TESTING HOMOGENEITY OF VARIANCE

Shuqiang Zhang, College of Education
University of Hawaii at Manoa, Honolulu, HI

ABSTRACT

Homogeneity of variance is a major assumption underlying the validity of many parametric tests such as the t-test and ANOVA. It also represents the null hypothesis in substantive studies that focus on cross- or within-group fluctuation in variability. This paper explains how 12 homogeneity of variance tests, many of which have no corresponding SAS® software options, can be conducted using SAS/STAT®.

INTRODUCTION

The statistical validity of many parametric tests such as the t-test and ANOVA depends on the extent to which the data conform to the assumption of homogeneity of variance. When the design involves groups that have very different variances, the p value accompanying the test statistic such as t or F may be too lenient or too harsh. Furthermore, substantive research often requires investigation of cross- or within-group fluctuation in variability. For example, in quality control research, it is often important to identify the treatment that outperforms other treatments in terms of consistency in product quality. In human performance studies, an increase or decrease in the dispersion of performance scores within the same group of subjects may shed light on how changing conditions affect human behavior.

Despite an acknowledged need for testing homogeneity of variance, such tests are often difficult to locate or missing from software packages. This paper explains how 12 homogeneity of variance tests may be performed for five types of research designs through SAS/STAT.

ONE-SAMPLE TEST OF HOMOGENEITY OF VARIANCE

A convenient chi-square test can determine whether the difference between a sample variance ($S^2$) and a known or posited population variance ($\sigma_0^2$) is large enough to reject the null hypothesis, $H_0: \sigma_1^2 = \sigma_0^2$. SAS/STAT does not have a special procedure or option for the test. However, once $S^2$ is obtained through PROC MEANS or PROC UNIVARIATE, the test can be done with minimal computation by hand.

$$X^2 = (n-1) S^2 / \sigma_0^2$$

where $n = $ sample size

$S^2 = $ sample variance

$\sigma_0^2 = $ known or posited population variance

The $X^2$ test has (n-1) degrees of freedom. The critical value for a chosen significance level can be found in the $X^2$ table available in most introductory statistics textbooks.

TWO-SAMPLE TEST OF HOMOGENEITY OF VARIANCE

This test, known as the folded form F-test, is automatically conducted whenever PROC TTEST is invoked. Until SAS 6.11®, it was the only readily accessible homogeneity of variance test in SAS/STAT. The folded form F-test uses the ratio of the larger sample variance to the smaller sample variance to test the null hypothesis, $H_0: \sigma_1^2 = \sigma_2^2$.

$$F' = S_1^2 / S_2^2$$

where $S_1^2 = $ the larger variance

$S_2^2 = $ the smaller variance

The test has $(n_1-1)$ and $(n_2-2)$ degrees of freedom for the numerator and denominator respectively. Because the larger variance is taken to be the numerator, $F'$ is always larger than unity. In other words, only one direction of the F distribution is considered. SAS/STAT adjusts for the directional tail and prints out the correct probability. Should anyone try to determine whether a significant difference exists by referring to the conventional F table, he or she needs to remember that the listed critical value at a significance level of 0.05 actually means a significance level of 0.10 in the case of the folded form F-test (Ferguson, 1981, 189-192).
HOMOGENEITY OF VARIANCE TEST FOR K>2 SAMPLES

Several methods can test the null, \( H_0: \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 \). Hartley's \( F_{\max} \) test of homogeneity of \( k \) variances is appropriate for the balanced one-way ANOVA. First, PROC MEANS is used to identify which of the \( k \) groups has the largest variance and which has the smallest. Then the ratio of the maximum variance to the minimum variance can be computed either by hand or through PROC TTEST. However, the purpose of the t-test is strictly for the folded form F-test explained above.

\[
F^* = \frac{S_{\text{max}}^2}{S_{\text{min}}^2}
\]

where \( S_{\text{max}}^2 \) = the maximum variance
\( S_{\text{min}}^2 \) = the minimum variance

The printed \( p \) of the folded form F-test should be ignored because it is based upon the assumption of two samples, not \( k \) samples. Instead, refer to the special table of critical values for Hartley's test (Kanji 1993, 182; Rosenthal & Rosnow, 1991, 608-609). The table specifies critical values for various \( k \) values and degrees of freedom (n-1). \( F^* \) exceeding the tabled critical value indicates rejection of the null. If the groups have unequal but very similar numbers of observations, the method can be applied with the harmonic mean serving as its adjusted degrees of freedom (Rosenthal & Rosnow, 1991, 338-339).

\[
\text{adjusted df} = k / \left( \frac{1}{n_1} + \frac{1}{n_2} + \ldots + \frac{1}{n_k} \right)
\]

Like Hartley's \( F_{\max} \) test, Cochran's \( G \) test is appropriate for the balanced ANOVA. And like Hartley's test, it is not directly available from SAS/STAT but can be performed with information teased out of PROC ANOVA or PROC GLM. PROC ANOVA or PROC GLM calculates the pooled mean square, also known as MS error or MS within. The MEANS statement under PROC ANOVA or PROC GLM automatically prints out the standard deviation of each group, therefore making it easy to identify the group with the maximum variance.

\[
G = \frac{S_{\text{max}}^2}{k} \text{ (MS error)}
\]

where \( S_{\text{max}}^2 \) = the maximum variance
\( k \) = number of groups
MS error = pooled mean square within

A special table of critical \( G \) values with respect to various combinations of \( k \) and (n-1) is needed to determine whether the null hypothesis can be rejected (Winer 1971, 876; Rosenthal & Rosnow, 1991, 610-611). The harmonic mean is used as the adjusted degrees of freedom if the groups have roughly the same number of observations.

When the groups have very different sample sizes, Hartley's and Cochran's tests are inappropriate. Levene's Test, which is essentially one-way ANOVA and therefore can be conducted through the regular PROC ANOVA or PROC GLM, is recommended. For each group, the mean is computed. Then the absolute difference between each individual score and the mean of the group to which the score belongs is calculated. Since the variance of each group is related to the sum of the absolute deviations from the group mean, testing the differences among the group means of the absolute deviations is tantamount to testing homogeneity of variance. SAS/STAT users who have access to SAS 6.11 or later releases can perform Levene's test through PROC ANOVA or PROC GLM:

\[
\text{PROC GLM;}
\text{CLASS GROUP;}
\text{MODEL SCORE = GROUP;}
\text{MEANS GROUP / HOVTEST = LEVENE (TYPE = ABS);}
\text{RUN;}
\]

The \( (\text{TYPE = ABS}) \) specification is recommended for highly skewed data. If no such lop-sidedness is observed, \( (\text{TYPE = SQUARE}) \) may be used. In the latter case, the \( F \) test is conducted on the squares of the deviations. By default, Levene's test uses squared deviations and therefore is more sensitive to extreme scores or outliers. Those who do not yet have access to SAS 6.11 can run the test by writing a short ANOVA program using the absolute deviations or squared deviations as the dependent variable. For instructions on how to use group means to calculate absolute deviations or squared deviations, see Cody and Smith (1997, 384-385).

An extension to Levene's test is the Brown-Forsythe test, which follows the same logic as Levene's test except that it adopts absolute deviations from the group median rather than the group mean. Research findings, although limited, seem to suggest that the Brown-Forsythe test is the best procedure to provide "power to detect variance differences while protecting the Type I error probability." (SAS Institute, 1997, 356). To call up this option, specify \( \text{HOVTEST = BF} \). Levene's and
the Brown-Forsythe tests are recommended because (a) they are conceptually straightforward even to those who have only the barest understanding of one-way ANOVA; (b) they are applicable to either balanced or unbalanced designs; (c) the two methods can be conducted, with minimal programming, through any software that runs one-way ANOVA; and (d) they are at least as robust as, if not more so than, their more complex alternatives.

Other alternatives available from PROC ANOVA or PROC GLM are Bartlett's \( X^2 \) test, the ANOVA-based O'Brien test and another ANOVA-based Welch's variance-weighted test. Their computational formulas are available from Latour (1992). It is not clear yet if those more complex procedures have definitive advantages over the conceptually straightforward Levene's test or the Brown-Forsythe test in real-life research situations. Some researchers warn against the use of Bartlett's test on the grounds that "it is likely to yield more significant results than it should" (Rosenthal & Rosnow, 1991, 339).

HOMOGENEITY OF VARIANCE TEST FOR TWO CORRELATED SAMPLES

Correlated samples are typically involved in pre-post designs or studies that match the two subjects in each pair on certain variables to control for confounding. Variance tests for such designs need take into consideration the correlation between the scores under the two conditions. It should be pointed out that the correlation has its own substantive meaning. A positive correlation plus an increase in variability indicates greater dispersion of prior differences. A positive correlation plus a decrease in variability means the narrowing of prior differences. However, when a negative correlation occurs, the researcher may have to reconsider the research question and search for other factors that may account for the reversal of the directions of individual differences. Zero correlation between two correlated samples is extremely rare. Should it occur, the matching process obviously has failed its purpose. The two groups might as well be treated as independent samples. The discussion below proceeds on the assumption that the correlation is positive. Other factors being equal, the higher the correlation, the more likely the null hypothesis will be rejected.

The \( t \)-test for the difference between the variances of two correlated samples is not available from SAS/STAT. Fortunately, it is simple enough for hand calculation. The test statistic follows the \( t \)-distribution with \( df = n-2 \).

\[
t = \frac{(S_1^2 - S_2^2)}{\left[\frac{(1-r^2)S_1^2S_2^2}{(n-2)}\right]^{1/2}}
\]

where \( S_1^2 \) = variance under one condition
\( S_2^2 \) = variance under the other condition
\( r \) = correlation between the conditions
\( n \) = sample size

A less known alternative to the above \( t \)-test is the \( F_r \) test, which follows the sampling distribution of the Pearson correlation coefficient \( r \) with \( df = n-2 \) (Kanji, 1993, 38). First, the ratio of the larger variance to the smaller variance is calculated (\( F' \)). Then \( F_r \) is computed using the following formula:

\[
F_r = \frac{(F' - 1)}{\left[(F' + 1)^2 - 4F'\right]^{1/2}}
\]

where \( F' \) = larger variance / smaller variance
\( r \) = correlation coefficient

Critical values for various degrees of freedom at the 0.05 or 0.01 level of significance are available from the \( r \) table in most introductory statistics textbooks.

HOMOGENEITY OF VARIANCE TEST FOR K>2 CORRELATED SAMPLES

Very little is known about the testing of the homogeneity of variance for the one-way repeated measures design, or complex ANOVA designs in general. One conceptually appealing suggestion is from Rosenthal and Rosnow (1991, 340), which extends Levene's test to repeated measures. For that reason, it may be referred to as the extended Levene's test for distinction. For each condition, individual scores are converted into absolute deviations from the mean under the condition. Then proceed with the repeated measures ANOVA using PROC ANOVA or PROC GLM. For instructions on how to run repeated measures ANOVA with or without the REPEATED statement, see Cody and Smith (1997, 182-189).

CONCLUSION

This paper has practically covered all the commonly needed homogeneity of variance tests. Twelve homogeneity of variance tests are specified for five types of designs. Of the 12 tests, six are directly available from SAS/STAT. Guidance is provided as
to how the remaining six tests may be conducted through regular SAS procedures such as TTEST, GLM or ANOVA or by simple hand calculation on the basis of easily obtainable descriptive statistics.

REFERENCES


SAS, SAS/STAT, and SAS 6.11 are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.