Chapter 402

Equality of Covariance

Introduction

Discriminant analysis, MANOVA, and other multivariate procedures assume that the individual group covariance matrices are equal (homogeneous across groups). This **NCSS** module lets you test this hypothesis using Box's M test, which was first presented by Box (1949). This module also performs Bartlett's univariate homogeneity of variance test for testing equality of variance among individual variables.

Box's M Test

The calculation of Box's M test proceeds as follows. Suppose you have *k* groups measured on each of *p* variables, with n_i observations per group. Represent the estimated within-group covariance as S_i (the divisor is $n_i - 1$). The calculations for Box's M and Bartlett's test are identical. Box's M is simply an extension of Bartlett's test to the multivariate case. To calculate Bartlett's test, set $p = 1$. The value of M is given by

$$M = (N - k)\log_e |S| - \sum_{i=1}^{k}(n_i - 1)\log_e |S_i|$$

where

$$N = \sum_{i=1}^{k} n_i$$

$$S = \frac{\sum_{i=1}^{k}(n_i - 1)S_i}{N - k}$$

We use the Chi-square and F-ratio to test the significance of the M value. These approximations are constructed as follows:

$$A_1 = \frac{2p^2 + 3p - 1}{6(p+1)(k-1)}\left[\sum_{i=1}^{k}\left(\frac{1}{n_i - 1}\right) - \frac{1}{N - k}\right]$$

$$v_1 = \frac{p(p+1)(k-1)}{2}$$

$$A_2 = \frac{(p-1)(p+2)}{6(k-1)}\left[\sum_{i=1}^{k}\left(\frac{1}{n_i - 1}\right)^2 - \frac{1}{(N-k)^2}\right]$$

If $A_2 - A_1^2 > 0$ then

$$v_2 = \frac{v_1 + 2}{A_2 - A_1^2}$$

$$b = \frac{v_1}{1 - A_1 - (v_1 / v_2)}$$

$$F_{v_1, v_2} = \frac{M}{b}$$

If $A_2 - A_1^2 < 0$ then

$$v_2 = \frac{v_1 + 2}{A_1^2 - A_2}$$

$$b = \frac{v_2}{1 - A_1 + (2 / v_2)}$$

$$F_{v_1, v_2} = \frac{v_2 M}{v_1 (b - M)}$$

$$\chi^2_{v_1} = M(1 - A_1)$$

Box's M test is very sensitive to non-normality, so that a significant value indicates either unequal covariance matrices or non-normality or both. Hence, it is important to establish multivariate normality before using Box's M test.

The Chi-square approximation should be used when all $n_i > 20$, $p < 6$, and $k < 6$. Otherwise, the F approximation is more accurate.

NCSS supplies both the multivariate Box's M test and the individual Bartlett's tests so that when Box's M test is significant, you can determine which variables contribute to the variance inequality.

Data Structure

The data given in the table below are the first eight rows (out of the 150 in the dataset) of the famous "iris data" published by Fisher (1936). These data are measurements in millimeters of sepal length, sepal width, petal length, and petal width of fifty plants for each of three varieties of iris: (1) Iris setosa, (2) Iris versicolor, and (3) Iris virginica.

We will test to see if the covariance matrices are equal across the three varieties of iris. Here Iris is the group variable while SepalLength, SepalWidth, PetalLength, and PetalWidth are the regular variables.

Fisher dataset (subset)

SepalLength	SepalWidth	PetalLength	PetalWidth	Iris
50	33	14	2	1
64	28	56	22	3
65	28	46	15	2
67	31	56	24	3
63	28	51	15	3
46	34	14	3	1
69	31	51	23	3
62	22	45	15	2

Missing Values

If missing values are found in any of the variables being used, the row is omitted.

Example 1 – Equality of Covariance Analysis

This section presents an example of how to run an analysis. The data used are shown in the table above and found in the Fisher dataset. In this example, we will test whether the covariance matrices of the four measurements (SepalLength, SepalWidth, PetalLength, and PetalWidth) are equal across the three iris varieties.

Setup

To run this example, complete the following steps:

1 Open the Fisher example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Fisher** and click **OK**.

2 Specify the Equality of Covariance procedure options

- Find and open the **Equality of Covariance** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings template. To load this template, click **Open Example Template** in the Help Center or File menu.

Option	Value
Variables Tab	
Y: Group Variable	**Iris**
X's: Independent Variables	**SepalLength-PetalWidth**
Reports Tab	
All Reports	**Checked** (Normally you would only view a few of these reports, but we are selecting them all so that we can document them.)
Report Options (*in the Toolbar*)	
Variable Labels	**Column Labels**
Data Labels	**Value Labels**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Group Means Report

Group Means

Variable	Iris Setosa	Versicolor	Virginica	Overall
Sepal Length	50.06	59.36	65.88	58.43333
Sepal Width	34.28	27.7	29.74	30.57333
Petal Length	14.62	42.6	55.52	37.58
Petal Width	2.46	13.26	20.26	11.99333
Count	50	50	50	150

This report shows the means of each of the variables across each of the groups. The last row shows the count (number of observations) in the group. Note that the column headings come from the use of value labels for the group variable.

Group Standard Deviations Report

Group Standard Deviations

Variable	Iris Setosa	Versicolor	Virginica	Overall
Sepal Length	3.524897	5.161712	6.358796	8.280662
Sepal Width	3.790644	3.137983	3.224966	4.358663
Petal Length	1.73664	4.69911	5.518947	17.65298
Petal Width	1.053856	1.977527	2.7465	7.622377
Count	50	50	50	150

This report shows the standard deviations of each of the variables across each of the groups. The last row shows the count or number of observations in the group.

Within Group Correlation\Covariance Matrices

Within-Group Correlation\Covariance For Iris = Total

Variable	Variable Sepal Length	Sepal Width	Petal Length	Petal Width
Sepal Length	26.50082	9.272109	16.75143	3.840136
Sepal Width	0.530236	11.53878	5.524354	3.27102
Petal Length	0.756164	0.377916	18.51878	4.266531
Petal Width	0.364506	0.470535	0.484459	4.188163

Within-Group Correlation\Covariance For Iris = Setosa

Variable	Variable Sepal Length	Sepal Width	Petal Length	Petal Width
Sepal Length	12.4249	9.921633	1.63551	1.033061
Sepal Width	0.742547	14.36898	1.169796	0.9297959
Petal Length	0.267176	0.177700	3.015918	0.6069388
Petal Width	0.278098	0.232752	0.331630	1.110612

Within-Group Correlation\Covariance For Iris = Versicolor

Variable	Variable Sepal Length	Sepal Width	Petal Length	Petal Width
Sepal Length	26.64326	8.518368	18.2898	5.577959
Sepal Width	0.525911	9.846939	8.265306	4.120408
Petal Length	0.754049	0.560522	22.08163	7.310204
Petal Width	0.546461	0.663999	0.786668	3.910612

Within-Group Correlation\Covariance For Iris = Virginica

Variable	Variable Sepal Length	Sepal Width	Petal Length	Petal Width
Sepal Length	40.43428	9.376327	30.32898	4.909388
Sepal Width	0.457228	10.40041	7.137959	4.762857
Petal Length	0.864225	0.401045	30.45877	4.882449
Petal Width	0.281108	0.537728	0.322108	7.543265

This report shows the within-group correlations in the lower-left portion of the matrix and the within-group covariances in the upper-right portion of the matrix. The within-group variances are displayed on the diagonal. The total within-group values are found by forming a weighted average of the group covariances, averaging across all groups.

The three individual-group reports show the correlations and covariances for each of the three iris varieties. These are the correlations and covariances that would be obtained if each group was analyzed separately. These are the group covariances that will be tested by Box's M test.

Bartlett-Box Homogeneity Tests

Bartlett-Box Homogeneity Tests

Variable	Bartlett Value	DF1	DF2	F Approx	F Prob	Chi2 Approx	Chi2 Prob
Sepal Length	16.1509	2	48620	8.01	0.000334	16.00	0.000335
Sepal Width	2.1100	2	48620	1.05	0.351514	2.09	0.351533
Petal Length	55.9252	2	48620	27.74	0.000000	55.42	0.000000
Petal Width	39.5688	2	48620	19.62	0.000000	39.21	0.000000
Box's M	146.6632	20	77567	7.05	0.000000	140.94	0.000000

This report gives Bartlett's test for each variable followed by the Box's M test for all variables together. These tests are used to determine whether the variances of each of the groups are close enough to each other so that they may be considered equal. For example, the first line of the report tests for equal group variances of sepal length (SepalLength). Since the probability levels are small (less than 0.01), we would assume that the variances are significantly different. As was mentioned earlier, this test is also sensitive to departures from normality, so a significant result should be interpreted to mean that the variances are different or the data is non-normal. You can run a normality test to check this assumption.

Notice that the probability levels of SepalWidth are 0.35151 and 0.35153. Hence, both tests indicate that the variances are essentially equal. This is the only variable that did not fail this test!

We should also make a point regarding sample size here. The size of the probability level is directly related to the size of the sample. This probability level is for statistical significance, which may or may not be related to practical significance. You will have to consider this by comparing the individual standard deviations from one of the prior reports.

Matrix Determinant Report

Matrix Determinant Section

Iris	Log of Covariance Determinant	Correlation Determinant
Setosa	5.353320	0.353359
Versicolor	7.546356	0.083594
Virginica	9.493622	0.137390
Pooled (Overall)	8.462142	0.199529

This report gives the logarithm (base e) of the determinant of each of the relevant covariance matrices and the determinant of each of the correlation matrices. This report is useful since Box's M test compares these values.

Eigenvalues of Covariance Matrices Report

Eigenvalues of Covariance Matrices

Number	Iris Setosa	Versicolor	Virginica	Overall
1	23.645569	48.787394	69.525484	44.356592
2	3.691873	7.238410	10.655123	8.618331
3	2.679640	5.477609	5.229543	5.535235
4	0.903326	0.979036	3.426585	2.236372

This report gives the eigenvalues of each of the individual covariance matrices followed by the eigenvalues of the within-group covariance matrix. Each column gives a set of eigenvalues.

This report is useful because the eigenvalues summarize the covariance matrix into a few values. By comparing the largest eigenvalues across all groups, you can determine which groups are different. Also, eigenvalues near zero indicate singularities in your data.

Eigenvalues of Correlation Matrices Report

Eigenvalues of Correlation Matrices Section

Number	Iris Setosa	Versicolor	Virginica	Overall
1	2.05854	2.926341	2.454737	2.503762
2	1.022178	0.5462747	0.9647126	0.7251373
3	0.6678202	0.3949976	0.4522719	0.5824012
4	0.2514613	0.1323871	0.1282783	0.1886997

This report gives the eigenvalues of each of the individual correlation matrices followed by the eigenvalues of the within-group correlation matrix. Each column gives a set of eigenvalues.

This report is useful because the eigenvalues summarize the correlation matrix into a few values. By comparing the largest eigenvalues across all groups, you can determine which groups are different. Also, eigenvalues near zero indicate singularities in your data.