Paper SD-81

ANOVA_HOV: A SAS® Macro for Testing Homogeneity of Variance in One-Factor ANOVA Models

Diep T. Nguyen, Thanh V. Pham, Patricia Rodríguez de Gil, Tyler Hicks, Yan Wang, Isaac Li,
Aarti Bellara, Jeanine L. Romano, Eun Sook Kim,
Harold Holmes, Yi-Hsin Chen, Jeffrey D. Kromrey,
University of South Florida

ABSTRACT

Variance homogeneity is one of the critical assumptions when conducting ANOVA as violations may lead to perturbations in Type I error rates. Previous empirical research suggests minimal consensus among studies as to which test is appropriate for a particular analysis. This paper provides a SAS macro for testing the homogeneity of variance assumption in one-way ANOVA models using fourteen different approaches. Using simulation methods, the fourteen tests were compared in terms of their Type I error rate and statistical power.

KEYWORDS: Analysis of variance, homogeneity of variance assumption, simulation study.

INTRODUCTION

In an ANOVA procedure, the assumption of homogeneity of variance (HOV) is that treatment variances are equal. That is,

$$H_0 = \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$$

Moderate deviations from the assumption of equal variances may not seriously affect the results in ANOVA (Glass, Peckham, & Sanders, 1972). Because the ANOVA procedure may be robust to small deviations from the HOV assumption, researchers may only need to be concerned about large deviations from the HOV assumption. However, the classic *F* test used to evaluate HOV is sensitive to departures of normality, for which researchers should turn to alternative tests when the assumption of normality is not met.

Considering the ongoing controversy on testing the homogeneity of variance and the minimal consensus among studies as to which test is appropriate for a particular analysis, two over-arching research questions guide this area of inquiry: Does it make sense to statistically test the homogeneity of variance assumption? Which method should we use for testing the homogeneity of variance assumption? The goal of the current paper is to provide a SAS macro for testing the homogeneity of variance assumption using fourteen different tests associated with ANOVA, some of which are not readily available in SAS. The paper provides the macro programming language as well as the output from an executed example of the macro. Brief results of a simulation study comparing the HOV tests are also presented. The following section describes statistical methods available in the macro for testing the HOV assumption and Table 1 presents the test statistics and equations for each method in detail.

STATISTICAL METHODS AVAILABLE IN SAS

BARLETT TEST. Bartlett (1937) proposed a special use of the chi-square test for testing the HOV assumption, under which the null hypothesis of equal variances will be rejected if the Bartlett's chi-square statistic is greater than the critical chi-square value with *df* = *k-1*; otherwise, the null hypothesis fails to be rejected.

LEVENE (ABSOLUTE AND SQUARED). Levene proposed to use the absolute residual values or squared residuals, which transforms the test of variances into a test of means that is relatively robust to the normality assumption. The *W* statistics of the absolute values and squared residuals are compared to the *F* critical value (F_{crit}) with *df* = k-1 and N-*k* for numerator and denominator, respectively. The null hypothesis is rejected if $W > F_{crit}$.

BROWN-FORSYTHE. Brown and Forsythe (1974) followed the idea of Levene's test but used the group median instead of the group mean in calculation of absolute deviation.

O'BRIEN. O'Brien (1979) proposed a test that transforms original scores so they would represent sample variances. The mean of the transformed values per group, $\overline{r_j} = \frac{\sum r_{ij}}{n_j} = {s_j}^2$, 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.

HOV Test	Test Statistic and Equation	Notation
Barlett	$\chi^2 = \frac{(N-k)log\left[\frac{\sum_{j=1}^{k}(n_j-1)S_j^2}{(N-k)}\right]-\sum_{j=1}^{k}(n_j-1)\log(S_j^2)}{1+\frac{\left(\sum_{j=1}^{k}\frac{1}{n_j-1}\right)-\frac{1}{(N-k)}}{3(k-1)}}$,	N = total sample size; N_j = group j sample size; k = number of groups; S_j^2 = group j variance.
Levene (Absolute and Squared)	$Z_{ij} = \left\|Y_{ij} - \bar{Y}_{.j}\right\| \; and \; Z_{ij} = (Y_{ij} - \bar{Y}_{.j})^2,$ $W = \frac{(N-k)\sum_{j=1}^{k} n_j(\bar{Z}_{.j}-\bar{Z}_{..})^2}{(k-1)\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Z_{ij}-\bar{Z}_{.j})^2}$	Y_{ij} = raw score; $\bar{Y}_{.j}$ = mean of the j^{th} group; $\bar{Z}_{.j}$ = group mean of Z_{ij}; $\bar{Z}_{..}$ = grand mean.
Brown-Forsythe (BF)	$z_{ij} = \left\|Y_{ij} - \tilde{Y}_{ji}\right\|,$ $W = \frac{(N-k)\sum_{j=1}^{k} n_j(\bar{Z}_{.j} - \bar{Z}_{..})^2}{(k-1)\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Z_{ij} - \bar{Z}_{.j})^2}$	$\tilde{Y}_j$ = median of group j; z_{ij} = transformed value of Y_{ij}; $\bar{Z}_{.j}$ = group mean of Z_{ij}; $\bar{Z}_{..}$ = grand mean.
O'Brien	$r_{ij}(w) = \frac{(w+n_j-2)n_j(Y_{ij}-\bar{Y}_{.j})^2 - ws_j{}^2(n_j-1)}{(n_j-1)(n_j-2)}$,	$s_j{}^2$ = the within-group unbiased estimate of variance for sample j; w ($0 \le w \le 1$) = weighting factor.
Ramsey	$b_2 = m_4/m_2{}^2,$ $b_{2j} = \frac{\frac{\sum\left(Y_{ij}-\bar{Y}_{.j}\right)^4}{n_j}}{\left[\frac{\sum\left(Y_{ij}-\bar{Y}_{.j}\right)^2}{n_j}\right]^2}.$	$m_r = \Sigma(Y_{ij} - \bar{Y}_{.j})^r / n_j$.
Cochran's C	$C = \frac{s_{max}^2}{\sum s_j^2},$ $Critical\ C = \frac{1}{1+\frac{F_{\frac{\alpha}{k},\ n-1,(k-1)(n-1)}}{k-1}}.$	n = number of observations in each group (for the balanced design); F = the critical value of F at α/k with $df = n-1, (k-1)(n-1)$.
G test	$G = \frac{v_{max}s_{max}{}^2}{\sum v_j s_j{}^2},$ $Critical\ G = \frac{1}{1+\frac{v_{pool}/(v_{max}-1)}{F_{\frac{\alpha}{k},\ \overline{n_j}-1,(k-1)(\overline{n_j}-1)}}}.$	v_{pool} = the pooled degrees of freedom; v_{max} = the degrees of freedom for the group within the largest variance; $\overline{n_j}$ = the mean number of observations in each group; $F_{\frac{\alpha}{k}}$ = the critical value of F at α/k with df = $\overline{n_j}-1,\ (k-1)(\overline{n_j}-1)$.
F-max	$F_{max} = \frac{s_{max}^2}{s_{min}^2}$	s_{max}= the largest group variance; s_{min}= the smallest group variance.
Z-variance	$z = \sqrt{2\chi^2} - \sqrt{2(df)-1}\ ,$ $\chi_{n-1}{}^2 = \frac{(n-1)S^2}{\sigma^2},$ $Z_j = \sqrt{\frac{c(n_j-1)S_j{}^2}{MS_w}} - \sqrt{c(n_j-1)-1},$ $F = \frac{\sum_{j=1}^{k} Z_j{}^2}{k-1}.$	s^2 = sample variance estimate; σ^2 = true population variance; $c = 2 + 1/n_j$; MS_w= pooled within-cells mean square across all groups (or cells in a more complex factorial design).
Modified Z-variance	$c = 2.0\left(\frac{2.9+.2/n_j}{K}\right)^{1.6(n_j-1.8K+14.7)/n_j},$ $K = \frac{\sum Z_{ij}^4}{n_j-2},$ $Z_{ij} = \frac{Y_{ij}-\bar{Y}_{.j}}{\sqrt{\frac{n_j-1}{n_j}S_j^2}}.$	K = the mean of the kurtosis indices from all groups.

Table 1. HOV Tests Statistics and Equations

STATISTICAL METHODS NOT AVAILABLE IN SAS

BOOTSTRAP BROWN-FORSYTHE TEST. Boos and Brownie (2004) recommended using the median version of Levene's test statistic (i.e., the Brown-Forsythe statistic), then obtaining the p-value via the bootstrap, which provided more power than the F distribution version.

RAMSEY CONDITIONAL TEST: BROWN-FORSYTHE OR O'BRIEN. Ramsey's (1994) conditional test is based on using Brown-Forsythe or O'Brien method, conditional on a test of kurtosis. Kurtosis ($b2$) for each of the *k* groups is computed by using Pearson's traditional measure; the kurtosis value for each group is then compared to critical values obtained from a table provided by Ramsey and Ramsey (1993). The O'Brien test will be applied if the data are platykurtic and the BF test will be used if the data are mesokurtic or leptokurtic.

COCHRAN'S *C* TEST. The Cochran's *C* (Cochran, 1941) test is a ratio of the largest group variance to the sum of sample variances. If the obtained value exceeds the critical value, the null hypothesis of variance homogeneity is rejected. The arithmetic and harmonic means are used for computations of Cochran's *C* test.

***G* TEST.** The *G* test 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. If the obtained value exceeds the critical value, the null hypothesis of variance homogeneity is rejected.

***F*-MAX.** Hartley (1950) developed the Hartley's or *F*-max test for comparing three or more group variances, which requires independent random samples of the same size from normally distributed populations (Ott & Longnecker, 2010). The value of *F*-max is compared to a critical value from the table containing the test sampling distribution. The arithmetic and harmonic means are used for computations of the *F*-max test.

***Z*-VARIANCE.** Overall and Woodward (1974) proposed the *Z*-variance test based on Fisher and Yates' formula (1963). It transforms the chi-square statistics with large *df* into standard normal deviate *z*-scores. It 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. To improve the performance of the *Z*-variance test when sample distributions are leptokurtic or skewed, Overall and Woodward (1976) developed the modified *Z*-variance approach to testing the HOV by implementing a *c* value based on sample size, skewness, and kurtosis.

THE ANOVA_HOV MACRO

```
* Critical values for F-max tests;
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;
Data cv05; Set cv05a;
    Array cc[11] c2 - c12;
    Do kval = 2 to 12;
      Cv05 = cc[kval-1];
      Output;
    End;
    Keep nuval kval cv05;
proc sort data=cv01; by nuval kval;
proc sort data=cv05; by nuval kval;
data fmaxcv; merge cv05 cv01; by nuval kval; run;
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
;
options mprint minoperator noquotelenmax;
%macro 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);
    %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;
    %do nn = 1 %to &n_groups;
data _null_; set cells;
    if &nn = _n_;
    call symput("size&nn",cs);
    %end;
* Bartlett, BF, Levene-absolute, Levene-square, and Obrien Tests;
    ods listing close;
proc glm data=&data;
    class &iv;
    model &dv = &iv / ss3;
    means &iv /  hovtest=bartlett;
    means &iv /  hovtest=bf;
    means &iv /  hovtest=levene(type=abs);
    means &iv /  hovtest=levene(type=square);
    means &iv /  hovtest=obrien (w=.50);
    ods output overallanova=anova1 hovftest = hov bartlett = bart (keep=df chisq
probchisq);
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;
data hovtest; length labl $45.; merge hoveff hoverr;
    n_groups = &n_groups;
    avg_cs = &avg_cs;
    if _n_ = 1 then labl = 'Brown and Forsythe';
    if _n_ = 2 then labl = "Levene's (absolute values)";
    if _n_ = 3 then labl = "Levene's (squared values)";
    if _n_ = 4 then labl = "O'Brien";
data bart; length labl $45.; set bart;
    labl = 'Bartlett';
    rename df = df_b chisq = value probchisq = p;
data sastests (keep= labl val pr); set hovtest bart;
    val = input(round(value,.0001), $15.);
    pr = input(round(p,.0001), $15.);
* Bootstrap BF;
data obtained;set hov;
*Retrieve sample BF value from previous PROC GLM ODS table;
    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);
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;
data outboot; set outboot; by replicate;
    if first.replicate then obsnum = 0; obsnum + 1;
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;
proc glm data=outboot outstat=stats;
    class &iv;
    model dev = &iv /ss3;
    means &iv / hovtest=bf;
    ods output hovftest = hov_boot; by replicate;
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));
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.);
*Ramsey's Conditional Test;
proc sort data=cells; by vargrp;
data kur1;    set cells;
*Division by zero here;
    b2 = 3*(cs-1)/(cs+1) + (cs-2)*(cs-3)/((cs+1)*(cs-1))*cg2;
    mnb2 = 3*(cs-1)/(cs+1);
    x = (b2 - mnb2)/sqrt(varb2);
    varb2 = 24*cs*(cs-2)*(cs-3)/((cs+1)*(cs+1)*(cs+3)*(cs+5));
    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;
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;
    if (cs <= 20) and (w2 > 0) and (w2 > uvalue05) then s = 1;
    if (cs > 20) and (b2 < 0) and (pbz_b2 < 0.05) then s = -1;
    if (cs > 20) and (b2 > 0) and (pbz_b2 < 0.05) then s = 1;
proc means data = kur3 noprint;
* Only one obs in kur3;
    output out=kur4 sum(s) = sums;
data kur5;
    if _n_ = 1 then set kur4;
    retain sums;
    set sastests;
    if (labl = 'Brown and Forsythe') or (labl = "O'Brien");
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";
```

```
*Cochran's Test;
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;
proc sort data  = ordermax; by desending vargrp desending nu;
data _null_; set ordermax;
    if _n_ = 1 then call symput('nu_j',nu); call symput('nu_pooled',nu_pooled);
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;
data getarray; set cells;
       obs = _n_;
       array nn[100] n1-n100;
       retain n1-n100;
       nn[obs] = cs;
       keep n1-n100;
data calc_harmonica; set getarray;
       har_avesize= harmean(of n1-n100);
       if _n_ = &n_groups;
proc means data=calc_harmonica noprint;
    var har_avesize;
    output out=calc_harm mean=;
data cochran_c_calc; set largevar;
    g_test = (big_ol_var*&nu_j)/g_gen;
    cochran_c = big_ol_var/sumvars;
data cochran_c_crit; merge largevar calc_harm (keep=har_avesize);
* Computation of critical values for Cochran C test of variances using arithmetic
mean;
    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));
data cca (keep=labl 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)) 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;
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)";
    if _n_ = 2 then labl = "Cochran's C test (with harmonic mean)";
    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";
    if value le p10 then pr = "p > .10";
    val = input(strip(put(value, 8.4)), $15.);
* F-max Test 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;
    if nu = 11 then nuval = 10;
    if 11 < nu < 15 then nuval = 12;
    if 15 <= nu < 20 then nuval = 15;
    if 20 <= nu < 30 then nuval = 20;
    if 30 <= nu < 60 then nuval = 30;
    if 60 <= nu then nuval = 60;
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 labl $ 45 pr $ 15;
    labl = 'F-max test (with arithmetic mean)';
    val = input(round(fmax,.0001), $15.);
    if fmax gt cv01 then pr = "p < .01";
    if fmax gt cv05 then pr = "p < .05";
    if fmax lt cv05 then pr = "p > .05";
    keep labl val pr;
* F-max Test 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;
    if nu = 11 then nuval = 10;
    if 11 < nu < 15 then nuval = 12;
    if 15 <= nu < 20 then nuval = 15;
    if 20 <= nu < 30 then  nuval = 20;
    if 30 <= nu < 60 then nuval = 30;
    if 60 <= nu then nuval = 60;
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 labl $ 45 pr $ 15;
     labl = 'F-max test (with harmonic mean)';
     val = input(round(fmax,.0001), $15.);
     if fmax gt cv01 then pr = "p < .01";
     if fmax gt cv05 then pr = "p < .05";
     if fmax lt cv05 then pr = "p > .05";
     keep labl val pr;
* Z-Variance Test;
data _null_; set anova1;
     if source = 'Error' then call symput('mse',ms);
     if source = 'Corrected Total' then call symput('totN',DF+1);
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;
proc means noprint data = z2 sum;
     var z_squared;
     output out = z2a sum = sum_z2;
data z2b (keep = labl val pr); set z2a; length labl $ 45;
     labl = "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.);
*Modified Z-Variance Test;
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;
*calculate 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;
     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 = labl val pr); set z2x; length labl $ 45;
     labl = "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.);
* Outputs of Testing HOV Methods;
```

```
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 substr(pr,1,2) = ' .';
    pr = LEFT(pr);
    drop check_p;
proc print data = toprint noobs label;
    title 'Tests of Homogeneity of Variance';
    label labl = '00'x val = '00'x pr = '00'x;
run;

%mend hov;
```

MACRO EXECUTION

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. The researchers need to specify all three parameters: the independent variable ("iv =" parameter), the dependent variable ("dv=" parameter), and the number of bootstraps used in macro ("nboots =" parameter).
The data TEST below with 20 observations on two variables (group and score) are used to illustrate the macro execution.

```
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;
```

In the macro call statement below, the data (TEST), the independent variable (group), the dependent variable (*Y* or score), and the number of bootstraps (50) are used to execute the macro.

```
%hov (data = test, iv = group, dv = Y, nboots = 50); run;
```

OUTPUT EXAMPLE OF MACRO

The output of the macro is shown in Output 1. The first part of the output shows the data information. The second part includes the obtained value and associated *p*-value for each HOV test. For the last 5 tests, only the alpha levels are reported because their p-values are not calculated.

Independent Variable:	group	
N of Groups:	4	
Dependent Variable:	Y	
Total N of Observations:	20	
Test	Obtained Value	p
Brown and Forsythe	0.9464	.4416
Levene's (absolute values)	1.9113	.1685
Levene's (squared values)	0.7581	.5338
O'Brien	0.5754	.6394
Bartlett	4.3665	.2245
Bootstrap Brown-Forsythe	0.9464	.4400
Ramsey's Conditional Test	0.9464	.4416
Z-variance	0.9573	.4118
MZ-variance test	0.1729	.9148
Cochran's C test (with arithmetic mean)	0.5410	p > .10
Cochran's C test (with harmonic mean)	0.5410	p > .10
G test	0.8549	p > .10
F-max test (with arithmetic mean)	24.0622	p < .05
F-max test (with harmonic mean)	24.0622	p > .05

Output 1. Sample output for the ANOVA_HOV macro

ACCURACY AND PRECISION OF FOURTEEN TESTING HOV APPROACHES

A simulation study was conducted to investigate the accuracy and precision of fourteen approaches for testing the homogeneity of variance assumption in one-way ANOVA models. Six design factors were included: (1) number of groups, (2) average number of observations per group, (3) sample size pattern, (4) variance pattern, (5) maximum group variance ratio, and (6) population distribution ($\gamma 1$ = 0.00 and $\gamma 2$ =0.00, $\gamma 1$ = 1.00 and $\gamma 2$ = 3.00, $\gamma 1$ = 1.50 and $\gamma 2$ =5.00, $\gamma 1$ = 2.00 and $\gamma 2$ = 6.00, $\gamma 1$ = 0.00 and $\gamma 2$ =25.00, as well as $\gamma 1$ = 0.00 and $\gamma 2$ = -1.00, where $\gamma 1$ and $\gamma 2$ represent skewness and kurtosis, respectively). Since the space is limited, sample patterns and variance patterns are available upon request. The performance of the fourteen testing HOV approaches was examined at different nominal alpha levels: .01, .05, .10, .15, .20, and .25. For each condition, 5,000 samples were generated. Type I error and statistical power were examined. Bradley's (1978) liberal criterion for robustness was set at ± 0.5α around the nominal alpha and eta-square (η^2) effect sizes were estimated to explore the significant impacts of the research design factors on the variability in the Type I error estimates. Finally, statistical power was estimated only for conditions in which Type I error was adequately controlled.

SIMULATION RESULTS

The distributions of Type I error rate estimates across all simulation conditions for the fourteen approaches of testing the HOV assumption at the nominal alpha level of .05 are presented in Figure 2. Among these approaches, Levene with the squared residuals, Brown-Forsythe, O'Brien, Ramsey's conditional procedure, and Bootstrap Brown-Forsythe tests were the five tests that controlled Type I error adequately. Equivalently, these five tests had the largest proportions of conditions meeting Bradley's criterion —Bradley's proportions hereafter— (.77, .72, .58, .58, and .52, respectively). As the nominal alpha level increased, the Bradley proportions for the fourteen tests increased as well. Estimated power of these five HOV tests is presented in Figure 3. As shown in Figure 3, the power differences among the five tests were very subtle; O'Brien had slightly less power and Bootstrap Brown-Forsythe had slightly greater power than other tests.

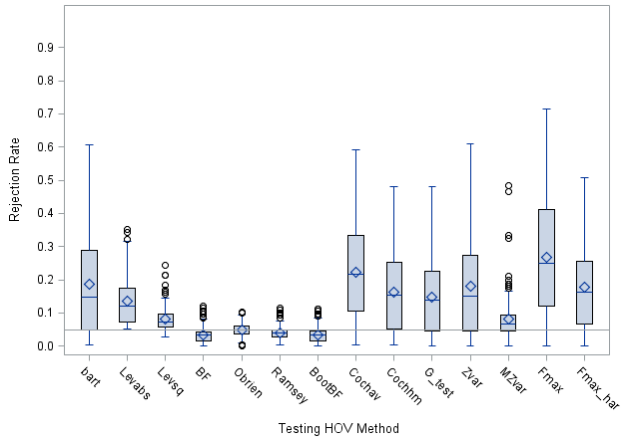

Figure 2. **Distributions of Type I Error rates estimates at .05 significance level**

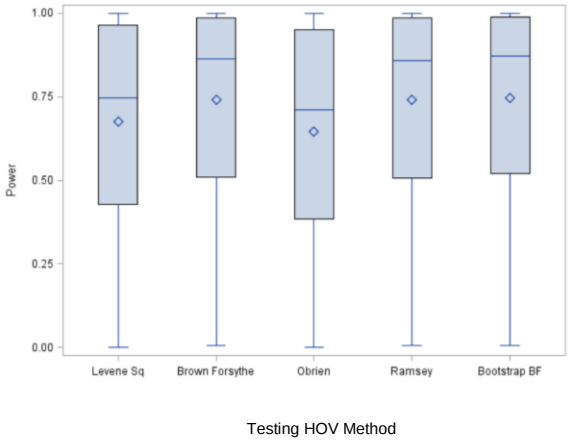

Figure 3. **Distributions of Estimated Power for the Five Best HOV tests**

Cell size and population shape were the two factors that had the largest impact on Type I error estimates. Bradley's proportions increased for the five tests having adequate type I error control as sample size increased. With small sample size (i.e., nj = 5), O'brien's test was the only test that had a Bradley's proportion greater than .60. When sample size was 10 in this study, only Ramsey's test reached a Bradley's proportion greater than .80. When sample size increased to 20, the Bradley's proportions were all greater than .80for the five tests with adequate Type I error control, with Ramsey's having a slightly higher Bradley's proportion.

O'Brien and Ramsey in general showed adequate Type I error control across various population shapes (i.e., all the Bradley's proportions greater than .50). Bradley's proportions decreased for Brown-Forsythe, Bootstrap Brown-Forsythe, Levene with the squared residual, and modified *Z*-variance tests depending on the degree of nonnormality. Brown-Forsythe and Bootstrap Brown-Forsythe became conservative when the distribution was extremely leptokurtic (i.e., $\gamma 1$ = 0.00 and $\gamma 2$ =25.00; skewness and kurtosis respectively). Levene test with the squared residual and modified *Z*-variance test became liberal when the population skewness and kurtosis increased. Finally, 12 out of the 14 tests had Bradley's proportions greater than .50 when the population was normally distributed. Among them, Bartlett, Cochran with the harmonic mean, and *Z*-variance tests had perfect Bradley's proportions of 1.0. For *F*-max test and Cochran *C* test with the arithmetic mean, Bradley's proportions were unacceptably low or simply zero regardless of distribution shapes.

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 especially regarding variance homogeneity as violations may impact Type I error rates. As noted, the O'Brien and Ramsey conditional procedure seemed to be the most robust to this violation. While O'Brien is available in by default in Base SAS, Ramsey's conditional procedure was not. This macro provides the researcher with the ability to conduct this procedure along with 13 other different methods from the literature.

REFERENCES

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Statistical Society, Series A, 160*, 268–282.

Boos, D. D. & Brownie, C. (2004). Comparing variances and other measures of dispersion. *Statistical Science*, *19*, 571-578.

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

Cochran, (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.

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

Glass, Peckham, & Sanders. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research, 42*(3), 237-288.

Hartley (1950). The maximum f-ratio as a short-cut test for heterogeneity of variance. *Biometrika, 37,* 308-312.

O'Brien, P. C., & Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, 549-556.

Ott, L. R. & Longnecker (2010). *An introduction to statistical methods and data analysis.* , Brooks/Cole, Cengage Learning, Belmont, CA.

Overall, J. E., & Woodward, J. A. (1974). A simple tests for heterogeneity of variance in complex factorial designs. *Psychometrika, 39*(3), 311-318.

Overall, J. E., & Woodward, J. A. (1976). A robust and powerful test for heterogeneity of variance. *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,* 231-241.

Ramsey, P. H. (1994). Testing variances in psychological and educational research. *Journal of Educational Statistics*, *19*, 23-42.

CONTACT INFORMATION

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

Diep T. Nguyen
Educational Measurement & Research program
College of Education, EDU 105
University of South Florida
4202 East Fowler Avenue, EDU 105
Tampa, FL 33620-5650
Phone: (813) 974-6064
Fax: (813) 974-5910
E-mail: diepnguyen@mail.usf.edu