Inference on the means of multiple multivariate samples
Part II - Multivariate Analysis of Variance (MANOVA)

The first step in the analysis of multiple samples is often the comparison of the sample means.
The method of analysis of variance (ANOVA) is designed to compare the means of multiple samples.

In the following we consider multivariate analysis of variance (MANOVA).

The notations become inevitably cumbersome for multiple samples of multivariate vectors, since the subscripts need to denote vector components, to label populations or the samples, as well as to index the observations.

1 Comparing mean vectors of several populations

Data: Multivariate data of several sample groups have the form

$$\begin{array}{l} \text{Sample 1 of size } n_1 \\ \\ \\ \\ \text{Sample 2 of size } n_2 \\ \\ \\ \\ \\ \\ \text{Sample } g \text{ of size } n_g \\ \\ \\ \\ \end{array} \begin{bmatrix} x_{1,11} & x_{1,12} & \cdots & x_{1,1p} \\ x_{1,21} & x_{1,22} & \cdots & x_{1,2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,n_1 1} & x_{1,n_1 2} & \cdots & x_{1,n_1 p} \\ \hline x_{2,11} & x_{2,12} & \cdots & x_{2,1p} \\ x_{2,21} & x_{2,22} & \cdots & x_{2,2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{2,n_2 1} & x_{2,n_2 2} & \cdots & x_{2,n_2 p} \\ \hline \\ \vdots & \vdots & \vdots & \vdots \\ \\ \hline x_{g,11} & x_{g,12} & \cdots & x_{g,1p} \\ x_{g,21} & x_{g,22} & \cdots & x_{g,2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{g,n_g 1} & x_{g,n_g 2} & \cdots & x_{g,n_g p} \end{bmatrix}$$

Here the multivariate data are treated as g groups of observed sample values from g distributions, respectively.

We have already studied inference methods such as Hotelling's T^2 for the one-sample case when $g = 1$ and the two-sample case when $g = 2$. Now we focus on the case $g \geq 3$.

Goal: Inference on the equality of the mean parameters of the g populations based on the observed g samples.

Recall that, in the univariate case when $p = 1$, the method of analysis of variance (ANOVA) is used to compare the means of several groups.

1.1 Review: Univariate analysis of variance one-factor model

Analysis of variance (ANOVA) is used to compare the means of several univariate samples.

Sample measurements: $X_{\{treatment\ \ell,\ observation\ j\}}$

As mentioned before, we follow the convention to name different samples or different populations as different treatment groups.

There are n_ℓ observations from the population or treatment group ℓ, $\ell = 1, \cdots, g$.

Under the normality assumption of the random errors, the ANOVA one-factor model can be written as

$$X_{\ell j} = \mu + \tau_\ell + e_{\ell j}, \qquad e_{\ell j} \sim N(0, \sigma^2), \quad \ell = 1, \cdots, g, \quad j = 1, \cdots, n_\ell,$$

where
μ is the overall mean parameter (or the common mean under certain parametrization),
τ_ℓ is the treatment effect parameter of the ℓth population,

Thus

$$\mathbb{E}(X_{\ell j}) = \mu + \tau_\ell, \qquad for \quad \ell = 1, \cdots, g, \quad j = 1, \cdots, n_\ell.$$

To avoid unidentifiability issue in the parameters, there should be constraints on the parameters.

Common constraints:

- The sum-zero constraint $\sum_{\ell=1}^{g} n_\ell \tau_\ell = 0$.
- $\tau_1 = 0$.
- $\tau_g = 0$

The most common constraint used in textbook derivations is the sum-zero constraint $\sum_{\ell=1}^{g} n_\ell \tau_\ell = 0$.

Be aware that software often automatically impose constraints, such as $\tau_1 = 0$ or $\tau_g = 0$, where the the group index $(i = 1, 2, \cdots, g)$ is automatically ordered, for example, by the alphabetical order of group names.

To detect differences in treatment effects among the groups, the first test of interest is

$$\begin{cases} H_0: & \tau_1 = \cdots = \tau_g = 0 \\ H_a: & \tau_i \neq 0, \text{ for some } i = 1, \cdots, g. \end{cases}$$

The test statistic is based on the analysis of variance by comparing variations within treatment groups and the variations between treatment groups.

Denote $\bar{x}_\ell, \bar{x}$ as the ℓth sample mean and the overall sample mean. Decompose the jth observation in the ℓth sample as

$$x_{\ell j} = \bar{x} + (\bar{x}_\ell - \bar{x}) + (x_{\ell j} - \bar{x}_\ell)$$

Equivalently,

$$(x_{\ell j} - \bar{x}) = (\bar{x}_\ell - \bar{x}) + (x_{\ell j} - \bar{x}_\ell)$$

Square both sides, then sum up over all observation j and all treatment group ℓ.

$$\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x})^2 = \sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} [(\bar{x}_\ell - \bar{x}) + (x_{\ell j} - \bar{x}_\ell)]^2$$

Since $\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\bar{x}_\ell - \bar{x})(x_{\ell j} - \bar{x}_\ell) = 0$, the equation becomes

$$\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x})^2 = \sum_{\ell=1}^{g} n_\ell (\bar{x}_\ell - \bar{x})^2 + \sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x}_\ell)^2$$

This is the fundamental equation of analysis of variance, often written in the shorthand form

$$SS_{total} = SS_{treatment} + SS_{error}$$

where the last term has various expressions, such as $SSE, SS_e, SS_{residual}$.

The fundamental ANOVA equation is commonly interpreted as

$$\sum (\text{total variation})^2 = \sum \begin{pmatrix} \text{"between-group"} \\ \text{treatment variation} \end{pmatrix}^2 + \sum \begin{pmatrix} \text{"within-group"} \\ \text{residual variation} \end{pmatrix}^2$$

The corresponding numbers of independent quantities of each term, which equal to the degrees of freedom, have the relation

$$\sum_{\ell=1}^g n_\ell \ -1 \quad = \quad (g-1) \quad + \quad \sum_{\ell=1}^g (n_\ell - 1)$$

The univariate analysis of variance (ANOVA) table below summarizes relations of the sums of squares, as well as the comparisons in terms of a ratio, which is of F distribution under the null hypothesis and normality assumption.

Source of variation	SS (sum of squares)	d.f.	Variance ratio (F-value)
Treatments	$SS_{trt} = \sum_{\ell=1}^g \sum_{j=1}^{n_\ell} (\bar{x}_\ell - \bar{x})^2$	$g-1$	$\frac{SS_{trt}/(g-1)}{SS_{res}/(\sum_{\ell=1}^g n_\ell - g)}$
Residuals	$SS_{res} = \sum_{\ell=1}^g \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x}_\ell)^2$	$\sum_{\ell=1}^g n_\ell - g$	
Total	$SS_{tot} = \sum_{\ell=1}^g \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x})^2$	$\sum_{\ell=1}^g n_\ell - 1$	

With the ANOVA model assumptions, under the null $H_0 : \tau_1 = \cdots = \tau_g = 0$,

$$\frac{SS_{trt}/(g-1)}{SS_{res}/(\sum_{\ell=1}^g n_\ell - g)} \sim F_{g-1, \sum_{\ell=1}^g n_\ell - g}$$

Therefore, at test level α, the null H_0 is rejected if

$$\frac{SS_{trt}/(g-1)}{SS_{res}/(\sum_{\ell=1}^g n_\ell - g)} > F_{g-1, \sum_{\ell=1}^g n_\ell - g}(\alpha)$$

1.2 MANOVA one-factor model

Multivariate analysis of variance (MANOVA) is for comparing the means of $g \geq 3$ multivariate samples, where each observation consists of measurements of $p > 1$ variables (or features).

Measurements on the jth observation in the ℓth treatment group is a p-vector

$$\boldsymbol{X}_{\ell j} = \boldsymbol{X}_{\{\text{treatment } \ell, \text{ observation } j\}} \ = \ \big(X_{\{\text{trt } \ell, \text{ obs } j, \text{ variable } 1\}}, \ \cdots, \ X_{\{\text{trt } \ell, \text{ obs } j, \text{ variable } p\}}\big)$$

For $p > 1$, the observations are p-vectors, the model becomes

$$\boldsymbol{X}_{\ell j} = \boldsymbol{\mu} + \boldsymbol{\tau}_\ell + \boldsymbol{e}_{\ell j}, \qquad \boldsymbol{e}_{\ell j} \sim N_p(0, \Sigma), \qquad j = 1, \cdots, n_\ell, \quad \ell = 1, \cdots, g.$$

Similar to the univariate case, $\boldsymbol{\mu}$ is the overall mean vector, $\boldsymbol{\tau}_\ell$ is the ℓth treatment effect vector. Thus

$$\mathbb{E}(\boldsymbol{X}_{kj}) = \boldsymbol{\mu} + \boldsymbol{\tau}_k$$

for any observation $j = 1, \cdots, n_k$ in treatment group k, for any $k = 1, \cdots, g$.

In order for the parameters to be identifiable, the parameters need to satisfy certain constraints.
In the following we use the sum-zero constraint

$$\sum_{\ell=1}^g n_\ell \boldsymbol{\tau}_\ell = 0$$

Analogous to the univariate case of $p = 1$, the hypothesis test of interest in MANOVA is

$$\begin{cases} H_0 : & \boldsymbol{\tau}_1 = \cdots = \boldsymbol{\tau}_g = 0_p \\ H_a : & \boldsymbol{\tau}_i \neq 0_p \ \text{ for some } i = 1, \cdots, g. \end{cases}$$

The test statistic is constructed based on the analysis of variance on the vector-valued observations, this time by comparing covariance matrices.

Each observed p-variate data point can be decomposed similarly,

$$\boldsymbol{x}_{\ell j} = \bar{\boldsymbol{x}} + (\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}}) + (\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}}_\ell)$$

which can be qualitatively stated as

$$\text{(observation)} = \text{(overall sample mean)} + \text{(estimated treatment effect)} + \text{(residual)}$$

Subtract $\bar{\boldsymbol{x}}$, the equation becomes a partition of variations.

$$(\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}}) = (\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}}) + (\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}}_\ell)$$

The variation of an individual vector-valued observation splits into two terms, similar to the variance decomposition for univariate observations. However we cannot square a vector.

Recall the vector transpose relation $(\boldsymbol{v} + \boldsymbol{w})' = \boldsymbol{v}' + \boldsymbol{w}'$,

$$(\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}})' = (\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})' + (\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}}_\ell)'$$

Multiplying the vector by its transpose, we obtain an **equation of $p \times p$ matrices**.

$$\begin{aligned}(\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}})(\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}})' = {} & (\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})' + (\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})(\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}}_\ell)' \\ & + (\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}}_\ell)(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})' + (\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}}_\ell)(\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}}_\ell)'\end{aligned}$$

Summing up over observation index j and treatment group index ℓ. Note that the cross product terms are zero matrices,

$$\sum_{\ell=1}^g \sum_{j=1}^{n_\ell} (\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})(\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}}_\ell)' = \sum_{\ell=1}^g (\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}}) \Big(\sum_{j=1}^{n_\ell} (\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}}_\ell)' \Big) = \sum_{\ell=1}^g (\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})(n\bar{\boldsymbol{x}}_\ell - n\bar{\boldsymbol{x}}_\ell)' = \boldsymbol{O}_{p \times p}$$

Similarly

$$\sum_{\ell=1}^g \sum_{j=1}^{n_\ell} (\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}}_\ell)(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})' = \boldsymbol{O}_{p \times p}$$

we are left with

$$\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}})(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}})' = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\bar{\boldsymbol{x}}_{\ell}-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_{\ell}-\bar{\boldsymbol{x}})' + \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}}_{\ell})(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}}_{\ell})' \tag{1}$$

The above equation (1) can be written as

$$\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}})(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}})' = \sum_{\ell=1}^{g} n_\ell(\bar{\boldsymbol{x}}_{\ell}-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_{\ell}-\bar{\boldsymbol{x}})' + \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}}_{\ell})(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}}_{\ell})' \tag{2}$$

The decomposition equation can be stated as

(**Total** sum of squares and cross products about the mean)

$\|$

(**treatment** or **between group** sum of squares and cross product)

$+$

(**residual** or **within group** sum of squares and product)

The decomposition is comparable to the partition of variance in the univariate ANOVA.

The component-wise numbers of independent quantities, the degrees of freedom, still satisfy the relation

$$\sum_{\ell=1}^{g} n_\ell - 1 \quad = \quad (g-1) \quad + \quad \sum_{\ell=1}^{g}(n_\ell - 1)$$

By convention, we denote the **between** group (or between population) sum of squares and cross products matrix as

$$\boldsymbol{B} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\bar{\boldsymbol{x}}_{\ell}-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_{\ell}-\bar{\boldsymbol{x}})' = \sum_{\ell=1}^{g} n_\ell(\bar{\boldsymbol{x}}_{\ell}-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_{\ell}-\bar{\boldsymbol{x}})'$$

and the **within** group sum of squares and cross products matrix as

$$\boldsymbol{W} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}}_{\ell})(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}}_{\ell})'$$

In fact $\boldsymbol{W}$ is related to the pooled covariance matrix. Since the sample covariance matrix for the ℓth group is

$$\boldsymbol{S}_\ell = \frac{1}{n_\ell - 1}\sum_{j=1}^{n_\ell}(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}}_{\ell})(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}}_{\ell})'$$

we can write the within group sum of squares and cross products matrix $\boldsymbol{W}$ as

$$\boldsymbol{W} = (n_1-1)\boldsymbol{S}_1 + \cdots + (n_g-1)\boldsymbol{S}_g$$

Thus

$$\boldsymbol{S}_{pool} = \frac{1}{\sum_{\ell=1}^{g}(n_\ell-1)}\left[(n_1-1)\boldsymbol{S}_1 + \cdots + (n_g-1)\boldsymbol{S}_g\right] = \frac{1}{n-g}\boldsymbol{W}$$

Correspondingly, there is a MANOVA table in appearance similar to the univariate ANOVA table. In the place of SS's in the MANOVA table are the "sums of squares and cross products", which are $p \times p$ matrices.

Source of variation	Matrix of sum of squares and cross-products	Degrees of freedom
Treatments	$\boldsymbol{B} = \sum_{\ell=1}^{g} n_\ell(\bar{\boldsymbol{x}}_{\ell}-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_{\ell}-\bar{\boldsymbol{x}})'$	$g-1$
Residuals	$\boldsymbol{W} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}}_{\ell})(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}}_{\ell})'$	$\sum_{\ell=1}^{g} n_\ell - g$
Total	$\boldsymbol{B}+\boldsymbol{W} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}})(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}})'$	$\sum_{\ell=1}^{g} n_\ell - 1$

Because the middle column entries are matrices, computational software does not print out MANOVA table in their matrix forms. Mostly often, only the determinant values of the matrices are displayed in MANOVA output.

The MANOVA table is a convenient summary of useful information needed in testing the hypothesis of no treatment effects:

$$H_0: \ \boldsymbol{\tau}_1 = \cdots = \boldsymbol{\tau}_g = 0_p$$

Analogous to the univariate situation, the hypothesis test should compare the relative "sizes" of variations due to the treatments with the variations due to random fluctuations.

The variation are characterized in the sums of squares and cross products matrices $\boldsymbol{B}$ and $\boldsymbol{W}$, respectively.

Since $\boldsymbol{B}$ and $\boldsymbol{W}$ are $p \times p$ matrices, the test statistic uses the highly capsulized measure, the determinant.

1.3 MANOVA tests

We introduce two popular test of multivariate analysis of variance.

- Wilk's lambda

  Wilk's lambda test considers the ratio of the determinant $\boldsymbol{W}$ and $\boldsymbol{B}+\boldsymbol{W}$:

  Wilks' lambda: $$\Lambda^* = \frac{\det(\boldsymbol{W})}{\det(\boldsymbol{B}+\boldsymbol{W})} = \frac{\left|\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}}_{\ell})(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}}_{\ell})'\right|}{\left|\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}})(\boldsymbol{x}_{\ell j}-\bar{\boldsymbol{x}})'\right|}$$

  In other words,

  $$\Lambda^* = \frac{\text{generalized variance of residual}}{\text{generalized variance of total}}$$

  The null distribution of Λ^* is related to F distribution with degrees depending on number of samples g, variable dimension p, and sample sizes $n_i, i = 1, \cdots, g$ (examples see remarks below).

- Bartlett's asymptotic approximation

  For $n = \sum n_\ell$ large, Bartlett gives a simple chi-square approximation.

  Bartlett's approximation: $$-\left(n - 1 - \frac{p+g}{2}\right)\ln\Lambda^* \quad \sim \ \chi^2_{p(g-1)} \quad under \quad H_o$$

Remarks on Wilks' and Barlett's test

- $\Lambda^* \in [0,1]$.
- We reject the null hypothesis that all group means are equal if the value of Λ^* is too small by our criteria.
- Wilks' lambda is equivalent to the likelihood ratio test statistic.
- Under the null Wilks' lambda is of its own Λ^*-distribution, which is derived from the ratio of two random matrices, each is of Wishart distribution (proof omitted).
- For multivariate normal samples, both $\boldsymbol{W}$ and $\boldsymbol{B}+\boldsymbol{W}$ are of Wishart distributions.
- The probability distribution of Wilks' Λ^* under H_0 is different for different variable dimension p, number of treatment groups g, and sample size $n = \sum_\ell n_\ell$.

  Some common combinations are given in Table 6.3 in the textbook by Johnson and Wichern.

  Note that the distributions are not for Λ^* directly. For example, for $g = 3$ and $p = 4$,

$$\frac{(\sum_\ell n_\ell - p - 2)}{p}\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} \sim F_{2p,2(\sum n_\ell - p - 2)} \qquad under\ H_o.$$

- The critical values of Wilks' test statistic are built in and provided by software such as R.
- For large sample size n, Bartlett's approximation is simpler and easier to calculate.
- Even for moderate sample size, it is good practice to check and compare both tests.

1.4 Basic ideas in test statistics of MANOVA

Common MANOVA tests in terms of determinants are related to matrix eigenvalues.

Let $\lambda_1, \cdots, \lambda_p$ be eigenvalues of matrix $\boldsymbol{W}^{-1}\boldsymbol{B}$.
Several common MANOVA test statistics, including Wilks' Λ^*, can be expressed in terms of the λ_i's.

The following are a few common MANOVA test statistics expressed in terms of the eigenvalues of $\boldsymbol{W}^{-1}\boldsymbol{B}$.

- Wilks' lambda

$$\Lambda^* = \frac{|\boldsymbol{W}|}{|\boldsymbol{B}+\boldsymbol{W}|} = \frac{1}{|\boldsymbol{W}^{-1}\boldsymbol{B}+I_p|} = \prod_{k=1}^{p}\frac{1}{1+\lambda_k}$$

- Hotelling-Lawley's test statistic

$$\text{Hotelling-Lawley's Trace} = trace(\boldsymbol{B}\boldsymbol{W}^{-1}) = \sum_{k=1}^{p}\lambda_k$$

- Pillai or Pillai-Barrlett test statistic

$$\text{Pillai's Trace} = trace(\boldsymbol{B}(\boldsymbol{B}+\boldsymbol{W})^{-1}) = trace(\boldsymbol{B}\boldsymbol{W}^{-1}(\boldsymbol{B}\boldsymbol{W}^{-1}+I_p)^{-1}) = \sum_{k=1}^{p}\frac{\lambda_k}{1+\lambda_k}$$

- Roy's test statistic (by one of several definitions)

$$\text{Roy's Largest Root} = \max_k\{\lambda_k\}$$

  which provides a conservative upper bound.

Remarks

- These common statistics are often included in software MANOVA output.
- Most of the statistics are related to F distributions, at least asymptotically.
- For large samples, the tests are nearly equivalent.
- In practice, the λ_i's are approximated by $\hat{\lambda}_i$'s, the eigenvalues of $\hat{\boldsymbol{W}}^{-1}\hat{\boldsymbol{B}}$, estimated from sample data.

1.5 Checking normality assumptions for multivariate sample

Recall in univariate case, to check if $\varepsilon_1, \cdots, \varepsilon_n$ are independent observations from a univariate normal distribution $N(\mu, \sigma^2)$, the normal Q-Q plot is a quick, visual diagnostic tool to check if the normality assumption is obviously violated.

The Q-Q plot (quantile-quantile plot) for the ε_i's is the plot of quantile points

$$\left(\Phi^{(-1)}\Big(\frac{i}{n+1}\Big),\ \varepsilon_{(i)}\right), \qquad i =, 1\cdots, n,$$

where

$$\Phi(z) = \int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$

is the cumulative distribution function of $N(0,1)$, and $\varepsilon_{(1)} \le \cdots \le \varepsilon_{(n)}$ are the ordered values of observed $\varepsilon_1, \cdots, \varepsilon_n$, called order statistics of $\varepsilon_1, \cdots, \varepsilon_n$.

The plot should resemble a straight line, if the ε_i's are indeed i.i.d. observations from a normal distribution. The idea or derivation of the method is based on order statistics (usually derived in STAT245x0 Mathematical Statistics Methods II).

In multivariate case, to check if observations $\boldsymbol{X}_i, i = 1, \cdots, n$ are indeed independent samples from a common multivariate normal $N_p(\boldsymbol{\mu}, \Sigma)$, a Chi-Squared Q-Q plot can be used to assess multivariate normality, based on the following fact.

For $\boldsymbol{X}_i$ i.i.d. $\sim N_p(\boldsymbol{\mu}, \Sigma)$, $i = 1, \cdots, n$, define

$$T_i^2 = (\boldsymbol{X}_i - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu}), \qquad i = 1, \cdots, n.$$

Then

$$T_i^2,\ i = 1, \cdots, n \qquad i.i.d. \quad \sim \chi_p^2$$

Analogous to the univariate normal Q-Q plot, quantiles of the T_i's vs. quantiles of χ^2_p should roughly form a line. This fact can be used to assess the normality assumption of the $\boldsymbol{X}_i$'s.

Since MANOVA tests are on the means, thanks to the central limit theorem, the normality assumption usually is relatively robust, that is, the MANOVA test still gives sensible results under moderate deviations from the normality assuption.

2 Post MANOVA

If the null hypothesis of MANOVA is rejected, naturally the next question is, which treatments have statistically significant effects?

Write the kth treatment mean as

$$\boldsymbol{\mu}_k = \boldsymbol{\mu} + \boldsymbol{\tau}_k = \text{ Overall mean } + \text{ Effect of treatment } k$$

To compare the effects of treatment k and treatment ℓ, the quantity of interests is the difference of the effect vectors

$$\boldsymbol{\tau}_k - \boldsymbol{\tau}_\ell$$

which is the same as the difference of the treatment-mean vectors

$$\boldsymbol{\mu}_k - \boldsymbol{\mu}_\ell$$

The sample estimator for $\mu_{\ell i}$, the mean of the ith component variable of the ℓth treatment group, is

$$\hat{\mu}_{\ell i} = \bar{X}_{\ell i} = \frac{1}{n_\ell}\sum_{j=1}^{n_\ell} X_{\ell j i}$$

the sample mean of group ℓ on the ith variable. Thus

$$\hat{\tau}_{ki} - \hat{\tau}_{\ell i} = \hat{\mu}_{ki} - \hat{\mu}_{\ell i} = \bar{X}_{ki} - \bar{X}_{\ell i}$$

A confidence interval for $\tau_{ki} - \tau_{\ell i}$ will have the form

$$\hat{\tau}_{ki} - \hat{\tau}_{\ell i} \pm c \times \sqrt{\widehat{var}(\hat{\tau}_{ki} - \hat{\tau}_{\ell i})}$$

where the multiplier c depends on the level and the type of the confidence interval.

Estimation of $var(\hat{\tau}_{ki} - \hat{\tau}_{\ell i})$

Assuming mutual independence among the g samples, so the observations from different samples (a.k.a. different treatment groups) are independent. For the ith component, the treatment effect difference between sample k and sample ℓ has variance

$$var(\hat{\tau}_{ki} - \hat{\tau}_{\ell i}) = var\left(\bar{X}_{ki} - \bar{X}_{\ell i}\right) = var\left(\bar{X}_{ki}\right) + var\left(\bar{X}_{\ell i}\right)$$

Under the assumption that the g populations are of equal covariance structure,

$$\Sigma_1 = \cdots = \Sigma_g = \Sigma = [\sigma_{k\ell}]_{k,\ell=1,\cdots,p}$$

The overall population variance of the ith component variable is σ_{ii}, the ith diagonal element of the common population covariance matrix Σ. Therefore the sample means have variance

$$var\left(\bar{X}_{ki}\right) = \frac{\sigma_{ii}}{n_k}, \qquad var\left(\bar{X}_{\ell i}\right) = \frac{\sigma_{ii}}{n_\ell}$$

The sample estimate of σ_{ii} is the ith diagonal element of the pooled sample covariance matrix

$$\boldsymbol{S}_{pool} = \frac{1}{\sum_{\ell=1}^{g}(n_\ell - 1)}\left[(n_1-1)\boldsymbol{S}_1 + \cdots + (n_g - 1)\boldsymbol{S}_g\right] = \frac{1}{n-g}\boldsymbol{W}$$

which gives

$$\widehat{var}(\hat{\tau}_{ki} - \hat{\tau}_{\ell i}) = \widehat{var}\left(\bar{X}_{ki}\right) + \widehat{var}\left(\bar{X}_{\ell i}\right) = \left(\frac{1}{n_k} + \frac{1}{n_\ell}\right)\hat{\sigma}_{ii} = \left(\frac{1}{n_k} + \frac{1}{n_\ell}\right)\frac{w_{ii}}{n-g}$$

with w_{ii} the ith diagonal element of $\boldsymbol{W}$, the within group sum of squares and cross products matrix used in MANOVA.

Derivation of the multiplier c

Rigorously speaking, a general comparison of component effects in terms of confidence intervals should require

$$\hat{\tau}_{ki} - \hat{\tau}_{\ell i} \pm c \times \sqrt{\widehat{var}(\hat{\tau}_{ki} - \hat{\tau}_{\ell i})}, \qquad \textbf{for all} \quad i = 1, \cdots, p; \quad k, \ell = 1, \cdots, g.$$

That is, the comparison needs simultaneous confidence intervals for all p components in all g samples, at a desired confidence level. For a given confidence level, the multiplier c depends on the type of the simultaneous confidence intervals.

Bonferroni simultaneous confidence intervals for effect differences (under equal covariance)

When the number of components p and the number of treatment groups g are small or moderate, the Bonferroni method is a convenient way to produce a set of simultaneous confidence intervals (admittedly that the method is on the conservative side, see remarks below).

By the Bonferroni method, the $(1-\alpha)100\%$ confidence interval for the ith component of the difference vector has the form

$$\hat{\tau}_{ki} - \hat{\tau}_{\ell i} \pm t_d(\alpha/2m)\sqrt{\widehat{var}(\hat{\tau}_{ki} - \hat{\tau}_{\ell i})}$$

where

$$m = p\binom{g}{2} = p\frac{g(g-1)}{2}$$

is the number of simultaneous confidence intervals. Therefore, for individual component confidence intervals, the halved test level is

$$\frac{\alpha}{2m} = \frac{\alpha}{pg(g-1)}$$

which is used to find the critical quantile $t_d(\alpha/2m)$, the multiplier.

> For example, if there are $p = 3$ variables and $g = 4$ groups, and we wish to test at the level $\alpha = 0.10$, we would have
>
> $$m = p\frac{g(g-1)}{2} = 18, \qquad \frac{\alpha}{m} = 0.006, \qquad \frac{\alpha}{2m} = 0.003$$
>
> In other words, in order to have 90% confidence level for all simultaneous confidence intervals, Bonferroni methods would require each individual confidence interval to be at confidence level of 99.7%.

Under the null hypothesis $\tau_{ki} - \tau_{\ell i} = 0$,

$$\frac{(\hat{\tau}_{ki} - \hat{\tau}_{\ell i}) - 0}{\sqrt{\widehat{var}(\hat{\tau}_{ki} - \hat{\tau}_{\ell i})}} \sim t_d$$

where the degrees of freedom d in the multiplier t_d is determined by the degrees of freedom of the variance estimator $\widehat{var}(\hat{\tau}_{ki} - \hat{\tau}_{\ell i})$. Recall that $n - g = \sum_{\ell=1}^{g} n_\ell - 1$ is the degrees of freedom of $\boldsymbol{W}$ and $\boldsymbol{S}_{pool}$, which gives the degrees of freedom of the t-test.

Therefore we have obtained Bonferroni simultaneous component-wise confidence intervals for $\boldsymbol{\tau}_k - \boldsymbol{\tau}_\ell$, the difference of treatment effects of group k and group ℓ, formulated as

$$\bar{x}_{ki} - \bar{x}_{\ell i} \pm t_{n-g}(\alpha/2m)\sqrt{s_{ii}\left(\frac{1}{n_k}+\frac{1}{n_\ell}\right)} \qquad i = 1, \cdots, p; \quad k, \ell = 1, \cdots, g$$

or

$$\bar{x}_{ki} - \bar{x}_{\ell i} \pm t_{n-g}(\alpha/pg(g-1))\sqrt{\frac{w_{ii}}{n-g}\left(\frac{1}{n_k}+\frac{1}{n_\ell}\right)} \qquad i = 1, \cdots, p; \quad k, \ell = 1, \cdots, g.$$

Remarks

In the simultaneous tests, when taking both dimension p and number of samples g into consideration, the Bonferroni method often gives confidence intervals too wide to be practical even for moderate number of variables p and number of samples g. The example above with $p = 3$ and $g = 4$ would required 99.7% confidence level for each of the 18 simultaneous confidence intervals. Such high confidence level is likely to result in wide, inconclusive confidence intervals.

3 Box's M-test for equality of covariance matrices

One of the assumptions in multivariate analysis of variance test is that the covariance matrices of the potentially different populations or treatment groups are the same. Box's M-test is to check the equal covariance assumption.

There are p variables and g populations of interest in the analysis.
Denote the covariance matrix of the p variables from population j as Σ_j, $j = 1, \cdots, g$.

Question: Are the variables in the g population groups sharing the same covariance structure?

$$\begin{cases} H_0: & \Sigma_1 = \Sigma_2 = \cdots = \Sigma_g = \Sigma \\ H_a: & \Sigma_i \neq \Sigma_j \quad for\ some\ \ i \neq j. \end{cases}$$

The most commonly used test statistic for the above hypotheses is Box's M-test.

Box's M-test

Box's M-test for equal covariance structure of several populations is a likelihood-ratio type of test.

Recall that, for a random sample of independent p-variate random vectors $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ of common distribution $N_p(\boldsymbol{\mu}, \Sigma)$, their joint likelihood function has the form

$$L(\boldsymbol{\mu}, \Sigma) = f(\boldsymbol{x}_1, \cdots, x_n; \boldsymbol{\mu}, \Sigma) = \prod_{j=1}^{n} f(\boldsymbol{x}_j) = \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} e^{-\frac{1}{2}\sum_{j=1}^{n}(\boldsymbol{x}_j - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}_j - \boldsymbol{\mu})}$$

We have derived (ref. week 2 lecture notes on multivariate normal distribution and chapter 4 in Johnson and Wichern) that, the maximum likelihood

$$\max_{\boldsymbol{\mu}, \Sigma} L(\boldsymbol{\mu}, \Sigma) = L(\hat{\boldsymbol{\mu}}, \hat{\Sigma}) = \frac{1}{(2\pi)^{np/2}|\boldsymbol{S}_n|^{n/2}} e^{-\frac{np}{2}}$$

The maximum is achieve at the overall sample mean vector $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}$ and the MLE sample covariance matrix estimator $\hat{\Sigma} = \boldsymbol{S}_n$, where $\boldsymbol{S}_n = \frac{n-1}{n}\boldsymbol{S}$, with $\boldsymbol{S}$ the sample covariance matrix of the n sample vectors.

Assume that the p variables are of multivariate normal distribution, the ℓth population has sample size n_ℓ and sample covariance matrix $\boldsymbol{S}_\ell$. Similar to the MLE for one sample multivariate normal variables, Box's M-test of equality among sample covariance matrices uses test statistic

$$\Lambda = \prod_{\ell=1}^{g}\left(\frac{|\boldsymbol{S}_\ell|}{|\boldsymbol{S}_{pool}|}\right)^{(n_\ell - 1)/2}$$

where

$$\boldsymbol{S}_{pool} = \frac{1}{n-g}\left[(n_1 - 1)\boldsymbol{S}_1 + \cdots + (n_g - 1)\boldsymbol{S}_g\right]$$

Box's test is based on an approximation that the sampling distribution of $\ln \Lambda$ is approximately of χ^2 distribution under the equal covariance matrix hypothesis H_0.

Box's M is defined as

$$M = -2\ln\Lambda = (n-g)\ln|\boldsymbol{S}_{pool}| - \sum_{\ell=1}^{g}\left[\ (n_\ell - 1)\ln|\boldsymbol{S}_\ell|\ \right].$$

To obtain the appropriate constant multiple for the χ^2 distribution, there is a needed intermediate value

$$u = \left(\sum_{\ell=1}^{g}\frac{1}{n_\ell - 1} - \frac{1}{n-g}\right)\frac{2p^2 + 3p - 1}{6(p+1)(g-1)}.$$

Under the hypothesis H_0 of equal covariance (derivation omitted),

$$(1-u)M \sim \chi^2_v \qquad (approximately)$$

with degrees of freedom

$$v = \text{d.f.} = p(p+1)(g-1)/2.$$

H_o is rejected if

$$(1-u)M > \chi^2_{p(p+1)(g-1)/2}(\alpha)$$

Box's M-test works well for small p and g (≤ 5) and moderate to large n_ℓ (≥ 20).

For large sample sizes, the MANOVA tests are quite robust to violations of normality assumptions, especially when sample sizes are equal.

Note: Relevant chapter in Johnson and Wichern: Chapter 6.