

**mvtest covariances** — Multivariate tests of covariances[Description](#)[Syntax](#)[Remarks and examples](#)[References](#)[Quick start](#)[Options for multiple-sample tests](#)[Stored results](#)[Also see](#)[Menu](#)[Options for one-sample tests](#)[Methods and formulas](#)

## Description

`mvtest covariances` performs one-sample and multiple-sample multivariate tests on covariances. These tests assume multivariate normality.

See [\[MV\] mvtest](#) for other multivariate tests. See [\[R\] sdtest](#) for univariate tests of standard deviations.

## Quick start

Test that the covariance matrix of `v1`, `v2`, `v3`, and `v4` is diagonal

```
mvtest covariances v1 v2 v3 v4
```

Test that the covariance matrix is spherical

```
mvtest covariances v1 v2 v3 v4, spherical
```

Test that the covariance matrix is compound symmetric

```
mvtest covariances v1 v2 v3 v4, compound
```

Test that the covariance matrix of the variables equals matrix `mymat`

```
mvtest covariances v1 v2 v3 v4, equals(mymat)
```

Test that the covariance matrix is block diagonal with `v1`, `v2`, and `v3` as block 1, `v4` as block 2, and `v5` and `v6` as block 3

```
mvtest cov v1 v2 v3 v4 v5 v6, block(v1 v2 v3 || v4 || v5 v6)
```

Box's  $M$  test that the covariance matrix of `v1`, `v2`, and `v3` is the same across the groups defined by `catvar`

```
mvtest cov v1 v2 v3, by(catvar)
```

## Menu

Statistics > Multivariate analysis > MANOVA, multivariate regression, and related > Multivariate test of means, covariances, and normality

## Syntax

### Multiple-sample tests

```
mvtest covariances varlist [if] [in] [weight], by(groupvars) [multisample_options]
```

### One-sample tests

```
mvtest covariances varlist [if] [in] [weight] [, one-sample_options]
```

<i>multisample_options</i>	Description
----------------------------	-------------

#### Model

* <i>by</i> ( <i>groupvars</i> ) <u>missing</u>	compare subsamples with same values in <i>groupvars</i> treat missing values in <i>groupvars</i> as ordinary values
--	--

\* *by*(*groupvars*) is required.

<i>one-sample_options</i>	Description
---------------------------	-------------

#### Options

<u>diagonal</u>	test that covariance matrix is diagonal; the default
<u>spherical</u>	test that covariance matrix is spherical
<u>compound</u>	test that covariance matrix is compound symmetric
<u>equals</u> ( <i>C</i> )	test that covariance matrix equals matrix <i>C</i>
* <u>block</u> ( <i>varlist</i> <sub>1</sub> [    ... ])	test that covariance matrix is block diagonal with blocks corresponding to <i>varlist</i> #

\* The full specification is block(*varlist*<sub>1</sub> [ || *varlist*<sub>2</sub> [ || ... ] ]).

*bootstrap*, *by*, *jackknife*, *rolling*, and *statsby* are allowed; see [U] 11.1.10 **Prefix commands**.

Weights are not allowed with the *bootstrap* prefix; see [R] **bootstrap**.

*aweight*s are not allowed with the *jackknife* prefix; see [R] **jackknife**.

*aweight*s and *fweight*s are allowed; see [U] 11.1.6 **weight**.

## Options for multiple-sample tests

### Model

*by*(*groupvars*) is required with the multiple-sample version of the test. Observations with the same values in *groupvars* form a sample. Observations with missing values in *groupvars* are ignored, unless the *missing* option is specified.

A modified likelihood-ratio statistic testing the equality of covariance matrices for the multiple independent samples defined by *by*() is presented along with an *F* and chi-squared approximation due to Box (1949). This test is also known as Box's *M* test.

*missing* specifies that missing values in *groupvars* are treated like ordinary values.

## Options for one-sample tests

### Options

`diagonal`, the default, tests the hypothesis that the covariance matrix is diagonal, that is, that the variables in *varlist* are independent. A likelihood-ratio test with first-order Bartlett correction is displayed.

`spherical` tests the hypothesis that the covariance matrix is diagonal with constant diagonal values, that is, that the variables in *varlist* are homoskedastic and independent. A likelihood-ratio test with first-order Bartlett correction is displayed.

`compound` tests the hypothesis that the covariance matrix is compound symmetric, that is, that the variables in *varlist* are homoskedastic and that every pair of two variables has the same covariance. A likelihood-ratio test with first-order Bartlett correction is displayed.

`equals(C)` specifies that the hypothesized covariance matrix for the  $k$  variables in *varlist* is  $C$ . The matrix  $C$  must be  $k \times k$ , symmetric, and positive definite. The row and column names of  $C$  are ignored. A likelihood-ratio test with first-order Bartlett correction is displayed.

`block(varlist1 [|| varlist2 [|| ...]])` tests the hypothesis that the covariance matrix is block diagonal with blocks *varlist*<sub>1</sub>, *varlist*<sub>2</sub>, etc. Variables in *varlist* not included in *varlist*<sub>1</sub>, *varlist*<sub>2</sub>, etc., are treated as an additional block. With this pattern, variables in different blocks are independent, but no assumptions are made on the within-block covariance structure. A likelihood-ratio test with first-order Bartlett correction is displayed.

## Remarks and examples

[stata.com](http://www.stata.com)

Remarks are presented under the following headings:

*One-sample tests for covariance matrices*  
*A multiple-sample test for covariance matrices*

### One-sample tests for covariance matrices

One-sample and multiple-sample tests for covariance matrices are provided by the `mvtest covariances` command. One-sample tests include the test that the covariance matrix of *varlist* is diagonal, spherical, compound symmetric, block diagonal, or equal to a given matrix.

#### ► Example 1

The gasoline-powered milk-truck dataset introduced in [example 1](#) of [\[MV\] mvtest means](#) has price per mile for fuel, repair, and capital. We test if the covariance matrix for these three variables has any special structure.

```
. use http://www.stata-press.com/data/r15/milktruck
(Milk transportation costs for 25 gasoline trucks (Johnson and Wichern 2007))
. mvtest covariances fuel repair capital, diagonal
Test that covariance matrix is diagonal
  Adjusted LR chi2(3) =    17.91
  Prob > chi2 =    0.0005
. mvtest covariances fuel repair capital, spherical
Test that covariance matrix is spherical
  Adjusted LR chi2(5) =    21.53
  Prob > chi2 =    0.0006
```

```
. mvtest covariances fuel repair capital, compound
Test that covariance matrix is compound symmetric
Adjusted LR chi2(4) =    11.29
Prob > chi2 =    0.0235
```

We reject the hypotheses that the covariance is diagonal, spherical, or compound symmetric.

We now test whether there is covariance between `fuel` and `repair`, with `capital` not covarying with these two variables. Thus we hypothesize a block diagonal structure of the form

$$\Sigma = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} & 0 \\ \sigma_{21} & \sigma_{22}^2 & 0 \\ 0 & 0 & \sigma_{33}^2 \end{pmatrix}$$

for the covariance matrix. The `block()` option of `mvtest covariances` provides the test:

```
. mvtest covariances fuel repair capital, block(fuel repair || capital)
Test that covariance matrix is block diagonal
Adjusted LR chi2(2) =    3.52
Prob > chi2 =    0.1722
```

We fail to reject the null hypothesis. The covariance matrix might have the block diagonal structure we hypothesized.

The same  $p$ -value could have been obtained from Stata's canonical correlation command:

```
. canon (fuel repair) (capital)
(output omitted)
```

See [MV] **canon**.

Now, in addition to hypothesizing that the covariance is block diagonal, we specifically hypothesize that the variance for `capital` is 10, the variance of `fuel` is three times that of `capital`, the variance of `repair` is two times that of `capital`, and that there is no covariance between `capital` and the other two variables, while there is a covariance of 15 between `fuel` and `repair`. We test that hypothesis by using the `equals()` option.

```
. mat B = (30, 15, 0 \ 15, 20, 0 \ 0, 0, 10)
. matrix list B
symmetric B[3,3]
   c1  c2  c3
r1  30
r2  15  20
r3   0   0  10
. mvtest covariances fuel repair capital, equals(B)
Test that covariance matrix equals matrix B
Adjusted LR chi2(6) =    5.48
Prob > chi2 =    0.4837
```

We fail to reject the null hypothesis; the covariance might follow the structure hypothesized.

## □ Technical note

If each block comprises a single variable, the test of independent subvectors reduces to a test that the covariance matrix is diagonal. Thus the following two commands are equivalent:

```
mvtest covariances x1 x2 x3 x4 x5, block(x1 || x2 || x3 || x4 || x5)
```

and

```
mvtest covariances x1 x2 x3 x4 x5, diagonal
```

□

## A multiple-sample test for covariance matrices

The `by()` option of `mvtest covariances` provides a modified likelihood-ratio statistic testing the equality of covariance matrices for the multiple independent samples defined by `by()`. This test is also known as Box's  $M$  test. There are both  $F$  and chi-squared approximations for the null distribution of the test.

### ▷ Example 2

We illustrate the multiple-sample test of equality of covariance matrices by using four psychological test scores on 32 men and 32 women (Rencher and Christensen 2012; Beall 1945).

```
. use http://www.stata-press.com/data/r15/genderpsych
(Four Psychological Test Scores, Rencher and Christensen (2012))
. mvtest covariances y1 y2 y3 y4, by(gender)
Test of equality of covariance matrices across 2 samples
    Modified LR chi2 =   14.5606
    Box F(10,18377.7) =     1.35      Prob > F =  0.1950
    Box chi2(10) =     13.55      Prob > chi2 =  0.1945
```

Both the  $F$  and the chi-squared approximations indicate that we cannot reject the null hypothesis that the covariance matrices for males and females are equal (Rencher and Christensen 2012, 269).

◀

Equality of group covariance matrices is an assumption of multivariate analysis of variance (see [MV] [manova](#)) and linear discriminant analysis (see [MV] [discrim lda](#)). Box's  $M$  test, produced by `mvtest covariances` with the `by()` option, is often recommended for testing this assumption.

## Stored results

`mvtest covariances` stores the following in `r()`:

### Scalars

<code>r(chi2)</code>	chi-squared
<code>r(df)</code>	degrees of freedom for chi-squared test
<code>r(p_chi2)</code>	significance
<code>r(F_Box)</code>	$F$ statistic for Box test ( <code>by()</code> only)
<code>r(df_m_Box)</code>	model degrees of freedom for Box test ( <code>by()</code> only)
<code>r(df_r_Box)</code>	residual degrees of freedom for Box test ( <code>by()</code> only)
<code>r(p_F_Box)</code>	significance of Box $F$ test ( <code>by()</code> only)

### Macros

<code>r(chi2type)</code>	type of model chi-squared test
--------------------------	--------------------------------

## Methods and formulas

When comparing the formulas in this section with those found in some multivariate texts, be aware of whether they define the sample covariance matrix with a divisor of  $N$  or  $N - 1$ . We use  $N$ . The formulas for several of the statistics are presented differently depending on your choice of divisor (but are still equivalent).

Methods and formulas are presented under the following headings:

- One-sample tests for covariance matrices*
- A multiple-sample test for covariance matrices*

### One-sample tests for covariance matrices

Let the sample consist of  $N$  i.i.d. observations,  $\mathbf{x}_i$ ,  $i = 1, \dots, N$ , from a  $k$ -variate multivariate normal distribution,  $MVN_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , with sample mean  $\bar{\mathbf{x}} = 1/N \sum_{i=1}^N \mathbf{x}_i$ , sample covariance matrix  $\mathbf{S} = 1/N \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$ , and sample correlation matrix  $\mathbf{R}$ .

To test that a covariance matrix equals a given matrix,  $H_0: \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ , `mvtest covariances` computes a likelihood-ratio test with Bartlett correction (Rencher and Christensen 2012, 260–261):

$$\chi_{\text{ovf}}^2 = (N - 1) \left\{ 1 - \frac{1}{6(N - 1) - 1} \left( 2k + 1 - \frac{2}{k + 1} \right) \right\} \\ \times \left\{ \ln |\boldsymbol{\Sigma}_0| - \ln \left| \frac{N}{N - 1} \mathbf{S} \right| + \text{trace} \left( \frac{N}{N - 1} \mathbf{S} \boldsymbol{\Sigma}_0^{-1} \right) - k \right\}$$

which is approximately  $\chi^2$  distributed with  $k(k + 1)/2$  degrees of freedom.

To test for a spherical covariance matrix,  $H_0: \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$ , `mvtest covariances` computes a likelihood-ratio test with Bartlett correction (Rencher and Christensen 2012, 261–262):

$$\chi_{\text{ovs}}^2 = \left\{ (N - 1) - \frac{2k^2 + k + 2}{6k} \right\} \left[ k \ln \{ \text{trace}(\mathbf{S}) \} - \ln |\mathbf{S}| - k \ln(k) \right]$$

which is approximately  $\chi^2$  distributed with  $k(k + 1)/2 - 1$  degrees of freedom.

To test for a diagonal covariance matrix,  $H_0: \boldsymbol{\Sigma}_{ij} = 0$  for  $i \neq j$ , `mvtest covariances` computes a likelihood-ratio test with first-order Bartlett correction (Rencher and Christensen 2012, 275):

$$\chi_{\text{ovd}}^2 = - \left( N - 1 - \frac{2k + 5}{6} \right) \ln |\mathbf{R}|$$

which is approximately  $\chi^2$  distributed with  $k(k - 1)/2$  degrees of freedom.

To test for a compound-symmetric covariance matrix,  $H_0: \boldsymbol{\Sigma} = \sigma^2 \{ (1 - \rho) \mathbf{I} + \rho \mathbf{1}\mathbf{1}' \}$ , that is, a covariance matrix with common variance  $\sigma^2$  and common correlation  $\rho$ , `mvtest covariances` computes a likelihood-ratio test with first-order Bartlett correction (Rencher and Christensen 2012, 263–264):

$$\chi_{\text{ovc}}^2 = \left\{ N - 1 - \frac{k(k + 1)^2(2k - 3)}{6(k - 1)(k^2 + k - 4)} \right\} \\ \times [k \ln(s^2) + (k - 1) \ln(1 - r) + \ln\{1 + (k - 1)r\} - \ln |\mathbf{S}|]$$

where

$$s^2 = \frac{1}{k} \sum_{j=1}^k s_{jj} \quad \text{and} \quad r = \frac{1}{k(k-1)s^2} \sum_{j=1}^k \sum_{h=1, h \neq j}^k s_{jh}$$

where  $s_{jh}$  is the  $(j, h)$  element of  $\mathbf{S}$ .  $\chi_{\text{ovc}}^2$  is approximately  $\chi^2$  distributed with  $k(k+1)/2 - 2$  degrees of freedom.

To test that a covariance matrix is block diagonal with  $b$  diagonal blocks and with  $k_j$  variables in block  $j$ , `mvtest covariances` computes a likelihood-ratio test with first-order Bartlett correction (Rencher and Christensen 2012, 271–272). Thus variables in different blocks are hypothesized to be independent.

$$\chi_{\text{ovb}}^2 = \left( N - 1 - \frac{2a_3 + 3a_2}{6a_2} \right) \left( \sum_{j=1}^b \ln |\mathbf{S}_j| - \ln |\mathbf{S}| \right)$$

where  $a_2 = k^2 - \sum_{j=1}^b k_j^2$ ,  $a_3 = k^3 - \sum_{j=1}^b k_j^3$ , and  $\mathbf{S}_j$  is the covariance matrix for the  $j$ th block.  $\chi_{\text{ovb}}^2$  is approximately  $\chi^2$  distributed with  $a_2/2$  degrees of freedom.

## A multiple-sample test for covariance matrices

Let there be  $m \geq 2$  independent samples with the  $j$ th sample containing  $N_j$  i.i.d. observations,  $\mathbf{x}_{ji}$ ,  $i = 1, \dots, N_j$ , from a  $k$ -variate multivariate normal distribution  $\text{MVN}_k(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ . The observed  $j$ th sample mean is  $\bar{\mathbf{x}}_j = 1/N_j \sum_{i=1}^{N_j} \mathbf{x}_{ji}$  and covariance is  $\mathbf{S}_j = 1/N_j \sum_{i=1}^{N_j} (\mathbf{x}_{ji} - \bar{\mathbf{x}}_j)(\mathbf{x}_{ji} - \bar{\mathbf{x}}_j)'$ . Let  $N = \sum_{j=1}^m N_j$ .

To test the equality of covariance matrices in  $m$  independent samples,  $H_0: \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \dots = \boldsymbol{\Sigma}_m$ , `mvtest covariances` computes a modified likelihood-ratio statistic, which is an unbiased variant of the likelihood-ratio statistic (Rencher and Christensen 2012, 266–268):

$$-2 \ln(M) = (N - m) \ln |\mathbf{S}_{\text{pooled}}| - \sum_{j=1}^m \left\{ (N_j - 1) \ln \left| \frac{N_j}{N_j - 1} \mathbf{S}_j \right| \right\}$$

where  $\mathbf{S}_{\text{pooled}} = \sum_{j=1}^m (N_j \mathbf{S}_j) / (N - m)$ . Asymptotically,  $-2 \ln(M)$  is  $\chi^2$  distributed. Box (1949, 1950) derived more accurate  $\chi^2$  and  $F$  approximations (Rencher and Christensen 2012, 267–268).

Box's  $\chi^2$  approximation is given by

$$\chi_{\text{mv}}^2 = -2(1 - c_1) \ln(M)$$

which is approximately  $\chi^2$  distributed with  $(m-1)k(k+1)/2$  degrees of freedom.

Box's  $F$  approximation is given by

$$F_{\text{mv}} = \begin{cases} -2b_1 \ln(M) & \text{if } c_2 > c_1^2 \\ \frac{2a_2 b_2 \ln(M)}{a_1 \{1 + 2b_2 \ln(M)\}} & \text{otherwise} \end{cases}$$

which is approximately  $F$  distributed with  $a_1$  and  $a_2$  degrees of freedom.

In the  $\chi^2$  and  $F$  approximations, we have

$$c_1 = \left\{ \sum_{j=1}^m (N_j - 1)^{-1} - (N - m)^{-1} \right\} \frac{2k^2 + 3k - 1}{6(k + 1)(m - 1)}$$

$$c_2 = \left\{ \sum_{j=1}^m (N_j - 1)^{-2} - (N - m)^{-2} \right\} \frac{(k - 1)(k + 2)}{6(m - 1)}$$

$a_1 = (m - 1)k(k + 1)/2$ ,  $a_2 = (a_1 + 2)/|c_2 - c_1^2|$ ,  $b_1 = (1 - c_1 - a_1/a_2)/a_1$ , and  $b_2 = (1 - c_1 + 2/a_2)/a_2$ .

## References

- Beall, G. 1945. Approximate methods in calculating discriminant functions. *Psychometrika* 10: 205–217.
- Box, G. E. P. 1949. A general distribution theory for a class of likelihood criteria. *Biometrika* 36: 317–346.
- . 1950. Problems in the analysis of growth and wear curves. *Biometrics* 6: 362–389.
- Johnson, R. A., and D. W. Wichern. 2007. *Applied Multivariate Statistical Analysis*. 6th ed. Englewood Cliffs, NJ: Prentice Hall.
- Rencher, A. C., and W. F. Christensen. 2012. *Methods of Multivariate Analysis*. 3rd ed. Hoboken, NJ: Wiley.

## Also see

- [MV] **candisc** — Canonical linear discriminant analysis
- [MV] **canon** — Canonical correlations
- [R] **correlate** — Correlations (covariances) of variables or coefficients
- [MV] **discrim lda** — Linear discriminant analysis
- [MV] **manova** — Multivariate analysis of variance and covariance
- [R] **sdtest** — Variance-comparison tests