Chapter 10. Canonical Analysis

Canonical analysis aims at finding the relation between two sets of variables, by singling out pairs of random variables with highest hierarchical correlations. In every pair, one is a linear combination of one set of variables and the other is that of the other set of variables. The following example is illustrates a typical scenario.

Example 10.1. ANALYSIS OF CHICKEN-BONE DATA There are totally $n = 276$ observations.

$$\text{Head: } \begin{cases} x_1^{(1)} = \text{skull length} \\ x_2^{(1)} = \text{skull breadth} \end{cases} \qquad \text{leg: } \begin{cases} x_1^{(2)} = \text{femur length} \\ x_2^{(2)} = \text{tibia breadth} \end{cases}$$

We have for this example ignored the other two variables about the wing. The sample correlation matrix is then

$$\mathbf{R} \quad = \quad \begin{array}{c} \\ x_1^{(1)} \\ x_2^{(1)} \\ x_1^{(2)} \\ x_2^{(2)} \end{array} \begin{array}{c} \begin{array}{cccc} x_1^{(1)} & x_2^{(1)} & x_1^{(2)} & x_2^{(2)} \end{array} \\ \begin{pmatrix} 1.000 & .505 & .569 & .602 \\ .505 & 1.000 & .422 & .467 \\ .569 & .422 & 1.000 & .926 \\ .602 & .467 & .926 & 1.000 \end{pmatrix} \end{array}$$

10.1 General description.

Suppose we have two random vectors in concern: $X^{(1)}$ and $X^{(2)}$ of p and q dimension respectively with $p \leq q$. Assume

$$\begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \sim MN\Big(0, \begin{pmatrix} \Sigma_1 1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\Big).$$

We wish to investigate the relations between this two vectors, one containing p variables and the other q variables. The general method is to find hierarchical pairs of variables, each pair consists of a linear combination of one set of variables and a linear combination of the other set of variables. The first pair explains most of the relation between the two sets of variables; the second pair explains most of the relation between the two sets of variables not explained by the first pair; the third pair explains most of the relation not explained by the first two pairs, and so on. These pairs of variables, denoted as

$$(U_1, V_1), ..., (U_p, V_p),$$

are called *canonical variates*, which are required to be orthogonal to each other, except for within pairs, and each with unit variance. Namely

$$\text{var}(U) = I_p, \;\; \text{var}(V) = I_q, \;\; \text{cov}(U_i, V_j) = 0 \text{ for } i \neq j,$$

where

$$U = \begin{pmatrix} U_1 \\ \vdots \\ U_p \end{pmatrix}, \qquad V = \begin{pmatrix} V_1 \\ \vdots \\ V_q \end{pmatrix}.$$

The correlation between canonical variates within the pairs, $\rho_1, ..., \rho_p$, are called *canonical correlation* where

$$\rho_i = Corr(U_i, V_i), \qquad i = 1, ..., p, \qquad \text{with } \rho_1 \geq .... \geq \rho_p \geq 0.$$

The canonical variates are linear combinations of the first or the second set of the variables:

$$U = \mathbf{A}_{p\times p} X^{(1)}, \qquad V = \mathbf{B}_{q\times q} X^{(1)}.$$

Then, the matrices **A** and **B**, playing the role of linear combinations, are called *canonical coefficients*.

In terms of explaining the relation between the two sets of variables, clearly the first pair of canonical variates is the most important, the second second important, and so on. Very often, the first few pairs already explain most of the relations. Then we wish to retain only the first few pairs. When this is the case, we shall use *matrices of errors* to evaluate whether the retained pairs of canonical variates indeed explain sufficient amount of the relations. Moreover, we also use *total variances explained* to evaluate how representative the retained canonical variates are for their set of variables.

10.2 A theorem about population canonical variates.

Theorem 10.1 *Suppose two random vectors $X^{(1)}$ and $X^{(2)}$ of p and q dimensions respectively,*

$$\mathrm{var}\begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} = \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Let $(\rho_1^2, \mathbf{e}_1), ..., (\rho_p^2, \mathbf{e}_p)$ be eigenvalue-eigenvector pairs of $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}$, with $\rho_1 \ge \rho_2 \ge ... \ge \rho_p \ge 0$. Then,

(i). $(\rho_1^2, \mathbf{f}_1), ..., (\rho_p^2, \mathbf{f}_p), (0, \mathbf{f}_{p+1}), ..., (0, \mathbf{f}_q)$ are eigenvalue-eigenvector pairs of $\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2}$, where

$$\mathbf{f}_k = \frac{1}{\rho_k}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}\mathbf{e}_k \qquad k = 1, ..., p.$$

(ii). Let

$$\mathbf{A}_{p\times p} = \begin{pmatrix} \mathbf{e}_1' \\ \vdots \\ \mathbf{e}_p' \end{pmatrix}\Sigma_{11}^{-1/2}, \;\; \mathbf{B}_{q\times q} = \begin{pmatrix} \mathbf{f}_1' \\ \vdots \\ \mathbf{f}_q' \end{pmatrix}\Sigma_{22}^{-1/2}, \;\; U = \begin{pmatrix} U_1 \\ \vdots \\ U_p \end{pmatrix} = \mathbf{A}X^{(1)} \text{ and } V = \begin{pmatrix} V_1 \\ \vdots \\ V_q \end{pmatrix} = \mathbf{B}X^{(2)}$$

Then, $U_k = \mathbf{e}_k'\Sigma_{11}^{-1/2}X^{(1)}$, $V_k = \mathbf{f}_k'\Sigma_{22}^{-1/2}X^{(2)}$, and $(U_1, V_1), ..., (U_p, V_p)$ are p pairs of canonical variates with canonical correlation $\rho_1, ..., \rho_p$.

A partial proof of this theorem is given below.

(i). Write, for $k = 1, ..., p$,

$$\begin{aligned} & \Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2}\mathbf{f}_k \\ = \;& \frac{1}{\rho_k}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}\mathbf{e}_k \\ = \;& \frac{1}{\rho_k}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}[\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}\mathbf{e}_k] \\ = \;& \frac{\rho_k^2}{\rho_k}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}\mathbf{e}_k \\ = \;& \rho_k^2\mathbf{f}_k \end{aligned}$$

(ii). Write

$$\begin{aligned} \mathrm{Cov}(U, V) &= \begin{pmatrix} \mathbf{e}_1' \\ \vdots \\ \mathbf{e}_p' \end{pmatrix}\Sigma_{11}^{-1/2}\mathrm{Cov}(X^{(1)}, X^{(2)})\Sigma_{22}^{-1/2}(\mathbf{f}_1\vdots\cdots\vdots\mathbf{f}_q) \\ &= \begin{pmatrix} \mathbf{e}_1' \\ \vdots \\ \mathbf{e}_p' \end{pmatrix}\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}(\mathbf{f}_1\vdots\cdots\vdots\mathbf{f}_q) \end{aligned}$$

$$= \begin{pmatrix} \rho_1 \mathbf{f}_1' \\ \vdots \\ \rho_p \mathbf{f}_p' \end{pmatrix} (\mathbf{f}_1 \vdots \cdots \vdots \mathbf{f}_q)$$

$$= \begin{pmatrix} \rho_1 & & & 0 & \cdots & 0 \\ & \ddots & & \vdots & \vdots & \vdots \\ & & \rho_p & 0 & \cdots & 0 \end{pmatrix}_{p\times q}.$$

Remark. We did not (and will not) prove

$$\begin{aligned} Corr(U_1, V_1) &= \max\{Corr(a'X^{(1)}, b'X^{(2)}) : a \in R^p, b \in R^q\} \\ Corr(U_2, V_2) &= \max\{Corr(a'X^{(1)}, b'X^{(2)}) : a'X^{(1)} \perp U_1, b'X^{(2)} \perp V_1, a \in R^p, b \in R^q\} \\ & \quad .... \end{aligned}$$

These are the mathematical statements that are interpreted as the k-th pair, (U_k, V_k), explains most of the relations not explained by the first $k-1$ pairs of canonical variates.

10.3. The sample analogues.

Let

$$\mathbf{S}_{(p+q)\times(p+q)} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}$$

be the sample variance matrix analogous to the population variance matrix

$$\Sigma_{(p+q)\times(p+q)} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Note that $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$ are sample variances of the first and the second set of variables. Let $(\hat\rho_1^2, \hat{\mathbf{e}}_1), ..., (\hat\rho_p^2, \hat{\mathbf{e}}_p)$ be eigenvalue-eigenvector pairs of $\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}$, with $\hat\rho_1 \geq \hat\rho_2 \geq ... \geq \hat\rho_p \geq 0$, and $(\hat\rho_1^2, \hat{\mathbf{f}}_1), ..., (\hat\rho_p^2, \hat{\mathbf{f}}_p), (0, \hat{\mathbf{f}}_{p+1}), ..., (0, \hat{\mathbf{f}}_q)$ are eigenvalue-eigenvector pairs of $\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2}$, where

$$\hat{\mathbf{f}}_k = \frac{1}{\hat\rho_k}\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}\hat{\mathbf{e}}_k \qquad k = 1, ..., p.$$

Let

$$\hat{\mathbf{A}}_{p\times p} = \begin{pmatrix} \hat{\mathbf{e}}_1' \\ \vdots \\ \hat{\mathbf{e}}_p' \end{pmatrix} \mathbf{S}_{11}^{-1/2} \qquad \hat{\mathbf{B}}_{q\times q} = \begin{pmatrix} \hat{\mathbf{f}}_1' \\ \vdots \\ \hat{\mathbf{f}}_q' \end{pmatrix} \mathbf{S}_{22}^{-1/2}$$

$$\hat{\mathbf{U}}' = \begin{pmatrix} \hat U_1 \\ \vdots \\ \hat U_p \end{pmatrix} = \hat{\mathbf{A}} \begin{pmatrix} \mathbf{x}_1^{(1)'} \\ \vdots \\ \mathbf{x}_p^{(1)'} \end{pmatrix} \qquad \text{and} \qquad \hat{\mathbf{V}}' = \begin{pmatrix} \hat V_1 \\ \vdots \\ \hat V_q \end{pmatrix} = \hat{\mathbf{B}} \begin{pmatrix} \mathbf{x}_1^{(2)'} \\ \vdots \\ \mathbf{x}_q^{(2)'} \end{pmatrix}$$

where, recall that $\mathbf{x}_k^{(1)}$ $\mathbf{x}_k^{(2)}$ is a n-vector representing n observations of the k-th variable in the first (second) set of variables. Then,

$$\begin{aligned} \hat U_k &= (\mathbf{x}_1^{(1)} \vdots \cdots \vdots \mathbf{x}_p^{(1)})\mathbf{S}_{11}^{-1/2}\hat{\mathbf{e}}_k, \qquad k = 1, ..., p; \\ \hat V_k &= (\mathbf{x}_1^{(2)} \vdots \cdots \vdots \mathbf{x}_q^{(2)})\mathbf{S}_{22}^{-1/2}\hat{\mathbf{f}}_k, , \qquad k = 1, ..., q. \end{aligned}$$

And $(\hat U_1, \hat V_1), ..., (\hat U_p, \hat V_p)$ are p pairs of canonical variates with canonical correlation. Note that $\hat U_i$ and $\hat V_i$ are n-vectors. Write

$$\begin{aligned} \hat{\mathbf{A}}_{p\times p}^{-1} &= \mathbf{S}_{11}^{1/2}(\hat{\mathbf{e}}_1 \vdots \cdots \vdots \hat{\mathbf{e}}_p) \equiv (\hat{\mathbf{a}}^{(1)} \vdots \cdots \vdots \hat{\mathbf{a}}^{(p)}) \\ \hat{\mathbf{B}}_{q\times q}^{-1} &= \mathbf{S}_{22}^{1/2}(\hat{\mathbf{f}}_1 \vdots \cdots \vdots \hat{\mathbf{f}}_q) \equiv (\hat{\mathbf{b}}^{(1)} \vdots \cdots \vdots \hat{\mathbf{b}}^{(q)}) \end{aligned}$$

where, for $i = 1, ..., p$ and $j = 1, ..., q$.

$$\hat{\mathbf{a}}^{(i)}_{p\times 1} = \mathbf{S}_{11}^{1/2}\hat{\mathbf{e}}_i \quad \text{and} \quad \hat{\mathbf{b}}^{(j)}_{q\times 1} = \mathbf{S}_{22}^{1/2}\hat{\mathbf{f}}_j.$$

Then,

$$\mathbf{S}_{11} = \hat{\mathbf{A}}^{-1}(\hat{\mathbf{A}}^{-1})' \quad \mathbf{S}_{22} = \hat{\mathbf{B}}^{-1}(\hat{\mathbf{B}}^{-1})' \text{ and } \mathbf{S}_{12} = \sum_{i=1}^{p} \hat{\rho}_i \hat{\mathbf{a}}^{(i)}(\hat{\mathbf{b}}^{(i)})',$$

since

$$\begin{aligned}
& \sum_{k=1}^{p} \hat{\rho}_k \hat{\mathbf{a}}^{(k)}(\hat{\mathbf{b}}^{(k)})' \\
= \; & (\hat{\mathbf{a}}^{(1)} \vdots \cdots \vdots \hat{\mathbf{a}}^{(p)}) \begin{pmatrix} \hat{\rho}_1(\hat{\mathbf{b}}^{(1)})' \\ \cdots \\ \hat{\rho}_p(\hat{\mathbf{b}}^{(p)})' \end{pmatrix} \\
= \; & \mathbf{S}_{11}^{1/2}(\hat{\mathbf{e}}_1 \vdots \cdots \vdots \hat{\mathbf{e}}_p) \begin{pmatrix} \hat{\rho}_1 \hat{\mathbf{f}}_1' \\ \cdots \\ \hat{\rho}_p \hat{\mathbf{f}}_p' \end{pmatrix} \mathbf{S}_{22}^{1/2} \\
= \; & \mathbf{S}_{11}^{1/2}(\hat{\mathbf{e}}_1 \vdots \cdots \vdots \hat{\mathbf{e}}_p) \begin{pmatrix} \hat{\rho}_1/\hat{\rho}_1 \hat{\mathbf{e}}_1' \\ \cdots \\ \hat{\rho}_p/\hat{\rho}_2 \hat{\mathbf{e}}_p' \end{pmatrix} \mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2}\mathbf{S}_{22}^{1/2} \\
= \; & \mathbf{S}_{11}^{1/2} I_p \mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2}\mathbf{S}_{22}^{1/2} \\
= \; & \mathbf{S}_{12}.
\end{aligned}$$

Suppose now we wish to retain the first r pairs of the canonical variates: $(U_1, V_1), ..., (U_r, V_r)$ with $r \leq p$. Then the matrices of errors are defined as

$$\begin{aligned}
\mathbf{S}_{11} - \sum_{i=1}^{r} \hat{\mathbf{a}}^{(i)}(\hat{\mathbf{a}}^{(i)})' &= \sum_{i=r+1}^{p} \hat{\mathbf{a}}^{(i)}(\hat{\mathbf{a}}^{(i)})', \\
\mathbf{S}_{22} - \sum_{i=1}^{r} \hat{\mathbf{b}}^{(i)}(\hat{\mathbf{b}}^{(i)})' &= \sum_{i=r+1}^{q} \hat{\mathbf{b}}^{(i)}(\hat{\mathbf{b}}^{(i)})', \\
\mathbf{S}_{12} - \sum_{i=1}^{r} \hat{\rho}_i \hat{\mathbf{a}}^{(i)}(\hat{\mathbf{b}}^{(i)})' &= \sum_{i=r+1}^{p} \hat{\rho}_i \hat{\mathbf{a}}^{(i)}(\hat{\mathbf{b}}^{(i)})'.
\end{aligned}$$

The first two matrices of errors are used to evaluate whether the variations inside each set of variables not explained by the retained r canonical variates are significant or not. The last is used to evaluate whether the remaining relations between the two sets of variables not explained by the r pairs of canonical variates are significant or not. If the matrices of errors have entries close to 0, they are considered as evidence of the sufficiency of the retained variates. If not, probably the number of retained canonical variates are too few.

When we retain r pairs of canonical variates, the proportion of (standardized) variance in the first set of variables explained by the retained r canonical variates is

$$R^2_{1,r} \equiv \frac{trace(\sum_{i=1}^{r} \hat{\mathbf{a}}^{(i)}(\hat{\mathbf{a}}^{(i)})')}{p},$$

and, likewise, the proportion of (standardize) variance in the first set of variables explained by the retained r canonical variates is

$$R^2_{2,r} \equiv \frac{trace(\sum_{i=1}^{r} \hat{\mathbf{b}}^{(i)}(\hat{\mathbf{b}}^{(i)})')}{q}.$$

Here $R^2_{1,r}$ ($R^2_{2,r}$) is a measurement of how representative the retained r canonical variates are for the first (second) set of variables. The higher the more representative.

10.4 An example.

Example 10.1 (Continued) We demonstrate the canonical analysis using the example of the chicken bone data. From the correlation matrix $\mathbf{R}$, we have

$$\mathbf{S}_{11} = \begin{pmatrix} 1.000 & 0.505 \\ 0.505 & 1.000 \end{pmatrix} \quad \mathbf{S}_{22} = \begin{pmatrix} 1.000 & 0.926 \\ 0.926 & 1.000 \end{pmatrix} \quad \mathbf{S}_{12} = \begin{pmatrix} 0.569 & 0.602 \\ 0.422 & 0.467 \end{pmatrix} = \mathbf{S}'_{21}$$

Here one might choose to write $\mathbf{S}_{ij}$ as $\mathbf{R}_{ij}$ since only correlation matrices rather than variance matrices are involved. All variables are clearly positively related with each other. The first set of variables represent head size and the second leg size. Here $p = q = 2$. With the matrix decomposition of $\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}$, one finds

(1). The canonical coefficients are

$$\hat{\mathbf{A}} = \begin{pmatrix} 0.781 & 0.345 \\ -0.856 & 1.106 \end{pmatrix} \quad \text{and} \quad \hat{\mathbf{B}} = \begin{pmatrix} 0.060 & 0.944 \\ -2.648 & 2.475 \end{pmatrix}$$

Then,

$$\hat{\mathbf{A}}^{-1} = \begin{pmatrix} 0.955 & -0.297 \\ 0.739 & 0.674 \end{pmatrix} \quad \text{and} \quad \hat{\mathbf{B}}^{-1} = \begin{pmatrix} 0.934 & -0.356 \\ 0.997 & 0.023 \end{pmatrix}$$

Hence,

$$\hat{\mathbf{a}}^{(1)} = \begin{pmatrix} 0.955 \\ 0.739 \end{pmatrix} \quad \hat{\mathbf{a}}^{(2)} = \begin{pmatrix} -0.297 \\ 0.674 \end{pmatrix} \quad \hat{\mathbf{b}}^{(1)} = \begin{pmatrix} 0.934 \\ 0.997 \end{pmatrix} \quad \hat{\mathbf{b}}^{(2)} = \begin{pmatrix} -0.356 \\ 0.023 \end{pmatrix}$$

(2). The canonical correlations are

$$\hat{\rho}_1 = 0.631 \quad \text{and} \quad \hat{\rho}_2 = 0.057.$$

The first is far larger than the second.

(3). The canonical variates, expressed as linear combination of the first or the second set of variables, are

$$\begin{aligned} \hat{U}_1 &= 0.781x_1^{(1)} + 0.345x_2^{(1)} & \hat{U}_2 &= -0.856x_1^{(1)} + 1.106x_2^{(1)} \\ \hat{V}_1 &= 0.060x_1^{(2)} + 0.944x_2^{(2)} & \hat{V}_2 &= -2.648x_1^{(2)} + 2.475x_2^{(2)} \end{aligned}$$

$\hat{U}_1$ may be understood as a weighted average of skull length and width, with more weight on length. $\hat{V}_1$ is basically tibia length, but we should also keep in mind that femur length is highly correlated with tibia length. $\hat{U}_2$ ($\hat{V}_2$) may be interpreted as contrasts between skull length and width (femur and tibia length).

(4). Suppose we only retain the first pair of the canonical variates. Then matrices of errors are

$$S_{11} - \hat{\mathbf{a}}^{(1)}(\hat{\mathbf{a}}^{(1)})' = \hat{\mathbf{a}}^{(2)}(\hat{\mathbf{a}}^{(2)})' = \begin{pmatrix} -0.297 \\ 0.674 \end{pmatrix} (-0.297 \;\; 0.674) = \begin{pmatrix} 0.088 & -0.2 \\ -0.2 & 0.454 \end{pmatrix}$$

$$S_{22} - \hat{\mathbf{b}}^{(1)}(\hat{\mathbf{b}}^{(1)})' = \hat{\mathbf{b}}^{(2)}(\hat{\mathbf{b}}^{(2)})' = \begin{pmatrix} -0.356 \\ 0.023 \end{pmatrix} (-0.356 \;\; 0.023) = \begin{pmatrix} 0.127 & -0.008 \\ -0.008 & 0.001 \end{pmatrix}$$

$$S_{12} - \hat{\rho}_1\hat{\mathbf{a}}^{(1)}(\hat{\mathbf{b}}^{(1)})' = \hat{\rho}_2\hat{\mathbf{a}}^{(2)}(\hat{\mathbf{b}}^{(2)})' = 0.057 \begin{pmatrix} -0.297 \\ 0.674 \end{pmatrix} (-0.356 \;\; 0.023) = \begin{pmatrix} 0.006 & -0.014 \\ -0.014 & 0.001 \end{pmatrix}.$$

The first matrix is not quite close to 0 while the second and the third are. The interpretation is $\hat{U}_1$ comparatively does not represent the head size quite well but $\hat{V}_1$ well represents the leg size. As the two variables about leg sizes are highly correlated with each other with correlation 0.926, it is reasonable that one variable ($\hat{V}_1$ here) is sufficient to represent both. This is not the case for the two variables of head size. More importantly, $(\hat{U}_1, \hat{V}_1)$ explains ideally the relations between the head and leg sizes.

(5). The proportion of variance in head size explained by $\hat{U}_1$ is

$$\frac{trace(\hat{\mathbf{a}}^{(1)}(\hat{\mathbf{a}}^{(1)})')}{2} = \frac{0.955^2 + 0.738^2}{2} = 0.728.$$

And the proportion of variance in leg size explained by $\hat{V}_1$ is

$$\frac{trace(\hat{\mathbf{b}}^{(1)}(\hat{\mathbf{b}}^{(1)})')}{2} = \frac{0.934^2 + 0.999^2}{2} = 0.935.$$

$\hat{V}_1$ represents leg size much more sufficiently than $\hat{U}_1$ represents head size. The interpretation agrees with that from matrices of errors.

Remark. Preferably for standardized observations, the canonical variates can be interpreted through the size of canonical coefficients. Another way is to interpret them from their correlation with the original variables they represent. Interpretation of the relations between the two sets of variables using retained pairs of canonical variates should be coupled with representativeness of these canonical variates.