

Using SAS® to Perform the Analysis of Means for Variances Test

Peter Wludyka, University of North Florida, Jacksonville, FL

ABSTRACT

The Analysis of Means for Variances (ANOMV) is a method for testing the equality of K variances from normal populations. The test can be performed by creating a decision chart that resembles a Shewhart control chart. Should any of the K sample variances (or equivalently, standard deviations) plot outside the decision limits the homogeneity of variances (HOV) hypothesis is rejected. Using the ANOMV decision chart allows practitioners to assess both statistical and practical significance. After a brief description of the ANOMV test, SAS code (a Macro) for generating an ANOMV decision chart and related output for balanced designs is presented. The MACRO will also produce an Analysis of Means (ANOM) decision chart.

INTRODUCTION

Often a researcher is interested in whether each of K populations has the same variance (standard deviation). For example, suppose that six different tools can be used to drill holes in a metal flange. Since there will always be some variability in the diameters of the holes (regardless of which tool is used) it is desirable to adopt a tool with "low variability". In order to compare the six tools with respect to variability an experiment can be performed. The experimental details are important, but for the methods presented in this paper it is required that six independent random samples of hole diameters be collected — one sample for each tool. The population of diameters associated with tool k has variance σ_k^2 . The general hypothesis to test is

$$H_0 : \sigma_1^2 = \dots = \sigma_K^2 \quad (1)$$

Hypothesis (1) will be called the Homogeneity of Variance (HOV) Hypothesis. Numerous statistical tests have been proposed for testing the HOV hypothesis. Under the circumstance that the populations being sampled from are normal one may choose among several general purpose tests including Bartlett's test, Hartley's test, and the Analysis of Means for Variances (ANOMV). See Wludyka and Nelson (1997) for a complete discussion. The key points are

- These three tests have roughly the same power.
- These are all-purpose tests that work well for all variance configurations.
- These tests should not be used when non-normality is suspected.

In particular they should not be used for kurtotic (fat-tailed) or skewed populations since samples from these type populations will lead to Type I error rates far in excess of the nominal rate (α). Under these circumstances other test methods should be used.

The main advantage in using the ANOMV test is that the test can be performed graphically, making assessment of both practical and statistical significance easier.

ANALYSIS OF MEANS TYPE TESTS

ANOM type tests are tests in which K statistics are plotted on a decision chart to determine whether to accept/reject the hypothesis that the K populations are identical with respect to some parameter. Typically the parameter is the means or the variance (standard deviation). The decision chart typically resembles a Shewhart control chart. Instead of control limits the decision chart has decision limits. The SAS Macro presented in this paper is for balanced designs (that is, the same size sample, n , is selected from each of the K populations).

THE ANALYSIS OF MEANS (ANOM)

ANOM is a test for comparing the means of K populations. In ANOM the sample means are plotted on an appropriately constructed decision chart. If one or more of the sample means plot outside either the upper or lower decision line the equal means hypothesis is

rejected. See for example Figure 2.

THE ANOMV TEST FOR BALANCED DESIGNS

When the data is balanced the ANOMV test is performed by calculating each of the K sample variances

$$S_k^2 = \frac{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}{(n-1)}$$

and plotting the sample variances (or standard deviations) on a decision chart. This chart (see Figure 1) has upper (UDL) and lower (LDL) decision lines which are used to perform the ANOMV test. The HOV hypothesis is rejected whenever one or more the variances plot outside the decision lines.

THE ANOMV MACRO

The ANOMV macro can be used to perform the ANOMV test and the ANOM test. The user specifies whether ANOMV or both tests are to be performed. Two macros can be used:

- %ANOMV, in which the ANOM and ANOMV critical values are read from files (which can be downloaded).
- %ANOMVM, in which the user supplies the ANOM and ANOMV critical values

ANOMV test results are presented in terms of the standard deviation. SAS source code for %ANOMV appears in this paper. It can also be downloaded. The SAS source code for %ANOMVM can be downloaded. Instructions for downloading follow.

USING %ANOMV

In order to use %ANOMV the user must have the following:

- source code for the %ANOMV macro
- files containing the critical values
 - ANOMV critical values
 - Large sample ANOMV critical values
 - ANOM critical values
- A data set containing the observations

The following parameters must be supplied to the %ANOMV macro.

- k , the number of populations being compared
- n , the sample size
- α , the level of significance
- ds , the name of the data set which contains two variables: a class variable defining the populations and an observation variable
- var , the name of the observation variable
- $classvar$, the name of the class variable
- $tops$, a variable indicating which tests are to be performed

DATA SET EXAMPLE

```
data example1;
  input tool @;
  do i=1 to 5;
    input diameter @;
    output;
  end;
  drop i;
  cards;
1 10 20 25 15 12
2 7 45 111 23 79
3 78 44 55 19 16
4 4 13 19 3 22
5 2 6 4 8 12
6 55 70 35 17 29
;
```

This data set corresponds to five diameter measurements ($n = 5$) on each of six tools.

%ANOMV EXAMPLE

Suppose that

```
%anomv(k=6,n=5,alpha=0.05,ds=example1,var=diameter,
classvar=tool,tops=2);
```

is invoked. Then there are six populations being compared, there is a balanced design with 5 observations per population, the level of significance is 5%, the data set containing the observations is 'example1', the observational variable in the data set is named 'diameter', the classification variable is 'tool', and assigning the value of 2 to tops requests that both ANOMV and ANOMV be performed.

%ANOMV GRAPHICS OUTPUT

The interpretation of the ANOMV decision chart is first made with respect to the nominal level of statistical significance. For the example in Figure 1 the significance level is 5%.

1. Variability in diameter for the six tools is different, since one or more (in this case two) standard deviations plot outside the decision lines.
2. Tool two exhibits greater variability than the average of the six.
3. Tool five exhibits less variability than the average of the six.

Next, an assessment of practical significance is made. That is a quasi-statistical question requiring subject matter knowledge. The ANOMV decision chart is helpful in assessing practical significance.

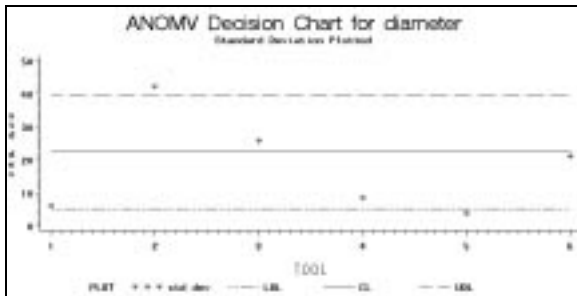


Figure 1: ANOMV Decision Chart

The interpretation of the ANOM decision chart in Figure 2 is that there are no statistically significant differences among the means (that is, none of the means differ significantly from the average for the six tools).

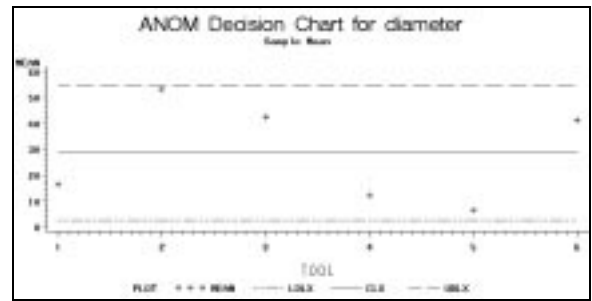


Figure 2: ANOM Decision Chart

%ANOMV TABULAR OUTPUT

Tabular output supplements the ANOMV Decision Chart. Observe that in ANOMV Decision Table (Table 2) two sample standard deviations plot outside the decision limits. The interpretation is the same as that arising from the decision chart. Most find the decision chart more user friendly.

Table 1: ANOMV Properties Table

Simultaneous ANOMV Test for Equality of $k = 6$ Variances
Conservative Wludyka & Nelson Critical Values for alpha 05

degrees of freedom	lower critical value	upper critical value
4	.009	0.5175

Table 2: ANOMV Decision Table

ANOMV Decision Table for alpha 05
class variable tool ($k = 6$) variable diameter ($n = 5$)
*** indicates standard deviation plots outside decision lines

TOOL	lower decision limit	diameter standard deviation	upper decision limit	reject
1	5.21543	6.1074	39.5479	
2	5.21543	42.1900	39.5479	***
3	5.21543	25.8515	39.5479	
4	5.21543	8.5849	39.5479	
5	5.21543	3.8471	39.5479	***
6	5.21543	21.1707	39.5479	

The output below (Tables 3 and 4) is for the ANOM test. It supplements the ANOM decision chart. The interpretation of Table 4 is the same as that for Figure 2.

Table 3: ANOM Properties Table

Simultaneous ANOM Test for Equality of k = 6 means Exact P. R. Nelson Critical Values for alpha 05	
degrees of freedom	ANOM critical value
24	2.83

Table 4: ANOM Decision Table

ANOM Decision Table for alpha 05 class variable tool (k = 6) variable diameter (n = 5) *** indicates mean plots outside decision lines					
TOOL	lower decision limit	diameter mean	grand mean	upper decision limit	reject
1	2.66991	16.4	28.6	54.5301	
2	2.66991	53.0	28.6	54.5301	
3	2.66991	42.4	28.6	54.5301	
4	2.66991	12.2	28.6	54.5301	
5	2.66991	6.4	28.6	54.5301	
6	2.66991	41.2	28.6	54.5301	

%ANOMV SAS SOURCE CODE

```

/*          This program performs
            ANOMV and ANOM
            ANOMV critical value files needed:
                low01, low05, low10
                up01, up05, up10 larges
            ANOM critical value files needed:
                h01 h05 h10
*/

goptions;
/*****
            INPUT DATA
            *****/
data example1;
    input tool @;
    do i=1 to 5;
        input diameter @;
        output;
    end;
    drop i;
    cards;
1 10 20 25 15 12
2 7 45 111 23 79
3 78 44 55 19 16
4 4 13 19 3 22
5 2 6 4 8 12
6 55 70 35 17 29
;
proc print;title 'data set';
run;

/*****
            DEFINE MACRO
            *****/

%macro anomv(
    k=,          /* the number of populations */

```

```

    n=,          /* the sample size per
population*/
    alpha=,      /* level of significance:
01,05,10 */
    ds=,         /* the data set with
observations */
    var=,        /* the variable name */
    classvar=,   /* variable name
populations */
    tops=);      /* 1 = anomv only, 2 = both
*/

/*****
            READ CRITICAL VALUES
            *****/
/

%if &n < 36 %then
    %do;
data cvlow;
    /* READ ANOMV CRITICAL VALUES */
    infile "c:\critvals\lower&alpha"
    dlm=',';

    do nuval=3 to 34;
    do kval =3 to 12;
        input lowcr @;
        if nuval=&n-1 and kval=&k then
output;

        end;
    end;
data cvup;
    infile "c:\critvals\upper&alpha" dlm
    ='',';

    do nuval=3 to 34;
    do kval =3 to 12;
        input upcr @;
        if nuval=&n-1 and kval=&k then
output;

        end;
    end;
data critvals;
    merge cvlow cvup; by nuval kval;
%end;

%if &n > 35 %then
    %do;
data critvals;
    infile "c:\critvals\larges"
    dlm=',';
    do jstep = 1, 2, 3;
        if jstep = 1 then alp =
'10';
        if jstep = 2 then alp =
'05';
        if jstep = 3 then alp =
'01';
        do kval = 3 to 12;
            input hls @;
            if kval = &k and alp =
&alpha then output;
        end;
    end;
    %end;
/* READ ANOM CRITICAL VALUES */
data hvals;
%if &tops > 1 %then
    %do;
    infile "c:\critvals\h&alpha" ;
    nu2 = &k*(&n-1);

```

```

do nu2val=1 to 20,24,30,40,60,120;
do kval =2 to 12;

    input h @;
    if nu2val < 20 then
    if nu2val=nu2 and kval=&k then output;
    if nu2val = 20 then
    if nu2 < 24 and nu2 > 19 and kval=&k
then output;
    if nu2val = 24 then
    if nu2 < 30 and nu2 > 23 and kval=&k
then output;
    if nu2val = 30 then
    if nu2 < 40 and nu2 > 29 and kval=&k
then output;
    if nu2val = 40 then
    if nu2 < 60 and nu2 > 39 and kval=&k
then output;
    if nu2val = 60 then
    if nu2 < 120 and nu2 > 59 and kval=&k
then output;
    if nu2val = 120 then
    if nu2 > 119 and kval=&k then output;

    end;
end;
%end;
%else
%do;
nuval=&n-1; kval=&k; h=1;output;
%end;
data critvals;
merge critvals hvals;

/*****
COMPUTE DECISION LIMITS
*****/

proc means data = &ds noprint;
by &classvar;var &var; output out =
stats1 std=stdx mean=mean;
proc print data=stats1; title 'stats1';
data stats2;
set stats1;
vars = stdx*stdx;
proc print data=stats2; title 'stats2';
proc means data=stats2 noprint;
var vars ; output out=stats2a mean=avgvar
;
proc print data=stats2a; title 'stats2a';
proc means data=stats2 noprint;
var mean ; output out=stats2b mean=gmean
;
proc print data=stats2b; title 'stats2b';
data stats3;
merge stats2a stats2b;
proc print data=stats3; title 'stats3';
data stats3a;
merge stats3 critvals;

do &classvar=1 to &k by .1;
%if &n < 36 %then
%do;
udl = sqrt(upcr*&k*avgvar);
cl = sqrt(avgvar);
ldl = sqrt(lowcr*&k*avgvar);
%end;
%else
%do;
udl = sqrt(avgvar+hls*avgvar*sqrt((2*&k-
1))/(&k*(&n-1))));
cl = sqrt(avgvar);
ldl = sqrt(avgvar-hls*avgvar*sqrt((2*&k-

```

```

1))/(&k*(&n-1))));
%end;
sediff = sqrt(avgvar)*sqrt((&k-
1))/(&n*&k));
udlx =
gmean+h*sqrt(avgvar)*sqrt((&k-1))/(&n*&k));
clx = gmean;
ldlx = gmean-
h*sqrt(avgvar)*sqrt((&k-1))/(&n*&k));
output;
end;
proc print data=stats3a;title 'stats3a';
data stats4;
merge stats1 stats3a; by
&classvar;
label stdx='std dev';
pop=&classvar;

proc print data=stats4; title 'stats4';

/*****
OUTPUT ANOMV DECISION CHART
*****/
proc gplot data=stats4 ;

plot stdx*&classvar=4
ldl*&classvar=1
cl*&classvar=2
udl*&classvar=3
/overlay
haxis=axis2
/* annotate=bars */
legend;
symbol1 c=BLUE,i=join, l=14,
v=none;
symbol2 c=BLUE, i=join, l=1,
v=none;
symbol3 c=BLUE, i=join, l=2
v=none;
symbol4 c=BLACK, i=none, v=star;
axis2 order=(1 to &k by 1)
offset=(2) label=(h=1.5);
title1 "ANOMV Decision Chart for
&var";
title2 'Standard Deviation
Plotted';

/*****
CREATE FILES FOR OUTPUT
*****/

data stats3b;
%if &n < 36 %then
%do;
merge stats3 critvals;
%end;
%else
%do;
merge stats3 critvals;
%end;
do &classvar=1 to &k ;
%if &n < 36 %then
%do;
udl = sqrt(upcr*&k*avgvar);
cl = sqrt(avgvar);
ldl = sqrt(lowcr*&k*avgvar);
%end;
%else
%do;
udl =
sqrt(avgvar+hls*avgvar*sqrt((2*&k-
1))/(&k*(&n-1))));
cl = sqrt(avgvar);

```

```

        ldl = sqrt(avgvar-hls*avgvar*sqrt((2*(&k-
1))/(&k*(&n-1))));
        %end;

        sediff = sqrt(avgvar)*sqrt((&k-
1))/(&n*&k));
        udlx = gmean+h*sqrt(avgvar)*sqrt((&k-
1))/(&n*&k));
        clx = gmean;
        ldlx = gmean-h*sqrt(avgvar)*sqrt((&k-
1))/(&n*&k));
        output;
    end;
data stats4a;
    merge stats1 stats3b; by &classvar;
    if (stdx > udl or stdx < ldl) then
reject1 = "****";
    else reject1 = ' ';
    if (mean > udlx or mean < ldlx) then
reject2 = "****";
    else reject2 = ' ';
proc print data=stats4a; title 'stats4a';

/*****
PRINT ANOMV PROPERTIES TABLE
*****/

%if &n < 36 %then
    %do;
proc print data=critvals label;

    title1 "Simultaneous ANOMV Test for
Equality of k = &k Variances";
    title2 "Conservative Wludyka and Nelson
Critical Values for alpha &alpha";
    id nuval;
    var lower upcr;
    label lower = 'lower critical value'
        upcr = 'upper critical value'
        nuval = 'degrees of freedom';
    %end;
    %else
    %do;
data stats2c;
    merge stats2a critvals;
    sigma1 = avgvar*sqrt((2*(&k-1))/(&k*(&n-
1))));

proc print data=stats2c label ;

    title1 "Large Sample Approximate ANOMV
Test for Equality of k = &k Variances";
    title2 "ANOM Critical Value has infinite
degrees of freedom and alpha &alpha";
    title3 "Class variable is &classvar and
variable is &var";
    id avgvar; var sigma1 hls;
    label
    avgvar = 'average of variances'
    hls = 'ANOM critical value'
    sigma1 = 'standard error';
    %end;

/*****
PRINT ANOMV DECISION TABLE
*****/
proc print data=stats4a label;
    title1 "ANOMV Decision Table for alpha
&alpha ";
    title2 "class variable &classvar (k =
&k) variable &var (n = &n) ";
    title3 ' *** indicates standard deviation

```

```

plots outside decision lines';
    id &classvar;
    var ldl stdx udl reject1;

    label
    stdx="&var standard deviation"
    ldl='lower decision limit'
    udl='upper decision limit'
    reject1 = 'reject';
%if &tops>=2 %then
    %do;

/*****
OUTPUT ANOM DECISION CHART
*****/
proc gplot data=stats4 ;

    plot mean*&classvar=4
    ldlx*&classvar=1 clx*&classvar=2
    udlx*&classvar=3 /overlay
    haxis=axis2
    legend;
    axis2 order=(1 to &k by 1)
offset=(2) label=(h=1.5);
    title1 "ANOM Decision Chart for
&var";
    title2 'Sample Means Plotted ';

/*****
PRINT ANOM PROPERTIES TABLE
*****/
proc print data=critvals label;

    title1 "Simultaneous ANOM Test for
Equality of k = &k means";
    title2 "Exact P. R. Nelson
Critical Values for alpha &alpha";
    id nu2val;
    var h ;
    label h = 'ANOM critical value'
        nu2val = 'degrees of
freedom';

/*****
PRINT ANOM DECISION TABLE
*****/
proc print data=stats4a label;
    title1 "ANOM Decision Table for
alpha &alpha ";
    title2 "class variable &classvar
(k = &k) variable &var (n = &n) ";
    title3 ' *** indicates mean plots
outside decision lines';
    id &classvar;
    var ldlx mean gmean udlx reject2;

    label
    mean="&var mean"
    ldlx='lower decision limit'
    udlx='upper decision limit'
    gmean = 'grand mean'
    reject2 = 'reject';
    %end;

run;
%mend anomv;
%anomv(k=6,n=5,alpha=05,ds=example1,
var=diameter,classvar=tool,tops=2);

```

ROBUST ANOMV

The ANOM test is (similarly to ANOVA, which has the same statistical assumptions) somewhat robust to non-normality. The ANOMV test is not. When non-normality is suspected, one easy to perform ANOM-type variance test that has been shown to be robust is ANOMV-LEV (see Bernard, (1999) for an example of this test and Monte Carlo results justifying the moniker "robust"). ANOMV-LEV is an ANOM version of Levene's test. To perform ANOMV-LEV replace each the original observations with the Absolute Deviations from the Median (ADM), where the median is the median of the sample for each group (population). Then apply ANOM (using the ANOMV macro) to the data set consisting of the ADM's. (Note that in the case where the sample size is odd, discard the zero ADM for each group and reduce the sample size by one (to $n-1$)). The resulting ANOM decision chart can be used to compare the average absolute deviations from the median for the K populations.

FORMULAS FOR ANOMV AND ANOM

Derivations for the ANOMV decision line formulas can be found in Wludyka and Nelson (1997). A good and easy to understand discussion of ANOM can be found in Ramig (1983). The critical values for ANOM can be found in Nelson (1993), which also has a useful discussion of some interesting applications of ANOM

ANOMV decision lines for a balanced design are

$$\begin{aligned} UDL &= U_{\alpha,k,v} k \bar{S}^2 \\ CL &= \bar{S}^2 \\ LDL &= L_{\alpha,k,v} k \bar{S}^2 \end{aligned} \quad (2)$$

where

$$\bar{S}^2 = \left(\sum_{i=1}^k S_i^2 \right) / k$$

is the average of the k sample variances. The upper decision limit critical value $U_{\alpha,k,v}$ and the lower decision limit critical value $L_{\alpha,k,v}$ can be found in tables provided by Wludyka and Nelson (1997). These values are functions of the level of significance α , the number of populations being compared K , and the degrees of freedom $v = n-1$. Recall that n is the sample size. Decision lines for the standard deviation are found by taking the square root of the variance decision lines (2).

LARGE SAMPLE ANOMV

For sample sizes greater than 35 the ANOM method can be used to produce approximate ANOMV decision lines.

$$\begin{aligned} UDL &= \bar{S}^2 + h_{\alpha,k,\infty} \hat{\sigma} \\ CL &= \bar{S}^2 \\ LDL &= \bar{S}^2 - h_{\alpha,k,\infty} \hat{\sigma} \end{aligned}$$

where $h_{\alpha,k,\infty}$ is an ANOM critical value (which can be found in P. R. Nelson's (1993) tables) with infinite degrees of freedom and

$$\hat{\sigma} = \bar{S}^2 \sqrt{2(k-1)/k(n-1)}$$

is an unbiased estimate for the standard error (the standard deviation of $S_i^2 - \bar{S}^2$).

DOWN LOADING FILES AND PROGRAMS

All critical value files and SAS source code programs for ANOMV can be down loaded from the University of North Florida Center for Research and Consulting in Statistics (CRCS) web page. These objects are in technical report #090199 entitled "ANOMV and ANOM Using SAS". Complete instructions for down loading are at the web site. Similarly, instructions for installing the critical values files are in the technical report. The address for the CRCS web page is

www.unf.edu/coas/math-stat/CRCS

CONCLUSION

An explanation of ANOMV, a test for comparing the variances of K populations based on independent samples of size n from normal populations, and SAS source code for a Macro to perform the test, have been presented.

REFERENCES

- Bernard, Anthony (1999), *Robust I-Sample Analysis of Means Type Randomization Tests for Variances.*, Masters Thesis, University of North Florida.
- Nelson, Peter R. (1993), "Additional Uses for the Analysis of Means and Extended Tables of Critical Values," *Technometrics*, 35, p61-71.
- Ramig, Pauline (1983), "Applications of the Analysis of Means," *Journal of Quality Technology*, 15, p19-25.
- Wludyka, Peter S., and Nelson, Peter R. (1997), "An Analysis-of-Means-Type Test for Variances from Normal Populations", *Technometrics*, 39, p274-285.

CONTACT INFORMATION

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

Peter Wludyka
University of North Florida
Center for Research and Consulting in Statistics
4567 St. Johns Bluff Road, South
Jacksonville, FL 32224-2645
Work Phone: (904)-620-1048
Fax: (904)-620-2818
Email: pwludyka@unf.edu