Most multivariate techniques make no assumptions about the data and therefore do not calculate significance levels. Actually, where multivariate hypothesis tests are provided, they are based on definite assumptions about the data, generally assuming a multivariate normal population, and also often making additional assumptions about the nature of the covariance matrix.

For these reasons the SIMFIT procedure for exhaustive analysis of a normal multivariate matrix provides options that are useful before proceeding to more specific techniques that depend on multivariate normality, e.g., MANOVA and some types of ANOVA. These multivariate normal analysis procedures can be used by selecting [Statistics] from the main SIMFIT menu, then [Data exploration] followed by [Exhaustive analysis of a multivariate normal matrix].

Example 1: The sample means and covariance matrix

Choosing the option to display the means and covariance matrix leads to the following results with test file `hotel.tf1`.

Variable	Mean	Std.err.	lower95%cl	upper95%cl
1	-0.53000	0.46261	-0.55983	-0.50017
2	-0.03000	0.38559	-0.05486	-0.00514
3	-0.59000	0.49091	-0.62165	-0.55835
4	3.10000	1.94622	2.97451	3.22549

Covariance matrix

2.14011			
-0.11878	1.48678		
-0.89411	0.79144	2.40989	
3.59222	1.88111	-4.60111	37.8778

Correlation matrix

1			
-0.066588	1		
-0.393709	0.418118	1	
0.398982	0.250668	-0.481584	1

CV matrix eigenvalues

0.683568
0.821943
0.775086
0.231793

Determinant	9.81414E+01

The column means $\bar{x}_j$, and m by m sample covariance matrix S displayed here are defined for a n by m data matrix x_{ij} with $n \geq 2, m \geq 2$ as

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$$

$$s_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

Example 2: Graphical test for multivariate normality

A graphical technique is provided for investigating if a data matrix with n rows and m columns, where $n >> m > 1$, is consistent with a multivariate normal distribution. For example, the next figure

Multivariate Plot: r = 0.754   **Multivariate Plot: r = 0.980**
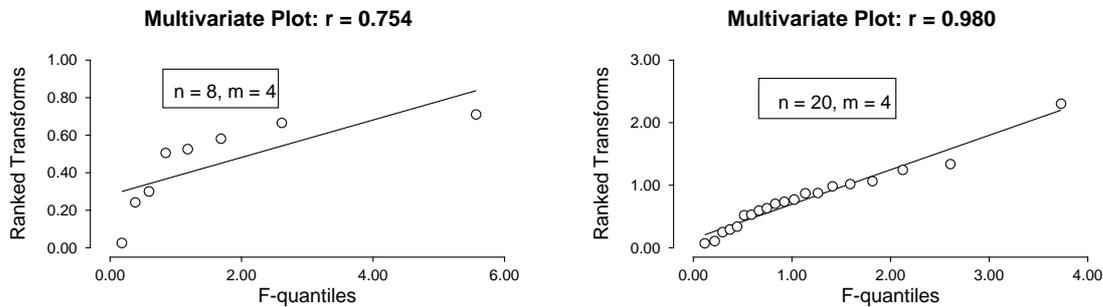
which shows plots for two random samples from a multivariate normal distribution. The plot uses the fact that, for a multivariate normal distribution with sample mean $\bar{x}$ and sample covariance matrix S,

$$(x - \bar{x})^T S^{-1} (x - \bar{x}) \sim \frac{m(n^2 - 1)}{n(n - m)} F_{m, n-m},$$

where x is a further independent observation from this population, so that the transforms plotted against the quantiles of an F distribution with m and $n - m$ degrees of freedom, i.e. according to the cumulative probabilities for $(i - 0.5)/n$ for $i = 1, 2, \ldots, n$ should be a straight line. It can be seen from the above figure that this plot is of little value for small values of n, say $n \approx 2m$ but becomes progressively more useful as the sample size increases, say $n > 5m$.

Example 3: Hotelling T^2 test H_0 : means = a supplied reference vector

It is possible to perform two variants of the Hotelling T^2 test, namely

- testing for equality of the mean vector with a specified reference vector of means, or

- testing for equality of all means without specifying a reference mean.

Dealing first with testing that a vector of sample means is consistent with a reference vector, consider the next table for a zero reference vector.

Hotelling one sample T^2 test

H_0: Column means = Expected values supplied (i.e. 0)

Number of rows = 10, Number of columns = 4
Hotelling T^2 = 7.43910
F Statistic (FTS) = 1.23985
Degrees of freedom ($d1, d2$) = 4, 6
$P(F(d1, d2) \geq FTS)$ = 0.386864

Column	Mean	Std.Err.	Expected	Delta	t	p
1	-0.53000	0.46261	0.0000	-0.53000	-1.14567	0.281481
2	-0.03000	0.38559	0.0000	-0.03000	-0.07780	0.939687
3	-0.59000	0.49091	0.0000	-0.59000	-1.20186	0.260086
4	3.10000	1.94622	0.0000	3.10000	1.59283	0.145662

This resulted when the test file `hotel.tf1` was analyzed using the Hotelling one sample test procedure. This tests the null hypothesis $H_0 : \mu = \mu_0$ against the alternative $H_1 : \mu \neq \mu_0$, where μ_0 is a known mean

vector and no assumptions are made about the covariance matrix Σ. Hotelling's T^2 is

$$T^2 = n(\bar{x} - \mu_0)^T S^{-1} (\bar{x} - \mu_0)$$

and, if H_0 is true, then an F test can be used since $(n - m)T^2/(m(n - 1))$ is distributed asymptotically as $F_{m,n-m}$. Users can input any reference mean vector μ_0 to test for equality of means but, when the data columns are all differences between two observations for the same subjects and the aim is to test for no significant differences, so that μ_0 is the zero vector, as with `hotel.tf1`, the test is a sort of higher dimensional analogue of the paired t test. The table also shows the results when t tests are applied to the individual columns of differences between the sample means $\bar{x}$ and the reference means μ_0, which is suspect because of multiple testing but, in this case, the conclusion is the same as the Hotelling T^2 test: none of the column means are significantly different from zero.

Example 4: Hotelling T^2 test H_0 : all means are equal

Now, turning to a test that all means are equal as displayed in the next table.

Hotelling one sample T^2 test

H_0: Column means are all equal

Number of rows	5	
Number of columns	4	
Hotelling T^2	170.474	
F Statistic (FTS)	28.4123	
Degrees of freedom ($d1, d2$)	3, 2	
$P(F(d1, d2) \geq FTS)$	0.034191	*Reject H_0 at 5% significance level*

This shows the results when the data in `anova6.tf1` are analyzed, and the theoretical background to this test is explained in the reference manual in connection with repeated measures analysis.

Example 5: Compound symmetry

Options are provided for investigating the structure of the covariance matrix. The sample covariance matrix and its inverse can be displayed along with eigenvalues and determinants, and there are also options to check if the covariance matrix has a special form, namely

- testing for compound symmetry,

- testing for spherical symmetry, and

- testing for spherical symmetry of the covariance matrix of orthonormal contrasts.

For instance, using the test file `hotel.tf1` produces the results in the next table.

Compound symmetry test

H_0: Covariance matrix has compound symmetry

Number of groups	1	
Number of variables (m)	4	
Sample size (n)	10	
Determinant of CV	98.14	
Determinant of S_0	1452.0	
$LRTS$ $(-2\log(\lambda))$	36.30	
Degrees of Freedom	8	
$P(\chi^2 \geq LRTS)$	0.0000	*Reject H_0 at 1% significance level*

3

This shows an application of a test for compound symmetry which is when a covariance matrix Σ has a special form with constant nonnegative diagonals and equal nonnegative off-diagonal elements as follows.

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \ldots & \rho \\ \rho & 1 & \ldots & \rho \\ \ldots & \ldots & \ldots & \ldots \\ \rho & \rho & \ldots & 1 \end{pmatrix}$$

This can be tested using estimates for the diagonal and off-diagonal elements σ^2 and $\sigma^2 \rho$ as follows

$$s^2 = \frac{1}{m} \sum_{i=1}^{m} s_{ii}$$

$$s^2 r = \frac{2}{m(m-1)} \sum_{i=2}^{m} \sum_{j=1}^{i-1} s_{ij}.$$

Example 6: Sphericity

The sphericity test, designed to test the null hypothesis $H_0 : \Sigma = kI$ against $H_1 : \Sigma \neq kI$. In other words, the population covariance matrix Σ is a simple multiple of the identity matrix, which is a central requirement for some analytical procedures.

Likelihood ratio sphericity test

H_0: Covariance matrix = k*Identity (for some $k > 0$)

Number of small eigenvalues	0 (i.e. $< 1.00E - 07$)
Number of variables (m)	4
Sample size (n)	10
Determinant of CV	98.1414
Mauchly W statistic	0.00676
$LRTS$ ($-2 \log(\lambda)$)	49.9740
Degrees of Freedom	9
$P(\chi^2 \geq LRTS)$	0.000000 *Reject H_0 at 1% significance level*

The Wilks generalized likelihood-ratio statistic is

$$L = \frac{|S|}{(s^2 - s^2 r)^{m-1} [s^2 + (m-1)s^2 r]},$$

where the numerator is the determinant of the covariance matrix estimated with ν degrees of freedom, while the denominator is the determinant of the matrix with average variance on the diagonals and average covariance as off-diagonal elements, and this is used to construct the test statistic $LRTS$

$$\chi^2 = -\left[\nu - \frac{m(m+1)^2(2m-3)}{6(m-1)(m^2+m-4)} \right] \log L$$

which, for large ν, has an approximate chi-squared distribution with $m(m+1)/2 - 2$ degrees of freedom.

If the sample covariance matrix S has eigenvalues α_i for $i = 1, 2, \ldots, m$ then, defining the arithmetic mean A and geometric mean G of these eigenvalues as

$$A = (1/m) \sum_{i=1}^{m} \alpha_i$$

$$G = (\prod_{i=1}^{m} \alpha_i)^{1/m},$$

the likelihood ratio test statistic

$$-2\log\lambda = nm\log(A/G)$$

is distributed asymptotically as χ^2 with $(m-1)(m+2)/2$ degrees of freedom. Using the fact that the determinant of a covariance matrix is the product of the eigenvalues while the trace is the sum, the Mauchly test statistic W can also be calculated from A and G since

$$W = \frac{|S|}{\{Tr(S)/m\}^m}$$
$$= \frac{\prod_{i=1}^m \alpha_i}{\{(\sum_{i=1}^m \alpha_i)/m\}^m}$$

so that $-2\log\lambda = -n\log W$.

Clearly, the test rejects the assumption that the covariance matrix is a multiple of the identity matrix in this case, a conclusion which is obvious from inspecting the sample covariance and correlation matrices. Since the calculation of small eigenvalues is very inaccurate when the condition number of the covariance matrix is appreciable, any eigenvalues less than the minimal threshold indicated are treated as equal to that threshold when calculating the test statistic.

Example 7: Helmert orthonormal contrasts

The next table results from analysis of `hotel.tf1` and full details will be found in the tutorial on repeat measures ANOVA or in the SimFiT reference manual.

Sphericity test on CV of Helmert orthonormal contrasts

H_0: Covariance matrix = k*Identity (for some $k > 0$)

Number of small eigenvalues	0 (i.e. $< 1.00E-07$)	
Number of variables (m)	4	
Sample size (n)	10	
Determinant of CV	41.5788	
Trace of CV	32.6105	
Mauchly W statistic	0.03237	
$LRTS$ ($-2\log(\lambda)$)	26.4909	
Degrees of Freedom	5	
$P(\chi^2 \geq LRTS)$	0.000072	*Reject H_0 at 1% significance level*
e (Geisser-Greenhouse)	0.402592	
e (Huynh-Feldt)	0.431104	
e (lower bound)	0.333333	