Tests on Covariance Matrices

1 Introduction

In multivariate statistics, the covariance matrix plays a central role. It encodes both the **variability** of each variable and the **covariability** (linear association) between pairs of variables. Formally, if we have a random vector

$$\mathbf{y} = (y_1, y_2, \ldots, y_p)^\top$$

with population mean vector μ and covariance matrix

$$\Sigma = \mathrm{Cov}(\mathbf{y}) = E\big[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^\top\big],$$

then Σ is a p×p symmetric, positive semi-definite matrix containing all the variances and covariances.

In practice, researchers often need to test **hypotheses about the structure of Σ**. For example:

- Is Σ equal to some hypothesized matrix Σ0?
- Is the covariance matrix **spherical**, i.e., proportional to the identity matrix?
- Do multiple groups share the same covariance matrix (homogeneity of covariance assumption in MANOVA)?
- Are subsets of variables independent (zero covariances)?

These questions arise in diverse applications: psychometrics (testing factor models), genetics (testing independence of traits), MANOVA (equal covariance matrices across groups), and repeated measures (sphericity assumptions).

2 Testing a Specified Pattern for Σ

2.1 Testing H0:Σ=Σ0

Suppose we want to test whether the population covariance matrix Σ equals a known, specified positive-definite matrix Σ0.

- Null hypothesis:

  H0:Σ=Σ0

- Alternative:

  Ha:Σ≠Σ0

This is the **multivariate generalization of the univariate chi-square test for variance**.

If $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ are i.i.d. multivariate normal $N_p(\mu, \Sigma)$, then the sample covariance matrix is

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^\top.$$

A likelihood ratio test statistic is

$$\Lambda = \frac{|S|^{n/2}}{|\Sigma_0|^{n/2}} \exp\left\{-\tfrac{n}{2}\mathrm{tr}(\Sigma_0^{-1}S)\right\}.$$

For large n, a transformation yields an approximate chi-square distribution:

$$-\left(n - 1 - \tfrac{2p+5}{6}\right)\ln\Lambda \;\sim\; \chi^2_{df}, \quad df = \frac{p(p+1)}{2}.$$

Interpretation: If the test statistic exceeds the chi-square critical value, we reject H0 and conclude that the covariance structure differs from Σ0.

Example: Testing if a bivariate covariance matrix equals the identity matrix (independent standard normals).

2.2 Testing Sphericity

Definition: A covariance matrix is **spherical** if

$$\Sigma = \sigma^2 I_p,$$

where all variables have equal variance and are uncorrelated.

- Null: $H_0 : \Sigma = \sigma^2 I.$
- Alternative: H_a : not spherical.

This test is important in **MANOVA** and **factor analysis** (Bartlett's test of sphericity).

Likelihood ratio test statistic:

$$\Lambda = \frac{|S|}{\left(\frac{\mathrm{tr}(S)}{p}\right)^p}.$$

Adjusted chi-square approximation:

$$-\left(n - 1 - \tfrac{2p+5}{6}\right) \ln \Lambda \quad \sim \quad \chi^2_{df}, \quad df = \frac{p(p-1)}{2}.$$

Interpretation: Rejection means the covariance matrix is not spherical—variables are either correlated or have unequal variances.

Example: In repeated measures ANOVA, sphericity implies equal variances of differences. Violations require corrections (Greenhouse–Geisser).

2.3 Testing Compound Symmetry

Often, we expect a pattern like:

$$\Sigma = \sigma^2 \left[(1-\rho)I + \rho J \right],$$

where I is the identity matrix and J is a matrix of all 1′s.

- This implies **equal variances** (σ^2) and **equal covariances** ($\rho\sigma^2$) across variables.
- Common in **longitudinal data** (e.g., test scores measured at multiple times with constant correlation).

Likelihood ratio test compares $|S|$ with the determinant under this structured form. Approximate chi-square with $df = \frac{p(p-1)}{2} - 1$.

3 Tests Comparing Covariance Matrices

3.1 Univariate Tests of Equality of Variances

For two groups, the classical **F-test**:

$$F = \frac{s_1^2}{s_2^2}, \quad \frac{s_1^2}{\sigma^2} \sim \chi^2_{n_1-1}, \quad \frac{s_2^2}{\sigma^2} \sim \chi^2_{n_2-1}.$$

But in practice, more robust alternatives are used (Levene’s test, Bartlett’s test).

3.2 Multivariate Tests of Equality of Covariance Matrices

Suppose g groups, with covariance matrices $\Sigma_1, \ldots, \Sigma_g$.

- Null: $H_0 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_g = \Sigma$.

The classical test is **Box’s M test**:

$$M = (N-g)\ln|S_p| - \sum_{i=1}^{g} (n_i - 1)\ln|S_i|,$$

where S_p is the pooled covariance.

For large n,

$$-2\ln\Lambda \approx M \sim \chi^2_{df}, \quad df = \frac{(g-1)p(p+1)}{2}.$$

Interpretation: If rejected, assumption of homogeneity of covariance is invalid—critical for MANOVA and discriminant analysis.

Caution: Box's M is sensitive to non-normality.

4 Tests of Independence

Here the question: **are subsets of variables independent?**

- Independence means corresponding off-diagonal covariances = 0.

4.1 Independence of Two Subvectors

Partition $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2')'$ with covariances

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Test $H_0 : \Sigma_{12} = 0.$

Likelihood ratio:

$$\Lambda = \frac{|\Sigma|}{|\Sigma_{11}||\Sigma_{22}|}.$$

In practice, replace with sample matrices.

4.2 Independence of Several Subvectors

Partition into k blocks. Hypothesis: blocks are mutually independent. Likelihood ratio derived similarly.

4.3 Test for Independence of All Variables

Special case: testing whether all off-diagonal elements of Σ are zero (i.e., covariance matrix is diagonal).

Equivalent to testing zero correlation among all variables.

Likelihood ratio reduces to:

$$\Lambda = \frac{|S|}{\prod_{i=1}^{p} s_{ii}}.$$

For large n,

$$-\left(n - 1 - \tfrac{2p+5}{6}\right) \ln \Lambda \sim \chi^2_{\frac{p(p-1)}{2}}.$$

Interpretation: If rejected, at least some variables are correlated.