/2332383.

Katz, G. S., Restori, A. F., & Lee, H. B. (2009). A Monte Carlo study comparing the Levene test to other homogeneity of variance tests. *North American Journal of Psychology*, *11*, 511-522.

't Lam, R. U. E. (2010). Scrutiny of variance results for outliers: Cochran's test optimized. *Analytica Chimica Acta*, *659*(1-2), 68-84. doi: 10.1016/j.aca.2009.11.032

Lee, H. B., Katz, G. S., & Restori, A. F. (2010). A Monte Carlo study of seven homogeneity of variance tests. *Journal of Mathematics and Statistics*, *6*(3), 359-366. doi: 10.3844/jmssp.2010.359.366

Levene, H. (1960). Robust tests for the equality of variance. In I. Olkin, Ed. *Contributions to Probability and Statistics*, pp. 278-292. Palo Alto, CA: Stanford Univ. Press.

Lim, T. S., & Loh, W. Y. (1996). A comparison of tests of equality of variances. *Computational Statistics and Data Analysis*, *22*(3), 287-301. doi: 10.1016/0167-9473(95)00054-2

O'Brien, R. G. (1979). A general ANOVA method for robust tests of additive models for variances. *Journal of the American Statistical Association*, *74*(368), 877-880. doi: 10.1080/01621459.1979.10481047

- O'Brien, R. G. (1981). A simple test for variance effects in experimental designs. *Psychological Bulletin*, 89(3), 570-574. doi: 10.1037/0033-2909.89.3.570
- Olejnik, S. F., & Algina, J. (1987). Type I error rates and power estimates of selected parametric and nonparametric tests of scale. *Journal of Educational Statistics*, 12, 45-61. doi: 10.2307/1164627
- Ott, L. R., & Longnecker, M. (2011). *An introduction to statistical methods and data analysis* (6<sup>th</sup> Ed.). Belmont, CA: Brooks/Cole, Cengage Learning.
- Overall, J. E., & Woodward, J. A. (1974). A simple tests for heterogeneity of variance in complex factorial designs. *Psychometrika*, 39(3), 311-318. doi: 10.1007/bf02291705.
- Overall, J. E., & Woodward, J. A. (1976). *A robust and powerful test for heterogeneity of variance*. Galveston, TX: University of Texas Medical Branch Psychometric Laboratory.
- Ramsey, P. H., & Ramsey, P. P. (1993). Updated version of the critical values of the standardized fourth moment. *Journal of Statistical Computation and Simulation*, 44(3-4), 231-241. doi: 10.1080/00949659308811460
- Ramsey, P. H. (1994). Testing variances in psychological and educational research. *Journal of Educational & Behavioral Statistics*, 19(1), 23-42. doi: 10.3102/10769986019001023.
- Sayago, A., & Asuero, A. G. (2004). Fitting straight lines with replicated observations by linear regression: Part II. Testing for homogeneity of Variances. *Clinical Reviews in Analytical Chemistry*, 34(3-4), 133-146). doi: 10.1080/10408340490888599.
- Snedecor, G. W., & Cochran, W. G. (1989). *Statistical methods* (8<sup>th</sup> ed.). Ames: Iowa State University Press.
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2014). *The Integrated postsecondary education data system (IPEDS)*. [Data file and code book]. Retrieved from <https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx>.
- Wang, Y., Rodriguez de Gil, P., Chen, Y.-H., Kromrey, J. D., Kim, E. S., Pham, T., Nguyen, D., & Romano, J. L. (2017). Comparing the performance of approaches for testing the homogeneity of variance assumption in one-factor ANOVA models. *Educational and Psychological Measurement*, 77(2), 305-329. doi: 10.1177/0013164416645162.

## Appendix A: The SAS Macro ANOVO\_HOV

```

data test;
  input group Y;
  datalines;
1      49.66
1      47.86
1      51.61
2      49.68
2      51.02
2      50.02
3      50.45
3      49.09
3      45.73
4      46.39
4      49.25
4      48.39
4      51.99
4      47.84
4      57.99
4      53.28
4      48.08
4      48.10
4      53.48
4      49.98
;;;
run;

```

```

* Critical values for the Fmax test;
data cv05a;
  Input nuval c2 c3 c4 c5 c6 c7 c8 c9 c10 c11 c12;
Cards;
2 39.0 87.5 142 202 266 333 403 475 550 626 704
3 15.4 27.8 39.2 50.7 62.0 72.9 83.5 93.9 104 114 124
4 9.60 15.5 20.6 25.2 29.5 33.6 37.5 41.1 44.6 48.0 51.4
5 7.15 10.8 13.7 16.3 18.7 20.8 22.9 24.7 26.5 28.2 29.9
6 5.82 8.38 10.4 12.1 13.7 15.0 16.3 17.5 18.6 19.7 20.7

```

```

7 4.99 6.94 8.44 9.70 10.8 11.8 12.7 13.5 14.3 15.1 15.8
8 4.43 6.00 7.18 8.12 9.03 9.78 10.5 11.1 11.7 12.2 12.7
9 4.03 5.34 6.31 7.11 7.80 8.41 8.95 9.45 9.91 10.3 10.7
10 3.72 4.85 5.67 6.34 6.92 7.42 7.87 8.28 8.66 9.01 9.34
12 3.28 4.16 4.79 5.30 5.72 6.09 6.42 6.72 7.00 7.25 7.48
15 2.86 3.54 4.01 4.37 4.68 4.95 5.19 5.40 5.59 5.77 5.93
20 2.46 2.95 3.29 3.54 3.76 3.94 4.10 4.24 4.37 4.49 4.59
30 2.07 2.40 2.61 2.78 2.91 3.02 3.12 3.21 3.29 3.36 3.39
60 1.67 1.85 1.96 2.04 2.11 2.17 2.22 2.26 2.30 2.33 2.36

```

```
;
```

```
data cv01a;
```

```
Input nuval c2 c3 c4 c5 c6 c7 c8 c9 c10 c11 c12;
```

```
Cards;
```

```

2 199 448 729 1036 1362 1705 2063 2432 2813 3204 3605
3 47.5 85 120 151 184 216 249 281 310 337 361
4 23.2 37 49 59 69 79 89 97 106 113 120
5 14.9 22 28 33 38 42 46 50 54 57 60
6 11.1 15.5 19.1 22 25 27 30 32 34 36 37
7 8.89 12.1 14.5 16.5 18.4 20 22 23 24 26 27
8 7.50 9.9 11.7 13.2 14.5 15.8 16.6 17.9 18.9 19.8 21
9 6.54 8.5 9.9 11.1 12.1 13.1 13.9 14.7 15.3 16.0 16.6
10 5.85 7.4 8.6 9.6 10.4 11.1 11.8 12.4 12.9 13.4 13.9
12 4.91 6.1 6.9 7.6 8.2 8.7 9.1 9.5 9.9 10.2 10.6
15 4.07 4.9 5.5 6.0 6.4 6.7 7.1 7.3 7.5 7.8 8.0
20 3.32 3.8 4.3 4.6 4.9 5.1 5.3 5.5 5.6 5.8 5.9
30 2.63 3.0 3.3 3.4 3.6 3.7 3.8 3.9 4.0 4.1 4.2
60 1.96 2.2 2.3 2.4 2.4 2.5 2.5 2.6 2.6 2.7 2.7

```

```
;
```

```
data cv01;
```

```
Set cv01a;
```

```
Array cc[11] c2 - c12;
```

```
Do kval = 2 to 12;
```

```
    Cv01 = cc[kval-1];
```

```
    Output;
```

```
End;
```

```
Keep nuval kval cv01;
```

```
Run;
```

## ANOVA\_HOV: A SAS MACRO

```
Data cv05;
  Set cv05a;
  Array cc[11] c2 - c12;
  Do kval = 2 to 12;
    Cv05 = cc[kval-1];
    Output;
  End;
Keep nuval kval cv05;
Run;

proc sort data=cv01; by nuval kval;
proc sort data=cv05; by nuval kval;

data fmaxcv;
  merge cv05 cv01;
  by nuval kval;
run;

* Critical values for Ramsey's conditional test;
data cvw2;
  input cs lvalue05 uvalue05;
  cards;
4 -1.22 0.83
5 -1.14 1.50
6 -1.19 1.91
7 -1.23 2.15
8 -1.20 2.28
9 -1.20 2.33
10 -1.19 2.40
11 -1.19 2.39
12 -1.16 2.40
13 -1.17 2.34
14 -1.15 2.32
15 -1.13 2.31
16 -1.13 2.27
17 -1.11 2.23
18 -1.10 2.22
```

```

19 -1.09 2.19
20 -1.08 2.16
;
run;

options mprint minoperator noquotelenmax ;

%macro ANOVA_HOV (data=, iv=, dv=, nboots=);

* Prepare tables;
proc sort data = &data;
    by &iv;
proc means noprint data = &data;
    class &iv;
    var &dv;
    output out = cells mean = cellmn n = cs median = m50 kurt=cg2
var=vargrp;
data grand (drop=&iv _:) cells;
    set cells;
    if _type_ = 0 then output grand;
    else output cells;

data _null_;
    set grand;
    call symputx('mean',cellmn);
    call symputx('variance',vargrp);

proc means noprint data = cells;
    var cs;
    output out = total n = howmany mean=avg_cs max=max_cs min=min_cs
sum=sum_cs; run;
data _null_;
    set total;
    call symputx('n_groups',howmany);
    call symputx('avg_cs',avg_cs);
    call symputx('sum_cs',sum_cs);
run;

```

## ANOVA\_HOV: A SAS MACRO

```
%put total is &sum_cs;

data head;
  format labl $45. val $15. pr $15.;
  labl = "Independent Variable:"; val = "&iv"; pr = " "; output;
  labl = "N of Groups:"; val = put("&n_groups", $10.) ; pr = " ";
output;
  labl = "Dependent Variable:"; val = "&dv"; pr = " "; output;
  labl = "Total N of Observations:"; val = put("&sum_cs", $10.); pr =
" "; output;
  labl = " "; val = " "; pr = " "; output;
  labl = " "; val = "Obtained"; pr = " "; output;
  labl = "Test"; val = "Value"; pr = "p"; output;
run;

%do nn = 1 %to &n_groups;
  data _null_;
    set cells;
    if &nn = _n_;
      call symput("size&nn",cs);
%end;
run;

* Bartlett, BF, Leveneabs, Levenesquare, Obrien;
ods listing close;

proc glm data=&data ;
  class &iv;
  model &dv = &iv / ss3 ;
  means &iv / hovtest=bartlett;
  means &iv / hovtest=bf;
  means &iv / hovtest=levене(type=abs);
  means &iv / hovtest=levене(type=square);
  means &iv / hovtest=obrien (w=.50);
  ods output overallanova=anova1 hovfttest = hov bartlett = bart (keep=df
chisq probchisq);
run;
```

```

data hoveff (keep= df fvalue probf rename=(df=df_b fvalue=value
probf=p)) hoverr (rename=(df=df_w) keep=df);
  set hov;
  if source = 'Error' then output hoverr;
  else output hoveff;
run;

```

```

data hovtest;
  length labl $45. ;
  merge hoveff hoverr;
  n_groups = &n_groups;
  avg_cs = &avg_cs;
  if _n_ = 1 then do;
    labl = 'Brown and Forsythe';
  end;
  else if _n_ = 2 then do;
    labl = "Levene's (absolute values)";
  end;
  if _n_ = 3 then do;
    labl = "Levene's (squared values)";
  end;
  else if _n_ = 4 then do;
    labl = "O'Brien";
  end;
run;

```

```

data bart;
  length labl $45.;
  set bart;
  labl = 'Bartlett';
  rename df = df_b chisq = value probchisq = p;
run;

```

```

data sastests (keep= labl val pr);
  set hovtest bart ;
  val = input(round(value,.0001), $15.);
  pr = input(round(p,.0001), $15.);
run;

```

## ANOVA\_HOV: A SAS MACRO

```
* Bootstrap BF;
data obtained; *Retrieve sample BF value from previous PROC GLM ODS
table;
    set hov;
    if method = 'BF' and source ne 'Error';
    fo = fvalue;
    keep fo;
proc sort data = cells; by &iv;
data deviates;
    merge &data cells;
    by &iv;
    dev = abs(&dv - m50);
run;

data outboot(drop=__i);
    do Replicate = 1 to &nboots;
        do __i = 1 to &sum_cs;
            p = int(1 + &sum_cs*(ranuni(0)));
            set deviates point=p;
            output;
        end;
    end;
    stop;
run;
data outboot;
    set outboot;
    by replicate;
    if first.replicate then obsnum = 0;
    obsnum + 1;
run;

data outboot;
    set outboot;
    * Assign group membership to observations in bootstrap samples,
maintaining group sizes in original sample;
    cumulate = 0;
    %do i = 1 %to &n_groups;
```

```

        cell&i = &&size&i;
        if obsnum gt cumulate then group = &i;
        cumulate = cumulate + cell&i;
    %end;
run;

proc glm data=outboot outstat=stats;
    class &iv;
    model dev = &iv /ss3;
    means &iv / hovtest=bf;
    ods output hovftest = hov_boot;
    by replicate;
run;
data hov_boot;
    set hov_boot;
    if upcase(source) = upcase("&iv");
    f_boot = fvalue;
    keep f_boot;
ods listing;
data stats;
    if _n_ = 1 then set obtained;
    retain fo;
    set hov_boot;
    if fo >= f_boot then count+1;
    p=strip(put(1-count/&nboots,8.4));
run;
data bf_final (keep = labl val pr);
    length labl $ 45;
    set stats;
    if _n_ = &nboots;
    labl = 'Bootstrap Brown-Forsythe';
    val = input(round(fo,.0001), $15.);
    pr = input(p, $15.);
run;

*Ramsey conditional H0V test;
proc sort data=cells ; by vargrp;

```

## ANOVA\_HOV: A SAS MACRO

```

data kur1; *Division by zero here;
  set cells;
  b2 = 3*(cs-1)/(cs+1) + (cs-2)*(cs-3)/((cs+1)*(cs-1))*cg2;
  mnb2 = 3*(cs-1)/(cs+1);
  varb2 = 24*cs*(cs-2)*(cs-3)/((cs+1)*(cs+1)*(cs+3)*(cs+5));
  x = (b2 - mnb2)/sqrt(varb2);
  moment=6*(cs*cs- 5*cs+2)/((cs+7)*(cs+9))*sqrt(6*(cs+3)*(cs+5)/(cs*(cs-
2)*(cs-3)));
  a = 6 + 8/moment*(2/moment+sqrt(1+4/(moment**2)));
  z_b2 = (1-2/(9*a)-((1-2/a)/(1+x*sqrt(2/(a-4))))**((1/3)))/sqrt(2/(9*a));
  pbz_b2 = 2*(1-probnorm(abs(z_b2)));
  w2 = (cs+1)*b2/(cs-1) - 3;
run;

proc sort data = kur1; by cs;
proc sort data = cvw2; by cs;
data kur2;
  merge cvw2 kur1;
  by cs;
data kur3;
  set kur2;
  if b2 ne .;
  s = 0;
  if (cs <= 20) and (w2 < 0) and (w2 < lvalue05) then s = -1;
  else if (cs <= 20) and (w2 > 0) and (w2 > uvalue05) then s = 1;
  else if (cs > 20) and (b2 < 0) and (pbz_b2 < 0.05) then s = -1;
  else if (cs > 20) and (b2 > 0) and (pbz_b2 < 0.05) then s = 1;
run;

proc means data = kur3 noprint; * Only one obs in kur3;
  output out=kur4 sum(s) = sums;
run;

data kur5;
  if _n_ = 1 then set kur4;
  retain sums;
  set sastests;

```

```

        if (labl = 'Brown and Forsythe') or (labl = "O'Brien");
run;

data ramseytest (keep = labl val pr);
  length labl $ 45;
  set kur5;
  if sums < 0 and labl = 'Brown and Forsythe' then delete;
  if sums > -1 and labl = "O'Brien" then delete;
  labl = "Ramsey's Conditional Test";
run;

*Cochran;
data ordermax;
  set cells;
  nu = cs - 1;
  nu_x_var = nu*vargrp;
  nu_pooled = &sum_cs - &n_groups;

proc univariate data = ordermax noprint;
  var nu_x_var;
  output out= get_g_den sum = s_den;
run;

proc sort data = ordermax;
  by descending vargrp descending nu;
data _null_;
  set ordermax;
  if _n_ = 1 then do;
    call symput('nu_j',nu);;
    call symput('nu_pooled',nu_pooled);;
  end;
run;

proc univariate data = ordermax noprint;
  var vargrp nu_x_var;
  output out = largevar max=big_ol_var big_nu_var min=sml_ol_var
min_nu_var sum =sumvars g_gen;

```

## ANOVA\_HOV: A SAS MACRO

```
run;

data getarray;
  set cells;
  obs = _n_;
  array nn[100] n1-n100;
  retain n1-n100;
  nn[obs] = cs;
  keep n1-n100;
run;

data calc_harmonica;
  set getarray;
  har_avesize= harmean(of n1-n100);
  if _n_ = &n_groups;
run;
proc means data=calc_harmonica noprint;
  var har_avesize;
  output out=calc_harm mean=;
run;

data cochran_c_calc;
  set largevar ;
  g_test = (big_ol_var*&nu_j)/g_gen;
  cochran_c = big_ol_var/sumvars;
run;

data cochran_c_crit;
  merge largevar calc_harm (keep=har_avesize) ;

* Computation of critical values for Cochran C test of variances using
arthaver;

* +-----+
  Researcher inputs these three values
* +-----+;
  k = &n_groups; * Number of Groups;
  n1 = &avg_cs; * N for arthm_ave;
```

```

n2= har_avesize; *N for harmonic average;

* +-----+
  Critical value calculations for arithmetic
  average and harmonic mean
* +-----+;

  dfna = n1 - 1;
  dfda = (n1 - 1)*(k - 1);
  dfnh = n2 - 1;
  dfdh = (n2 - 1)*(k - 1);
  crit_f_a01 = finv(1-.01/k,dfna,dfda);
  crit_f_a05 = finv(1-.05/k,dfna,dfda);
  crit_f_a10 = finv(1-.10/k,dfna,dfda);
  crit_f_h01 = finv(1-.01/k,dfnh,dfdh);
  crit_f_h05 = finv(1-.05/k,dfnh,dfdh);
  crit_f_h10 = finv(1-.10/k,dfnh,dfdh);

  cochran_crit__arith_ave01 = 1/(1 + ((k-1)/crit_f_a01));
  cochran_crit__arith_ave05 = 1/(1 + ((k-1)/crit_f_a05));
  cochran_crit__arith_ave10 = 1/(1 + ((k-1)/crit_f_a10));
  cochran_crit__har_ave01 = 1/(1 + ((k-1)/crit_f_h01));
  cochran_crit__har_ave05 = 1/(1 + ((k-1)/crit_f_h05));
  cochran_crit__har_ave10 = 1/(1 + ((k-1)/crit_f_h10));

  crit_F_g01 = finv(1-.01/k,&nu_j,&nu_pooled - &nu_j);
  crit_F_g05 = finv(1-.05/k,&nu_j,&nu_pooled - &nu_j);
  crit_F_g10 = finv(1-.10/k,&nu_j,&nu_pooled - &nu_j);
  g_test_crit01 = 1/(1 + ((&nu_pooled/&nu_j - 1)/crit_f_g01));
  g_test_crit05 = 1/(1 + ((&nu_pooled/&nu_j - 1)/crit_f_g05));
  g_test_crit10 = 1/(1 + ((&nu_pooled/&nu_j - 1)/crit_f_g10));

run;

data cca (keep=lab1 cochran_c cochran_crit__arith_ave01
cochran_crit__arith_ave05 cochran_crit__arith_ave10
rename=(cochran_c=value
cochran_crit__arith_ave01=p01 cochran_crit__arith_ave05=p05
cochran_crit__arith_ave10=p10))

```

## ANOVA\_HOV: A SAS MACRO

```

        cch      (keep=labl      cochran_c      cochran_crit__har_ave01
cochran_crit__har_ave05 cochran_crit__har_ave10
                rename=(cochran_c=value      cochran_crit__har_ave01=p01
cochran_crit__har_ave05=p05 cochran_crit__har_ave10=p10))
        ccg (keep=labl g_test g_test_crit01 g_test_crit05 g_test_crit10
                rename=(g_test=value      g_test_crit01=p01
g_test_crit05=p05 g_test_crit10=p10));
    length labl $ 45;
    merge cochran_c_calc cochran_c_crit;
run;
```

```

data cctest (keep = labl val pr);
    set cca cch ccg;
    length pr $ 15;
    if _n_ = 1 then labl = "Cochran's C test (with arithmetic mean)";
    else if _n_ = 2 then labl = "Cochran's C test (with harmonic mean)";
    else if _n_ = 3 then labl = "G test";
    if value gt p01 then pr = "p < .01";
    if value gt p05 then pr = "p < .05";
    if value gt p10 then pr = "p < .10";
    else if value le p10 then pr = "p > .10";
    val = input(strip(put(value, 8.4)), $15.);
run;
```

```

* Fmax with arithmetic mean;
data fmax_cal;
    set largevar;
    fmax = big_ol_var/sml_ol_var;
    k = &n_groups; * Number of Groups;
    kval = k;
    n1 = &avg_cs;
    nu = round(n1 - 1);
    if nu < 11 then nuval = nu;
    else if nu = 11 then nuval = 10;
    else if 11 < nu < 15 then nuval = 12;
    else if 15 <= nu < 20 then nuval = 15;
    else if 20 <= nu < 30 then nuval = 20;
    else if 30 <= nu < 60 then nuval = 30;
```

```

else if 60 <= nu then nuval = 60;
run;
proc sort data=fmax_cal; by nuval kval;
proc sort data=fmaxcv; by nuval kval;
data fmaxcv1;
merge fmaxcv fmax_cal (in=a);
by nuval kval;
if a;run;
data fmax;
set fmaxcv1;
length lab1 $ 45 pr $ 15 ;
lab1 = 'F-max test (with arithmetic mean)';
val = input(round(fmax,.0001), $15.);

if fmax gt cv01 then pr = "p < .01";
else if fmax gt cv05 then pr = "p < .05";
else if fmax lt cv05 then pr = "p > .05";
keep lab1 val pr;
run;

* fmax with harmonic mean;
data fmax_cal_har;
merge largevar calc_harm (keep=har_avesize);
fmax = big_ol_var/sml_ol_var;
k = &n_groups; * Number of Groups;
kval = k;
n1 = har_avesize; * group size (harmonic mean);
nu = round(n1 - 1);

if nu < 11 then nuval = nu;
else if nu = 11 then nuval = 10;
else if 11 < nu < 15 then nuval = 12;
else if 15 <= nu < 20 then nuval = 15;
else if 20 <= nu < 30 then nuval = 20;
else if 30 <= nu < 60 then nuval = 30;
else if 60 <= nu then nuval = 60;
run;

```

## ANOVA\_HOV: A SAS MACRO

```
proc sort data=fmax_cal_har; by nuval kval;
proc sort data=fmaxcv; by nuval kval;
data fmaxcv1_har;
  merge fmaxcv fmax_cal_har (in=a);
  by nuval kval;
  if a;
data fmax_har ;
  set fmaxcv1_har;
  length lab1 $ 45 pr $ 15 ;
  lab1 = 'F-max test (with harmonic mean)';
  val = input(round(fmax,.0001), $15.);

  if fmax gt cv01 then pr = "p < .01";
  else if fmax gt cv05 then pr = "p < .05";
  else if fmax lt cv05 then pr = "p > .05";
  keep lab1 val pr;
run;

* Z variance;

data _null_;
  set anova1;
  if source = 'Error' then call symput('mse',ms);
  if source = 'Corrected Total' then call symput('totN',DF+1);
run;

data z2;
  set cells;
  c= 2 + (1/cs);
  z=sqrt((c*(cs-1)*vargrp)/&mse)-sqrt(c*(cs-1)-(c/2));
  z_squared=z**2;
  keep z_squared;
run;
proc means noprint data = z2 sum;
  var z_squared;
  output out = z2a sum = sum_z2;
```

```

data z2b (keep = lab1 val pr);
  set z2a;
  length lab1 $ 45 ;
  lab1 = "Z-variance";
  value = sum_z2/(&n_groups-1);
  df_b = &n_groups-1;
  df_w = &totn - &n_groups;
  p= 1-probchi(sum_z2,df_b);
  val = input(round(value,.0001), $15.);
  pr = input(round(p,.0001), $15.);
run;

* Modified Z variance;

proc means noprint data=&data;
  var &dv;
  by &iv;
  output out=tt var=var_group n=n_group mean=mean_group;

data ones (drop=_type_ _freq_);
  merge &data tt;
  by &iv;

* calc kurtosis;
data temp;
  set ones;
  by &iv;
  z=(&dv - mean_group) / sqrt ((n_group - 1) * var_group / n_group);
  z_4power=z**4;
  if first.&iv then sumz4 = 0;
  sumz4 + z_4power ;
  if last.&iv then output;
  keep &iv var_group n_group mean_group sumz4;

data temp ;
  set temp;
  K_i=sumz4/(n_group - 2);
proc means noprint data = temp;

```

## ANOVA\_HOV: A SAS MACRO

```

var K_i;
output out = temp2 mean = K;

data temp3;
if _N_ = 1 then set temp2;
retain K;
set temp;
c= 2*((2.9+0.2/n_group)/K)**(1.6*(n_group-1.8*K+14.7)/n_group);
z=sqrt( (c*(n_group-1)*var_group) / &mse ) - sqrt (c*(n_group-1)-
(c/2) );
z_squared=z**2;
keep z_squared;

proc means noprint data = temp3 sum;
var z_squared;
output out = z2x sum = sum_z2;

data z2y (keep = lab1 val pr);
set z2x;
length lab1 $ 45 ;
lab1 = "MZ-variance test";
value = sum_z2/(&n_groups-1);
df_b = &n_groups-1;
p = 1 - probchi(sum_z2,df_b);
val = input(round(value,.0001), $15.);
pr = input(round(p,.0001), $15.);
run;

* Results;
data toprint;
set head sastests bf_final ramseytest z2b z2y cctest fmax fmax_har;
check_p = substr(pr,1,2);
if check_p = '0.' then do;
    substr(pr,1,2) = ' .';
end;
pr = LEFT(pr);
drop check_p;
run;

```

```
proc print data = toprint noobs label;  
  title 'Tests of Homogeneity of Variance';  
  label labl = '00'x val = '00'x pr = '00'x;  
run;  
  
%mend ANOVA_HOV;
```