Université Libre de Bruxelles

Devoir Maison 2022/2023

Statistique Mathématique II

Professeure:
Sophie Dabo

Élève:
Márcio Reverbel

May 21, 2023

Contents

List of Figures

List of Tables

Devoir Maison 2022/2023

1 ACP sur la matrice des distances

On observe un $p-$vecteur aléatoire quantitatif sur n individus, soit $\mathbf{X}_i$, $i = 1, \ldots, n$ la $i-$ème observations de dimension p. On note $X_{i,k}$ la $k-$ème composante du vecteur $\mathbf{X}_i$, $k = 1, \ldots, p$. Notons $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)'$ la matrice $(n \times p)$ des observations. On suppose que les composantes de $\mathbf{X}_i$ sont centrées. Soit $\mathbf{D}$ la matrice diagonale des poids $p_i = 1/n$. On munit $\mathbb{R}^p$ d'une métrique $\mathbf{M}$; $\|x\|^2_{\mathbf{M}} = x'\mathbf{M}x$, $x \in \mathbb{R}^p$. Soit $\mathcal{D} = (d^2_{ij})_{i,j=1,\ldots,n}$ la matrice $n \times n$ des carrés des distances entre les n individus (d_{ij} est la distance entre $\mathbf{X}_i$ et $\mathbf{X}_j$, $d_{ii} = 0$) :

$$d^2_{ij} = (\mathbf{X}_i - \mathbf{X}_j)'\mathbf{M}(\mathbf{X}_i - \mathbf{X}_j) = \|\mathbf{X}_i - \mathbf{X}_j\|^2_{\mathbf{M}}$$

Posons

$$d^2_{i.} = \sum_{j=1}^{n} p_j d^2_{ij},\ d^2_{.j} = \sum_{i=1}^{n} p_i d^2_{ij},\ d^2_{..} = \sum_{i=1}^{n} p_i d^2_{i.}.$$

Soit $I_g = \sum_{i=1}^{n} p_i \|\mathbf{X}_i\|^2_{\mathbf{M}}$, dit inertie du nuage de points des observations.
Posons $\mathbf{W} = (w_{ij} =< \mathbf{X}_i, \mathbf{X}_j >_{\mathbf{M}} = \mathbf{X}_i'\mathbf{M}\mathbf{X}_j)_{i,j}$ la matrice des produits scalaires.

1.1 Montrer que la matrice de variance-covariance empirique des $\mathbf{X}_i$ est $\mathbf{S} = \mathbf{X}'\mathbf{D}\mathbf{X}$.

$$\begin{aligned}
\mathbf{S} &= \frac{1}{n}\sum_{i=1}^{n}\left[(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})'\right], \text{ avec les } \mathbf{X}_i \text{ centrées (alors } \overline{\mathbf{X}} = \mathbf{0}_p). \\
\Longrightarrow \mathbf{S} &= \frac{1}{n}\sum_{i=1}^{n}[\mathbf{X}_i\mathbf{X}_i'] = \frac{1}{n}\mathbf{X}'\mathbf{X} = \mathbf{X}'\frac{1}{n}\mathbf{X} = \mathbf{X}'\left(\frac{1}{n}\mathbf{I}_n\right)\mathbf{X} = \mathbf{X}'\mathbf{D}\mathbf{X}
\end{aligned}$$

1.2 Montrer que $\forall i = 1, ...n,\ d_{i.}^2 = \|\mathbf{X}_i\|_{\mathbf{M}}^2 + I_g$.

$$\begin{aligned}
d_{i.}^2 &= \sum_{j=1}^n p_j d_{ij}^2 = \sum_{j=1}^n p_j \|\mathbf{X}_i - \mathbf{X}_j\|_{\mathbf{M}}^2 \\
&= \sum_{j=1}^n p_j (\mathbf{X}_i - \mathbf{X}_j)'\mathbf{M}(\mathbf{X}_i - \mathbf{X}_j) \\
&= \sum_{j=1}^n p_j (\mathbf{X}_i'\mathbf{M}\mathbf{X}_i - \mathbf{X}_j'\mathbf{M}\mathbf{X}_i - \mathbf{X}_i'\mathbf{M}\mathbf{X}_j + \mathbf{X}_j'\mathbf{M}\mathbf{X}_j) \\
&= \sum_{j=1}^n p_j \mathbf{X}_i'\mathbf{M}\mathbf{X}_i - \sum_{j=1}^n p_j \mathbf{X}_j'\mathbf{M}\mathbf{X}_i - \sum_{j=1}^n p_j \mathbf{X}_i'\mathbf{M}\mathbf{X}_j + \sum_{j=1}^n p_j \mathbf{X}_j'\mathbf{M}\mathbf{X}_j \\
&= \sum_{j=1}^n p_j \|\mathbf{X}_i\|_{\mathbf{M}}^2 - \overline{\mathbf{X}}'\mathbf{M}\mathbf{X}_i - \mathbf{X}_i'\mathbf{M}\overline{\mathbf{X}}' + \sum_{j=1}^n p_j \|\mathbf{X}_j\|_{\mathbf{M}}^2 \\
&= \|\mathbf{X}_i\|_{\mathbf{M}}^2 - 0 - 0 + I_g, \text{puisque } \overline{\mathbf{X}} = 0_p \\
&= \|\mathbf{X}_i\|_{\mathbf{M}}^2 + I_g
\end{aligned}$$

1.3 En déduire que $d_{..}^2 = 2I_g$

$$\begin{aligned}
d_{..}^2 &= \sum_{i=1}^n p_i d_{i.}^2 = \sum_{i=1}^n p_i (\|\mathbf{X}_i\|_{\mathbf{M}}^2 + I_g) \\
&= \sum_{i=1}^n p_i \|\mathbf{X}_i\|_{\mathbf{M}}^2 + \sum_{i=1}^n p_i I_g = I_g + I_g = 2I_g
\end{aligned}$$

1.4 Montrer que $w_{ij} = -\frac{1}{2}(d_{ij}^2 - d_{i.}^2 - d_{.j}^2 + d_{..}^2)$

Notons que, analoguement au procédiment dans l'exercice 1.2, nous avons que:

$$d_{.j}^2 = \sum_{i=1}^n p_i d_{ij}^2 = I_g + \|\mathbf{X}_j\|_{\mathbf{M}}^2$$

En plus, nous savons que M, en se traitant d'une métrique, est symétrique et définie positive, de sorte que:

$$< \mathbf{X}_i, \mathbf{X}_j >_{\mathbf{M}} = \mathbf{X}_i'\mathbf{M}\mathbf{X}_j = (\mathbf{X}_j'\mathbf{M}'\mathbf{X}_i)' = (\mathbf{X}_j'\mathbf{M}\mathbf{X}_i)' = \mathbf{X}_j'\mathbf{M}\mathbf{X}_i = < \mathbf{X}_j, \mathbf{X}_i >_{\mathbf{M}},$$

puisque $\mathbf{X}_j'\mathbf{M}\mathbf{X}_i$ est un scalaire (et le transpose d'un scalaire est lui même).

Alors, soit $w_{ij} =< \mathbf{X}_j, \mathbf{X}_i >_{\mathbf{M}}$. Nous avons que:

$$\begin{aligned}
d_{ij}^2 &= (\mathbf{X}_i - \mathbf{X}_j)'\mathbf{M}(\mathbf{X}_i - \mathbf{X}_j) = \|\mathbf{X}_i - \mathbf{X}_j\|_{\mathbf{M}}^2 \\
&= \|\mathbf{X}_i\|_{\mathbf{M}}^2 - < \mathbf{X}_i, \mathbf{X}_j >_{\mathbf{M}} - < \mathbf{X}_j, \mathbf{X}_i >_{\mathbf{M}} + \|\mathbf{X}_j\|_{\mathbf{M}}^2 \\
&= \|\mathbf{X}_i\|_{\mathbf{M}}^2 - 2 < \mathbf{X}_i, \mathbf{X}_j >_{\mathbf{M}} + \|\mathbf{X}_j\|_{\mathbf{M}}^2 \\
&= \|\mathbf{X}_i\|_{\mathbf{M}}^2 - 2w_{ij} + \|\mathbf{X}_j\|_{\mathbf{M}}^2,
\end{aligned}$$

$$\begin{aligned}
\Longrightarrow d_{ij}^2 - d_{i.}^2 - d_{.j}^2 &= \|\mathbf{X}_i\|_{\mathbf{M}}^2 - 2w_{ij} + \|\mathbf{X}_j\|_{\mathbf{M}}^2 - (I_g + \|\mathbf{X}_i\|_{\mathbf{M}}^2) - (I_g + \|\mathbf{X}_j\|_{\mathbf{M}}^2) \\
&= -2w_{ij} - 2I_g \\
\Longrightarrow d_{ij}^2 - d_{i.}^2 - d_{.j}^2 + d_{..}^2 &= -2w_{ij}
\end{aligned}$$

$$\Longrightarrow w_{ij} = -\frac{1}{2}(d_{ij}^2 - d_{i.}^2 - d_{.j}^2 + d_{..}^2)$$

Supposons dans la suite que $\mathbf{M} = \mathbf{I}_p$ et que l'ACP du nuage de $\mathbf{X}$ donne p axes principaux normés $(u_k)_{k=1,\ldots,p}$ de valeurs propres correspondants λ_k. Notons v_k les composantes principales associées.

1.5 Exprimer W en fonction de $\mathcal{D}$

Notons d'abord que:

$$\mathbf{D}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}\mathcal{D} = \begin{pmatrix} \sum_{i=1}^n \frac{1}{n} d_{i1}^2 & \cdots & \sum_{i=1}^n \frac{1}{n} d_{in}^2 \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n \frac{1}{n} d_{i1}^2 & \cdots & \sum_{i=1}^n \frac{1}{n} d_{in}^2 \end{pmatrix} = \begin{pmatrix} d_{.1}^2 & \cdots & d_{.n}^2 \\ \vdots & \ddots & \vdots \\ d_{.1}^2 & \cdots & d_{.n}^2 \end{pmatrix},$$

$$\mathcal{D}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}\mathbf{D} = \begin{pmatrix} \sum_{j=1}^n \frac{1}{n} d_{1j}^2 & \cdots & \sum_{j=1}^n \frac{1}{n} d_{1j}^2 \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^n \frac{1}{n} d_{nj}^2 & \cdots & \sum_{i=1}^n \frac{1}{n} d_{nj}^2 \end{pmatrix} = \begin{pmatrix} d_{1.}^2 & \cdots & d_{1.}^2 \\ \vdots & \ddots & \vdots \\ d_{n.}^2 & \cdots & d_{n.}^2 \end{pmatrix},$$

et que

$$\mathbf{D}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}\mathcal{D}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}\mathbf{D} = \mathbf{D}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}\begin{pmatrix} d_{1.}^2 & \cdots & d_{1.}^2 \\ \vdots & \ddots & \vdots \\ d_{n.}^2 & \cdots & d_{n.}^2 \end{pmatrix} = \begin{pmatrix} d_{..}^2 & \cdots & d_{..}^2 \\ \vdots & \ddots & \vdots \\ d_{..}^2 & \cdots & d_{..}^2 \end{pmatrix}.$$

Alors, nous avons que:

$$\mathbf{W} = \frac{1}{2}(\mathcal{D} - \mathcal{D}\ 1_{n\times n}\ \mathbf{D} - \mathbf{D}\ 1_{n\times n}\ \mathcal{D} + \mathbf{D}\ 1_{n\times n}\ \mathcal{D}\ 1_{n\times n}\ \mathbf{D}),$$

où $1_{n\times n}$ est la matrice $n \times n$ de 1s.

1.6 Montrer que $\mathbf{XS}u_k = \lambda_k v_k$. Que peut-on en déduire?

$\mathbf{S}u_k = \lambda_k u_k$, par la définition de valeurs propres et vecteurs propres, avec u_k le vecteur propre et λ_k le valeur propre associé. Et comme, $\mathbf{S}$ est symétrique définie positive, on sait qu'il y a p solutions pour cette équation à valeurs propres dans $\mathbb{R}$

et vecteurs propres dans $\mathbb{R}^p$.

Alors:

$$\begin{aligned}
\mathbf{S}u_k &= \lambda_k u_k \\
\Longrightarrow \mathbf{X}\mathbf{S}u_k &= \mathbf{X}\lambda_k u_k \\
\Longrightarrow \mathbf{X}\mathbf{S}u_k &= \lambda_k \mathbf{X} u_k \\
\\
\Longrightarrow \mathbf{X}\mathbf{S}u_k &= \lambda_k v_k,
\end{aligned}$$

puisque $\mathbf{X}u_k = v_k$, la composante principale associée à λ_k.

En plus, si $\mathbf{M} = \mathbf{I}_p$, v_k est un vecteur de produits scalaires entre chaque observation $\mathbf{X}_i$, $1 \leq i \leq n$, et le vecteur u_k:

$$v_k = \begin{pmatrix} < \mathbf{x}_1, \mathbf{u}_k > \\ \vdots \\ < \mathbf{X}_n, \mathbf{u}_k > \end{pmatrix}$$

Alors, chaque composante i de v_k represente le scalaire associé à u_k de la projection orthogonale de l'observation $\mathbf{X}_i$ sur le vecteur u_k:

$$\pi_k(\mathbf{X}_i) = \frac{< \mathbf{X}_i, \mathbf{u}_k >}{\|u_k\|^2} u_k =< \mathbf{X}_i, \mathbf{u}_k > u_k$$

puisque $\|u_k\|^2 = 1$.

1.7 Montrer que, toujours si $\mathbf{M} = \mathbf{I}_p$, v_k est également vecteur propre de WD.

Notons d'abord que, sous la métrique $\mathbf{M} = \mathbf{I}_p$, nous avons:

$$\mathbf{W} = \begin{pmatrix} < \mathbf{X}_1, \mathbf{X}_1 > & \dots & < \mathbf{X}_1, \mathbf{X}_n > \\ \vdots & \ddots & \vdots \\ < \mathbf{X}_n, \mathbf{X}_1 > & \dots & < \mathbf{X}_n, \mathbf{X}_n > \end{pmatrix} = \mathbf{X}\mathbf{X}'$$

Alors,

$$\begin{aligned}
\mathbf{X}\mathbf{S}u_k &= \lambda_k v_k \\
\Longrightarrow \mathbf{X}\mathbf{X}'\mathbf{D}\mathbf{X}u_k &= \lambda_k v_k \\
\Longrightarrow \mathbf{W}\mathbf{D}\mathbf{X}u_k &= \lambda_k v_k \\
\\
\Longrightarrow \mathbf{W}\mathbf{D}v_k &= \lambda_k v_k,
\end{aligned}$$

encore une fois la définition de vecteur propre et valeur propre, cette fois pour la matrice **WD**. Alors, chaque λ_k qui est valeur propre de **S** est aussi valeur propre de **WD**, mais associé au vecteur propre v_k. Observons que **WD** est aussi symétrique et à valeur et vecteur propres réels.

1.8 Soit le vecteur $f_k \in \mathbb{R}^n$ dont la composante numéro i est $f_{ik} = \sqrt{p_i}v_{ik}$. En déduire que la matrice WD admet pour vecteur propre f_k avec valeur propre associé à λ_k.

Notons que $\frac{1}{\sqrt{n}} = \sqrt{p_i}$, $\forall\ i$. De manière similaire au exercice précedent, nous avons:

$$\begin{aligned}
\mathbf{S}u_k &= \lambda_k u_k \\
\Longrightarrow \mathbf{D}^{\frac{1}{2}}\mathbf{X}\mathbf{S}u_k &= \mathbf{D}^{\frac{1}{2}}\mathbf{X}\lambda_k u_k \\
\Longrightarrow \mathbf{D}^{\frac{1}{2}}\mathbf{X}\mathbf{X}'\mathbf{D}\mathbf{X}u_k &= \lambda_k \mathbf{D}^{\frac{1}{2}}\mathbf{X}u_k \\
\Longrightarrow \mathbf{D}^{\frac{1}{2}}\mathbf{W}\mathbf{D}^{\frac{1}{2}}\mathbf{D}^{\frac{1}{2}}v_k &= \lambda_k v_k \\
\Longrightarrow \mathbf{W}\mathbf{D}\frac{1}{\sqrt{n}}v_k &= \lambda_k \frac{1}{\sqrt{n}}v_k \\
\Longrightarrow \mathbf{W}\mathbf{D}f_k &= \lambda_k f_k
\end{aligned}$$

Alors, f_k est le vecteur propre associé à λ_k pour la matrice **WD**.

1.9 Montrer que le vecteur $\left(\sqrt{p_i}\right)_{i=1,\dots,n}$ est vecteur propre de WD associé à la valeur propre 0.

Observons que (pour toute métrique M):

$$\begin{aligned}
\mathbf{WD}\ 1_n &= \begin{pmatrix} <\mathbf{X}_1,\mathbf{X}_1>_\mathbf{M} & \cdots & <\mathbf{X}_1,\mathbf{X}_n>_\mathbf{M} \\ \vdots & \ddots & \vdots \\ <\mathbf{X}_n,\mathbf{X}_1>_\mathbf{M} & \cdots & <\mathbf{X}_n,\mathbf{X}_n>_\mathbf{M} \end{pmatrix} \begin{pmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix} \\
&= \begin{pmatrix} <\mathbf{X}_1, \sum_{i=1}^n \frac{1}{n}\mathbf{X}_i>_\mathbf{M} \\ \vdots \\ <\mathbf{X}_n, \sum_{i=1}^n \frac{1}{n}\mathbf{X}_i>_\mathbf{M} \end{pmatrix} \\
&= \begin{pmatrix} <\mathbf{X}_1, \overline{\mathbf{X}}>_\mathbf{M} \\ \vdots \\ <\mathbf{X}_n, \overline{\mathbf{X}}>_\mathbf{M} \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},
\end{aligned}$$

puisque $\overline{\mathbf{X}} = \mathbf{0}_p$. Alors, $\mathbf{WD}\ 1_n = 0\ 1_n = \mathbf{0}_n$. Et, évidemment, $\mathbf{WD}1_n = \mathbf{0}_n$ implique $\mathbf{WD}\ 1_n\mathbf{D}^{\frac{1}{2}} = \mathbf{0}_n$, puisque $\mathbf{WD}\ 1_n\mathbf{D}^{\frac{1}{2}} = 0\ 1_n\mathbf{D}^{\frac{1}{2}} = \mathbf{0}_n\mathbf{D}^{\frac{1}{2}} = \mathbf{0}_n$.

Alors, $1_n\mathbf{D}^{\frac{1}{2}}$ est vecteur propre associé à la valeur propre 0. (Notez que $\|1_n\mathbf{D}^{\frac{1}{2}}\| = 1$.)

En plus, pour $\mathbf{M} = \mathbf{I}_p$, nous avons que:

$$\begin{aligned}
\mathrm{tr}(\mathbf{S}) &= \mathrm{tr}(\boldsymbol{\beta}\boldsymbol{\Lambda}_p\boldsymbol{\beta}') = \mathrm{tr}(\boldsymbol{\Lambda}_p\boldsymbol{\beta}'\boldsymbol{\beta}) = \mathrm{tr}(\boldsymbol{\Lambda}_p) \\
\mathrm{tr}(\mathbf{S}) &= \mathrm{tr}(\mathbf{X}'\mathbf{D}\mathbf{X}) = \mathrm{tr}(\mathbf{X}\mathbf{X}'\mathbf{D}) = \mathrm{tr}(\mathbf{W}\mathbf{D}) \\
&= \mathrm{tr}(\boldsymbol{\theta}\boldsymbol{\Lambda}_n\boldsymbol{\theta}') = \mathrm{tr}(\boldsymbol{\Lambda}_n\boldsymbol{\theta}'\boldsymbol{\theta}) = \mathrm{tr}(\boldsymbol{\Lambda}_n)
\end{aligned}$$

Considerons $n > p$ (sans perde de géneralité): Alors, $\mathrm{tr}(\boldsymbol{\Lambda}_n) = \mathrm{tr}(\boldsymbol{\Lambda}_p)$. En sachant que les p valeurs propres de $\mathbf{S}$ sont aussi valeurs propres de $\mathbf{WD}$ (Ex. 1.7 et 1.8), et que $\mathbf{WD}$ admet seulement de valeurs propres réels et positives (puisque $\mathbf{WD}$ est symétrique et semi-définie positive*), alors toutes les autres valeurs propres de $\boldsymbol{\Lambda}_n$ doivent être 0:

$\mathrm{tr}(\boldsymbol{\Lambda}_p) = \mathrm{tr}(\boldsymbol{\Lambda}_n) = \mathrm{tr}(\boldsymbol{\Lambda}_p) + \lambda_{p+1} + ... + \lambda_n$, avec $\lambda_i \geq 0 \; \forall \; i, 1 \leq i \leq n$.
$\implies \mathrm{tr}(\boldsymbol{\Lambda}_n) - \mathrm{tr}(\boldsymbol{\Lambda}_p) = \lambda_{p+1} + ... + \lambda_n = 0$, avec $\lambda_1 > 0, ..., \lambda_n > 0$.
$\implies \lambda_{p+1} = \lambda_{p+2} = ... = \lambda_n = 0$.

* $a'\mathbf{WD}a = \frac{1}{n}a'\mathbf{X}\mathbf{X}'a = \frac{1}{n}(\mathbf{X}'a)'\mathbf{X}'a = \frac{1}{n}\|\mathbf{X}'a\|^2 \geq 0 \; \forall \; a \in \mathbb{R}^n$.

1.10 Montrer que $\sum_{i=1}^{n} f_{ik}^2 = \lambda_k$ et, pour tout $k \neq l$, $\sum_{i=1}^{n} f_{ik}f_{il} = 0$.

Sous $\mathbf{M} = \mathbf{I}_p$:

$$\begin{aligned}
\sum_{i=1}^{n} f_{ik}^2 &= < f_k, f_k > \\
&= < \mathbf{D}^{\frac{1}{2}}\mathbf{X}u_k, \mathbf{D}^{\frac{1}{2}}\mathbf{X}u_k > \\
&= u_k'\mathbf{X}'\mathbf{D}\mathbf{X}u_k \\
&= u_k'\mathbf{S}u_k \\
&= u_k'\lambda_k u_k \\
&= \lambda_k < u_k, u_k > \\
&= \lambda_k\|u_k\|^2 \\
&= \lambda_k,
\end{aligned}$$

puisque $\|u_k\|^2 = 1$.

$$\begin{aligned}
\sum_{i=1}^{n} f_{ik}^2 &= < f_k, f_l > \\
&= < \mathbf{D}^{\frac{1}{2}}\mathbf{X}u_k, \mathbf{D}^{\frac{1}{2}}\mathbf{X}u_l > \\
&= u_k'\mathbf{X}'\mathbf{D}\mathbf{X}u_l \\
&= u_k'\mathbf{S}u_l \\
&= u_k'\lambda_l u_l \\
&= \lambda_l < u_k, u_l > \\
&= 0,
\end{aligned}$$

puisque u_k et u_l sont orthogonaux ($< u_k, u_l >= 0, \forall k \neq j$).

1.11 Application sous Python

Soit un nuage de points de 3 individus tel que

$$d_{12}^2 = d_{23}^2 = 1,\ d_{13}^2 = 2,\ p_i = 1/3,\ i = 1, \ldots, 3$$

Déterminer **WD**, les valeurs propres λ_k et les vecteurs propres propres f_k associés.

Obs: Vous pouvez trouver la résolution sur l'archive Exercise_1_11.ipynb.

2 Application sur données réelles avec Python

On considère 11 pôles de dépenses d'un Etat (répartitions des dépenses en pourcentages) entre plusieurs années successives. On note X la matrice des données dont les pôles de dépenses (en colonne): PVP : pouvoirs publics; AGR : agriculture; CMI : commerce et industrie; TRA : travail; LOG : logement et aménagement du territoire; EDU : éducation; ACS : action sociale; ACO : anciens combattants; DEF : défense; DET : dette; DIV : divers.

2.1 Effectuer sur ces données une Analyse en Composantes Principales.

Obs: Vous pouvez trouver l'analyse sur l'archive
STAT_MATH_-_Devoir_Exercice2_ACP_sur_DataDepenses.ipynb.

For this analysis, we use a dataset which expresses expenditure across 11 different departments for some state in percentage of total expenditure. We begin by looking at the summary statistics for our dataset (Table 1).

	count	mean	std	min	25%	50%	75%	max
PVP	24.0	12.212500	2.238267	7.6	10.575	12.60	13.425	18.0
AGR	24.0	1.995833	1.681221	0.3	0.800	1.40	2.650	6.0
CMI	24.0	3.937500	4.579806	0.1	0.400	1.30	7.350	16.5
TRA	24.0	8.320833	2.520866	4.5	6.675	8.00	9.150	15.3
LOG	24.0	3.958333	4.271841	0.5	0.675	1.85	6.200	15.8
EDU	24.0	9.941667	5.335600	2.1	7.325	8.70	10.600	23.8
ACS	24.0	4.816667	3.482087	0.5	1.800	4.55	6.800	11.3
ACO	24.0	4.275000	4.244203	0.0	0.000	3.80	5.450	13.4
DEF	24.0	30.258333	7.466733	18.8	25.925	29.15	37.025	42.4
DET	24.0	19.141667	12.455972	3.5	6.350	19.30	26.450	41.6
DIV	24.0	1.183333	1.047841	0.0	0.000	1.40	2.025	3.0

Table 1: Descriptive statistics for the data.

The summary statistics show how variability in our dataset is mostly concentrated in two variables. The debt (DET) expenditure concentrates enourmous variability, as can be seen by it's large standard deviations and interquartile range, while the variability of the defense department's expenses (DEF) is also quite high. This can be seen more clearly in Figure 1, which shows a heatmap of the variance-covariance matrix in our dataset, with almost all of the variance concentrated on the DET variable, with DEF coming in second place. Thus, this can be seen as an indication that, to better understand how our data varies, it will be preferable to perform a principal component analysis over the standardized dataset.

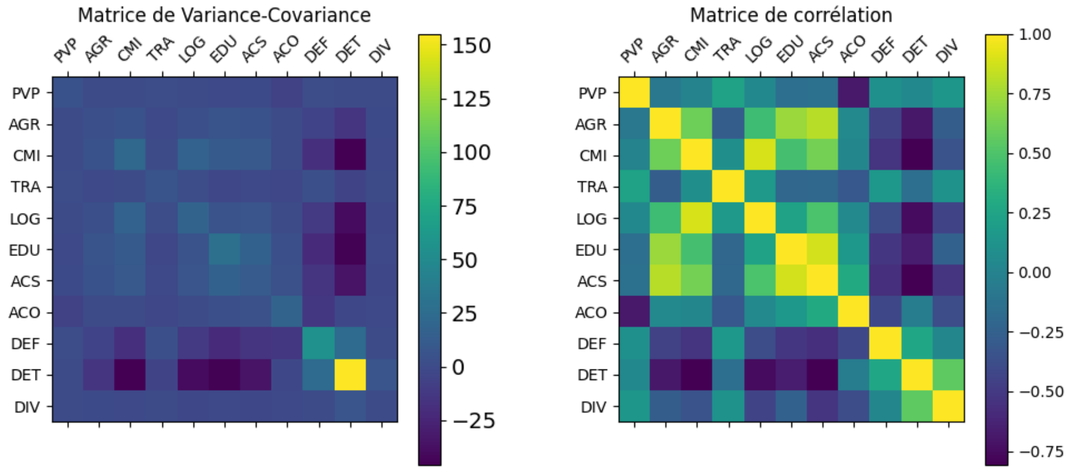

Figure 1: Heatmaps of the variance-covariance matrix (left) and correlation matrix (right).

Indeed, the heatmap of the correlation matrix shows a much richer scenario for our analysis. Otherwise, if we used the non-standardized data (only centered), we would arrive at one principal component which explains the vast majority of our data and which correlates strongly with the DEF variable. This would not result in an added understanding of the data, and the interpretation of the components would be more difficult due to the difference in scale between these variables.

The correlation matrix is represented in more detail in Figure 2, in the form of a clustermap. As can be seen, our variables can be sorted into two large groups, and then each of those groups further divided into two groups each, based on how they correlate with eachother. We name these groups as such:

- Group A: ACO.
- Group B: PVP, TRA.
- Group C: AGR, CMI, LOG, EDU and ACS.
- Group D: DET, DEF, DIV.

We can expect our PCA to add understanding in terms of how these clusters are represented by the principal components, among other things. The fact that groups of variables are strongly correlated means that we can expect to perform a principal component analysis to reduce the dimensionality of our dataset while accounting for most of its original variability.

When performing the principal component analysis, we effectively calculate the eigenvalues and associated eigenvectors of our correlation matrix. This is identical to calculating the eigenvectors and eigenvalues of our centered and standardized dataset. PCA allows us to express our data in terms of a different orthogonal basis and which maximizes the variability expressed by each subsequent component of the

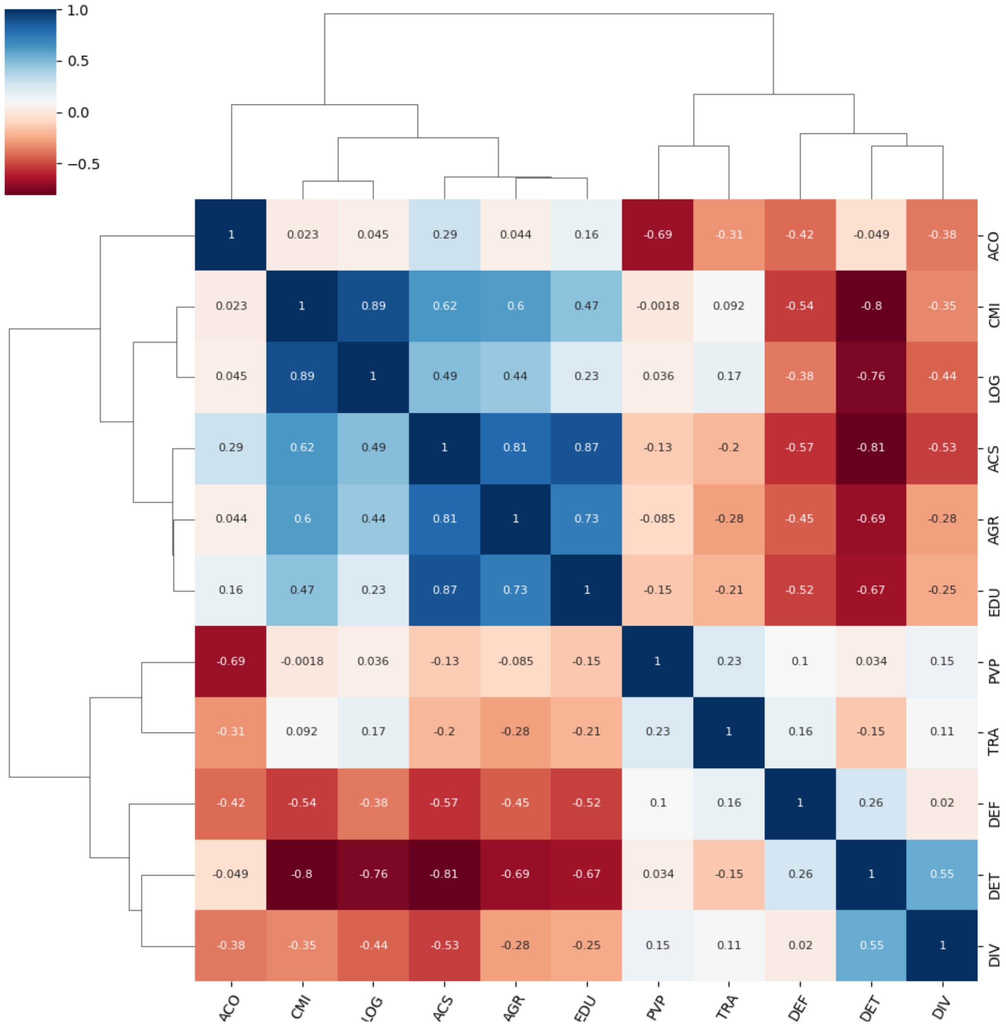

Figure 2: Cluster map of the correlation matrix.

basis. Since the eigenvalues represent the portion of the variance captured by the projection of the data onto the associated unit eigenvector, we order the components from highest to lowest eigenvalue. Each principal component is thus calculated by projecting our data over its associated eigenvector. Then we can chose how many components we wish to retain, reducing the dimensionality of our dataset while minimizing the information (variability) that is lost. Table 2 shows our dataset as represented by the first three principal components (calculated over the standardized dataset) [1]. The scatterplots in Figure 3 allow us to visualize the dataset through the retained principal components.

[1] The principal components were calculated using the sklearn and prince packages in Python.

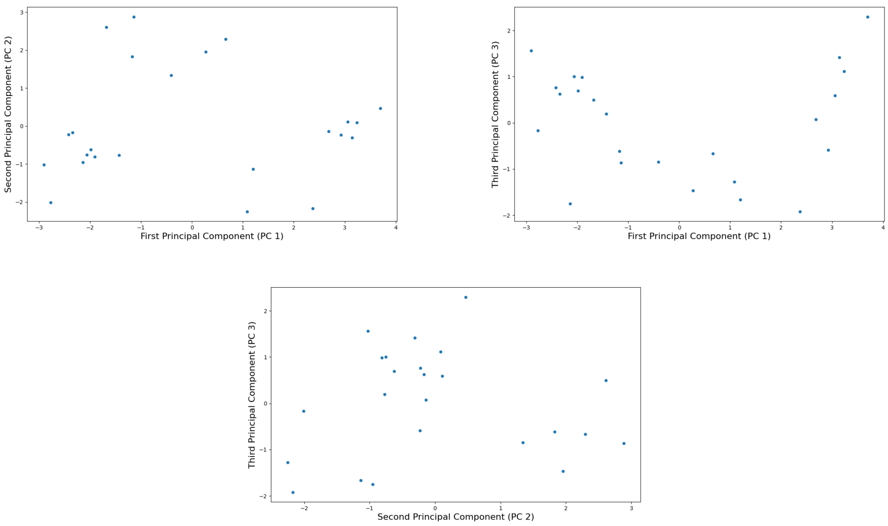

Figure 3: Scatterplots for the first three principal components.

	PC 1	PC 2	PC 3
0	-2.900539	-1.024429	1.564588
1	-2.767389	-2.011953	-0.169510
2	-2.416316	-0.224014	0.765716
3	-2.056634	-0.755155	1.006817
4	-2.337858	-0.167246	0.622559
5	-1.985142	-0.626137	0.692399
6	-1.907355	-0.812222	0.986687
7	-1.431071	-0.768419	0.194171
8	-2.139175	-0.955910	-1.747200
9	-1.142910	2.883950	-0.865633
10	-1.674088	2.610955	0.499305
11	-1.173432	1.831198	-0.610408
12	0.270638	1.959320	-1.461990
13	0.659049	2.296209	-0.662722
14	-0.402398	1.342964	-0.850175
15	1.081281	-2.251166	-1.276465
16	2.372809	-2.175400	-1.917469
17	1.203777	-1.134323	-1.662421
18	2.927966	-0.230691	-0.589851
19	2.686197	-0.140184	0.071391
20	3.054717	0.110790	0.586404
21	3.143013	-0.311233	1.412376
22	3.695610	0.466653	2.297145
23	3.239249	0.086445	1.114289

Table 2: Principal component scores for the first three principal. components.

2.2 Combien d'axes retiendriez-vous pour cette analyse? Pourquoi?

The amount of principal components to retain is generally dependant on a few criteria:

- As a rule of thumb, we chose a number of principal components such that the amount of variance explained is greater than 80%.
- We retain those components whose eigenvalues are greater than the mean eigenvalue, known as the Kaiser rule.
- If there is a large difference between the variances associated to two subsequent principal components, q and q+1, we take the first q components.
- Taking into account the correlation between the principal components and the initial data, we exclude the first axis which does not correlate with the variables, as well as subsequent components which contain less variability. This ensures interpretability of the components.

component	eigenvalue	% of variance	% of variance (cumulative)
0	4.973	45.21%	45.21%
1	2.050	18.64%	63.85%
2	1.290	11.73%	75.57%
3	0.993	9.03%	84.61%
4	0.708	6.44%	91.04%
5	0.559	5.08%	96.12%
6	0.204	1.86%	97.98%
7	0.125	1.14%	99.12%
8	0.062	0.56%	99.68%
9	0.035	0.32%	100.00%

Table 3: Principal components' variance and cumulative variance.

According to the first criteria, we would want to retain the first 4 principal components in our analysis, which preserves 84.61% of the variance in the data. However, Kaiser's rule would tell us to preserve the first 3 principal components. The third criteria doesn't really fit here. One could interpret it in such a way the first five principal components. But then that amount of components doesn't help with the interpretability of our analysis. Finally, when checking the correlations between the principal components and the variables (Table 4), we can see strong correlations present in the first two components. However, the third component doesn't correlate strongly with any variable, the highest being a correlation of approximately 0.54 with DIV (diverse expenditures).

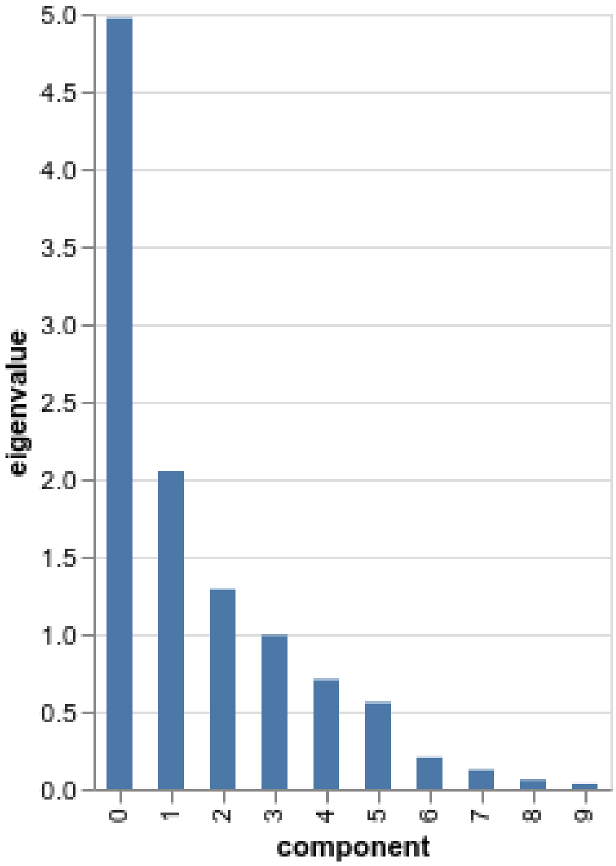

Figure 4: Variance retained by principal component.

component variable	0	1	2	3
PVP	-0.173578	-0.739763	0.341469	-0.107558
AGR	0.818517	-0.005918	0.366433	-0.153632
CMI	0.833594	-0.340222	-0.140809	0.258018
TRA	-0.137166	-0.630559	-0.376015	0.281110
LOG	0.721613	-0.397716	-0.385161	0.207956
EDU	0.786829	0.136504	0.424637	-0.115890
ACS	0.933259	0.100484	0.166134	-0.150740
ACO	0.288961	0.807563	-0.374504	0.202154
DEF	-0.612343	-0.216146	-0.259903	-0.636766
DET	-0.888749	0.301471	0.160800	0.179310
DIV	-0.548122	-0.112364	0.536572	0.504656

Table 4: Correlation between the principal components and the original variables.

Since the four criteria do not lead to the same conclusion, we must find a compromise. In this case, we opt to preserve 3 principal components, under the following arguments: The first 3 principal components retain 75.57% of the variance in the data, which is close enough to 80%, while being considerably easier to visualise and interpret than an analysis in 4 dimensions. Three components satisfies the Kaiser rule, and an argument can be made for the Kaiser rule being particularly interesting when working with standardised data (such as our case). The mean eigenvalue is simply 1, making it easy to calculate. Finally, when considering the third principal component: While a 0.54 correlation is not strong, it is still relevant considering that there are 11 dimensions involved: For each original variable, the sum of the square of its correlations with each principal component is 1. In that light, 0.54 isn't too bad, and so the third principal component accounts for more than 25% of the correlation between DIV and the principal components.

A strong argument could be made for retaining the fourth principal component, under the reasoning that it's eigenvalue is very close to 1, and that there are relevant correlations between this component and the DEF and DIV variables. Additionally, it would satisfy the first criteria. However, in this case we opt for fewer components to favour the interpretability of our analysis.

Therefore, we retain 3 first principal components, which represent 75.57% of the variance found in our data.

2.3 Donner une interprétation globale des dépenses sur les axes retenus.

To organise our interpretation, we will first organise the 11 variables in 4 groups, according to the clustermap for the correlation:

- Group A: ACO.
- Group B: PVP, TRA.
- Group C: AGR, CMI, LOG, EDU and ACS.
- Group D: DET, DEF, DIV.

The three biplots [2], Figure 5, Figure 6 and Figure 7, help us to see how the three principal components encapsulate the data. We can see that the first principal component has strong positive correlation with Group C, and strong negative correlation with Group D. A negative PC 1 value would indicate large government debt and military spending for that observation, and a smaller budget dedicated to Group C's expenses. A large PC 1 value indicates the opposite, while a value close to zero indicates an observation with close to average spending for both categories C and D.

The second principal component has a strong positive correlation with Group A, and strong negative correlation with Group B. It also correlates moderately with some variables from the other groups (positively with DET, negatively with LOG and CMI). That means that a high value in the PC2 would indicate high ACO values, low PVP and TRA values, as well as slightly higher DET values and slightly

[2]Note: The scale of the biplots is different to that of the scatterplots, since it corresponds to the loading values, and not the principal component scores.

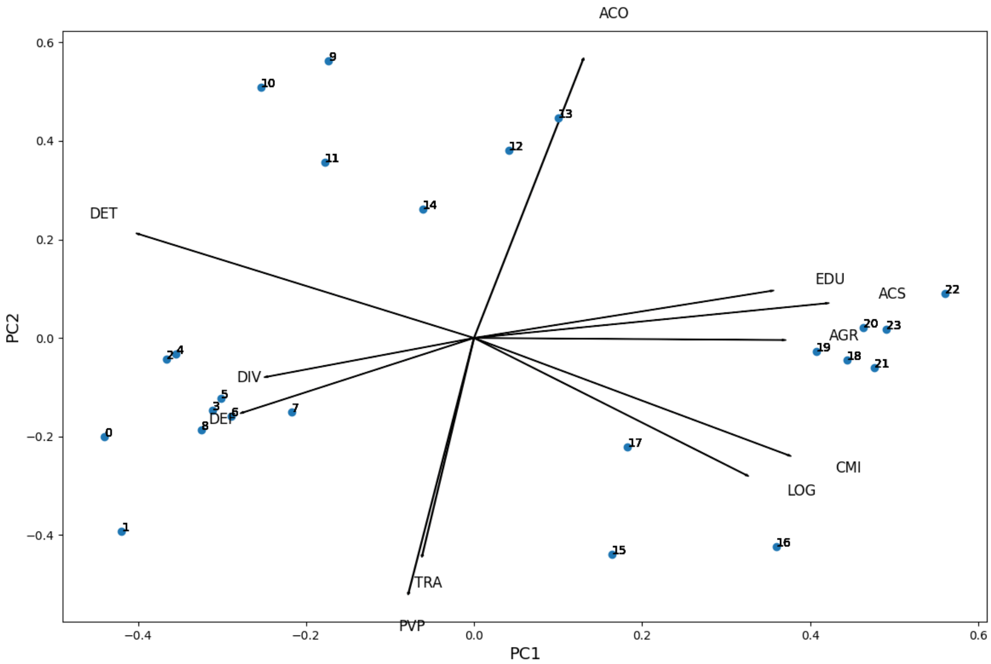

Figure 5: Biplot: PC1 vs PC2

lower CMI and LOG values. Thus, the second principal component introduces some variability into Group C and Group D.

Finally, the third principal component has a strong positive correlation with DIV (Group D), and moderate positive correlation with PVP (Group B), EDU and AGR (from Group C). It has moderate negative correlation with TRA (from Group B), LOG (from Group C) and ACO (from Group A). Thus, the third principal component introduces even more variability between variables within a group, such that a high value of PC3 would indicate higher DIV, slightly higher PVP, EDU and AGR, and slightly lower TRA, LOG and ACO.

Thus, for example, a very high LOG value would be possible with high PC1, a low PC2 and low PC3 values.

It becomes risky to conclude anything beyond that which the data allows us to see. Nonetheless, a few interesting things might be worth looking into (especially if we would have access to a larger dataset):

The first component (PC 1) represents the balance between state investments in social and economic issues and costs made to cover debt and the military expenses (DIV expenses are very small), which are sunk costs that do not stimulate the economy or promote long-term growth. Therefore, the first component can be seen as how much the state invests in its own social and economic development. A high value of PC1 means it invests a lot, while a low PC1 value means it little its own welfare.

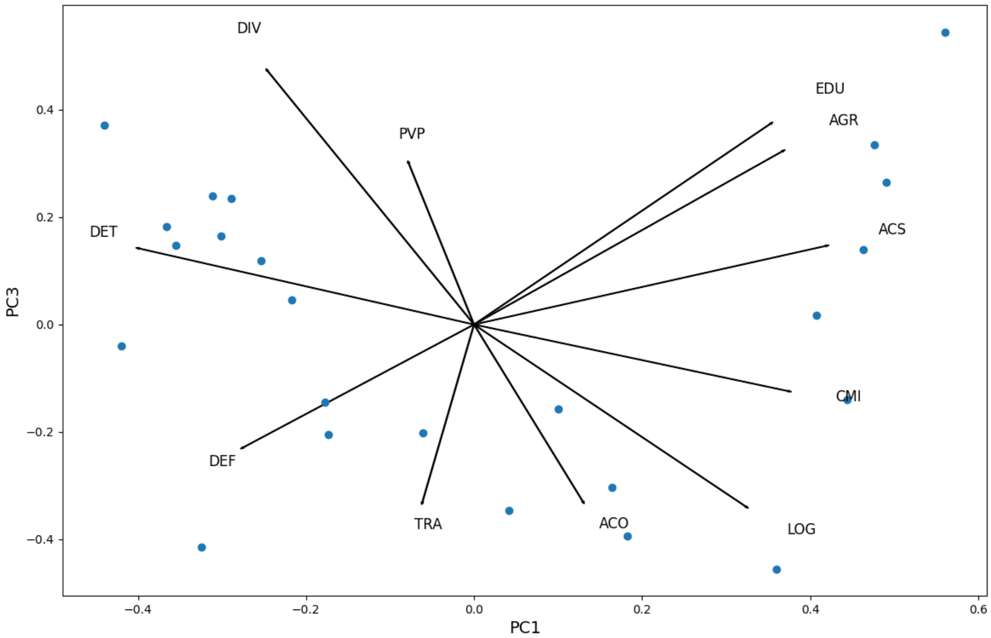

Figure 6: Biplot: PC1 vs PC3

Additionally, when looking at the transformed data, we see a very clear trend over time (remember that the observations are for subsequent years): The first component increases progressively (See Figure 5 and Figure 6), as we can see that the initial observations are more present in the negative end of this component, and the last observations are concentrated on the positive end. We can thus see how state investment in social and economic development issues increased with time, while state debt and military spending decreased in the long term. So the first component correlates with time itself, and it merges in one variable the linear trends present over all variables. This would mean that ACO, TRA and PVP don't present linear trends in the long term, and fluctuate differently. Of course, this conclusion is, as mentioned above, risky to make with a low amount of data, as well as without being certain of a dependence in time between the variables represented by PC1.

The variables most strongly represented in the second component are ACO and PVP. It's important to notice how ACO jumps from 0% to 10% of the budget on the 10th year, and there it remains for 5 years, before falling again. Thus the second principal component could be seen as capturing fluctuations due to electoral results which change governance. This agrees with our previous observation regarding the first component: If the first component captures long-term trends, the second component, being uncorrelated with the first, would thus represent more short-term fluctuations.

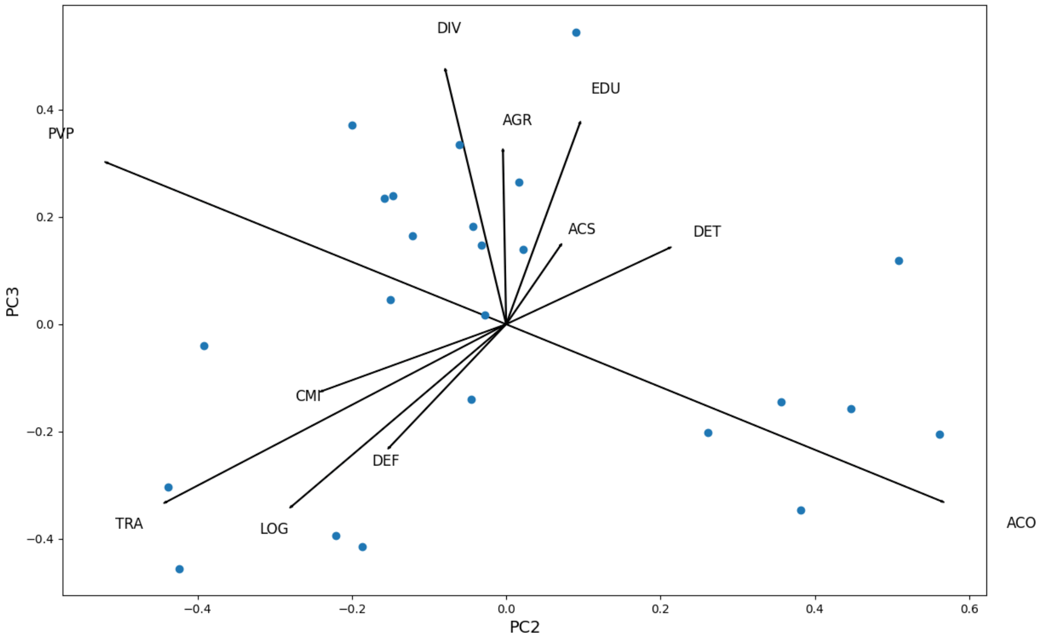

Figure 7: Biplot: PC2 vs PC3

3 Convergence des vecteurs propres d'une matrice de variance-covariance empirique

Soit $\mathbf{X}_1, \ldots, \mathbf{X}_n$ des p-vecteurs aléatoires Gaussiens i.i.d. d'espérances nulles et de matrices de variance-covariance $\mathbf{\Sigma} = \mathbf{I}_p$. Soit $\hat{\mathbf{S}}$ la matrice de variance-covariance empirique des $\hat{\mathbf{X}}_i$, $i = 1, \ldots, n$ et sa décomposition spectrale $\hat{\mathbf{S}} = \hat{\boldsymbol{\beta}}\hat{\mathbf{\Lambda}}\hat{\boldsymbol{\beta}}'$. Ici, $\hat{\mathbf{\Lambda}}$ est une matrice diagonale dont les p éléments diagonaux sont bien ordonnés et $\hat{\boldsymbol{\beta}} \in \mathcal{SO}_p$. Nous rappelons que par le théorème central limite multivarié, nous avons que $\sqrt{n}(\hat{\mathbf{S}} - \mathbf{I}_p) = O_{\mathrm{P}}(1)$.

3.1 Montrer que $\sqrt{n}(\hat{\mathbf{\Lambda}} - \mathbf{I}_p) = O_{\mathrm{P}}(1)$ pour $n \to \infty$.

$$\begin{aligned} \sqrt{n}(\hat{\mathbf{S}} - \mathbf{I}_p) &= O_{\mathrm{P}}(1) \\ \sqrt{n}(\hat{\mathbf{S}} - \mathbf{I}_p) &= \sqrt{n}(\hat{\boldsymbol{\beta}}\hat{\mathbf{\Lambda}}\hat{\boldsymbol{\beta}}' - \hat{\boldsymbol{\beta}}\mathbf{I}_p\hat{\boldsymbol{\beta}}') = \hat{\boldsymbol{\beta}}\sqrt{n}(\hat{\mathbf{\Lambda}} - \mathbf{I}_p)\hat{\boldsymbol{\beta}}' \end{aligned}$$

Aussi, nous avons que:

$$\hat{\boldsymbol{\beta}}'\sqrt{n}(\hat{\mathbf{S}} - \mathbf{I}_p)\hat{\boldsymbol{\beta}} = \sqrt{n}(\hat{\mathbf{\Lambda}} - \mathbf{I}_p)$$

Alors, on sait que

$$\|\hat{\beta}\sqrt{n}(\hat{\mathbf{\Lambda}}-\mathbf{I}_p)\hat{\beta}'\| \leq \underbrace{\|\hat{\beta}\|}_{1}\,\|\sqrt{n}(\hat{\mathbf{\Lambda}}-\mathbf{I}_p)\|\,\underbrace{\|\hat{\beta}'\|}_{1} \implies \|\sqrt{n}(\hat{\mathbf{S}}-\mathbf{I}_p)\| \leq \|\sqrt{n}(\hat{\mathbf{\Lambda}}-\mathbf{I}_p)\|$$

et que

$$\|\hat{\beta}'\sqrt{n}(\hat{\mathbf{S}}-\mathbf{I}_p)\hat{\beta}\| \leq \underbrace{\|\hat{\beta}'\|}_{1}\,\|\sqrt{n}(\hat{\mathbf{S}}-\mathbf{I}_p)\|\,\underbrace{\|\hat{\beta}\|}_{1} \implies \|\sqrt{n}(\hat{\mathbf{\Lambda}}-\mathbf{I}_p)\| \leq \|\sqrt{n}(\hat{\mathbf{S}}-\mathbf{I}_p)\|$$

$$\implies \|\sqrt{n}(\hat{\mathbf{S}}-\mathbf{I}_p)\| = \|\sqrt{n}(\hat{\mathbf{\Lambda}}-\mathbf{I}_p)\|$$

Comme $\sqrt{n}(\hat{\mathbf{S}}-\mathbf{I}_p)$ est bornée un probabilité, cet à dire:

Soit $\varepsilon > 0, \exists M > 0$ et $N \in \mathbb{N}$ tel que $\mathbb{P}(\|\sqrt{n}(\hat{\mathbf{S}}-\mathbf{I}_p)\| > M) < \varepsilon, \forall n > N$, alors on sait aussi que:

Soit $\varepsilon > 0, \exists M > 0$ et $N \in \mathbb{N}$ tel que $\mathbb{P}(\underbrace{\|\sqrt{n}(\hat{\mathbf{\Lambda}}-\mathbf{I}_p)\|}_{=\|\sqrt{n}(\hat{\mathbf{S}}-\mathbf{I}_p)\|} > M) < \varepsilon, \forall n > N.$

$$\implies \sqrt{n}(\hat{\mathbf{\Lambda}}-\mathbf{I}_p) = O_{\mathrm{P}}(1).$$

3.2 Montrer que pour toute matrice $\mathbf{\Theta} \in \mathcal{SO}_p$, $\mathbf{\Theta}\sqrt{n}(\hat{\mathbf{S}}-\mathbf{I}_p)\mathbf{\Theta}' \stackrel{\mathcal{D}}{=} \sqrt{n}(\hat{\mathbf{S}}-\mathbf{I}_p)$.

$$\begin{aligned}
\Theta\sqrt{n}(\hat{\mathbf{S}}-\mathbf{I}_p)\Theta' &= \sqrt{n}(\Theta\,\hat{\mathbf{S}}\,\Theta'-\mathbf{I}_p)\\
&= \sqrt{n}\left(\Theta\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})(X_i-\overline{X})'\right]\Theta'-\mathbf{I}_p\right)\\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\Theta(X_i-\overline{X})[\Theta(X_i-\overline{X})]'-\mathbf{I}_p\right)
\end{aligned}$$

Alors, $\Theta\sqrt{n}(\hat{\mathbf{S}}-\mathbf{I}_p)\Theta'$ est resultat d'une rotation de X tel que $Y = \Theta X$

$$X \sim \mathcal{N}_p(O,\mathbf{I}_p) \implies f^X(x) = \frac{1}{2\pi^{\frac{p}{2}}}\exp\left(-\frac{1}{2}\|x\|^2\right)$$

Changement de variables:

$$\begin{aligned}
f^Y(y) &= \left(\frac{1}{2\pi}\right)^{\frac{p}{2}}\exp\left(-\frac{1}{2}\|\Theta' y\|^2\right)\underbrace{\left|D_{\phi^{-1}(y)}\right|}_{1}\\
&= \left(\frac{1}{2\pi}\right)^{\frac{p}{2}}\exp\left(-\frac{1}{2}\|y\|^2\right),
\end{aligned}$$

où $\left|D_{\phi^{-1}(y)}\right|$ est le valeur absolue du déterminant de la Jacobienne de ϕ^{-1} en y, et où nous avons utilisé les resultats suivants:

- $Y = \Theta X \implies X = \Theta' Y$
  $\implies \left|D_{\phi^{-1}(y)}\right| = \left|Det\left(\frac{dX}{dY}\right)\right| = |Det\,(\Theta')| = 1$
- $\|\Theta' y\|^2 = y'\Theta\Theta' y = \|y\|^2$

Alors, nous arrivons que:

$$Y \sim \mathcal{N}_p(O, \mathbf{I}_p)$$

$\implies \Theta\mathbf{X} \stackrel{\mathcal{D}}{=} \mathbf{X}$ ($\mathbf{X}$ est invariant par rotation).

$$\implies \frac{1}{n}\sum_{i=1}^{n}\Theta(X_i - \overline{X})(X_i - \overline{X})'\Theta' \stackrel{\mathcal{D}}{=} \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})'$$

$$\implies \sqrt{n}(\Theta\,\hat{\mathbf{S}}\,\Theta' - \mathbf{I}_p) \stackrel{\mathcal{D}}{=} \sqrt{n}(\hat{\mathbf{S}} - \mathbf{I}_p)$$

$$\implies \Theta\sqrt{n}(\hat{\mathbf{S}} - \mathbf{I}_p)\Theta' \stackrel{\mathcal{D}}{=} \sqrt{n}(\hat{\mathbf{S}} - \mathbf{I}_p)$$

3.3 Montrer qu'il existe $\tilde{\boldsymbol{\Theta}} \in \mathcal{SO}_p$ tel que $\tilde{\boldsymbol{\Theta}}\sqrt{n}(\hat{\boldsymbol{\Lambda}} - \mathbf{I}_p)\tilde{\boldsymbol{\Theta}}'$ et $\sqrt{n}(\hat{\boldsymbol{\Lambda}} - \mathbf{I}_p)$ ne convergent pas vers la même distribution lorsque $n \to \infty$.

Indice: Rappelez-vous que les valeurs propres de $\hat{\boldsymbol{\Lambda}}$ sont *ordonnées*.

Notons que $\hat{\Lambda}$ a ces valeurs propres ordonées: $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_p$

Observons que pour

$$\tilde{\Theta} = \left(\begin{array}{cc|cccc} 0 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 \\ \hline 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & & \ddots & \\ 0 & 0 & 0 & 0 & \dots & 1 \end{array}\right),$$

nous avons que:

$$\tilde{\Theta}\hat{\Lambda}\tilde{\Theta}' = \begin{pmatrix} \hat{\lambda}_2 & 0 & 0 & \dots & 0 \\ 0 & \hat{\lambda}_1 & 0 & \dots & 0 \\ 0 & 0 & \hat{\lambda}_3 & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \hat{\lambda}_p \end{pmatrix}$$

Prenons l'évènement

$$A = \left\{ \begin{pmatrix} a & 0 & 0 & \dots & 0 \\ 0 & b & 0 & \dots & 0 \\ 0 & 0 & c_1 & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 0 & c_{p-2} \end{pmatrix} \mid a > b \right\}$$

Comme $\hat{\lambda}_1 > \hat{\lambda}_2$, alors $\mathbb{P}[\sqrt{n}(\hat{\Lambda} - I_p) \in A] = 1$, et $\mathbb{P}[\sqrt{n}(\tilde{\Theta}\hat{\Lambda}\tilde{\Theta}' - I_p) \in A] = 0$.

Donc, nous avons que

$$\mathbb{P}[\lim_{n\to\infty} \sqrt{n}(\hat{\Lambda} - I_p) \in A] = \lim_{n\to\infty} \mathbb{P}[\sqrt{n}(\hat{\Lambda} - I_p) \in A] = 1$$

et

$$\mathbb{P}[\lim_{n\to\infty} \sqrt{n}(\tilde{\Theta}\hat{\Lambda}\tilde{\Theta}' - I_p) \in A] = \lim_{n\to\infty} \mathbb{P}[\sqrt{n}(\tilde{\Theta}\hat{\Lambda}\tilde{\Theta}' - I_p) \in A] = 0$$

Alors $\tilde{\Theta}\sqrt{n}(\hat{\Lambda} - I_p)\tilde{\Theta}'$ et $\sqrt{n}(\hat{\Lambda} - I_p)$ ne convergent pas vers le même ditribuition.

3.4 Montrer que les trois premières questions impliquent qu'il n'existe pas de matrice $\boldsymbol{\beta}$ telle que $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + o_{\mathrm{P}}(1)$ pour $n \to \infty$.

Raisonnement par l'absurde:

Soit $\hat{\beta}\hat{\Lambda}\hat{\beta}' = \hat{S}$. Supposons que $\hat{\beta} \xrightarrow{\mathcal{D}} \beta$.

Choisissons Θ tel que $\beta'\Theta\beta = \tilde{\Theta}$ (la matrice de l'exercice précedent) et $\mathbf{Y}$ tel que $\sqrt{n}(\hat{\Lambda} - I_p) \xrightarrow{\mathcal{D}} \mathbf{Y}$.

Alors on a

$$\begin{aligned} \Theta\sqrt{n}(\hat{S} - I_p)\Theta' &\overset{\mathcal{D}}{=} \sqrt{n}(\hat{S} - I_p) \qquad (Ex.\ 3.2) \\ \Theta\hat{\beta}\sqrt{n}(\hat{\Lambda} - I_p)\hat{\beta}'\Theta' &\overset{\mathcal{D}}{=} \hat{\beta}\sqrt{n}(\hat{\Lambda} - I_p)\hat{\beta}' \\ \hat{\beta}'\Theta\hat{\beta}\sqrt{n}(\hat{\Lambda} - I_p)(\hat{\beta}'\Theta\hat{\beta})' &\overset{\mathcal{D}}{=} \sqrt{n}(\hat{\Lambda} - I_p) \end{aligned}$$

Par le lemme de Slutzky, on a que, pour $\hat{\beta}'\Theta\hat{\beta} \xrightarrow{P} \tilde{\theta}$ et $\sqrt{n}(\hat{\Lambda} - I_p) \xrightarrow{\mathcal{D}} \mathbf{Y}$,

$$(\hat{\beta}'\Theta\hat{\beta})\sqrt{n}(\hat{\Lambda} - I_p)(\hat{\beta}'\Theta\hat{\beta})' \xrightarrow{\mathcal{D}} \tilde{\Theta}\mathbf{Y}\tilde{\Theta}',$$

de sorte que $\tilde{\Theta}\mathbf{Y}\tilde{\Theta}' \overset{\mathcal{D}}{=} \mathbf{Y}$, puisque $\hat{\beta}'\Theta\hat{\beta}\sqrt{n}(\hat{\Lambda} - I_p)(\hat{\beta}'\Theta\hat{\beta})' \overset{\mathcal{D}}{=} \sqrt{n}(\hat{\Lambda} - I_p)$.

Ceci est une contradiction, puisqu'on a vu dans l'exercice précedent que

$\tilde{\Theta}\sqrt{n}(\hat{\Lambda} - I_p)\tilde{\Theta}'$ ne converge pas vers la même loi que $\sqrt{n}(\hat{\Lambda} - I_p)$.

$$\implies \hat{\beta} \overset{\mathcal{D}}{\nrightarrow} \beta$$

$$\implies \hat{\beta} \overset{P}{\nrightarrow} \beta$$

3.5 En considérant le résultat obtenu au point 4, est-il pertinent de donner une interprétation du type de celle donnée dans l'exercice 2.3 aux différents vecteurs propres contenus dans $\hat{\boldsymbol{\beta}}$?

Au exercice 3.4, nous avons montré que $\hat{\beta}$ n'est pas un estimateur convergent de β. Ainsi, dans le cas spécifique où $\mathbf{\Sigma} = \mathbf{I}_p$, nous pouvons conclure que l'ACP n'est pas un outil très intéressant, puisque chaque composante $\hat{\beta}_i$ ne contribuera pas à expliquer plus de la vraie variance que les autres vecteurs propres $\hat{\beta}_j, \forall j < i$.

Visuellement, nous constatons qu'une grande quantité de données, lorsque $\mathbf{\Sigma} = \mathbf{I}_p$, remplit une sphère p-dimensionnelle, de telle sorte qu'aucune base orthogonale ne permettra à un axe de capturer plus de variance qu'un autre axe d'une autre base orthogonale. Petites variations dans les données nous conduiront à des estimations différentes des vecteurs propres, aucun desquels n'aura de signification.