mvtest correlations — Multivariate tests of correlations

Description

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

See [MV] **mvtest** for more multivariate tests.

Quick start

Lawley's test that the correlations of variables `v1`, `v2`, `v3`, and `v4` are equal

```
mvtest correlations v1 v2 v3 v4
```

Jennrich's test that the correlation matrix equals hypothesized matrix `mymat`

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

Jennrich's test for equal correlations with samples defined by `catvar`

```
mvtest correlations v1 v2 v3 v4 v5, by(catvar)
```

Menu

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

Syntax

Multiple-sample tests

mvtest correlations *varlist* [*if*] [*in*] [*weight*] , by(*groupvars*) [*multisample_options*]

One-sample tests

mvtest correlations *varlist* [*if*] [*in*] [*weight*] [, *one-sample_options*]

multisample_options	Description
Model	
* by(*groupvars*)	compare subsamples with same values in *groupvars*
missing	treat missing values in *groupvars* as ordinary values

*by(*groupvars*) is required.

one-sample_options	Description
Options	
compound	test that correlation matrix is compound symmetric (equal correlations); the default
equals(*C*)	test that correlation matrix equals matrix *C*

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**.
aweights are not allowed with the jackknife prefix; see [R] **jackknife**.
aweights and fweights 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 each sample. Observations with missing values in *groupvars* are ignored, unless the missing option is specified. A Wald test due to Jennrich (1970) is displayed.

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

Options for one-sample tests

Options

compound, the default, tests the hypothesis that the correlation matrix of the variables is compound symmetric, that is, that the correlations of all variables in *varlist* are the same. Lawley's (1963) χ^2 test is displayed.

equals(*C*) tests the hypothesis that the correlation matrix of *varlist* is *C*. The matrix *C* should be $k \times k$, symmetric, and positive definite. *C* is converted to a correlation matrix if needed. The row and column names of *C* are immaterial. A Wald test due to Jennrich (1970) is displayed.

Remarks and examples

Remarks are presented under the following headings:

One-sample tests for correlation matrices

Both one-sample and multiple-sample tests of correlation matrices are provided with the `mvtest correlations` command. The one-sample tests include Lawley's (1963) test that the correlation matrix is compound symmetric (that is, all correlations are equal), and the Wald test proposed by Jennrich (1970) that the correlation matrix equals a given correlation 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 correlations between these three variables are equal (that is, the correlation matrix is compound symmetric) using the `compound` option of `mvtest correlations`.

```
. use https://www.stata-press.com/data/r19/milktruck
(Milk transportation costs for 25 gasoline trucks (Johnson and Wichern 2007))
. mvtest correlations fuel repair capital, compound
Test that correlation matrix is compound symmetric (all correlations equal)
        Lawley chi2(2) =      7.75
           Prob > chi2 =    0.0208
```

We reject the null hypothesis and conclude that there are probably differences in the correlations of the three cost variables.

◁

▷ Example 2

Using the `equals()` option of `mvtest correlations`, we test the hypothesis that fuel and repair costs have a correlation of 0.75, while the correlation between capital and these two variables is zero.

```
. matrix C = (1, 0.75, 0 \ 0.75, 1, 0 \ 0, 0, 1)
. matrix list C
symmetric C[3,3]
     c1   c2   c3
r1    1
r2  .75    1
r3    0    0    1
. mvtest correlations fuel repair capital, equals(C)
Test that correlation matrix equals specified pattern C
      Jennrich chi2(3) =      4.55
           Prob > chi2 =    0.2077
```

We fail to reject this null hypothesis.

◁

A multiple-sample test for correlation matrices

A multiple-sample test of equality of correlation matrices is provided by the `mvtest correlations` command with the `by()` option defining the multiple samples (groups).

▷ Example 3

Psychological test score data are introduced in example 2 of [MV] **mvtest covariances**. We test whether the correlation matrices for the four test scores are the same for males and females.

```
. use https://www.stata-press.com/data/r19/genderpsych
(Four psychological test scores, Rencher and Christensen (2012))
. mvtest correlations y1 y2 y3 y4, by(gender)
Test of equality of correlation matrices across samples
      Jennrich chi2(6) =      5.01
           Prob > chi2 =    0.5422
```

We fail to reject the null hypothesis of equal correlation matrices for males and females.

◁

Stored results

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

Scalars

`r(chi2)`	χ^2 statistic
`r(df)`	degrees of freedom for χ^2 test
`r(p_chi2)`	p-value for χ^2 test

Macros

`r(chi2type)`	type of model χ^2 test

Methods and formulas

Methods and formulas are presented under the following headings:

One-sample tests for correlation matrices
A multiple-sample test for correlation matrices

One-sample tests for correlation matrices

Let the sample consist of N i.i.d. observations from a k-variate multivariate normal distribution $\text{MVN}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with sample correlation matrix $\mathbf{R}$.

To test that a correlation matrix equals a given matrix, $\mathbf{R}_0$, `mvtest correlations` computes a Wald test proposed by Jennrich (1970):

$$\chi^2_{\text{ocf}} = \frac{1}{2}\text{trace}(\mathbf{ZZ}) - \text{diagonal}(\mathbf{Z})' \left(\mathbf{I} + \mathbf{R}_0 \bullet \mathbf{R}_0^{-1}\right)^{-1} \text{diagonal}(\mathbf{Z})$$

where $\mathbf{Z} = \sqrt{N}\mathbf{R}_0^{-1}(\mathbf{R} - \mathbf{R}_0)$ and $\bullet$ denotes the Hadamard product. χ^2_{ocf} is asymptotically χ^2 distributed with $k(k-1)/2$ degrees of freedom.

To test that the correlation matrix is compound symmetric, that is, to test that all correlations are equal, the likelihood-ratio test is somewhat cumbersome. Lawley (1963) offers an asymptotically equivalent test that is computationally simple (Johnson and Wichern 2007, 457–458):

$$\chi^2_{\rm occ} = \frac{N-1}{(1-\overline{R})^2} \left\{ \sum_{i=2}^{k} \sum_{j=1}^{i-1} (R_{ij} - \overline{R})^2 - u \sum_{h=1}^{k} (\overline{R}_h - \overline{R})^2 \right\}$$

where

$$\overline{R} = \frac{2}{k(k-1)} \sum_{i=2}^{k} \sum_{j=1}^{i-1} R_{ij}$$

$$\overline{R}_h = \frac{1}{k-1} \sum_{i=1;i\neq h}^{k} R_{ih}$$

$$u = \frac{(k-1)^2 \left\{1-(1-\overline{R})^2\right\}}{k-(k-2)(1-\overline{R})^2}$$

and R_{ij} denotes element (i,j) of the $k \times k$ correlation matrix $\mathbf{R}$. $\chi^2_{\rm occ}$ is asymptotically χ^2 distributed with $(k-2)(k+1)/2$ degrees of freedom. Aitkin, Nelson, and Reinfurt (1968) study the quality of this χ^2 approximation for k up to six and various correlations, and conclude that the approximation is adequate for N as small as 25.

A multiple-sample test for correlation matrices

Let there be $m \geq 2$ independent samples with the jth sample containing N_j i.i.d. observations from a k-variate multivariate normal distribution, $\text{MVN}_k(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, with sample correlation matrix $\mathbf{R}_j$, $j=1,\dots,m$. Let $N = \sum_{j=1}^{m} N_j$.

To test for the equality of correlation matrices across m independent samples, `mvtest correlations` computes a Wald test proposed by Jennrich (1970):

$$\chi^2_{\rm mc} = \sum_{j=1}^{m} \left\{ \frac{1}{2}\text{trace}\left(\mathbf{Z}_j^2\right) - \text{diagonal}\left(\mathbf{Z}_j\right)' \left(\mathbf{I} + \overline{\mathbf{R}} \bullet \overline{\mathbf{R}}^{-1}\right)^{-1} \text{diagonal}\left(\mathbf{Z}_j\right) \right\}$$

where $\overline{\mathbf{R}} = 1/N \sum_{j=1}^{m} N_j \mathbf{R}_j$, $\mathbf{Z}_j = \sqrt{N_j}\, \overline{\mathbf{R}}^{-1} \left(\mathbf{R}_j - \overline{\mathbf{R}}\right)$, and $\bullet$ denotes the Hadamard product. $\chi^2_{\rm mc}$ is asymptotically χ^2 distributed with $(m-1)k(k-1)/2$ degrees of freedom.

References

Aitkin, M. A., W. C. Nelson, and K. H. Reinfurt. 1968. Tests for correlation matrices. *Biometrika* 55: 327–334. https://doi.org/10.1093/biomet/55.2.327.

Jennrich, R. I. 1970. An asymptotic χ^2 test for the equality of two correlation matrices. *Journal of the American Statistical Association* 65: 904–912. https://doi.org/10.1080/01621459.1970.10481133.

Johnson, R. A., and D. W. Wichern. 2007. *Applied Multivariate Statistical Analysis*. 6th ed. Englewood Cliffs, NJ: Prentice Hall.

Lawley, D. N. 1963. On testing a set of correlation coefficients for equality. *Annals of Mathematical Statistics* 34: 149–151. https://doi.org/10.1214/aoms/1177704249.

Rencher, A. C., and W. F. Christensen. 2012. *Methods of Multivariate Analysis*. 3rd ed. Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118391686.

Also see

[MV] **canon** — Canonical correlations

[R] **correlate** — Correlations of variables