11 One-Way Multivariate Analysis of Variance (MANOVA)

A preliminary to Canonical Variates Analysis (CVA)

11.1 Introduction

One-way MANOVA extends the two-sample Hotelling T^2 test to K groups ($K > 2$).

Once again, we start by considering the *univariate* case and one-way ANOVA. Suppose that we have a sample of size n_k from each of $k = 1, \ldots, K$ populations with a common variance σ^2, but possibly different means $\mu_1, \ldots, \mu_k$.

Let x_{kj} be the observation for the jth individual in sample k.

Then,

$$\overline{x} = \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} x_{kj}$$

where

$$n = \sum_{k=1}^{K} n_k$$

and

$$\overline{x}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} x_{kj}, k = 1, \ldots, K.$$

Note also that

$$\overline{x} = \frac{1}{n} \sum_{k=1}^{n} n_k \overline{x}_k \quad \text{is a } \textit{weighted} \text{ average of the } \overline{x}_k.$$

To test $H_0 : \mu_1 = \ldots \mu_k$ versus $H_1 : \mu_k \neq \mu_\ell$ for some $k \neq \ell$ we refer to the following ANOVA table.

ANOVA Table

Source	SS	df	MS	VR
Between Groups	B	$K-1$	$B/(K-1)$	$(B/(K-1))/(W/(n-K))$
Within Groups	W	$n-K$	$W/(n-K)$	
	T	$n-1$		

where the sums of squares (SS) terms are given by,

$$T = \sum_{k=1}^{K} \sum_{j=1}^{n_k} (x_{kj} - \overline{x})^2$$

$$W = \sum_{k=1}^{K} \sum_{j=1}^{n_k} (x_{kj} - \overline{x}_k)^2$$

$$B = \sum_{k=1}^{K} \sum_{j=1}^{n_k} (\overline{x}_k - \overline{x})^2 = \sum_{k=1}^{K} n_k (\overline{x}_k - \overline{x})^2.$$

Assuming normality, so that the x_{kj} are independently drawn from $\mathrm{N}(\mu_k, \sigma^2)$, then under the null hypothesis, $\mathrm{VR} \sim \mathrm{F}(K-1, n-K)$.

Now consider the *multivariate* case.

Let

$$\mathbf{X}_k = \begin{pmatrix} \mathbf{x}_{k1}^T \\ \mathbf{x}_{k2}^T \\ \vdots \\ \mathbf{x}_{kn_k}^T \end{pmatrix}$$

be an $(n_k \times p)$ data matrix from population k. Thus the $(n \times p)$ data matrix

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_K \end{pmatrix}$$

divides into K groups assumed to come from populations with a common covariance matrix $\mathbf{\Sigma}$, but possibly different mean vectors $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K$.

$\mathbf{x}_{kj}$ is the $(p \times 1)$ observation vector for the jth individual in sample k.
Then,

$$\overline{\mathbf{x}}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} \mathbf{x}_{kj}$$

is the $(p \times 1)$ vector of variable means in sample k.

$$\overline{\mathbf{x}} = \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \mathbf{x}_{kj}$$

is the overall $(p \times 1)$ vector of variable means.

The total, within-samples and between-samples SSP (sums of squares and products) matrices are given respectively by

$$\mathbf{T} = \sum_{k=1}^{K} \sum_{j=1}^{n_k} (\mathbf{x}_{kj} - \overline{\mathbf{x}})(\mathbf{x}_{kj} - \overline{\mathbf{x}})^T$$

$$\mathbf{W} = \sum_{k=1}^{K} \sum_{j=1}^{n_k} (\mathbf{x}_{kj} - \overline{\mathbf{x}}_k)(\mathbf{x}_{kj} - \overline{\mathbf{x}}_k)^T$$

$$\mathbf{B} = \sum_{k=1}^{K} n_k (\overline{\mathbf{x}}_k - \overline{\mathbf{x}})(\overline{\mathbf{x}}_k - \overline{\mathbf{x}})^T$$

Assuming multivariate normality, so that the $\mathbf{x}_{kj}$ are independently drawn from $\mathrm{N}_p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$, we can test

$$H_0 : \boldsymbol{\mu}_1 = \ldots, = \boldsymbol{\mu}_K \text{ versus } H_1 : \boldsymbol{\mu}_k \neq \boldsymbol{\mu}_\ell \text{ for some } k \neq \ell.$$

This is the one-way <u>m</u>ultivariate <u>an</u>alysis <u>o</u>f <u>va</u>riance (MANOVA).

MANOVA Table

Source	SSP Matrix	df	?
Between Groups	$\mathbf{B}$	$K-1$	
Within Groups	$\mathbf{W}$	$n-K$	
Total	$\mathbf{T}$	$n-1$	

How do we compare the between-groups and within-groups SSP matrices, $\mathbf{B}$ and $\mathbf{W}$?

11.2 Likelihood Ratio Test (LRT) Construction

The LRT test statistic for testing

$$H_0 : \boldsymbol{\mu}_1 = \ldots, = \boldsymbol{\mu}_K \text{ versus } H_1 : \boldsymbol{\mu}_k \neq \boldsymbol{\mu}_\ell \text{ for some } k \neq \ell$$

is derived in Mardia, Kent and Bibby (1979) (see pp 138-139, and results from p 108). The resulting statistic is

$$\begin{aligned}\Lambda &= |\mathbf{W}|/|\mathbf{T}| \\ &= |\mathbf{W}|/|\mathbf{W}+\mathbf{B}| \\ &= |\mathbf{I}+\mathbf{W}^{-1}\mathbf{B}|^{-1}.\end{aligned}$$

Note: $|\mathbf{I}+\mathbf{W}^{-1}\mathbf{B}|^{-1} = \prod_{i=1}^{p}(1+\lambda_i)^{-1}$, where the λ_i are eigenvalues of $\mathbf{W}^{-1}\mathbf{B}$. The matrix $\mathbf{W}^{-1}\mathbf{B}$ is the multivariate generalization of the SS ratio from the univariate ANOVA.

Under H_0 (if $n \geq p+K$) Λ has a Wilk's Lambda distribution. That is

$$\Lambda = |\mathbf{I}+\mathbf{W}^{-1}\mathbf{B}|^{-1} = \prod_{i=1}^{p}(1+\lambda_i)^{-1} \sim \Lambda(p, n-K, K-1),$$

the Wilk's Lambda distribution with degrees of freedom p, $n-K$ and $K-1$. For certain special cases, this reduces to an F distribution (see Mardia, Kent and Bibby). For example, for three groups use

$$\frac{1-\sqrt{\Lambda}(p,m,2)}{\sqrt{\Lambda}(p,m,2)} \sim \frac{p}{m-p+1}\mathrm{F}(2p, 2(m-p+1)).$$

Also, since λ is a LRT statistic, we expect an asymptotic χ^2 distribution under H_0. Bartlett (1947) showed that asymptotically (as $r = n - K \to \infty$)

$$-\left\{r - \frac{p-s+1}{2}\right\}\ln\Lambda(p,r,s) \sim \chi^2(ps).$$

11.3 Union-Intersection Test (UIT) Construction

Let $y = \mathbf{a}^T\mathbf{x}$, where $\mathbf{x}$ is the general $(p \times 1)$ observation vector.

If H_0 is true, then so is $H_{0\mathbf{a}} : \mathbf{a}^T\boldsymbol{\mu}_1 = \ldots = \mathbf{a}^T\boldsymbol{\mu}_K$, and

$$H_0 = \bigcap_{\mathbf{a}} H_{0\mathbf{a}}$$

To test $H_{0\mathbf{a}}$, we use the usual *univariate* one-way ANOVA F statistic, $F_{\mathbf{a}}$. Hence, to test H_0 use

$$\max_{\mathbf{a}} F_{\mathbf{a}} = \max_{\mathbf{a}} \frac{\text{between-groups MS for } \mathbf{a}^T\mathbf{x}}{\text{within-groups MS for } \mathbf{a}^T\mathbf{x}} = \frac{\text{between-groups sample variance of } \mathbf{a}^T\mathbf{x}}{\text{within-groups sample variance of } \mathbf{a}^T\mathbf{x}}.$$

That is, to derive the UIT statistic, we need to find the linear compound $\ell^T\mathbf{x}$ for which the ratio of between-groups to within-groups variance (the $F-$statistic) is maximized. This is called the *first canonical variate*. In finding it we shall show that this maximum ratio is equal to

$$\frac{n-K}{K-1}\lambda_1$$

i.e. it is proportional to the *largest* eigenvalue of $\mathbf{W}^{-1}\mathbf{B}$.

- **Note**: $\mathbf{W}^{-1}\mathbf{B}$ is not (necessarily) symmetric, so that it is not immediately obvious that its eigenvalues are therefore real or that if real, they are positive. In addition, it is not true that the eigenvectors corresponding to the distinct eigenvalues of $\mathbf{W}^{-1}\mathbf{B}$ are orthogonal. (Real eigenvalues and real, orthogonal eigenvectors are properties of real, symmetric matrices). However, the matrix $\mathbf{W}^{-\frac{1}{2}}\mathbf{B}\mathbf{W}^{-\frac{1}{2}}$ has the same eigenvalues as $\mathbf{W}^{-1}\mathbf{B}$ and its eigenvectors $\boldsymbol{\gamma}$ give us the corresponding eigenvectors, $\mathbf{W}^{-\frac{1}{2}}\boldsymbol{\gamma}$ of $\mathbf{W}^{-1}\mathbf{B}$.

  Clearly, $\mathbf{W}^{-\frac{1}{2}}\mathbf{B}\mathbf{W}^{-\frac{1}{2}}$ is symmetric ($\mathbf{B}$ and $\mathbf{W}^{-\frac{1}{2}}$ are symmetric) and p.s.d. ($\mathbf{B}$ is p.s.d.).

We shall assume that the symmetric SSP matrix $\mathbf{W}$ is of full rank and therefore positive definite and invertible, so that $\mathbf{W}^{-\frac{1}{2}}$ exists.

If $y = \ell^T\mathbf{x}$, then

$$\text{Between-groups variance of } y = \ell^T\mathbf{B}\ell/(K-1)$$

and

$$\text{Within-groups variance of } y = \ell^T\mathbf{W}\ell/(n-K)$$

Thus it is required to maximize the ratio

$$R(\ell) = \frac{\ell^T\mathbf{B}\ell/(K-1)}{\ell^T\mathbf{W}\ell/(n-K)}$$

by a choice of ℓ.

Since $R(\ell)$ is invariant under changes of scale in ℓ, we can normalize by requiring that

$$\ell^T\mathbf{W}\ell \;=\; (n-K)$$

i.e. by requiring $\ell^T\left(\frac{1}{n-K}\mathbf{W}\right)\ell \;=\; \ell^T\mathbf{S}_U\ell = 1$, so that the within-groups sample variance of $\ell^T\mathbf{x}$ is equal to 1.

Thus it is required to maximize

$$R(\ell) \;=\; \ell^T\mathbf{B}\ell/(K-1) \tag{1}$$

subject to,

$$\ell^T\mathbf{W}\ell \;=\; (n-K). \tag{2}$$

Let $\mathbf{W}^{\frac{1}{2}}$ be the symmetric square root of $\mathbf{W}$ and let

$$\mathbf{u} \;=\; \frac{1}{\sqrt{n-K}}\mathbf{W}^{\frac{1}{2}}\ell \tag{3}$$

so that,

$$\ell \;=\; \sqrt{n-K}\,\mathbf{W}^{-\frac{1}{2}}\mathbf{u} \tag{4}$$

and maximizing (1) subject to (2) is equivalent to maximizing

$$\frac{n-K}{K-1}\mathbf{u}^T\mathbf{W}^{-\frac{1}{2}}\mathbf{B}\mathbf{W}^{-\frac{1}{2}}\mathbf{u} \tag{5}$$

subject to

$$(n-K)\mathbf{u}^T\mathbf{W}^{-\frac{1}{2}}\mathbf{W}\mathbf{W}^{-\frac{1}{2}}\mathbf{u} \;=\; n-K,$$

that is subject to

$$\mathbf{u}^T\mathbf{u} \;=\; 1. \tag{6}$$

Forming the Lagrangian,

$$L(\mathbf{u}, \alpha) \;=\; \frac{n-K}{K-1}\mathbf{u}^T\mathbf{W}^{-\frac{1}{2}}\mathbf{B}\mathbf{W}^{-\frac{1}{2}}\mathbf{u} - \alpha(\mathbf{u}^T\mathbf{u} - 1)$$

we obtain,

$$\frac{\partial L}{\partial \mathbf{u}} \;=\; 2\,\frac{n-K}{K-1}\mathbf{W}^{-\frac{1}{2}}\mathbf{B}\mathbf{W}^{-\frac{1}{2}}\mathbf{u} \;-\; 2\alpha\mathbf{u}$$

and equating to $\mathbf{0}$ gives

$$\left(\frac{n-K}{K-1}\mathbf{W}^{-\frac{1}{2}}\mathbf{B}\mathbf{W}^{-\frac{1}{2}} \;-\; \alpha\mathbf{I}\right)\mathbf{u} \;=\; \mathbf{0} \tag{7}$$

with

$$\mathbf{u}^T\mathbf{u} \;=\; 1. \tag{8}$$

Consider the symmetric p.s.d. matrix $\mathbf{W}^{-\frac{1}{2}}\mathbf{B}\mathbf{W}^{-\frac{1}{2}}$. It has p non-negative eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$, where some of these may be zero ($\text{rank}(\mathbf{W}^{-\frac{1}{2}}\mathbf{B}\mathbf{W}^{-\frac{1}{2}}) \leq \min(p, K-1)$).

Let $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p$ be the corresponding set of orthogonal eigenvectors. Then the eigenvalues of $\dfrac{n-K}{K-1}\mathbf{W}^{-\frac{1}{2}}\mathbf{B}\mathbf{W}^{-\frac{1}{2}}$ are

$$\frac{n-K}{K-1}\lambda_1 \;\geq\; \frac{n-K}{K-1}\lambda_2 \;\geq\; \ldots \;\geq\; \frac{n-K}{K-1}\lambda_p \;\geq\; 0$$

with the same eigenvectors $\boldsymbol{\gamma}_1, \ldots \boldsymbol{\gamma}_p$. From (7) and (8) it is clear that α is an eigenvalue of $\dfrac{n-K}{K-1}\mathbf{W}^{-\frac{1}{2}}\mathbf{B}\mathbf{W}^{-\frac{1}{2}}$ and $\mathbf{u}$ is the corresponding standardized eigenvector.

Pre-multiplying (7) by $\mathbf{u}^T$ gives

$$\frac{n-K}{K-1}\mathbf{u}^T\mathbf{W}^{-\frac{1}{2}}\mathbf{B}\mathbf{W}^{-\frac{1}{2}}\mathbf{u} = \alpha$$

and therefore we need to choose $\alpha = \dfrac{n-K}{K-1}\lambda_1$, the largest eigenvalue of $\dfrac{n-K}{K-1}\mathbf{W}^{-\frac{1}{2}}\mathbf{B}\mathbf{W}^{-\frac{1}{2}}$ and $\mathbf{u} = \boldsymbol{\gamma}_1$, the corresponding eigenvector. Hence, from (4), $\ell = \ell_1$ where $\ell_1 = \sqrt{n-K}\mathbf{W}^{-\frac{1}{2}}\boldsymbol{\gamma}_1$ is the eigenvector of $\mathbf{W}^{-1}\mathbf{B}$ corresponding to its largest eigenvalue λ_1 and scaled so that

$$\ell_1^T\left(\frac{1}{n-K}\mathbf{W}\right)\ell_1 \;=\; \boldsymbol{\gamma}_1^T\mathbf{W}^{-\frac{1}{2}}\mathbf{W}\mathbf{W}^{-\frac{1}{2}}\boldsymbol{\gamma}_1 \;=\; \boldsymbol{\gamma}_1^T\boldsymbol{\gamma}_1 \;=\; 1.$$

The maximized ratio of between to within-groups sample variance is given by

$$\frac{n-K}{K-1}\lambda_1.$$

Thus the UIT statistic is $\frac{n-K}{K-1}\lambda_1$ and can be tested for significance from charts of the upper percentage points of the *largest-root distribution*. If the UIT statistic is adopted - and it is significant - we have immediately the particular combination of original variables giving $F_{\mathbf{a}}^{\max}$. This is the first *canonical variate*.

11.4 The Canonical Variate Space

As with principal components, we can extract further canonical variates. We define the r'th canonical variate as that compound, $y_r = \ell^T\mathbf{x}$, which has the maximum *between-groups* sample variance, subject to the *within-groups* sample variance of y_r being 1, and subject to the *within-groups* sample correlations of y_r with y_1, y_2, …, y_{r-1} being zero. Consider

$$\ell_r = \sqrt{n-K}\mathbf{W}^{-\frac{1}{2}}\boldsymbol{\gamma}_r, \;\; r = 1, \ldots, p. \tag{9}$$

These are the eigenvectors of $\mathbf{W}^{-1}\mathbf{B}$ corresponding to the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$. Let $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p)$ and $\mathbf{L} = (\ell_1, \ldots, \ell_p)$. Clearly $\boldsymbol{\Gamma}^T\boldsymbol{\Gamma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T = \mathbf{I}$, since the $\boldsymbol{\gamma}_r$ are mutually orthogonal. Then

$$\mathbf{L} = \sqrt{n-K}\mathbf{W}^{-\frac{1}{2}}\boldsymbol{\Gamma}. \tag{10}$$

Transforming from the general vector $\mathbf{x}$ of original variables to the general vector $\mathbf{y}$ of canonical variables, we have

$$\mathbf{y} = \mathbf{L}^T\mathbf{x} \tag{11}$$

Then the within-groups sample covariance matrix for $\mathbf{y}$ is given by

$$
\begin{aligned}
\frac{1}{n-K}\mathbf{W}_{\mathbf{y}} &= \frac{1}{n-K}\mathbf{L}^T\mathbf{W}\mathbf{L} \\
&= \frac{1}{n-K}(n-K)\mathbf{\Gamma}^T\mathbf{W}^{-\frac{1}{2}}\mathbf{W}\mathbf{W}^{-\frac{1}{2}}\mathbf{\Gamma} \\
&= \mathbf{I}
\end{aligned}
$$

That is, the within-groups sample covariance matrix of the canonical variables is $\mathbf{I}$, corresponding to the scaling of eigenvectors $\ell_1, \ldots, \ell_p$ such that

$$
\frac{1}{n-K}\mathbf{L}^T\mathbf{W}\mathbf{L} = \mathbf{I}\,. \tag{12}
$$

Thus the canonical variates $\mathbf{y}$ are uncorrelated within groups and have within-group variances 1.

Now consider the between-groups sample covariance matrix for $\mathbf{y}$. This is given by

$$
\begin{aligned}
\frac{1}{K-1}\mathbf{B}_{\mathbf{y}} &= \frac{1}{K-1}\mathbf{L}^T\mathbf{B}\mathbf{L} \\
&= \frac{n-K}{K-1}\mathbf{\Gamma}^T\mathbf{W}^{-\frac{1}{2}}\mathbf{B}\mathbf{W}^{-\frac{1}{2}}\mathbf{\Gamma} \\
&= \frac{n-K}{K-1}\mathbf{\Gamma}^T\mathbf{\Gamma}\mathrm{diag}(\lambda_r) \\
&= \mathrm{diag}\left(\frac{n-K}{K-1}\lambda_r\right).
\end{aligned}
$$

Thus the y_r are also uncorrelated *between* groups, and the between-group variances are $\frac{n-K}{K-1}\lambda_r$. So the rth canonical variate, $y_r = \ell_r^T\mathbf{x}$, $r = 1, \ldots, p$ has the maximum between-groups variance such that the within-groups variance is 1 and such that the within-groups correlations with $y_1, \ldots, y_{r-1}$ are zero. This maximum is $\frac{n-K}{K-1}\lambda_r$.

It is not difficult to show that $(\lambda_1 + \ldots + \lambda_r)/(\lambda_1 + \ldots + \lambda_p)$ represents the proportion of between-group variation accounted for by the first r canonical variates.

Exercise: As with y_1, derive y_2 directly by maximization, adding the constraint that y_2 must be uncorrelated within-groups with y_1.

In general we use the canonical variables to plot the group centroids (or means) with respect to the new canonical axes.
Let

$$\overline{\mathbf{y}}_k = \mathbf{L}^T \overline{\mathbf{x}}_k, \quad k = 1, \ldots, K \tag{13}$$

Then, $\overline{\mathbf{y}}_k$ is the $(p \times 1)$ vector of canonical variate means for sample k.

What do distances in canonical variate space represent?

To examine this we need to look at the $(p \times p)$ matrix $\mathbf{L}\mathbf{L}^T$.

$$\begin{aligned} \mathbf{L}\mathbf{L}^T &= (n-K)\mathbf{W}^{-\frac{1}{2}}\mathbf{\Gamma}\mathbf{\Gamma}^T\mathbf{W}^{-\frac{1}{2}} \\ &= (n-K)\mathbf{W}^{-1} \\ &= \left(\frac{1}{n-K}\mathbf{W}\right)^{-1} = \mathbf{S}_U^{-1}. \end{aligned} \tag{14}$$

Now, suppose we consider a pair of samples, k_1 and k_2. Then the squared Euclidean distance between $\overline{\mathbf{y}}_{k_1}$ and $\overline{\mathbf{y}}_{k_2}$ is given by

$$\begin{aligned} (\overline{\mathbf{y}}_{k_1} - \overline{\mathbf{y}}_{k_2})^T(\overline{\mathbf{y}}_{k_1} - \overline{\mathbf{y}}_{k_2}) &= (\overline{\mathbf{x}}_{k_1} - \overline{\mathbf{x}}_{k_2})^T\mathbf{L}\mathbf{L}^T(\overline{\mathbf{x}}_{k_1} - \overline{\mathbf{x}}_{k_2}) \\ &= (\overline{\mathbf{x}}_{k_1} - \overline{\mathbf{x}}_{k_2})^T\mathbf{S}_U^{-1}(\overline{\mathbf{x}}_{k_1} - \overline{\mathbf{x}}_{k_2}) \\ &= D^2(\overline{\mathbf{x}}_{k_1}, \overline{\mathbf{x}}_{k_2}) \end{aligned}$$

the squared Mahalanobis distance between the two samples.

Confidence Circles

Assume now that $\mathbf{x}_{kj} \sim \mathrm{N}_p(\boldsymbol{\mu}_k, \mathbf{\Sigma})$, $k = 1, \ldots, K$, $j = 1, \ldots, n_k$. Transforming to canonical variates we have

$$\underset{(s\times 1)}{\mathbf{y}_{kj}} = \underset{(s\times p)}{\mathbf{L}^T} \underset{(p\times 1)}{\mathbf{x}_{kj}},$$

where $\mathbf{L} = (\ell_1, \ldots, \ell_s)$ and the $\ell_r = \sqrt{n-K}\mathbf{W}^{-\frac{1}{2}}\boldsymbol{\gamma}_r$, $r = 1, \ldots, s$, are the eigenvectors of $\mathbf{W}^{-1}\mathbf{B}$ corresponding to the non-zero eigenvalues. The dimensionality $s = \mathrm{rank}(\mathbf{W}^{-1}\mathbf{B}) \leq \min(p, K-1))$.

If the degrees of freedom $(n-K)$ are large, so that

$$\mathbf{S}_U = \frac{1}{n-K}\mathbf{W} \approx \mathbf{\Sigma}$$

then

$$\mathbf{L}^T\boldsymbol{\Sigma}\mathbf{L} \approx (n-K)^{-1}\mathbf{L}^T\mathbf{W}\mathbf{L} = \frac{1}{n-K}\mathbf{W}_{\mathbf{y}} = \underset{(s\times s)}{\mathbf{I}}$$

from (12). It follows that

$$\mathbf{y}_{kj} \overset{\text{approx}}{\sim} \mathrm{N}_s(\mathbf{L}^T\boldsymbol{\mu}_k, \mathbf{I}) \tag{15}$$

and

$$\overline{\mathbf{y}}_k \overset{\text{approx}}{\sim} \mathrm{N}_s(\mathbf{L}^T\boldsymbol{\mu}_k, n_k^{-1}\mathbf{I}). \tag{16}$$

Then, writing $\boldsymbol{\mu}_k^* = \mathbf{L}^T\boldsymbol{\mu}_k$, we have

$$\sqrt{n_k}(\overline{\mathbf{y}}_k - \boldsymbol{\mu}_k^*) \overset{\text{approx}}{\sim} \mathrm{N}_s(\mathbf{0}, \mathbf{I}) \tag{17}$$

so that

$$n_k(\overline{\mathbf{y}}_k - \boldsymbol{\mu}_k^*)^T(\overline{\mathbf{y}}_k - \boldsymbol{\mu}_k^*) \overset{\text{approx}}{\sim} \chi^2(s). \tag{18}$$

It follows that a $100(1-\alpha)\%$ confidence region for the true mean vector $\boldsymbol{\mu}_k^*$ (relative to canonical axes) is given by the interior of the hypersphere of radius $\sqrt{\chi^2_{s;\alpha}/n_k}$ centred at $\overline{\mathbf{y}}_k$. Thus if we plot the canonical variate mean vectors with respect to the first two canonical axes, we can calculate (and impose on the plot) approximate circular confidence regions about the $\overline{\mathbf{y}}_k$ for the positions of the $\boldsymbol{\mu}_k^*$ of radius $\sqrt{\chi^2_{2;\alpha}/n_k}$. (Note: $\chi^2_{0.95}(2) = 5.99$, so that 95% confidence circles for the $\boldsymbol{\mu}_k^*$ have radii $(2.45/\sqrt{n_k})$ in two dimensions). This will give some idea of the 'significance' of the distances between samples that we observe.

Notes:

1. Canonical variates may be used purely as an exploratory, hypothesis generating tool to investigate the differences between samples. (If we are able to assume multivariate normality, then we can carry out a MANOVA test and if $H_0 : \boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_K$ is rejected, then we have some evidence that there is something to investigate).

2. Confidence circles are fairly robust to non-normality. (Central Limit Theorem.)

3. Canonical variates are scale invariant, unlike principal components. (They are unaffected by a change in scale of the original variables).

4. To assess the importance of the original variables in discriminating between samples, we can look at the coefficients of these variables in the canonical variates. Note that the coefficients on different variables are not comparable since they depend on the scale on which the original variables were measured. However, because of Note 3, we can rescale the original variables to have within-groups sample variance 1, and the corresponding canonical variate coefficients are then comparable.

Shrews Example (See Mardia, Kent and Bibby, p345)

White-toothed shrews of the genus Crocidura occur in the Channel and Scilly Isles of the British Isles and in the French Mainland. Data consist of p=10 quantitative measurements on each of n=399 skulls obtained from K=10 localities.

1	Tresco	T	Scilly Isles
2	Bryher	B	
3	St Agnes	Ag	
4	St Martin's	Mn	
5	St Mary's	My	
6	Sark	S	Channel Islands
7	Jersey	J	
8	Alderney	Al	
9	Guernsey	G	
10	Cap Gris Nez	C	French Mainland

The sample sizes for the data from the localities are 144, 16, 12, 7, 90, 25, 6, 26, 53 and 20 respectively.

The canonical variates were calculated and the means of the 10 groups of shrews were plotted with respect to the first two canonical variates axes (out of 9). The first two axes accounted for 93.7% of the between-sample variation. The confidence circles are 99% confidence regions for the true canonical variate means with respect to the first two canonical axes.

Figure 1. Shrews Data: Plot of first two canonical variates

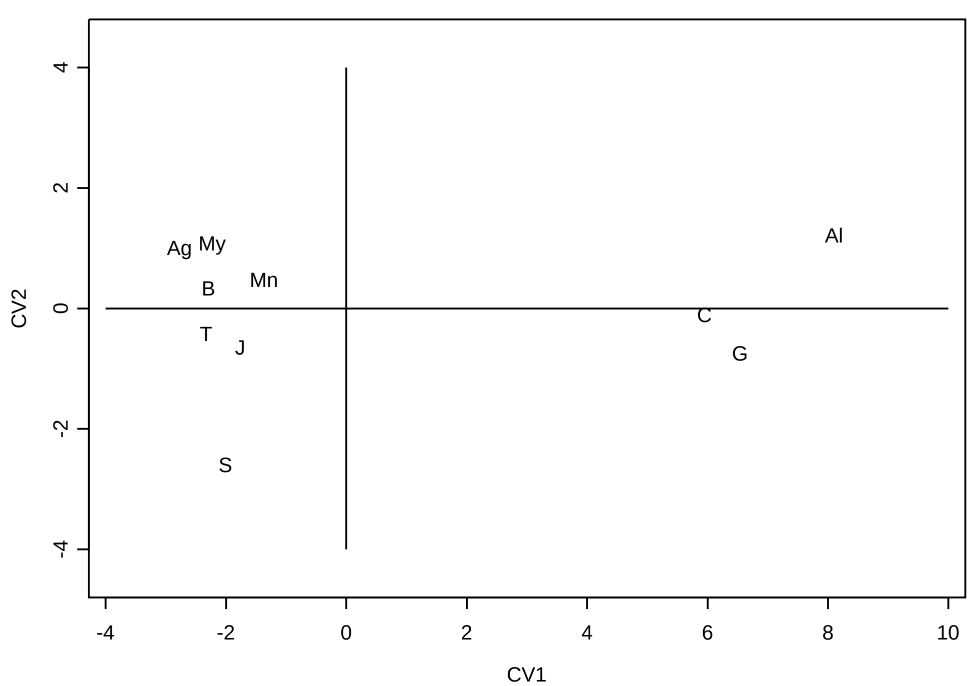

Figure 2. Shrews Data: CV Plot with 99% confidence circles