République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université de Batna 2

Faculté de Mathématiques & Informatique
Département de Mathématiques

Cours

Analyse des Données

Destiné aux étudiants de L2 SAD

Chargé de cours :
M^{elle} TAMITI Kenza

Année Universitaire : 2019/2020

Avant propos

Ces chatiptres s'adresse aux étudiants de SAD, d'Informatiques ainsi qu'aux étudiants d'économie et gestion et à tous ceux désirant s'initier à l'analyse des données.

Il est conseillé aux lecteurs d'avoir de bonnes connaissances d'Algèbre linéaire (Calculmatricielle, diagonalisaton, normes matricielles) et quelques notions d'Analyse mathématiques (espace métrique, produit scalaire, recherche d'extrema).

.

Introduction

Le but des statistiques est de dégager les significations de données, numériques ou non, obtenues lors de l'étude d'un certain phénomène donné. Les statistiques peuvent être vue en fonction de l'objectif fixé. Des méthodes statistiques sont employées pour explorer des données (approches desriptives et graphiques) ou pour prédire un certain comportement (approches décisionnelles, inferentielles et prédictives).

La statistique mathématiques ou inductive est un ensemble de méthodes permettant de faire des prévisions et des interpolations à partir de caractères observés sur chaque individu de la population à étudier. L'analyse des données est un ensemble de méthodes qui s'inscrivent dans le cadre de la statistique exploratoire multidimentionelle et peuvent également servir la statistique prédictive. Le développement des méthodes d'analyse des données a commencé durant les années 50 avec le développement de l'informatique et le stockage des données, qui depuis n'a cessé de croître.

Actuellement, l'analyse des données fais toujours l'objet de recherche pour s'adapter à tous type de données et faire face aux considérations de traitements en temps réel. Les méthodes développées sont souvent, intégré avec des méthodes issues de l'informatique et de l'intelligence artificielle dans le data mining (en français "fouille de données" ou encore extraction de l'information à partir de données).

Domaines d'application

Aujourd'hui, les méthodes d'analyse des données sont employées dans de nombreux domaines tel que le marketing dans la gestion de la clientèle, le sondage dans l'analyse d'enquêtes, on peut également, citer la recherche documentaire très utile spécialement dans la recherche avec internet.

Le très grand nombre de données en méthodologie a été une des premières motivations pour développer les méthodes d'analyse des données. En effet, tout domaine scientifique qui doit gérer un grand nombre de données de type varié ont recours à ces approches (comme écologie, linguistique, économie etc.), ainsi que tout domaine industriel(tel que les assurances, banques, téléphonie etc.). Ces dernières sont également utilisées en traitement de signal et image et en ingénierie mécanique.

Les données

Nous considérons tout à bord une population de taille $n < \infty$ sur la quelle se fera l'étude d'un certain phénomène. La population est décrite par un ensemble de p caractères (ou variables) qui peuvent être de type quantitatifs ou qualitatifs. Les données se présente sous forme d'un tableau ou de matrices à n lignes et p colonnes.

$$X_{(n,p)} = \begin{pmatrix} x_{11} & . & . & . & x_{1p} \\ & & . & & \\ & & . & & \\ & & x_{ij} & & \\ & & . & & \\ & & . & & \\ x_{n1} & . & . & . & x_{np} \end{pmatrix}$$

Cette représentation de données peut faciliter la lecture des tableaux de petite dimension. Cependant, dés que la taille de la population n est grande ou le nombre de caractères p est élevé les techniques simples des statistiques descriptives ne suffisent plus.
En règle générale, il est difficile lorsque le volume des données est important, de tirer des enseignements utiles sans traiter au préalable.

Les objectifs

L'analyse des données regroupe deux familles de méthodes suivant l'objectif fixé.

1. Une première famille de méthodes cherche à représenter de grands ensembles de données par peu de caractères ie. chercher les dimensions pertinentes de ces données. Parmi ces méthodes nous citons l'analyse en composantes principales (ACP) et (ACP) normée qui sera présenté dans cet ouvrage, analyse des correspondances (AFC), analyse des correspondances multiples AFC multiples ou encore l'analyse canonique.

2. Une deuxième famille de méthodes cherche à classer les données de manière automatique tel que la classification par partition, classification hiérarchique ou analyse discriminante.

Les logiciels

Dans certains logiciels sont intégrés les méthodes d'analyse de données. On peut citer les logiciels SAS, splus, R, XIStat, UniWinplus, Stalab et SPAD.

CHAPITRE 1

Préliminaires

Introduction

Dans ce chapitre, nous présentons dans une première quelques définitions et outils de base utilisé en analyse des données.

L'analyse statistique multivariée consiste à analyser et comprendre des données de grandes dimension.
L'analyse factorielle est une famille de méthodes géométriques dont les objectifs sont :

Résumer l'information contenues dans un tableau de données. En remplaçons ce dernier par un tableau de plus faible dimension.

Visualiser cette information. En éliminant les redondances d'information contenues dans le tableau de départ.

De l'analyse factorielle, plusieurs variantes se sont développées, tel que l'analyse en composantes principale (ACP) et ACP normée, l'analyse factorielle des correspondances AFC, l'analyse factorielle des correspondances multiple AFCM et l'analyse discriminante.

1.1 Population et variables

Définition 1.1.1 *On appelle population, un ensemble fini de personnes ou d'objets, noté par I sur lequel se fera l'étude. Un élément de la population est dit individu et le cardinal de I est dit taille de la population, notée n.*

Définition 1.1.2 *Une variable ou caractère est une application qui à chaque individu associe un élément d'un ensemble appelé ensemble des modalités du caractère.*
On note

$$\begin{aligned} X \quad : &\quad I \to E \\ : &\quad I_i \to X(I_i) = X_i \end{aligned}$$

Si $E \subset \mathbb{R}$ *la variable* X *est dite variable quantitative.*
Si $E \subset \mathbb{R}^p$ *la variable* X *est dite vectorielle.*
Si E *échappe à la mesure,* X *est dite variable qualitative.*

1.2 Tableau des données

Soient I une population de taille n, $X^1; ...; X^p$ p variables quantitatives observées sur les individus de I. On construit un tableau de données quantitatives à partir de la population et les p variables, on pose

$$X = \begin{pmatrix} x_1^1 & ... & x_1^p \\ . & & . \\ . & x_i^j & . \\ . & & . \\ x_n^1 & ... & x_n^p \end{pmatrix}$$

Où x_i^j est la valeur du $i^{ème}$ individu par la variable X^j.

Notations

On note par

$$X^j = \begin{pmatrix} x_1^j \\ . \\ . \\ . \\ x_n^j \end{pmatrix} \in \mathbb{R}^n,$$

la $j^{ème}$ variabe mesurée sur les n individus.

Et par

$${}^tX_i = \begin{pmatrix} x_i^1 \\ . \\ . \\ . \\ x_i^p \end{pmatrix} \in \mathbb{R}^p,$$

le $i^{ème}$ individu observée sur les p variables.

On suppose que chaque individu est muni d'un poids p_i.

$$N(I) = \{({}^tX_i, p_i) / {}^tX_i \in \mathbb{R}^p \; pour \; i = 1, ..., n\}$$

le nuage des individus.

$$N(J) = \{X^j / X^j \in \mathbb{R}^n \ pour \ j = 1, ..., p\}$$

le nuage des variables.

$$z = \begin{pmatrix} 1 \\ . \\ . \\ . \\ 1 \end{pmatrix} \in \mathbb{R}^n$$

$$D_p = \begin{pmatrix} p_1 & & & & \\ & . & & & \\ & & . & & \\ & & & . & \\ & & & & p_n \end{pmatrix}$$

la matrice diagonale des poids.

1.3 Eléments descriptifs du nuage des variables

Moyenne empirique

Pour $j = 1, ..., p$

$$\overline{X}^j = \sum_{i=1}^{n} p_i x_i^j$$

On peut facilement vérifier que

$$\overline{X}^j = \left\langle X^j, z \right\rangle_{D_p} =^t X^j D_p z.$$

La moyenne empirique est donc la projection D_p-orthogonale du point X^j de $\mathbb{R}^n$sur la droite portée par le vecteur z.

La variance

$$Var(X^j) = \sum_{i=1}^{n} p_i(x_i^j - \overline{X}^j)^2$$

On a aussi

$$\begin{aligned} Var(X^j) &= \left\langle X^j - \overline{X}^j z, X^j - \overline{X}^j z \right\rangle_{D_p} \\ &= {}^t\left(X^j - \overline{X}^j z\right) D_p \left(X^j - \overline{X}^j z\right) \\ &= \left\| X^j - \overline{X}^j z \right\|_{D_p}^2 \\ &= \sigma^2 \end{aligned}$$

La variance est le carré de la norme relativement à la métrique D_p du vecteur centré

$$\widetilde{X}^j = X^j - \overline{X}^j z.$$

L'ecart-type σ_j est sa norme.

Covariance

$$cov(X^j, X^{j'}) = \sum_{i=1}^{n} p_i(x_i^j - \overline{X}^j)(x_i^{j'} - \overline{X}^{j'})$$

On a

$$\begin{aligned} cov(X^j, X^{j'}) &= \left\langle X^j - \overline{X}^j z, X^{j'} - \overline{X}^{j'} z \right\rangle_{D_p} \\ &= {}^t\left(X^j - \overline{X}^j z\right) D_p \left(X^{j'} - \overline{X}^{j'} z\right) \end{aligned}$$

La covariance est la projection du vecteur centré $\widetilde{X}^j = X^j - \overline{X}^j z$ sur la droite portée par le vecteur centré $\widetilde{X}^{j'} = X^{j'} - \overline{X}^{j'} z$.

Coefficient de corrélation linéaire

$$\begin{aligned}
\rho_{jj^i} &= \frac{cov(X^j, X^{j'})}{\sqrt{Var(X^j)Var(X^{j'})}} \\
&= \frac{\left\langle X^j - \overline{X}^j z, X^{j'} - \overline{X}^{j'} z \right\rangle_{D_p}}{\left\| X^j - \overline{X}^j z \right\|_{D_p} \left\| X^{j'} - \overline{X}^{j'} z \right\|_{D_p}} \\
&= \frac{\left\| X^j - \overline{X}^j z \right\|_{D_p} \left\| X^{j'} - \overline{X}^{j'} z \right\|_{D_p}}{\left\| X^j - \overline{X}^j z \right\|_{D_p} \left\| X^{j'} - \overline{X}^{j'} z \right\|_{D_p}} \cos\theta \\
&= \cos\theta
\end{aligned}$$

θ est l'angle entre les vecteurs centrés $\widetilde{X}^j = X^j - \overline{X}^j z$ et $\widetilde{X}^{j'} = X^{j'} - \overline{X}^{j'} z$.

Matrice des variances-covariances

La matrice des variances-covariances est la matrice symétrique suivante

$$V =^t \widetilde{X} D_p \widetilde{X}$$

$\widetilde{X}$ est la matrice centrée donnée par

$$\widetilde{X} = \left(\widetilde{x}_i^j\right)_{\substack{1\leq i\leq n \\ 1\leq j\leq p}} = \left(x_i^j - \overline{X}^j\right)_{\substack{1\leq i\leq n \\ 1\leq j\leq p}}$$

Matrice des corrélations

La matrice des corrélations est la matrice symétrique suivante

$$R =^t \widetilde{\widetilde{X}} D_p \widetilde{\widetilde{X}}$$

Avec

$$\widetilde{\widetilde{X}} = \left(\frac{x_i^j - \overline{X}^j}{\sigma_j}\right)_{\substack{1\leq i\leq n \\ 1\leq j\leq p}}$$

est la matrice centrée-réduite.

On pose

$$D_{\frac{1}{\sigma}} = \begin{pmatrix} \frac{1}{\sigma_1} & & & & \\ & . & & & \\ & & . & & \\ & & & . & \\ & & & & \frac{1}{\sigma p} \end{pmatrix}$$

On aura

$$R = D^t_{\frac{1}{\sigma}} \widetilde{X} D_p \widetilde{X} D_{\frac{1}{\sigma}}$$

1.4 Eléments descriptifs du nuage des individus

On muni $\mathbb{R}^p$ de la métrique identité I_p

Centre de gravité

$$g = \sum_{i=1}^{n} p_i^t X_i = \begin{pmatrix} \overline{X}^1 \\ . \\ . \\ . \\ \overline{X}^j \end{pmatrix}$$

Inertie au point $\alpha \in \mathbb{R}^p$

Définition 1.4.1 *On appelle inertie au point $\alpha \in \mathbb{R}^p$ la quantité*

$$I_\alpha = \sum_{i=1}^{n} p_i \left\| {}^t X_i - \alpha \right\|^2_{I_p}.$$

Remarque 1.4.1 *L'inertie est minimale au centre de gravité. Inertie au centre de gravité est donnée par*

$$I_g = \sum_{i=1}^{n} p_i \left\| {}^t X_i - g \right\|^2_{I_p}.$$

En posant ${}^t\widetilde{X}_i ={}^t X_i - g$, on aura

$$\begin{aligned} I_g &= \sum_{i=1}^{n} p_i \left\| {}^t\widetilde{X}_i \right\|_{I_p}^2 \\ &= \sum_{i=1}^{n} p_i \widetilde{X}_i^t \widetilde{X}_i. \end{aligned}$$

Comme $\widetilde{X}_i\ {}^t\widetilde{X}_i$ est un scalaire, alors

$$\widetilde{X}_i^t \widetilde{X}_i = tr\left(\widetilde{X}_i^t \widetilde{X}_i\right) = tr\left({}^t\widetilde{X}_i\ \widetilde{X}_i\right),$$

et

$$I_g = \sum_{i=1}^{n} p_i tr\left({}^t\widetilde{X}_i\ \widetilde{X}_i\right) = tr\left(\sum_{i=1}^{n} p_i^t\ \widetilde{X}_i \widetilde{X}_i\right) = tr(V).$$

CHAPITRE 2

Analyse Factorielle Générale AFG

Introduction

On considèez un tableau de données quantitatives $X_{(n,p)}$ tel que :

$n =$ nombre d'individus.

$p =$ nombre de variables supposées corrélées.

Le problème est de réduire le nombre de p variables initiales.

L'**Analyse Factorielle Générale** $(A.F.G)$ permet de resoudre ce type de problème, en synthetisant, visualisant et éléminant les redendances d'informations

$$\begin{array}{ccc} p & & nouvelles \\ variables & A.F.G & variables \\ initiales & & indépendante \\ X^1, ..., X^p & \Longrightarrow & Y^1, ..., Y^p \end{array}$$

Remarque 2.0.2 *Le tableau X est un tableau à deux entiers*

$$X = (X^1, ..., X^p)$$

$X^j \in \mathbb{R}$ est la $j^{ème}$ colonne de X

$$X = \begin{pmatrix} X_1 \\ . \\ . \\ . \\ X_n \end{pmatrix}$$

${}^tX_i \in \mathbb{R}^p$, X_i est la $i^{ème}$ ligne de X.

2.1 A.F.G de tableau X

On considère le nuage des p points de $\mathbb{R}^n$

$$N(J) = \left\{ X^j / \ X^j \in \mathbb{R}^n \right\}.$$

À chaque variable V^j on associe un vecteur de $\mathbb{R}^n$

$$V^j \longmapsto X^j = \begin{pmatrix} x_1^j \\ . \\ . \\ . \\ x_n^j \end{pmatrix}$$

qui sont les mesures des n individues. Par la suite,on posera $V^j = X^j$ l'espace de variable est $\mathbb{R}^n$.

Le nuage des n points de $\mathbb{R}^p$

$$N(I) = \left\{ {}^tX_1, ..., {}^tX_n \ / \ {}^tX_i \in \mathbb{R}^p \right\}.$$

À chaque individu ω_i on associe le vecteur X_i

$$\omega_i \longmapsto X_i = \left(x_i^1, ..., x_i^p \right)$$

${}^tX_i$ est la mesure du $i^{ème}$ individus par les p variables

L'espace des individus est $\mathbb{R}^p$.

On dit que : une ligne du tableau X est un **individu** et une colonne de X est une **variable**.

On se pose sur $\mathbb{R}^p$

Soit le nuage $N(I) \subset \mathbb{R}^p$ reduit le nombre de variable.

Pour $i = 1, ..., n$

$$\begin{array}{ccc} & & X_i / \ {}^tX_i \in \mathbb{R}^p \\ & \nearrow & \\ I_i & & \\ & \searrow & \\ & & \omega_i \in \mathbb{R}^q \end{array}$$

Cela revient à projeter chaque point de $\mathbb{R}^p$ dans un sous espace de dimension q $(q < p)$. Le problème revient alors à chercher le sous espace de $\mathbb{R}^p$ de dim q.

En utilisant l'A.F.G, on cherche dabord un s-e.v de $\mathbb{R}^p$ de dim1 qui ajuste au mieux les n points de $\mathbb{R}^p$, on mesure l'information apportée par ce s-e.v si cette dernière est considerable, ie: la perte d'information est négligeable, on remplace les p variables initiales $(X_i, i = \overline{1,n})$ par une seule nouvelle variable $(\omega_i, i = \overline{1,n})$ sinon on cherche un s.e.v de dim2.

Recherche du s.e.v de dim1

On note par E_1 le s-e.v de $\mathbb{R}^p$, de dim1, et par$\{$ u$\}$ base de E_1, on suppose que $\|$ u$\| = 1$, $u \in E \subset \mathbb{R}^p$.

$\|.\|_p$ la norme de $\mathbb{R}^p$ relativement à la métrique identité.

Le problème revient à trouver u $\in \mathbb{R}^p$.

Remarque 2.1.1 *E_1 est une droite de $\mathbb{R}^p$, de vecteur directeur normé U*
Chaque point ${}^tX_i \in \mathbb{R}^p$ est projeté sur E_i au point ω_i, u $\in E_1$

$$\begin{aligned} \omega_i &= \left\langle {}^tX_i, \mathrm{u}\right\rangle_{I_p} \\ &= {}^t\mathrm{u}\,{}^tX_i \\ &= X_i\,\mathrm{u}. \end{aligned}$$

Avec

$$\left\|{}^tX_i\right\|_{I_p}^2 = \omega_i^2 + \varepsilon_i^2$$
$$\varepsilon_i^2 = \left\|{}^tX_i\right\|_{I_p}^2 - \omega_i^2$$

Pour trouver u*, on utilise le critère des moindres carrés qui consiste à minimiser la somme des erreurs quadratique*

$$\begin{aligned} \min_{\|\mathrm{u}\|_p^2=1} \sum_{i=1}^n \varepsilon_i^2 &\Leftrightarrow min_{\|\mathrm{u}\|_p^2=1} \sum_{i=1}^n \left[\left\|{}^tX_i\right\|_{I_p}^2 - \omega_i^2\right] \\ &\Leftrightarrow \max_{\|\mathrm{u}\|_{I_p}^2=1} \sum_{i=1}^n \omega_i^2 \\ &\Leftrightarrow \max_{\|\mathrm{u}\|_p^2=1} \|\omega\|_{I_p}^2 \qquad \omega = \begin{pmatrix} \omega_1 \\ . \\ . \\ . \\ \omega_n \end{pmatrix}. \end{aligned}$$

Le lagrangier

$$L\left(\omega\right)=\|\omega\|_{I_p}^{2}-\lambda\left(\|\,\mathrm{u}\|^{2}-1\right)$$

$$\omega_i = X_i\,\mathrm{u} \qquad \omega = \begin{pmatrix} X_1 u \\ . \\ . \\ . \\ X_n u \end{pmatrix} = X\,\mathrm{u}$$

$$\begin{aligned}
\|\omega\|_{I_p}^{2} &= {}^{t}\omega I_p\omega \\
&= {}^{t}\left(X\,\mathrm{u}\right)\left(X\,\mathrm{u}\right) \\
&= {}^{t}\,\mathrm{u}^{t}XX\,\mathrm{u}
\end{aligned}$$

$$L(\,\mathrm{u}) =^{t}\ \mathrm{u}^{t}XX\,\mathrm{u}-\lambda\left({}^{t}\,\mathrm{u}\,\mathrm{u}-1\right)$$

$$\max L(\,\mathrm{u}) \Longleftrightarrow \begin{Bmatrix} \frac{\partial \mathrm{L}}{\partial\,\mathrm{u}} = 0 \\ \frac{\partial^2 \mathrm{L}}{\partial\,\mathrm{u}}\ |_{\hat{u}} < 0 \end{Bmatrix}$$

Rappel 1

1. $f(\,\mathrm{u}) = \|\,\mathrm{u}\|^{2} = \langle\,\mathrm{u},\,\mathrm{u}\rangle$
   2. $h \cdot \|h\| \to 0$
   3.

$$\begin{aligned}
f(\,\mathrm{u},h)-f(\,\mathrm{u}) &= \langle\,\mathrm{u}+h,\,\mathrm{u}+h\rangle-\langle\,\mathrm{u},\,\mathrm{u}\rangle \\
&= \langle\,\mathrm{u},\,\mathrm{u}\rangle+\langle h,\,\mathrm{u}\rangle+\langle\,\mathrm{u},h\rangle+\langle h,h\rangle-\langle\,\mathrm{u},\,\mathrm{u}\rangle \\
&= 2\,\langle h,\,\mathrm{u}\rangle+\|h\|^{2} \\
&= 2^{t}\,\mathrm{u}h+o(h)
\end{aligned}$$

   4. $\frac{\partial}{\partial\,\mathrm{u}}({}^{t}\,\mathrm{u}\,\mathrm{u}) = 2^{t}\,\mathrm{u}$

   $\frac{\partial}{\partial\,\mathrm{u}}\left({}^{t}\,\mathrm{u}^{t}XX\,\mathrm{u}\right) = 2^{t}\,\mathrm{u}^{t}XX$

5.

$$\begin{aligned}
\frac{\partial}{\partial\,\mathrm{u}}\mathrm{L}(\mathrm{u}) &= 2^t\,\mathrm{u}^tXX - 2\lambda^t\,\mathrm{u} = 0 \Leftrightarrow^t\ \mathrm{u}^tXX = \lambda^t\,\mathrm{u}\\
&\Leftrightarrow\ {}^t\left({}^t\,\mathrm{u}^tXX\right) =^t \left(\lambda^t\,\mathrm{u}\right)\\
&\Leftrightarrow\ \left({}^tXX\right)\,\mathrm{u} = \lambda\,\mathrm{u}.
\end{aligned}$$

u est un vecteur propre nomré de la matrice ${}^tXX$ associe à la valeur propre λ.

6.

$$\begin{aligned}
\max_{\|\mathrm{u}\|^2=1}{}^t\omega\omega &\Leftrightarrow \max_{\|\mathrm{u}\|^2=1}{}^t\mathrm{u}({}^tXX\,\mathrm{u})\\
&\Leftrightarrow \max_{\|\mathrm{u}\|^2=1}{}^t\mathrm{u}(\lambda\,\mathrm{u})_{\|\mathrm{u}\|^2_p}\\
&\Leftrightarrow \max_{\|\mathrm{u}\|^2=1}\lambda({}^t\,\mathrm{u}\,\mathrm{u})\\
&\Leftrightarrow \max\lambda.
\end{aligned}$$

Conclusion

le s.e.v de dim1 qui ajuste au mieux les n points de $\mathbb{R}^p$ est la droite passant par l'origine et engendrée par le vecteur directeur normé u qui est vecteur propre de la matrice ${}^tXX$ associé à la plus grande valeur propre.

Rappel 2

1.

$$\left(X^1,...,X^p\right) \rightsquigarrow \omega = \begin{pmatrix}\omega_1\\ .\\ .\\ .\\ \omega_n\end{pmatrix}$$

ω_i est la projection du $i^{ème}$ individu sur le s-e.v de dim1 , $E = \triangle\,\mathrm{u}$

2. $\omega_i = \langle \, {}^tX_i, \, \mathrm{u} \rangle = X_i \, \mathrm{u}$

$$\omega = \begin{pmatrix} \omega_1 \\ . \\ . \\ . \\ \omega_n \end{pmatrix} \mathrm{u} = X \, \mathrm{u}$$

3. $\omega = X \, \mathrm{u} = (X^1, ..., X^p) \begin{pmatrix} \mathrm{u}_1 \\ . \\ . \\ . \\ \mathrm{u}_p \end{pmatrix} = \sum_{j=1}^{p} \mathrm{u}_j \; X^j.$

u vecteur propre de ${}^tXX$, $u \in \mathbb{R}$

Remarque 2.1.2 *La nouvelle variable ω est une combinaison lineaire des p variables initiales $X^1, ..., X^p$ dont les coefficients sont les composantes $\overrightarrow{vp}$ u de ${}^tXX$ associe à la plus grande valeur propre λ.*

Définition 2.1.1 *On appelle axe factoriel le sous espace vectoriel de dim1 engendré par le vecteur u de la matrice ${}^tXX$ à une valeur propre λ.*
Le premier axe factoriel est l'axe factoriel qui correspend à la plus grand valeur propre.

Proposition 2.1.1 *les valeurs propre de ${}^tXX$ sont non négatives.*

Démonstration. soit λ une valeur propre de ${}^tXX$ associe au $\overrightarrow{vp}$ u
On a

$$\begin{aligned} {}^tXX \, \mathrm{u} &= \lambda \, \mathrm{u} \Leftrightarrow^t \mathrm{u}^t XX = \lambda^t \, \mathrm{u} \, \mathrm{u} \\ &\Leftrightarrow \; {}^t(\mathrm{u}X)(X \, \mathrm{u}) = \lambda^t \, \mathrm{u} \, \mathrm{u} \\ &\Leftrightarrow \; \|X \, \mathrm{u}\|^2 = \lambda \, \| \, \mathrm{u}\|^2 . \end{aligned}$$

Comme $\|X \, \mathrm{u}\|^2 \geq 0$ et $\| \, \mathrm{u}\|^2 \geq 0$ alors $\lambda \geq 0$. ■

Remarque 2.1.3 *On suppose que l'information fournie par ω est insuffisante. Il est necessaire donc de concstruire deux nouvelles variables $\omega^{(1)}$ et $\omega^{(2)}$ qui remplaçent les p initiales.*

Proposition 2.1.2 *Le s-e.v de dim2 qui ajuste au mieux les n points de $\mathbb{R}^2$ $({}^tX_i, i = 1, ..., n)$ contient le s.e.v de dim1 qui ajuste au mieux ces n points.*

Démonstration. On note par $E_1 = \triangle\, \mathrm{u}_1$ le s.e.v de $\mathbb{R}^p$ de dim1 $\|\,\mathrm{u}\|^2 = 1$.
où u_1 est le vecteur propre de ${}^tXX$ associe à la grande valeur propre λ_n.
Par E_2 le s-e.v de $\mathbb{R}^p$ de dim2 qui ajuste au mieux les points du nuage $N(I)$ et par $\{V_1, V_2\}$ une base de E_2.
On suppose que $\|V_1\|^2 = \|V_2\|^2 = 1$.
On suppose que $\mathrm{u}_1 \neq V_1$.
On a $E_2 = (\triangle V_1, \triangle V_2)$
On pose $E = (\triangle\, \mathrm{u}_1, \triangle V_1)$
$\triangle\, \mathrm{u}_1$ est le s.e.v qui ajuste au mieux les points de $N(I)$ et $\triangle V_1$ est le s.e.v qui ajuste les points de $N(I)$ on a alors E est meilleurs que E_2 contradiction avec le fait que $E_2 \Rightarrow$ $\mathrm{u}_1 = V_1 \Rightarrow E_1 \subset E_2$
Pour avoir E_2, il suffit de completer la base $\{\,\mathrm{u}_1\}$ par un vecteur $v \perp \mathrm{u}_1$
En reprenant les mêmes étapes précédentes V_1 sera le vecteur propre de ${}^tXX$ associe à la deuxième valeur propre λ_2. ■

Généralisation

Le s-e.v de dim q $(q < p)$ qui ajuste au mieux les points de $N(I)$ est le s-e.v donné par la base $\{\,\mathrm{u}_1, ..., \mathrm{u}_q\}$ où u_k est un $\overrightarrow{vp}$ de ${}^tXX$ associe à la $k^{ème}$ valeur propre λ_k.
$\lambda_1 > \lambda_2 > .... > \lambda_q$.

${}^tXX$ est la matrice à diagonaliser.

$$\left(X^1, ..., X^p\right) \rightarrow \left(\omega^{(1)}, ..., \omega^{(q)}\right)$$

avec

$$\omega^{(k)} = \sum_{j=1}^{p} \mathrm{u}_k^j X^j$$

$$\mathrm{u}_k = \begin{pmatrix} \mathrm{u}_k^1 \\ . \\ . \\ . \\ .\ \mathrm{u}_k^p \end{pmatrix}$$

est le $\overrightarrow{vp}$ associe à λ_k de ${}^tXX$

les nouvelles variables ω sont non corrélées $\left\langle \omega^{(k)}, \omega^{(k')} \right\rangle = 0$

■**si** $n < p$

on se pose sur $\mathbb{R}^n$

On considère le nuage des points de $\mathbb{R}^n$

$$N(J) = \{X^j / X^j \ \in \mathbb{R}^n, j = 1, ..., p\}$$

De la même manière que précédement, chaque point de $N(J)$ est projeté sur un s-e.v de de dimension $1, 2, ..., p$.

Recherche de s.e.v de dim1

On note par s le vecteur de la base du s-e.v de dim 1 sur $\triangle s$.

On suppose que s est normé, $\|s\|^2 = 1$.

Pour trouver s on utilise le critère des Moindre Carrée (MC) qui consiste à minimiser la somme des erreurs quadratique ε_j est l'erreur dûe à la projection $\min_{\|s\|^2=1} \sum_{j=1}^{p} \varepsilon_i^2$.
On a v^j projection de X^j sur $\triangle s$

$$V^j = \left\langle X^j, s \right\rangle =^t X^j s$$

$$
\begin{aligned}
V_{(p,1)} &= \begin{pmatrix} v^1 \\ . \\ . \\ . \\ v^p \end{pmatrix} \\
&= \begin{pmatrix} {}^tX^1 s \\ . \\ . \\ . \\ {}^tX^p s \end{pmatrix} \\
&= \begin{pmatrix} {}^tX^1 \\ . \\ . \\ . \\ {}^tX^p \end{pmatrix} s \\
&= {}^tX_{(p,n)} s_{(n,1)}
\end{aligned}
$$

vecteur des p projections

$$
\begin{aligned}
\varepsilon_i^2 &= \left\|X^j\right\|^2 - V^{j2} = \min_{\|s\|^2=1} \sum_{j=1}^{p} \left\|X^j\right\|^2 - V^{j2} \\
&\Leftrightarrow \max_{\|s\|^2=1} \sum_{j=1}^{p} (V^j)^2 \\
&\Leftrightarrow \max_{\|s\|^2=1} {}^tVV \\
&\Leftrightarrow \max_{\|s\|^2=1} {}^t\left({}^tXs\right)\left({}^tXs\right) \\
&\Leftrightarrow \max_{\|s\|^2=1} {}^tsX{}^tXs......(1).
\end{aligned}
$$

On a L$(s) = {}^ts\ X\ {}^tX\ s - \lambda\left({}^ts\ s - 1\right)$

D'où

$$\begin{aligned}
\frac{\partial}{\partial s}\text{Ł}(s) &= 0 \Leftrightarrow 2^t s X^t X - 2\lambda^t s \\
&\Leftrightarrow {}^t s X\ {}^t X = \lambda^t s \\
&\Leftrightarrow {}^t\left({}^t s\ X\ {}^t X\right) =^t \left(\lambda^t s\right) \\
&\Leftrightarrow \left(X\ {}^t X\right) s = \lambda s.
\end{aligned}$$

s est un $\overrightarrow{vp}$ de $X\ {}^t X$ associe à la valeur propre λ.

En utilisant (1), on aura

$$\max_{\|s\|^2=1}^{t} sX^t Xs \Leftrightarrow \max_{\|s\|^2=1}^{t} s(\lambda s) \Leftrightarrow \max \lambda.$$

Remarque 2.1.4 *Les propositions donées sur l'espace $\mathbb{R}^P$, reste vrais pour $\mathbb{R}^n$.*
$\triangle s$ est dite axe factoriel.

Conclusion

Le s.e.v de $\mathbb{R}^n$ de dim 1, qui ajuste au mieux les n points de $N(J)$ est la droite passant par l'origine et engendrée par le $\overrightarrow{vp}$ de $X\ {}^t X$ associe à la valeur propre λ.

2.2 Relation entre les espaces propres de $\mathbb{R}^n$ et $\mathbb{R}^p$:

Proposition 2.2.1 *Les matrices $X\ {}^t X$ et ${}^t X\ X$ ont les mêmes valeurs propre λ.*

Démonstration. 1/ Soit λ une valeur propre de ${}^t X\ X$ associeé au $\overrightarrow{vp}$ u.
Montrons que λ est aussi la valeur propre de $X\ {}^t X$.
On a

$${}^t X\ X\,\mathrm{u} = \lambda\,\mathrm{u}$$

D'où

$$X\ {}^t X(X\,\mathrm{u}) = \lambda(X\,\mathrm{u})$$

$X\,\mathrm{u} \in \mathbb{R}^n$ est un $\overrightarrow{vp}$ de $X\ {}^t X$ associé à la valeur propre λ.
2/ Soit λ une valeur propre de $X\ {}^t X$ associe au $\overrightarrow{vp}\, s$

Montrons que λ est aussi la valeur propre de de ${}^tX\ X$

On a

$$X\ {}^tXs = \lambda s$$

D'où

$${}^tXX({}^tXs) = \lambda({}^tXs)$$

λ est une valeur propre de ${}^tX\ X$ associeé au $\overrightarrow{vp}$ ${}^tX_{(p,n)}s_{(n,1)} \in \mathbb{R}^p$. ■

Proposition 2.2.2 ***(Formule de transmission)***

Soit u *est un* $\overrightarrow{vp}$ *de* ${}^tX\ X$ *associe à la valeur propre* λ *et* s *un* $\overrightarrow{vp}$ *de* $X\ {}^tX$ *associe à la même valeur propre* λ.

On a alors

$$\left\{\begin{matrix} \mathrm{u}_{(p,1)} = \frac{1}{\sqrt{\lambda}}{}^tX_{(p,n)}s_{(n,1)} \\ s_{(n,1)} = \frac{1}{\sqrt{\lambda}}X_{(n,p)}\,\mathrm{u}_{(p,1)} \end{matrix}\right\}.$$

Démonstration. 1/

i) $\frac{1}{\sqrt{\lambda}}{}^tX\ s\ \overrightarrow{vp}$ de ${}^tX\ X$ associe à la valeur propre λ.

ii) $\left\|\frac{1}{\sqrt{\lambda}}{}^tXs\right\|^2 = 1$

i)${}^tX\ X\left(\frac{1}{\sqrt{\lambda}}{}^tXs\right) = \lambda\left(\frac{1}{\sqrt{\lambda}}{}^tXs\right)$?

En effet,

$$\begin{aligned} {}^tX\ X\left(\frac{1}{\sqrt{\lambda}}{}^tXs\right) &= \frac{1}{\sqrt{\lambda}}{}^tX\left(X\ {}^tXs\right) \\ &= \frac{1}{\sqrt{\lambda}}{}^tX\left(\lambda s\right) \\ &= \lambda\left(\frac{1}{\sqrt{\lambda}}{}^tXs\right). \end{aligned}$$

ii) On a

$$\begin{aligned} \left\|\frac{1}{\sqrt{\lambda}}{}^tXs\right\|^2 &= \left\langle \frac{1}{\sqrt{\lambda}}{}^tXs\ , \frac{1}{\sqrt{\lambda}}{}^tXs \right\rangle \\ &= \frac{1}{\lambda}{}^t({}^tXs)({}^tXs) \\ &= \frac{1}{\lambda}{}^ts(X{}^tXs) \\ &= \frac{1}{\lambda}{}^ts\lambda s \\ &= \|s\|^2 = 1. \end{aligned}$$

2/

i) $\frac{1}{\sqrt{\lambda}}X\,\mathrm{u}$ $\overrightarrow{vp}$ de $X\ {}^tX$ associe à la valeur propre λ

$$\begin{aligned} X^tX\frac{1}{\sqrt{\lambda}}X\,\mathrm{u} &= \frac{1}{\sqrt{\lambda}}X({}^tX\ X\,\mathrm{u}) \\ &= \frac{1}{\sqrt{\lambda}}X(\lambda\,\mathrm{u}) \\ &= \lambda(\frac{1}{\sqrt{\lambda}}X\,\mathrm{u}) \end{aligned}$$

ii)

$$\begin{aligned} \left\|\frac{1}{\sqrt{\lambda}}{}^tXs\right\|^2 &= \frac{1}{\lambda}{}^t(X\,\mathrm{u})(X\,\mathrm{u}) \\ &= \frac{1}{\lambda}{}^t\mathrm{u}({}^tXX\,\mathrm{u}) = \frac{1}{\lambda}{}^t\mathrm{u}(\lambda\,\mathrm{u}) = \|\,\mathrm{u}\|^2 = 1 \end{aligned}$$

* Si on se place sur $\mathbb{R}^p$ la matrice à diagolaniser est ${}^tX_{(p,n)}X_{(n,p)}$.

*Si on se place sur $\mathbb{R}^n$ la matrice à diagolaniser est $X_{(n,p)}\ {}^tX_{(p,n)}$. ■

2.3 Projection simultanée

I) Si $n \geq p$ on se place sur $\mathbb{R}^p$, on diagonalise ${}^tXX$ on note par u_k les $\overrightarrow{vp}$ normés de ${}^tXX$ associé aux valeurs propre λ_k et par s_k les $\overrightarrow{vp}$ normés de X^tX associé aux mêmes valeurs propre λ_k.

2.3.1 Projection des Individus

Pour $i = 1, ..., n$ la projection du $i^{ème}$ individus sur le $k^{ème}$ axe factoriel est

$$\omega^k = X\,\mathrm{u}_k$$

$$\omega_i^{(k)} = \left\langle {}^tX_i\ ,\ \mathrm{u}_k \right\rangle = X_i\,\mathrm{u}_k$$

$$\omega^{(k)} = \begin{pmatrix} \omega_1^{(k)} \\ . \\ . \\ . \\ \omega_n^{(k)} \end{pmatrix} = \begin{pmatrix} X_1 \\ . \\ . \\ . \\ X_n \end{pmatrix} \mathrm{u}_k = X\,\mathrm{u}_k$$

est le vecteur projection des n individus.

2.3.2 Projection des variables

Pour $j = 1, ..., p$, la projection de la $j^{ème}$ variable X^j sur la $k^{ème}$ axe factoriel ($\triangle\, \mathrm{u}_k$) est

$$\begin{aligned} v^{j^{(k)}} &= \left\langle X^j, s_k \right\rangle \\ &= {}^tX^j s_k \\ &= {}^tX^j \left(\frac{1}{\sqrt{\lambda}} X\, \mathrm{u}_k\right) \end{aligned}$$

$$\begin{aligned} V^k &= \begin{pmatrix} v^{1^k} \\ . \\ . \\ . \\ v^{p^k} \end{pmatrix} \\ &= {}^tX s_k \\ &= {}^tX\left(\frac{1}{\sqrt{\lambda}} X\, \mathrm{u}_k\right) \\ &= \frac{1}{\sqrt{\lambda_k}}({}^tXX\, \mathrm{u}_k) = \frac{1}{\sqrt{\lambda_k}}(\lambda_k\, \mathrm{u}_k) = \sqrt{\lambda_k}\, \mathrm{u}_k \end{aligned}$$

$II)$ Si $n \geq p$ on se place sur $\mathbb{R}^n$, on diagonalise X^tX c-à-d on calcule les $\overrightarrow{vp}$ normés s_k

Projection des Individus

Pour $i = 1, ..., n$ on a $\quad \omega_i^{(k)} = X_i \;\; \mathrm{u}_k$

$$\begin{aligned}
\omega^{(k)} &= \begin{pmatrix} \omega_1^{(k)} \\ . \\ . \\ . \\ \omega_n^{(k)} \end{pmatrix} = X\,\mathrm{u}_k \\
&= X(\frac{1}{\sqrt{\lambda_k}}{}^t X s_k) \\
&= \frac{1}{\sqrt{\lambda_k}}(X^t X s_k) = \frac{1}{\sqrt{\lambda_k}}(\lambda_k s_k) = \sqrt{\lambda_k} s_k
\end{aligned}$$

Projection des variables

On a

$$v^{j^{(k)}} = {}^t X^j s_k$$

et

$$V^k = \begin{pmatrix} v^{1^k} \\ . \\ . \\ . \\ v^{p^k} \end{pmatrix} =^t X \ s_k$$

projection des p variables sur $\triangle s_k$

2.4 Reconstitution du tableau initial

On a s_k un vecteur propre normé de $X\ {}^t X$ associé aux valeur propre λ_k et u_k un vecteur propre normé de ${}^t X\ X$ associé à la même valeur propre λ_k. On a alors

$$X_{(n,p)} = \sum_{k \in K} \sqrt{\lambda_k} s_k^t u_k$$

2.4.1 Inertie

Définition 2.4.1 *On appelle inertie expliquée par un axe factoriel* $\triangle \mathrm{u}_k$ *ou par* $(\triangle s_k)$ *la quantité donnée par*

$$I_k = \frac{\lambda_k}{\sum_{l=1}^{p} \lambda_l}$$

Cette quantité représente la quantité d'information apportée par l'axe $\triangle \mathrm{u}_k$ *ou par* $(\triangle s_k)$.

Définition 2.4.2 *On appelle inertie expliquée par le s-e.v de dim k la quantité*

$$I_{1,\ldots,k} = \frac{\lambda_1 + \ldots + \lambda_k}{\sum_{l=1}^{p} \lambda_l}$$

Propriété 2.4.1 *On a*

$$0 \leq I_{1,\ldots,k} \leq 1$$

Plus I *est proche de* 1, *plus la perte d'information est petite.*

En pratique, dès que $I_{1,\ldots,k} \geq 75\%$, *on considère la dimension du nouveau tableau est* (n, k).

CHAPITRE 3

Analyse en Composantes Principale A.C.P

Introduction

Dans le chapitre précédent, nous avons donné une méthode générale pour réduire un tableau de données quantitatives.

Les nouvelles variables $\omega^{(1)}, ..., \omega^{(q)}$ sont des combinaisons linéaires des p variables initiales:

$$\omega^{(k)} = \sum_{l=1}^{p} \mathrm{u}_{k,l} X^{l} \ , \ \ \mathrm{u}_k = \begin{pmatrix} \mathrm{u}_{k,1} \\ . \\ . \\ . \\ \mathrm{u}_{k,p} \end{pmatrix}.$$

Deux problèmes peuvent se poser:

1. Certaines variables X^j, prennent des valeurs trés petites devant les autres (ie: la difference entre les moyennes et considérables). Dans ce cas l'importance de ces derniérs sera négligée.

2. Les variables sont exprimée avec des unités $\neq$ au quel cas ω^k n'auras pas de sens.

Avec **Analyse en Composantes Principale** *A.C.P*, on resoud le problème de grandeur. Et avec **l'A.C.P normée**, on resoud le problème des échèlles de mesure.

L'A.C.P consiste à centrer le tableau initial, et muni chaque individues d'un poids

$$p_i (\sum p_i = 1).$$

Appliquer l'A.F.G du tableau centre lorsque $\mathbb{R}^n$ est muni de la metrique des poids

$$D_p = \begin{pmatrix} p_1 & 0 \\ & \\ 0 & p_n \end{pmatrix},$$

et $\mathbb{R}^p$ est muni de la metrique identité

$$I_{(p,p)} = \begin{pmatrix} 1 & 0 \\ & \\ 0 & 1 \end{pmatrix}.$$

On note par $\widetilde{X}$ le tableau centre $\overline{X}^j = \sum_{i=1}^n p_i x_i^j$ la moyenne de X^j.

$$\widetilde{X}_{(n,p)} = \begin{pmatrix} x_1^1 - \overline{X}^1 & ... & x_1^p - \overline{X}^p \\ . & & . \\ . & & . \\ . & & . \\ x_n^1 - \overline{X}^1 & ... & x_n^p - \overline{X}^p \end{pmatrix}$$

3.1 A.F.G du tableau X

On se pose sur $(\mathbb{R}^p, I_p)$

On considère le nuage de n points de $\mathbb{R}^p$

$$N(I) = \{({}^tX_i, p_i)/{}^tX_i \in \mathbb{R}^p, 0 \leq p_i \leq 1\}.$$

Chaque point de $N(I)$ est projecter sur s.e.v de $\mathbb{R}^p$.

Nous allons procéder de la même manière que dans le cas de l'A.F.G.

Nous allons d'abord construire un s-e.v de $\mathbb{R}^p$, de dim1, qui ajuste au mieux les problèmes de $N(I)$, si l'information apporté par ce s-e.v est insuffiante on construit un s-e v de dim2,...etc, jusqu'à récupérer un max d'information.

Construction du s-e v de dim1

On note u le vecteur de la base de s-e v de dim1, $\triangle\, \mathrm{u}_1$ est la droite de $\mathbb{R}^p$.

On suppose $\|\mathrm{u}\|_I^2 = 1 =^t \mathrm{u}\,\mathrm{u}$ chaque point de $N(I)$ est projecté sur $\triangle\, \mathrm{u}$.

Soit ε_i est l'erreur dûe à la projection, d'aprés le critère du M.C on minimise

$$\min_{\|\mathrm{u}\|_I^2} = \sum_{i=1}^n p_i \varepsilon_i^2$$

On a

$$\left\| {}^t\widetilde{X_i} \right\|_{I_p}^2 = \omega_i^2 + \varepsilon_i^2 \Rightarrow \varepsilon_i^2 = \left\| {}^t\widetilde{X_i} \right\|_{I_p}^2 - \omega_i^2.$$

Ce qui revient

$$\max_{\|\mathrm{u}\|^2_{I_p}} \sum_{i=1}^{n} p_i \omega_i^2 = C^{ste} - \omega_i^2 \qquad , \omega = \begin{pmatrix} \omega_1 \\ . \\ . \\ . \\ \omega_n \end{pmatrix}$$

On a

$$\begin{aligned} \|\omega\|^2_{D_p} &= {}^t\omega D_p \omega = (\omega_1, ..., \omega_n) \begin{pmatrix} p_1 & & & 0 \\ & . & & \\ & & . & \\ & & & . \\ 0 & & & p_n \end{pmatrix} \begin{pmatrix} \omega_1 \\ . \\ . \\ . \\ \omega_n \end{pmatrix} \\ &= (p_1\omega_1, ..., p_1\omega_n) \begin{pmatrix} \omega_1 \\ . \\ . \\ . \\ \omega_n \end{pmatrix} \\ &= \sum_{i=1}^{n} p_i \omega_i^2. \end{aligned}$$

Ainsi le problème (PB) devient

$$PB \Leftrightarrow \max_{\|\mathrm{u}\|^2_{I_p}} {}^t\omega D_p \omega$$

Avec

$$\omega_i = \left\langle \mathrm{u}, {}^t\widetilde{X_i}^p \right\rangle_{I_p} = \widetilde{X}_i\,\mathrm{u} \Rightarrow \omega = \widetilde{X}\,\mathrm{u}$$

D'où

$$\max_{\|\mathrm{u}\|^2_{I_p}} {}^t\left(\widetilde{X}\,\mathrm{u}\right) D_p \left(\widetilde{X}\,\mathrm{u}\right) \Leftrightarrow \max_{\|\mathrm{u}\|^2_{I_p}} {}^t\mathrm{u}^t\widetilde{X} D_p\,\mathrm{u}\widetilde{X}............(1)$$

Le lagrangier du pb d'optimisation à une contrainte

On a $Z(\mathrm{u}) = {}^t\,\mathrm{u}\;{}^t\widetilde{X} D_p\;\widetilde{X}\;\;\mathrm{u} - \lambda\,({}^t\,\mathrm{u}\,\mathrm{u} - 1)$

D'où

$$\begin{aligned}
\frac{\partial}{\partial\,\mathrm{u}} Z(\mathrm{u}) &= 0 \Longleftrightarrow 2^t\,\mathrm{u}^t\widetilde{X} D_p \widetilde{X} - 2\lambda^t\,\mathrm{u} = 0 \\
&\Longleftrightarrow {}^t\,\mathrm{u}^t\widetilde{X} D_p \widetilde{X} = \lambda^t\,\mathrm{u} \\
&\Longleftrightarrow {}^t\left({}^t\,\mathrm{u}^t\widetilde{X} D_p \widetilde{X}\right) =^t \left(\lambda^t\,\mathrm{u}\right) \\
&\Longleftrightarrow {}^t\widetilde{X} D_p \widetilde{X}\,\mathrm{u} = \lambda\,\mathrm{u}
\end{aligned}$$

u est un $\overrightarrow{vp}$ de ${}^t\widetilde{X} D_p\ \widetilde{X}$ associé à la valeur propre λ.

D'aprés (1) on a

$$\max_{\|\,\mathrm{u}\|^2=1}{}^t\,\mathrm{u}({}^t\widetilde{X} D_p \widetilde{X})\,\mathrm{u}) \Leftrightarrow \max_{\|\,\mathrm{u}\|^2=1}{}^t\,\mathrm{u}(\lambda\,\mathrm{u}) \Leftrightarrow \max\lambda$$

$V =^t \widetilde{X} D_p \widetilde{X}$ est la matrice de Variance-Covariance .

Conclusion

Le s.e.v de dim1 qui ajuste au mieux les points de$N(I)$ est la droite passant par l'origine et engendré par le $\overrightarrow{vp}$ normé de la matrice des Variance-Covariance associé à la plus grande valeur propre λ.

u_1 de $V =^t \widetilde{X} D_p \widetilde{X}$ associé + grande vp λ_1.

La1^{ère} nouvelle variable dite aussi la première composante principale $\omega^{(1)}$ telleque

$$\omega_i^{(1)} = \left\langle {}^t\widetilde{X}_i\ ,\ \mathrm{u}_1 \right\rangle_{I_p} = \widetilde{X}_i\,\mathrm{u}_1.$$

Et donc

$$\omega^{(1)} = \begin{pmatrix} \omega_1 \\ . \\ . \\ . \\ \omega_n \end{pmatrix} = \widetilde{X}\,\mathrm{u}_1.$$

La moyenne de la $1^{ère}$ composante principale est:

$$\begin{aligned}
\overline{\omega}^1 &= \sum_{i=1}^{n} p_i \omega_i^{(1)} \\
&= \sum_{i=1}^{n} p_i (\widetilde{x}_i^1, ..., \widetilde{x}_i^p) \begin{pmatrix} \mathrm{u}_{11} \\ . \\ . \\ . \\ \mathrm{u}_{1p} \end{pmatrix} \\
&= (\sum_{i=1}^{n} p_i \widetilde{x}_i^1, ..., \sum_{i=1}^{n} p_i \widetilde{x}_i^p) \begin{pmatrix} \mathrm{u}_{11} \\ . \\ . \\ . \\ \mathrm{u}_{1p} \end{pmatrix} \\
&= (\overline{\widetilde{X}^1}, ..., \overline{\widetilde{X}^p}) \begin{pmatrix} \mathrm{u}_{11} \\ . \\ . \\ . \\ \mathrm{u}_{1p} \end{pmatrix} = (0, ..., 0) \begin{pmatrix} \mathrm{u}_{11} \\ . \\ . \\ . \\ \mathrm{u}_{1p} \end{pmatrix} = 0
\end{aligned}$$

La variance:

$$\begin{aligned}
\sigma^2_{\omega^{(1)}} &= var(\omega^{(1)}) \\
&= \sum_{i=1}^{n} p_i (\omega_i^{(1)} - \overline{\omega}^{(1)})^2 \\
&= \sum_{i=1}^{n} p_i \left(\omega_i^{(1)}\right)^2 =^t \omega^{(1)} D_p \omega^{(1)} \\
&= \left\|\omega^{(1)}\right\|^2_{D_p} =^t (\widetilde{X}\, \mathrm{u}_1) D_p (\widetilde{X}\, \mathrm{u}_1) \\
&= {}^t\, \mathrm{u}_1 ({}^t \widetilde{X} D_p \widetilde{X})\, \mathrm{u}_1 =^t \mathrm{u}_1 V\, \mathrm{u}_1 \\
&= {}^t\, \mathrm{u}_1 \lambda\, \mathrm{u}_1 = \lambda_1 \parallel \mathrm{u}_1 \|^2_{I_p} = \lambda_1.
\end{aligned}$$

Remarque 3.1.1 *Toutes les propositions vues au chapitre A.F.G sont vérifiés dans le cas de l'A.C.P.*

Généralisation

Le s.e.v de dim k qui ajuste au mieux les n points de $\mathbb{R}^p$ est le s.e engendré par les $\overrightarrow{vp}$ normés de V associé aux k plus grandes valeurs propres $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_k$.

Les composantes principales sont $\omega^{(k)}, ..., \omega^{(k)}$ telle que $\frac{\lambda_k}{\sum \lambda_k} \geq 0,75$ et $\overline{\omega}^{(1)}, ..., \overline{\omega}^{(k)} = 0$, la vairance respectivement $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_k$.

On se pose sur $\mathbb{R}^n$, muni de la métrique des poids D_p

On considère le nuage des points de $\mathbb{R}^n$

$$N(J) = \{\widetilde{X}^1, ..., \widetilde{X}^p / \widetilde{X}^j \in \mathbb{R}^n\}$$

On note par $\triangle s$ le s-e.v de $\mathbb{R}^n$ de dim $1, s \in \mathbb{R}^n$ chaque point v^j.

$v^j = \left\langle \widetilde{X}^j, s \right\rangle_{D_p}$ projection D_p orthogonale $={}^t\widetilde{X}^j D_p\ s$.

$$v = \begin{pmatrix} v^1 \\ . \\ . \\ . \\ v^j \end{pmatrix} ={}^t \widetilde{X}^j D_p s.$$

Par le critère des moindre carrées qui consiste à minimiser $\sum_{j=1}^p \varepsilon_j^2 = \|\varepsilon\|_{I_p}^2$

$$\varepsilon_j^2 = \left\| {}^t\widetilde{X}^j \right\|_{D_p}^2 - (v^j)^2.$$

$$\begin{aligned} \min_{\|s\|^2=1} \sum_{j=1}^p \varepsilon_j^2 &\Leftrightarrow \max_{\|s\|=1} \sum_{j=1}^p (v^j)^2 \\ &\Leftrightarrow \max_{\|s\|=1}{}^t vv \\ &\Leftrightarrow \max_{\|s\|=1}{}^t \left({}^t\widetilde{X} D_p s\right) \left({}^t\widetilde{X} D_p s\right) \\ &\Leftrightarrow \max_{\|s\|=1}{}^t s D_p \widetilde{X}^t \widetilde{X} D_p s..............(1) \end{aligned}$$

On a

$$L(s) =^t sD_p\widetilde{X}^{\,t}\widetilde{X}D_p s - \lambda\left({}^t sD_p s - 1\right)$$

D'où

$$\begin{aligned} \frac{\partial}{\partial s}L(s) &= 0 \Longleftrightarrow 2^t sD_p\widetilde{X}D_p^t\widetilde{X} - 2\lambda^t sD_p = 0 \\ &\Leftrightarrow {}^t s\widetilde{X}D_p^t\widetilde{X} = \lambda^t sD_p \\ &\Leftrightarrow (\widetilde{X}^t\widetilde{X}D_p)s = \lambda s \end{aligned}$$

s est un $\overrightarrow{vp}$ de la matrice $\widetilde{X}\ {}^t\widetilde{X}D_p$ associé à la vp λ.

D'apres (1), on a

$$\begin{aligned} \max_{\|s\|^2=1}{}^t sD_p(\widetilde{X}^{\,t}\widetilde{X}D_p)s &\Leftrightarrow \max_{\|s\|^2=1}{}^t sD_p(\lambda s) \\ &\Leftrightarrow \max_{\|s\|^2=1}\lambda\,\|s\|_{D_p}^2 = \max\lambda. \end{aligned}$$

Conclusion Le s.e.v de $\mathbb{R}^n$ de dim1, est la droite passant par l'origine et engendrée par le $\overrightarrow{vp}$ normé s de $\widetilde{X}\ {}^t\widetilde{X}D_p$ associé à la plus grand vp λ.

Le s.e.v de $\mathbb{R}^n$ de dim k qui ajuste au mieux les p point de R^n est l'espace engendré par les $\overrightarrow{vp}$ normés $s_1, ..., s_k$ (orthogonaux) de $\widetilde{X}\ {}^t\widetilde{X}D_p$ associe aux k plus grand vp$\lambda_1 \geq ... \geq \lambda_k$.

Proposition 3.1.1 *Les matrices ${}^t\widetilde{X}\ D_p\widetilde{X}$ et $\widetilde{X}\ {}^t\widetilde{X}D_p$ ont les mêmes $\overrightarrow{vp}$ s*

Relation entre les espaces propres de $\mathbb{R}^n$ et $\mathbb{R}^p$:

On note par u un $\overrightarrow{vp}$ normé de V associe à λ, et par s un $\overrightarrow{vp}$ normé de $\widetilde{X}\ {}^t\widetilde{X}D_p$ associe à la même vp λ.

Proposition 3.1.2

$$\left\{\begin{array}{l} \mathrm{u}_{(p,1)} = \frac{1}{\sqrt{\lambda}}{}^t\widetilde{X}D_p s \\ s_{(n,1)} = \frac{1}{\sqrt{\lambda}}\widetilde{X}\ \mathrm{u} \end{array}\right\}.......(*)$$

Le système $(*)$ est dite formule de transmition .

Projection simultanée:

1− Si $n \geq p$ on se place sur l'espace $\mathbb{R}^p$, la matrice à diagonaliser est

$$V =^t \widetilde{X} D_p \widetilde{X}\ .$$

Projection des individus

La projection du $i^{ème}$ individus sur le $k^{ème}$ axe factoriel est $\omega_i^{(k)} = \widetilde{X}_i\, \mathrm{u}_k$

$$\omega^{(k)} = \begin{pmatrix} \omega_1^{(k)} \\ . \\ . \\ . \\ \omega_n^{(k)} \end{pmatrix} = \widetilde{X}\, \mathrm{u}_k$$

projection des individus au $k^{ème}$ nouvelle variable

Projection des variables

la projection de la $j^{ème}$ variable sur le $k^{ème}$ axe factoriel est $v^j =^t \widetilde{X} D_p s_k$

$$\begin{aligned} v^{(k)} &= \begin{pmatrix} v^{1^{(k)}} \\ . \\ . \\ . \\ v^{p^{(k)}} \end{pmatrix} = \begin{pmatrix} {}^t\widetilde{X}^1 \\ . \\ . \\ . \\ {}^t\widetilde{X}^p \end{pmatrix} D_p s_k \\ &= {}^t\widetilde{X} D_p s_k =^t \widetilde{X} D_p (\frac{1}{\sqrt{\lambda_k}} \widetilde{X}\, \mathrm{u}_k) \\ &= \frac{1}{\sqrt{\lambda_k}}{}^t \widetilde{X} D_p \widetilde{X}\, \mathrm{u}_k \end{aligned}$$

D'où: $v^{(k)} = \sqrt{\lambda_k}\, \mathrm{u}_k$ et $v^{j(k)} = \sqrt{\lambda_k}\, \mathrm{u}_{k,j}$

2− Si $n < p$ on se pose sur l'espace des variables $\mathbb{R}^p$, la matrice à diagonaliger est $\widetilde{X}$ ${}^t\widetilde{X} D_p$.

Projection des individus

La projection du $i^{ème}$ individus sur le $k^{ème}$ axe factoriel est $\omega_i^{(k)} = \widetilde{X}_i \, \mathrm{u}_k$

$$\begin{aligned}
\omega^{(k)} &= \begin{pmatrix} \omega_1^{(k)} \\ . \\ . \\ . \\ \omega_n^{(k)} \end{pmatrix} = \widetilde{X} \, \mathrm{u}_k \\
&= \widetilde{X}(\frac{1}{\sqrt{\lambda_k}} {}^t\widetilde{X} D_p s_k) \\
&= \frac{1}{\sqrt{\lambda_k}} \widetilde{X}^t \widetilde{X} D_p s_k = \sqrt{\lambda_k} s_k = \sqrt{\lambda_k} s_{k,i}
\end{aligned}$$

Projection des variables

la projection de la$j^{ème}$ variable sur le $k^{ème}$ axe factoriel est $v^{j(k)} =^t \widetilde{X}^j D_p s_k$

$$v^{(k)} = \begin{pmatrix} {}^t\widetilde{X}^1 \\ . \\ . \\ . \\ {}^t\widetilde{X}^p \end{pmatrix} D_p s_k =^t \widetilde{X} D_p s_k.$$

3.2 Aide à l'interprétation les contributions:

3.2.1 Contributions absolues

Contributions absolues des individus

Définition 3.2.1 *On appelle contribution absolue du $i^{ème}$ individu pour la contribution du $k^{ème}$ axe factoriel, la quantitié donnée par*

$$C_a^{(k)}(i) = p_i \frac{\left|\omega_i^{(k)}\right|^2}{\lambda_k}.$$

La contribution absolue du $i^{ème}$ individus pour la contribution du s.e.v de dim k est

$$C_a^{(1,\dots,k)}(i) = \sum_{l=1}^{k} C_a^{(l)}(i).$$

La contribution absolue de deux individus i_1 et i_2 pour construire le $k^{ème}$ axe factoriel est donnée par

$$C_a^{(k)}(i_1, i_2) = C_a^{(k)}(i_1) + C_a^{(k)}(i_2).$$

Propriété 3.2.1 $\forall i = 1, ..., q \quad , \forall k$

1/ $0 \leq C_a^{(k)}(i) \leq 1$

2/ $\sum_{i=1}^{n} C_a^{(k)}(i) = 1$

3.2.2 Contributions absolues des variables

Définition 3.2.2 *On appelle contribution absolue du $j^{ème}$ variable pour la contribution du $k^{ème}$ axe factoriel, la quantitié donnée par*

$$C_a^{(k)}(j) = p_i \frac{\left|v^{j(k)}\right|^2}{\lambda_k}.$$

La contribution absolue du $j^{ème}$ variable pour la contribution du s.e.factoriel$(\triangle\, \mathrm{u}_1, ..., \triangle\, \mathrm{u}_k)$ *est*

$$C_a^{(1,\dots,k)}(j) = \sum_{l=1}^{k} C_a^{(l)}(j).$$

La contribution absolue de deux variables j_1 et j_2 pour construire le $k^{ème}$ axe factoriel est donnée par

$$C_a^{(k)}(j_1, j_2) = C_a^{(k)}(j_1) + C_a^{(k)}(j_2).$$

Propriété 3.2.2 $\forall j = 1, ..., q \quad , \forall k \leq p$

1/ $0 \leq C_a^{(k)}(j) \leq 1$

2/ $\sum_{j=1}^{p} C_a^{(k)}(j) = 1$

Remarque 3.2.1 *La contribution absolue représente la part de l'individu ou la variable pour construire les axes, les plans ou les s.e factoriels*

3.2.3 Contributions relatives

Contributions relatives des individus

Définition 3.2.3 *On appelle contribution relative du* 1^{er} *inidividu pour construire le* $k^{ème}$ *axe factoriel est donnée par*

$$C_r^{(k)}(i) = \frac{\left|\omega_i^{(k)}\right|^2}{\left\|{}^t\widetilde{X}_i\right\|_{I_P}^2}.$$

La contribution relative de deux individus i_1 *et* i_2 *pour construire le* $k^{ème}$ *axe factoriel est donnée par*

$$C_r^{(k)}(i_1, i_2) = C_r^{(k)}(i_1) + C_r^{(k)}(i_2).$$

La contribution absolue du $i^{ème}$ *individus pour la contribution du s.e.v de dim* k *est*

$$C_r^{(1,\dots,k)}(i) = \sum_{l=1}^{k} C_r^{(l)}(i).$$

Contributions relatives des variables:

Définition 3.2.4 *On appelle contribution relative du* 1^{er} *variable pour construire le* $k^{ème}$ *axe factoriel est donnée par*

$$C_r^{(k)}(j) = \frac{\left|v^{j(k)}\right|^2}{\left\|{}^t\widetilde{X}^j\right\|_{D_P}^2}.$$

La contribution relative de deux variables j_1 *et* j_2 *pour construire le* $k^{ème}$ *axe factoriel est donnée par*

$$C_r^{(k)}(j_1, j_2) = C_r^{(k)}(j_1) + C_r^{(k)}(j_2).$$

La contribution relative du $j^{ème}$ *variable pour la contribution du s.e.factoriel* $(\triangle\, \mathrm{u}_1, ..., \triangle\, \mathrm{u}_k)$ *est*

$$C_r^{(1,\dots,k)}(j) = \sum_{l=1}^{k} C_k^{(l)}(j).$$

La contribution relative mesure la qualité de représentation de la variable sur l'axe ou s.e factoriel . Plus la contribution est proche de 1 plus la variable est représenté sur l'axe ou le s.e factoriel.

Remarque 3.2.2 *Généralement, en pratique dèsque* $C_a \geq 0,75$ *(respectivement* $C_r \geq 0,75$*) l'individu ou la variable a bien participé ou construit l'axe ou le s.e factoriel (resp l'individu ou la variable est bien representé sur l'axe ou le s.e factoriel).*

CHAPITRE 4

Analyse en composantes principales normée (ACP Normée)

4.1 Introduction

l'objectif de ce chapitre est d'étudier la méthode classiquement utilisée pour d'écrire et visualiser des données multivariées issues des variables continues : Analyse en composantes principales normée.

Les techniques d'analyse descriptive sont utilisées, notamment pour visualiser des données dans un sous espace représentatif, pour détecter des groupes d'individus et /ou variables, des valeurs aberrantes ou pour aider au choix de variables.

Ces méthodes permettent aussi de répondre aux questions de type : Quels individus se ressemblent du point de vue des différentes variables considérées. Ou bien, quelles sont les variables qui sont semblables du point de vue des individus de la population étudiée.

L'analyse en composantes principales est un outil de réduction de dimension tout en retirant la redondance d'informations apportées par plusieurs variables corrélées. Les variables initiales seront alors remplacées par un nombre réduit de nouvelles variables construites comme combinaisons linéaires des variables initiales. Ces dernières sont alors indépendantes les unes des autres et peuvent être classées par ordre d'importance.

L'ACP normée, contrairement à l'ACP, permet d'éliminer la contrainte d'échelles de mesure des variables observées, dans le cas où ces dernières ne sont pas exprimées dans la même échelle de mesure.

Soit un tableau de données quantitatives X de dimension (n, p) (n individus et p variables). $\underset{(n,p)}{X} = (x_{ij})_{\substack{1\leq i\leq n\\ 1\leq j\leq p}}$

On suppose que chaque individu est muni d'un poids p_i et on note par $\widetilde{X}$ le tableau centré-réduit. $\underset{(n,p)}{\widetilde{X}} = (\widetilde{x}_{ij})_{\substack{1\leq i\leq n\\ 1\leq j\leq p}}$

où $\widetilde{x}_{ij} = \frac{x_{ij}-\overline{X^j}}{\sigma_j}$.

Le tableau $\widetilde{X}$ est dit à deux entrées :

$$\widetilde{X} = \begin{cases} \left[\widetilde{X}^1, \cdots, \widetilde{X}^P\right] \\ \begin{bmatrix} \widetilde{X}_1 \\ \vdots \\ \widetilde{X}_n \end{bmatrix} \end{cases}$$

Où $\widetilde{X}^j \in \mathbb{R}^n$ représente la $j^{éme}$ colonne de $\widetilde{X}$. Et ${}^t\widetilde{X}_i \in \mathbb{R}^p$, $\widetilde{X}_i$ représente la i^{éme} ligne de $\widetilde{X}$.

Remarque 2.1.1 *la matrice $\widetilde{X}$ peut être assimilée à l'application linéaire définie comme suit : On note par $(e_1, \cdots e_n)$ la base canonique de $\mathbb{R}^n$ et par $\left(e'_1, \cdots e'_p\right)$ la base canonique de $\mathbb{R}^p$*

$$\widetilde{X} : \mathbb{R}^p \to \mathbb{R}^n$$

Pour tout $j = 1, \cdots, p$,

$$\widetilde{X}\left(e'_j\right) = \widetilde{X}^j = \sum_{i=1}^{n} \widetilde{x}_{ij} e_i$$

$$^t\widetilde{X} : \mathbb{R}^n \to \mathbb{R}^p$$

Pour tout $i = 1, \cdots, n$

$$^t\widetilde{X}\,(e_i) = \widetilde{X}_i = \sum_{j=1}^{p} \widetilde{x}_{ij} e'_j$$

Remarque 4.1.2 *On a deux espaces vectoriels* $\mathbb{R}^n$ *et* $\mathbb{R}^p$*. L'approche consiste à se placer sur l'un des deux espaces vectoriels et projeter les points (de* $\mathbb{R}^n$ *ou* $\mathbb{R}^p$*) sur un sous espaces et établir des formules de transitions permettant de passer à l'analyse sur l'autre espace.*

4.2 Position du problème

4.2.1 On se place sur $\mathbb{R}^p$ muni de la métrique identité $(\mathbb{R}^p, \mathbb{I}_p)$

On considère le nuage des n points de $\mathbb{R}^p$ noté $N(I) = \left\{ \left(^t\widetilde{X}_i, p_i \right) / ^t\widetilde{X}_i \in \mathbb{R}^p, i = 1, \cdots, n \right\}$.

Le problème est de déterminer un sous espace vectoriel de $\mathbb{R}^p$ de dimension $1, 2, \cdots$,ou r $(r < p)$ qui ajuste aux mieux les n points du nuage $N(I)$, au sens du critère des moindres carrés. La méthode consiste à chercher le sous espace vectoriel E_1 de $\mathbb{R}^p$ de dimension 1 qui ajuste au mieux le nuage. Si l'information apportée par ce sous espace vectoriel est considérable, on remplace les p variables initiales par une seule nouvelle variable. Sinon, on cherche un sous espace E_2 de dimension 2. Et, ainsi de suite jusqu'à ce que la perte d'information soit négligeable.

Recherche du sous espace vectoriel de $\mathbb{R}^p$ de dimension 1 qui ajuste aux mieux les n points du nuage $N(I)$

Remarque 4.2.1 *Un sous espace vectoriel de* $\mathbb{R}^p$ *de dimension* 1 *est engendré par un seul vecteur, qu'on peut choisir normé.*

On note par u_1 le vecteur normé qui engendre E_1. Alors, $E_1 = \Delta u_1$ est une droite passant par l'origine de $\mathbb{R}^p$ et portée par le vecteur u_1. Schématiquement, cela se présente comme suit

$\omega_i^{(1)}$ est la projection orthogonale de ${}^t\widetilde{X}_i$ sur Δu_1. On a alors,

$$\omega_i^{(1)} = \left\langle {}^t\widetilde{X}_i, u_1 \right\rangle_{\mathbb{I}_p} = \widetilde{X}_i u_1 \ \in \mathbb{R}^p$$

Il est facile de voir, que le vecteur des n projections

$$\omega^{(1)} = \begin{pmatrix} \omega_1^1 \\ \vdots \\ \omega_n^1 \end{pmatrix} = \widetilde{X} u_1$$

Proposition 4.2.1 *Pour tout* $i = 1, \cdots, n$

$$\widetilde{X}_i = \underline{X}_i D_{\frac{1}{\sigma}}$$

avec $\widetilde{X}_i = (\widetilde{x}_i^1, \cdots, \widetilde{x}_i^p)$, $\widetilde{x}_i^j = \frac{x_i^j - \overline{X}^j}{\sigma_j}$ *est la donnée centrée réduite et* $\underline{X}_i = \left(x_i^1 - \overline{X}^1, \cdots, x_i^p - \overline{X}^p\right)$

.

Démonstration.

$$\begin{aligned} \underline{X}_i D_{\frac{1}{\sigma}} &= \left(x_i^1 - \overline{X}^1, \cdots, x_i^p - \overline{X}^p\right) \begin{pmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma p} \end{pmatrix} \\ &= \left(\frac{x_i^1 - \overline{X}^1}{\sigma_1}, \cdots, \frac{x_i^p - \overline{X}^p}{\sigma_p}\right) \\ &= \widetilde{X}_i \end{aligned}$$

■

Proposition 4.2.2 *On note par $\underline{X}$ le tableau centré. On alors,*

$$\widetilde{X} = \underline{X}D_{\frac{1}{\sigma}}$$

Démonstration. On a

$$\underline{X} = \begin{pmatrix} \underline{X}_1 \\ \vdots \\ \underline{X}_n \end{pmatrix}$$

donc

$$\begin{aligned} \underline{X}D_{\frac{1}{\sigma}} &= \begin{pmatrix} \underline{X}_1 \\ \vdots \\ \underline{X}_n \end{pmatrix} D_{\frac{1}{\sigma}} = \begin{pmatrix} \underline{X}_1 D_{\frac{1}{\sigma}} \\ \vdots \\ \underline{X}_n D_{\frac{1}{\sigma}} \end{pmatrix} \\ &= \begin{pmatrix} \widetilde{X}_1 \\ \vdots \\ \widetilde{X}_n \end{pmatrix} = \widetilde{X} \end{aligned}$$

■

Corollaire 4.2.1 *Le vecteur des n projections $\omega^{(1)} = \widetilde{X}u_1 = \underline{X}D_{\frac{1}{\sigma}}u_1$*

Remarque 4.2.2 *Déterminer l'axe Δu_1 qui ajuste au mieux les points du nuage revient à déterminer le vecteur normé u_1.*

On utilise le critère des moindres carrés, qui consiste à minimiser les erreurs quadratiques moyennes sous la contrainte $\|u_1\|^2_{\mathbb{I}_p} = 1$.

Problème d'optimisation s'écrit

$$\min_{\|u_1\|^2_{\mathbb{I}_p}=1} \sum_{i=1}^{n} p_i \varepsilon_i^2$$

D'après le théorème de Pythagore on a,

$$\left\| {}^t\widetilde{X}_i \right\|_{\mathbb{I}_p}^2 = \omega_i^2 + \varepsilon_i^2$$

Le problème est alors équivalent à maximiser l'inertie

$$\begin{aligned}
\min_{\|u_i\|_{\mathbb{I}_p}^2=1} \sum_{i=1}^{n} p_i \left(\left\| {}^t\widetilde{X}_i \right\|_{\mathbb{I}_p}^2 - \omega_i^2 \right) &\Leftrightarrow \max_{\|u_i\|_{\mathbb{I}_p}^2=1} \sum_{i=1}^{n} p_i \omega_i^2 \\
&\Leftrightarrow \max_{\|u_i\|_{\mathbb{I}_p}^2=1} \|\omega\|_{D_p}^2 \\
&\Leftrightarrow \max_{\|u_i\|_{\mathbb{I}_p}^2=1} {}^t\omega D_p \omega \\
&\Leftrightarrow \max_{\|u_i\|_{\mathbb{I}_p}^2=1} {}^t\left(\widetilde{X} u_1\right) D_p \left(\widetilde{X} u_1\right) \\
&\Leftrightarrow \max_{\|u_i\|_{\mathbb{I}_p}^2=1} {}^t u_1 \, {}^t\widetilde{X} D_p \widetilde{X} u_1 \\
&\Leftrightarrow \max_{\|u_i\|_{\mathbb{I}_p}^2=1} I_{\Delta u^\perp} \qquad (4.2.1)
\end{aligned}$$

Remarque 4.2.3 *La matrice $R = {}^t\widetilde{X} D_p \widetilde{X} = D_{\frac{1}{\sigma}}^t \underline{X} D_p \underline{X}$ représente la matrice des corrélations.*

Résolution du problème d'optimisation On pose $H(u_1) = {}^t u_1 R u_1$ on cherche un extremum de H sous la contrainte $\|u_1\|_{\mathbb{I}_p}^2 = {}^t u_1 u_1 = 1$. Le Lagrangien s'écrit,

$$\mathcal{L}(u_1) = {}^t u_1 R u_1 - \lambda \left({}^t u_1 u_1 - 1 \right)$$

La condition nécessaire

$$\begin{aligned}
\frac{\partial \mathcal{L}(u_1)}{\partial u_1} &= 0 \Leftrightarrow 2 {}^t u_1 R - 2\lambda {}^t u_1 u_1 = 0 \\
&\Leftrightarrow {}^t u_1 R = \lambda {}^t u_1 u_1 \\
&\Leftrightarrow R u_1 = \lambda u_1
\end{aligned}$$

u_1 est alors un vecteur propre de la matrice des corrélations R associé à la valeur propre λ.

On alors $H(u_1) = {}^t u_1 R u_1 = \lambda\, {}^t u_1 u_1 = \lambda$ et

$$\max_{\|u_i\|^2_{\mathbb{I}_p}=1} H(u_1) = \max_{\|u_i\|^2_{\mathbb{I}_p}=1} \lambda$$

Le maximum est donc atteint au point $u_1 \in \mathbb{R}^p$, qui est vecteur propre de R associé à la plus grande valeur propre λ.

Conclusion : Le sous espace vectoriel $E_1 = \Delta u_1$ de $\mathbb{R}^p$ de dimension 1 qui ajuste au mieux les points du nuage $N(I)$, est la droite engendrée par le vecteur propre normé u_1 de la matrice des corrélations R associé à la plus grande valeur propre λ_1. $\left(\lambda_1 = \max_{1\leq j\leq p} \lambda j\right)$

Proposition 4.2.3 *1. Les valeurs propres de la matrices R sont non négatives.*

2. Si λ_m est la première valeur propre nulle $(\lambda_1 \geq \cdots \geq \lambda_{m-1} > 0)$, alors $rg(R) = m-1$

Démonstration.

1. Soit λ une valeur propre de R associée à un vecteur propre u,

$$\begin{aligned} Ru &= \lambda u \Leftrightarrow {}^t u R u = \lambda {}^t u\ u \\ &\Leftrightarrow {}^t u {}^t \widetilde{X} D_p \widetilde{X} u = \lambda \\ &\Leftrightarrow \left\langle u\widetilde{X}, \widetilde{X}u \right\rangle_{D_p} = \left\| \widetilde{X}u \right\|^2_{D_p} = \lambda \geq 0 \end{aligned}$$

2. R est une matrice carrée définie positive car

   . ${}^t R = R$

   . $\forall u \in \mathbb{R}^p$, $\langle u, u\rangle_R = {}^t u R u = \left\langle u\widetilde{X}, \widetilde{X}u \right\rangle_{D_p} = \left\| \widetilde{X}u \right\|^2_{D_p} \geq 0$

   R est donc diagonalisable ie. il existe une matrice inversible. P et une matrice diagonale des valeurs propres D telle que

$$R = P^{-1} D P$$

$$D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_{m-1} \end{pmatrix}$$ et $P = (u_1, \cdots, u_{m-1})$, u_i est un vecteur propre de R associé à λ_j

Comme P est inversible. alors $rg\left(P\right) = m-1$, et comme $\lambda_j \neq 0$ alors $rg\left(D\right) = m-1$

$$rg\left(R\right) = rg\left(P^{-1}DP\right) = rg\left(P^{-1}PD\right) = rg\left(D\right) = m-1$$

■

Projection des individus sur Δu_1

A chaque individu I_i on associe un vecteur de $\mathbb{R}^p$: $I_i \rightarrow {}^t\widetilde{X}_i = \begin{pmatrix} \widetilde{x}_i^p \\ \vdots \\ \widetilde{x}_i^p \end{pmatrix} \in \mathbb{R}^p$

A près réduction, on associe à I_i un vecteur de $\mathbb{R}$: $I_i \rightarrow \omega_i^{(1)} = \widetilde{X}_i u_1$ sa projection sur Δu_1.

Le vecteur des projections des n individus est noté $\omega^{(1)} = \begin{pmatrix} \omega_n^{(1)} \\ \vdots \\ \omega_n^{(1)} \end{pmatrix} = \widetilde{X} u_1$

Définition 4.2.1 *La nouvelle variable $\omega^{(1)}$ est dite composante principale.*

Recherche du sous espace vectoriel de $\mathbb{R}^p$ de dimension 2 qui ajuste aux mieux les n points du nuage $N\left(I\right)$

Proposition 4.2.4 *Le sous espace vectoriel E_2 de $\mathbb{R}^p$ de dimension 2 qui ajuste au mieux les points du nuage $N\left(I\right)$ contient E_1.*

Démonstration. Le sous espace vectoriel E_2 est engendré par deux axes passant par l'origine de $\mathbb{R}^p$. $E_2 = (\Delta x, \Delta y)$ où $\{x, y\}$ est une base de E_2. On suppose que $E_1 \nsubseteq E_2$. On alors, $E_1 \neq \Delta x$ et $E_1 \nparallel \Delta x$

$$E_1 \neq \Delta y \text{ et } E_1 \nparallel \Delta y$$

Considérons le sous espace F_2 de $\mathbb{R}^p$ de dimension 2 engendré par E_1 et Δx, comme $E_1 = \Delta u_1$ alors $\{u_1, x\}$ est une base de F_2 et la somme des projections des n points sur E_1 est meilleur que celle sur Δy et la projection des n points sur F_2 est donc meilleur que sur E_2.

Contradiction avec le fait que E_2 est le meilleur sous espace qui ajuste les points du nuage $N(I)$. D'où $E_1 \subset E_2$. ■

Remarque 4.2.4 *Le résultats de la proposition précédente se généralise facilement au cas $E_{r-1} \subset E_r$.*

Avec E_k le sous espace vectoriel de $\mathbb{R}^p$ de dimension k qui ajuste au mieux les points du nuage $N(I)$.

Remarque 4.2.5 *La recherche de E_2 revient à la recherche d'un sous espace vectoriel de dimension 1 orthogonal à E_1*

On note par Δu_2 ce sous espace de dimension 1 orthogonale à $E_1 = \Delta u_1$, on a alors $\langle u_1, u_2 \rangle_{\mathbb{I}_p} = 0$. On suppose que $\|u_2\|^2_{\mathbb{I}_p} = 1$.

Problème d'optimisation s'écrit

$$\min_{\|u_2\|^2_{\mathbb{I}_p}=1} \sum_{i=1}^{n} p_i \varepsilon_i^2 \tag{4.2.2}$$

En procédant de la même manière que précédemment (2.2.1), on en déduit que u_2 est un vecteur propre normé de la matrice des corrélations R associé à la deuxième plus grande valeur propre λ_2.

Généralisation Le sous espace vectoriel de dimension r qui ajuste au mieux les n points du nuage $N(I)$ est l'espace engendré par les r vecteurs propres normés, $u_1, \cdots, u_p$ de la matrice des corrélations R associés aux r plus grandes valeurs propres $\lambda_1 \geq \cdots \geq \lambda_r$ respectivement.

Définition 4.2.2 *On appelle r ième axe factoriel, l'axe passant par l'origine et engendré par le vecteur propre normé u_r de la matrice R associé à la r ième plus grande valeur propre λ_r.*

Définition 4.2.3 *Les vecteurs projections des n points du nuage $N(I)$ sur les axes factoriels sont dits facteurs.*

Définition 4.2.4 *Le plan $(\Delta u_1, \Delta u_2)$ est dit premier plan factoriel, le plan $(\Delta u_1, \Delta u_3)$ est dit deuxième plan factoriel et $(\Delta u_2, \Delta u_3)$ est dit troisième plan factoriel.*

Remarque 4.2.6 *L'oeil humain ne sachant pas regarder dans un espace de dimension supérieur à trois, on se restreint généralement en pratique, à deux ou trois axes factoriels. D'une part la représentation est plus lisible, d'autre part, on verra ultérieurement que l'information recueillie sur les deux ou trois premiers axes factoriels est plus importante, car cette information est directement liée aux plus grandes valeurs propres.*

Moyennes et variances des composantes principales

Proposition 4.2.5 *les composantes principales sont :*

1. *Centrées*
2. *De variances égales aux valeurs propres.*

Démonstration. Soit $\omega^{(r)} = \begin{pmatrix} \omega_1^{(r)} \\ \vdots \\ \omega_n^{(r)} \end{pmatrix}$ la r ième composante principale

1.

$$\begin{aligned}
\overline{\omega}^{(r)} &= \sum_{i=1}^{n} p_i \omega_i^{(r)} \\
&= \sum_{i=1}^{n} p_i \left(\widetilde{X}_i u_r \right) \\
&= \left(\sum_{i=1}^{n} p_i \widetilde{X}_i \right) u_r \\
&= \left(\sum_{i=1}^{n} p_i \widetilde{x}_i^1, \cdots, \sum_{i=1}^{n} p_i \widetilde{x}_i^p \right) u_r \\
&= (0, \cdots, 0)\, u_r = 0
\end{aligned}$$

2.

$$\begin{aligned}
Var\left(\omega^{(r)}\right) &= \left\langle \omega^{(r)}, \omega^{(r)} \right\rangle_{D_p} \\
&= {}^t\omega^{(r)} D_p \omega^{(r)} \\
&= {}^t\left(\widetilde{X} u_r \right) D_p \left(\widetilde{X} u_r \right) \\
&= {}^t u_r \, {}^t\widetilde{X} D_p \widetilde{X} u_r \\
&= {}^t u_r \lambda_r u_r \\
&= \lambda_r
\end{aligned}$$

■

4.2.2 On se place sur $\mathbb{R}^n$ muni de la métrique des poids $(\mathbb{R}^n, D_p)$

Considérons le tableau initial centré-réduit

$$\widetilde{X} = \left(\widetilde{X}^1, \cdots, \widetilde{X}^p \right)$$

Remarque 4.2.7 *1.* $\overline{\widetilde{X}}^j = 0, \; j = 1, \cdots, p$

2. $Var\left(\widetilde{X}^j\right) = Var\left(\frac{X^j - \overline{X}\mathbf{1}}{\sigma_j}\right) = \frac{1}{\sigma_j^2} Var\left(X^j\right) = 1$

On note par $N(J) = \left\{ \widetilde{X}^j / \widetilde{X}^j \in \mathbb{R}^n \text{ pour } j = 1, \cdots p \right\}$ le nuage des p points de $\mathbb{R}^n$.

On procède de manière analogue au cas où on se place sur $\mathbb{R}^p$. Le problème d'optimisation lié à la recherche d'un sous espace vectoriel de $\mathbb{R}^n$ de dimension 1, noté Δs_1, qui ajuste au mieux les p points du nuage $N(J)$, consiste à minimiser la somme des erreurs quadratiques sous la contrainte que le vecteur directeur s_1 est normé. $\left(\|s_1\|^2_{D_p} = 1 \right)$.

Problème d'optimisation s'écrit

$$\min_{\|s_1\|^2_{\mathbb{D}_p}=1} \sum_{j=1}^{p} \varepsilon_j^2$$

D'après le théorème de Pythagore on a,

$$\left\| \widetilde{X}^j \right\|^2_{D_p} = \left(\psi_j^{(1)} \right)^2 + \varepsilon_j^2$$

$\psi_j^{(1)}$ est la projection D_p-orthogonale de la $j^{ième}$ variable centrée-réduite $\widetilde{X}^j$ sur Δs_1.

$$\psi_j^{(1)} = \left\langle \widetilde{X}^j, s_1 \right\rangle_{D_p} = {}^t\widetilde{X}^j D_p s_1$$

On note par $\psi^{(1)} = \begin{pmatrix} \psi_1^{(1)} \\ \vdots \\ \psi_p^{(1)} \end{pmatrix} = {}^t\widetilde{X} D_p s_1$ le vecteur des projections des p variables.

Le problème est alors équivalent à

$$\begin{aligned}
\min_{\|s_1\|^2_{D_p}=1} \sum_{j=1}^{p} \left(\left\| \widetilde{X}^j \right\|^2_{D_p} - \psi_i^2 \right) &\Leftrightarrow \max_{\|s_1\|^2_{D_p}=1} \sum_{j=1}^{p} \left(\psi_i^{(1)} \right)^2 \\
&\Leftrightarrow \max_{\|s_1\|^2_{D_p}=1} \left\| \psi^{(1)} \right\|^2_{\mathbb{I}_p} \\
&\Leftrightarrow \max_{\|s_1\|^2_{D_p}=1} {}^t\psi^{(1)} \psi^{(1)} \\
&\Leftrightarrow \max_{\|s_1\|^2_{D_p}=1} {}^t\left({}^t\widetilde{X} D_p s_1 \right) \left({}^t\widetilde{X} D_p s_1 \right) \\
&\Leftrightarrow \max_{\|s_1\|^2_{D_p}=1} {}^t s_1 \; D_p \; \widetilde{X} \; {}^t\widetilde{X} D_p s_1
\end{aligned}$$

Résolution du problème d'optimisation On pose $H(u_1) = {}^t s_1 \ D_p \ \widetilde{X} \ {}^t\widetilde{X} D_p s_1$ on cherche un extremum de H sous la contrainte $\|s_1\|^2_{D_p} = {}^t s_1 D_p s_1 = 1$. Le Lagrangien s'écrit,

$$\mathcal{L}(s_1) = \ {}^t s_1 \ D_p \widetilde{X} \ {}^t\widetilde{X} D_p s_1 - \lambda \left({}^t s_1 D_p s_1 - 1\right)$$

La condition nécessaire

$$\begin{aligned}
\frac{\partial \mathcal{L}(s_1)}{\partial s_1} &= 0 \Leftrightarrow 2^t s_1 \ D_p \widetilde{X} \ {}^t\widetilde{X} D_p - 2\lambda^t s_1 D_p = 0 \\
&\Leftrightarrow \ {}^t s_1 \ D_p \widetilde{X} \ {}^t\widetilde{X} = \lambda^t s_1 \\
&\Leftrightarrow \ \widetilde{X} \ {}^t\widetilde{X} D_p s_1 = \lambda s_1
\end{aligned}$$

s_1 est alors un vecteur propre de la matrice $\widetilde{X} \ {}^t\widetilde{X} D_p$ associé à la valeur propre λ.

On alors $H(s_1) = {}^t s_1 D_p \widetilde{X} \ {}^t\widetilde{X} D_p s_1 = \lambda \ {}^t s_1 D_p s_1 = \lambda$ et

$$\max_{\|s_1\|^2_{D_p}=1} H(s_1) = \max_{\|s_1\|^2_{D_p}=1} \lambda$$

Le maximum est donc atteint au point $s_1 \in \mathbb{R}^n$, qui est vecteur propre de $\widetilde{X} \ {}^t\widetilde{X} D_p$ associé à la plus grande valeur propre λ.

Conclusion : Le sous espace vectoriel Δs_1 de $\mathbb{R}^n$ de dimension 1 qui ajuste au mieux les points du nuage $N(J)$, est la droite engendrée par le vecteur propre normé s_1 de la matrice $\widetilde{X} \ {}^t\widetilde{X} D_p$ associé à la plus grande valeur propre λ_1. $\left(\lambda_1 = \max\limits_{1 \leq j \leq n} \lambda j\right)$

Remarque 4.2.8
1. *Les propositions (1) et (2) et les remarque (1) et (2) précédentes sont vérifiées lorsqu'on remplace R par $\widetilde{X} \ {}^t\widetilde{X} D_p$.*
2. *Les axes factoriels sont les droites de $\mathbb{R}^n$ passant par l'origine et engendrées par les r vecteurs propres normées $s_1, \cdots, s_r$ de la matrice $\widetilde{X} \ {}^t\widetilde{X} D_p$ associés aux r plus grandes valeurs propres respectives $\lambda_1 \geq \cdots \geq \lambda_r$.*

4.3 Relation entre les sous espaces propres de $\mathbb{R}^p$ et de $\mathbb{R}^n$

Proposition 4.3.1 *Les matrices R et $\widetilde{X}\,{}^t\widetilde{X}D_p$ ont les mêmes valeurs propres.*

Démonstration. $\Rightarrow$ / Soit λ une valeur propre de $R = {}^t\widetilde{X}D_p\widetilde{X}$ associée au vecteur propre u. Montrons que λ est encore valeur propre de $\widetilde{X}\,{}^t\widetilde{X}D_p$.

On a,

$$ {}^t\widetilde{X}D_p\widetilde{X}u = \lambda u \Longleftrightarrow \widetilde{X}{}^t\widetilde{X}D_p\left(\widetilde{X}u\right) = \lambda\left(\widetilde{X}u\right) $$

λ est donc valeur propre de $\widetilde{X}\,{}^t\widetilde{X}D_p$ associée au vecteur propre $\widetilde{X}u$.

$\Leftarrow$ / Soit λ une valeur propre de $\widetilde{X}\,{}^t\widetilde{X}D_p$ associée au vecteur propre s. Montrons que λ est encore valeur propre de ${}^t\widetilde{X}D_p\widetilde{X}$.

On a,

$$ \widetilde{X}\,{}^t\widetilde{X}D_ps = \lambda s \Longleftrightarrow {}^t\widetilde{X}D_p\widetilde{X}\left({}^t\widetilde{X}D_ps\right) = \lambda\left({}^t\widetilde{X}D_ps\right) $$

λ est donc valeur propre de ${}^t\widetilde{X}D_p\widetilde{X}$ associée au vecteur propre ${}^t\widetilde{X}D_pu$. ■

Proposition 4.3.2 *Formule de transition*

$\forall r \geq 1$, *on note par u_r le vecteur propre de ${}^t\widetilde{X}D_p\widetilde{X}$ et s_r le vecteur propre de $\widetilde{X}\,{}^t\widetilde{X}D_p$ associés à la même valeur propre λ_r. On a alors les deux formules de transition suivantes :*

1. $u_r = \frac{1}{\sqrt{\lambda_r}}{}^t\widetilde{X}D_ps_r$

2. $s_r = \frac{1}{\sqrt{\lambda_r}}\widetilde{X}u_r$

Démonstration.

1. La démonstration de 1 revient à vérifier les deux points suivants :

1 a/ $\frac{1}{\sqrt{\lambda_r}}{}^t\widetilde{X}D_p s_r$ est un vecteur propre de R.

1 b/ $\left\|\frac{1}{\sqrt{\lambda_r}}{}^t\widetilde{X}D_p s_r\right\|^2_{\mathbb{I}_p} = 1.$

1 a/

$$
\begin{aligned}
R\left(\frac{1}{\sqrt{\lambda_r}}{}^t\widetilde{X}D_p s_r\right) &= \frac{1}{\sqrt{\lambda_r}}{}^t\widetilde{X}D_p\left(\widetilde{X}^t\widetilde{X}D_p s_r\right) \\
&= \frac{1}{\sqrt{\lambda_r}}{}^t\widetilde{X}D_p\left(\lambda_r s_r\right) \\
&= \lambda_r\left(\frac{1}{\sqrt{\lambda_r}}{}^t\widetilde{X}D_p s_r\right)
\end{aligned}
$$

1 b/

$$
\begin{aligned}
\left\|\frac{1}{\sqrt{\lambda_r}}{}^t\widetilde{X}D_p s_r\right\|^2_{\mathbb{I}_p} &= {}^t\left(\frac{1}{\sqrt{\lambda_r}}{}^t\widetilde{X}D_p s_r\right)\left(\frac{1}{\sqrt{\lambda_r}}{}^t\widetilde{X}D_p s_r\right) \\
&= \frac{1}{\lambda_r}\left({}^t s D_p\widetilde{X}\right)\left({}^t\widetilde{X}D_p s_r\right) \\
&= \frac{1}{\lambda_r}{}^t s D_p\left(\widetilde{X}^t\widetilde{X}D_p s_r\right) \\
&= \frac{1}{\lambda_r}{}^t s D_p\left(\lambda_r s_r\right) \\
&= \|s_r\|^2_{D_p} = 1
\end{aligned}
$$

La démonstration de la formule 2 s'obtient exactement de la même manière. ■

4.4 Représentation simultanée des individus et variables sur les axes factoriels

Comme il existe des formules de transition entre les deux espaces $\mathbb{R}^p$ et $\mathbb{R}^n$, l'ACP normée permet une interprétation simultanée des nuages $N(I)$ et $N(J)$ et de les représenter simultanément dans les plans factoriels. Notons que les nuages $N(I)$ et $N(J)$ ne sont en réalités pas dans les mêmes espaces qui ont des dimensions différentes. Cette représentation simultanée est essentiellement graphique. L'inertie totale des nuages $N(I)$ et $N(J)$ est la même et est égale à

$$\lambda = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\widetilde{x}_i^j\right)^2$$

qui représente la variance ou dispersion totale.

4.4.1 Cas où on se place sur $\mathbb{R}^p$

Si $n \geq p$, on se place sur $(\mathbb{R}^p, \mathbb{I}_p)$, car la matrice à diagonaliser R est de dimension (p, p).

Soient $u_1, \cdots, u_r$ les r vecteurs propres normés de R associés aux r plus grandes valeurs propres $\lambda_1 \geq \cdots \geq \lambda_r$ respectivement. $\Delta u_1, \cdots, \Delta u_r$ les r premiers axes factoriels.

Projection des individus

La projection du $i^{ème}$ individu sur le $k^{ème}$ axe factoriel Δu_k est donnée par

$$\omega_i^{(k)} = \widetilde{X}_i u_k$$

Le vecteur des projections des n individus est

$$\omega^{(k)} = \begin{pmatrix} \omega_1^{(k)} \\ \vdots \\ \omega_n^{(k)} \end{pmatrix} = \widetilde{X} u_k$$

Projection des variables

La projection de la $j^{ème}$ variable sur le $k^{ème}$ axe factoriel Δs_k est donnée par

$$\psi_j^{(k)} = {}^t\widetilde{X}^j D_p s_k$$

Le vecteur des projections des p variables est

$$\begin{aligned}\psi^{(k)} &= \begin{pmatrix}\psi_1^{(k)}\\ \vdots \\ \psi_p^{(k)}\end{pmatrix} = {}^t\widetilde{X}D_p s_k \\ &= {}^t\widetilde{X}D_p\left(\frac{1}{\sqrt{\lambda_k}}\widetilde{X}u_k\right) \\ &= \frac{1}{\sqrt{\lambda_k}}(\lambda_k u_k) \\ &= \sqrt{\lambda_k}u_k\end{aligned}$$

Ainsi la projection de la $j^{\text{ème}}$ variable sur le $k^{\text{ème}}$ axe factoriel Δu_k est donnée par le $j^{\text{ème}}$ composante de $\sqrt{\lambda_k}u_k \in \mathbb{R}^p$.

4.4.2 Cas où on se place sur $\mathbb{R}^n$

Si $n \leq p$, on se place sur $(\mathbb{R}^n, D_p)$, car la matrice à diagonaliser $\widetilde{X}\,{}^t\widetilde{X}D_p$ est de dimension (n, n).

Soient $s_1, \cdots, s_r$ les r vecteurs propres normés de $\widetilde{X}\,{}^t\widetilde{X}D_p$ associés aux r plus grandes valeurs propres $\lambda_1 \geq \cdots \geq \lambda_r$ respectivement. $\Delta s_1, \cdots, \Delta s_r$ les r premiers axes factoriels.

Projection des individus

La projection du $i^{\text{ème}}$ individu sur le $k^{\text{ème}}$ axe factoriel Δu_k est donnée par

$$\omega_i^{(k)} = \widetilde{X}_i u_k$$

Le vecteur des projections des n individus est

$$\begin{aligned}
\omega^{(k)} &= \begin{pmatrix} \omega_1^{(k)} \\ \vdots \\ \omega_n^{(k)} \end{pmatrix} = \widetilde{X} u_k \\
&= \widetilde{X} \left(\frac{1}{\sqrt{\lambda_k}} \, {}^t\widetilde{X} D_p s_k \right) \\
&= \frac{1}{\sqrt{\lambda_k}} (\lambda_k s_k) \\
&= \sqrt{\lambda_k} s_k
\end{aligned}$$

Ainsi la projection du $i^{ème}$ individu sur le $k^{ème}$ axe factoriel Δs_k est donnée par la $i^{ème}$ composante de $\sqrt{\lambda_k} s_k \in \mathbb{R}^n$.

Projection des variables

La projection de la $j^{ème}$ variable sur le $k^{ème}$ axe factoriel Δs_k est donnée par

$$\psi_j^{(k)} = {}^t\widetilde{X}^j D_p s_k$$

Le vecteur des projections des p variables est

$$\psi^{(k)} = \begin{pmatrix} \psi_1^{(k)} \\ \vdots \\ \psi_p^{(k)} \end{pmatrix} = {}^t\widetilde{X} D_p s_k$$

4.5 Aide à l'interprétation

A partir des relations données précédemment, on définit quelques règles d'interprétation

. Un individu sera du côté des variables pour les quelles il a de fortes valeurs, inversement, il sera du côté opposé des variables pour lesquelles il a de faibles valeurs.

. Plus les valeurs d'un individu sont fortes pour une variable, plus il sera éloigné de l'origine suivant l'axe factoriel qui décrit au mieux cette variable.

. Deux individus qui sont à une même extrémité d'un axe et sont éloignés de l'origine sont semblables (proches).

. Deux variables très corrélées positivement sont de même côté sur un axe.

Les axes factoriels donnent les images approchées des deux nuages $N(I)$ des individus et $N(J)$ des variables. Il est donc nécessaire de définir des indicateurs (ou contributions) pour mesurer la qualité de l'approximation (ou représentation).

Définition 4.5.1 *On appelle inertie expliquée par le $r^{ième}$ axe factoriel, la quantité*

$$I_r = \frac{\lambda_r}{\sum\limits_{k \geq 1} \lambda_k}$$

Remarque 4.5.1 *On sait que $tr(R) = \sum\limits_{k \geq 1} \lambda_k$ où R est la matrice des corrélations alors $tr(R) = p$ et*

$$I_r = \frac{\lambda_r}{p}$$

Remarque 4.5.2 *En pratique l'inertie est exprimée en pourcentage $100 I_r \%$, et représente la quantité d'information recueilli sur l'axe factoriel.*

Définition 4.5.2 *L'inertie expliquée par le plan factoriel $(\Delta_r, \Delta_{r'})$ est donnée par*

$$I_{r,r'} = I_r + I_{r'} = \frac{\lambda_r + \lambda_{r'}}{p}$$

L'inertie expliquée par le sous espace factoriel de dimension r est donnée par

$$I_{1,\cdots,r} = \frac{\lambda_1 + \cdots + \lambda_r}{p}$$

Remarque 4.5.3 *En pratique, en générale, dés que le pourcentage d'inertie expliquée par le sous espace factoriel de dimension r dépasse 75%, on dit qu'on a une bonne visualisation. On remplace les p variables initiales par les r composantes principales.*

4.5.1 Interprétation des nouvelles variables

Corrélation entre les nouvelles et les variables initiales;

Soit la $k^{ème}$ nouvelle variable (ou composante principale) $\omega^{(k)}$ et $\widetilde{X}^j$ la $j^{ème}$ variable initiale centrée-réduite.

Proposition 4.5.1 *La corrélation entre* $\omega^{(k)}$ *et* $\widetilde{X}^j$ *est égale à* $\sqrt{\lambda_k} u_k^j = \psi_j^{(k)}$. *Où* u_k^j *représente la* $j^{ème}$ *composante du vecteur propre* u_k .

Démonstration.

$$Corr\left(\omega^{(k)}, \widetilde{X}^j\right) = \frac{Covar\left(\widetilde{X}^j, \omega^{(k)}\right)}{\sqrt{Var\left(\omega^{(k)}\right)}\sqrt{Var\left(\widetilde{X}^j\right)}}$$

$Var\left(\omega^{(k)}\right) = \lambda_k$ et $Var\left(\widetilde{X}^j\right) = 1$

$$\begin{aligned} Covar\left(\widetilde{X}^j, \omega^{(k)}\right) &= \left\langle \widetilde{X}^j, \omega^{(k)} \right\rangle_{D_p} \\ &= {}^t\widetilde{X}^j D_p \omega^{(k)} \\ &= {}^t\widetilde{X}^j D_p \widetilde{X} u_k \end{aligned}$$

or ${}^t\widetilde{X}^j D_p \widetilde{X} u_k$ est la $j^{ème}$ composante de la matrice Ru_k et $Ru_k = \lambda_k u_k$.

Alors,

$$Corr\left(\omega^{(k)}, \widetilde{X}^j\right) = \sqrt{\lambda_k} u_k^j = \psi_j^{(k)}$$

∎

Cercle des corrélations

Pour tout couple de composantes principales $\left(\omega^{(k)}, \omega^{(k')}\right)$, dans un cercle de centre $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, on représente les points $M_j = \begin{pmatrix} Corr\left(\omega^{(k)}, \widetilde{X}^j\right) \\ Corr\left(\omega^{(k')}, \widetilde{X}^j\right) \end{pmatrix}$.

Plus les points $M_j, j = 1 \cdots, p$ sont éloignés du centre, plus les variables $\widetilde{X}^j$ sont fortement corrélées avec les composantes principales. Donc bien représentées dans le plan $\left(\Delta\omega^{(k)}, \Delta\omega^{(k')}\right)$.

4.5.2 Les contributions

Qualité de représentation d'un élément sur l'axe factoriel

Un élément représente soit un individu ou une variable.

Définition 4.5.3 *On appelle contribution relative du $i^{\text{ème}}$ individu sur le $k^{\text{ème}}$ axe factoriel, la quantité donnée par*

$$C_r^k(i) = \frac{\left|\omega_i^{(k)}\right|}{\left\|{}^t\widetilde{X}_i\right\|_{\mathbb{I}_p}^2} = \cos^2(\theta)$$

où θ est l'angle formé par ${}^t\widetilde{X}_i$ et u_k.

On appelle contribution relative de la $j^{\text{ème}}$ variable sur le $k^{\text{ème}}$ axe factoriel, la quantité donnée par

$$C_r^k(j) = \frac{\left|\psi_j^{(k)}\right|}{\left\|\widetilde{X}^j\right\|_{D_p}^2} = \cos^2(\theta)$$

où θ est l'angle formé par $\widetilde{X}^j$ et s_k.

La contribution relative d'un élément représente la qualité de visualisation de l'élément sur l'axe factoriel. Elle est donnée par le rapport de l'inertie de la projection sur l'axe et l'inertie totale. Ainsi, si la contribution relative est proche de 1, l'élément est proche de l'axe et donc, proche du plan de projection contenant l'axe. La projection de l'élément est donc proche de la valeur réelle. On dit que l'élément est très bien représenté sur l'axe.

Définition 4.5.4 *On appelle contribution relative du $i^{ème}$ individu sur le sous espace factoriel de dimension r $(\Delta u_1, \cdots, \Delta u_r)$, la quantité donnée par*

$$C_r^{1,\cdots,r}(i) = \sum_{1 \leq k \leq r} C_r^k(i) = \cos^2(\theta)$$

où θ est l'angle formé par ${}^t\widetilde{X}_i$ et $(\Delta u_1, \cdots \Delta u_r)$.

On appelle contribution relative de la $j^{ème}$ variable sur le sous espace factoriel de dimension r $(\Delta s_1, \cdots, \Delta s_r)$, la quantité donnée par

$$C_r^{1,\cdots,r}(j) = \sum_{1 \leq k \leq r} C_r^k(j) = \cos^2(\theta)$$

où θ est l'angle formé par $\widetilde{X}^j$ et $(\Delta s_1, \cdots \Delta s_r)$.

Qualité de représentation du nuage des individus

La qualité de représentation du nuage est donnée par le pourcentage d'inertie associée à l'axe, ie. le rapport d'inertie de la projection du nuage sur l'axe et l'inertie totale du nuage. Donnée par

$$QLTn = \sum_{1 \leq j \leq n} C_r^k(i)$$

Cette quantité mesure l'importance d'un axe factoriel. Bien sûr, les premiers axes auront plus d'importance que les suivant. Nous jugons ces pourcentages en fonction de la taille du tableau. Par exemple, 10% est une valeur faible pour un tableau qui comporte 10 variables, mais c'est une valeur forte pour dans le cas de 100 variables.

Contribution d'un élément à l'inertie d'un axe ou contribution absolue

Définition 4.5.5 *On appelle contribution absolue du $i^{ème}$ individu à l'inertie du $k^{ème}$ axe factoriel, la quantité donnée par*

$$C_a^k(i) = p_i \frac{\left|\omega_i^{(k)}\right|}{\lambda_k}$$

On appelle contribution absolue de la $j^{\text{ème}}$ variable à l'inertie du $k^{\text{ème}}$ axe factoriel, la quantité donnée par

$$C_a^k(j) = \frac{\left|\psi_j^{(k)}\right|}{\lambda_k}$$

La contribution absolue est le rapport de l'inertie de la projection de l'élément sur l'axe et l'inertie de l'ensemble des points du nuage. Elle représente la part de l'élément pour construire l'axe factoriel et permet de mettre en évidence le sous ensemble des éléments ayant participés essentiellement à la construction de l'axe.

Définition 4.5.6 *On appelle contribution absolue de deux individus i_1 et i_2 à l'inertie du $k^{\text{ème}}$ axe factoriel, la quantité donnée par*

$$C_a^{k,}(i_1, i_2) = C_a^k(i_1) + C_a^k(i_2)$$

On appelle contribution absolue de deux variables j_1 et j_2 à l'inertie du $k^{\text{ème}}$ axe factoriel, la quantité donnée par

$$C_a^k(j_1, j_2) = C_a^k(j_1) + C_a^k(j_2)$$

Remarque 4.5.4 *1.*

2. $\forall i = 1, \cdots, n$ *et pour tout* $k \geq 1$*, et* $0 \leq C_a^k(i) \leq 1$.

3. $\forall j = 1, \cdots, p$ *et pour tout* $k \geq 1$*, et* $0 \leq C_a^k(j) \leq 1$.

4. $\sum\limits_{1 \leq i \leq n} C_a^k(i) = 1$ *et* $\sum\limits_{1 \leq j \leq n} C_a^k(j) = 1$

En effet,

$$\begin{aligned}
\sum_{1 \leq i \leq n} C_a^k(i) &= \sum_{1 \leq i \leq n} p_i \frac{\left|\omega_i^{(k)}\right|}{\lambda_k} \\
&= \frac{1}{\lambda_k} \left\|\omega^{(k)}\right\|_{D_p}^2 \\
&= \frac{1}{\lambda_k} Var\left(\omega^{(k)}\right) \\
&= 1
\end{aligned}$$

et

$$\begin{aligned}\sum_{1\leq j\leq p} C_a^k(i) &= \sum_{1\leq j\leq p} \frac{\left|\psi_i^{(k)}\right|}{\lambda_k} \\ &= \frac{1}{\lambda_k}\left\|\psi^{(k)}\right\|_{\mathbb{I}_p}^2 \\ &= \frac{1}{\lambda_k}\,{}^t\left(\psi^{(k)}\right)\left(\psi^{(k)}\right) \\ &= \frac{1}{\lambda_k}\,{}^t\left({}^t\widetilde{X}D_p s_k\right)\left({}^t\widetilde{X}D_p s_k\right) \\ &= \frac{1}{\lambda_k}\,{}^t s_k D_p \widetilde{X}\,{}^t\widetilde{X} D_p s_k \\ &= \|s_k\|_{D_p}^2 \\ &= 1\end{aligned}$$

Définition 4.5.7 *On appelle contribution absolue du $i^{ème}$ individu à l'inertie du sous espace factoriel de dimension r $(\Delta u_1, \cdots, \Delta u_r)$, la quantité donnée par*

$$C_a^{1,\cdots,r}(i) = \sum_{1\leq k\leq r} C_r^k(i)$$

On appelle contribution absolue de la $j^{ème}$ variable à l'inertie du sous espace factoriel de dimension r $(\Delta s_1, \cdots, \Delta s_r)$, la quantité donnée par

$$C_a^{1,\cdots,r}(j) = \sum_{1\leq k\leq r} C_a^k(j)$$

Pour aider l'interprétation, nous devons s'appuyer sur les points suivants :

- Choisir le nombre d'axes. Notons que le choix du nombre d'axes reste un problème car il n'y a pas de solution rigoureuse. Les valeurs propres permettent de choisir le nombre d'axes de sorte que le pourcentage d'inertie cumulée soit supérieur ou égale à 75% environ, ou de sorte que les valeurs propres soient supérieurs où égales à 1, ou encore si on note un saut important dans l'histogramme des valeurs propres. Le nombre d'axes ne doit pas être trop grand.

- Etudier les indicateurs de la qualité des approximations.

- Interpréter les facteurs simultanément : En calculant

  a/ les contributions des individus

  b/ les coordonnés des individus et variables

CHAPITRE 5

Analyse discriminante

Introduction

L'analyse discriminante est une méthode statistique qui consiste à construire une variable qualitative à k modalités (ou catégories) en utilisant p variables explicatives (qualitatives).

On considère un ensemble Ω (ensemble des individus) muni d'une partion apriori $P(p_1, ..., p_k)$ en k classes. Chaque individu $\omega \in \Omega$ est caractèrisé par un ensemble de p variable descriptives (ou predicteurs).(variable quantitative). Le but général, de l'analyse discriminante est de séparer au maximum les k classes apriori sur Ω à partir des p predicteurs.

Exemple 1:Analyse linguistique.

Soit Ω l'ensemble des députés élus en 1981. La partition p sur Ω comprend deux classes: les députés de droite et les députés de gauche. Les descripteurs sont les fréquences d'utilisation de 53 mots figurant dans les discours des députés. le but de l'etude est de voir si les mots utilisés permettent de retrouver la séparation droite-gauche.

Il s'agit ici d'une étude **descriptive** .Le but n'est pas de découvrir la tendnace des députés mais de faire une étude politico linguistique.

Exemple 2:Prévision de risque d'avalanche

Soit Ω l'ensmble des stations de ski des Alpes. La partiton apriori de Ω comprend 3 classes: Avalanche à la station , couloir d'avalanche,et pas d'avalanche. les descripteurs sont au nombre de 274 et concernent des mes météorologiques.

Il s'agit d'analyse descriptive à but **prévisionnel** :à partir de l'etude du passé le but est de determiner chaque jour la station a haut risque d'avalanche.

Le but de l'analyse previsionnelle ou decisionnelle est de construire à partir d'un echantillion d'individus connus des règles de decisions statistiques.Ces règles serviront à affecter dans le futur, les individus à une classe apriori et deveront minimiser les erreurs de prévision ou de diagnostie.

En résume: affecter un individus à une classe apriori au moyen de fonctions des variables explicatives avec une probabilité d'erreur minimale.

Les fonctions peuvent être explicite(lineaire, quadratique ou logique) ou implicite (estimation de densité).

Dans ce cour nous nouslimitons à **l'analyse factorielle discriminante,** qui consiste à rechercher un ensemble de s $(s < p)$ variables de classement (axe dicriminants) qui soient des combinaisons lineaires des p variables initiales les plus dicriminantes possibles au sens de certains critères à fixer.

Ce qui signifie que les valeurs des individus d'une même classe par ces variables dis-rciminantes sont le plus concentré possible et celles prises par les individus des classes $\neq$ sont les plus éloignée possibles.

Un critère naturel utilisant les données est d'examiner les variances intraclasse (dans la classe) et variance interclasses (entre classe).

5.1 Notations et definitions:

Soit Ω l'ensemble des individus et P $(P_1, ..., P_k)$ la partition definie a priori sur Ω.On dispose d'un echantillion de taille n (échantillion d'apprentissage) des élements de Ω. On disigne par X, la matrice (n, p) représentant le tableau de données par les p variables descriptibles.

$X = (X_i^j \;\; ; i = 1, ..., n \; ; j = 1, ..., p)$ où X_i^j est la valeur prise par le $i^{ème}$ individus pour la $j^{ème}$variable

A chaque individus ω_i on associe un vecteur

$$\underline{X}_i = \begin{pmatrix} X_i^1 \\ . \\ . \\ . \\ X_i^p \end{pmatrix} \in \mathbb{R}^p$$

qui correspont à la $i^{ème}$ ligne de X. $\mathbb{R}^p$ est l'espace des individus.

Et à chaque variable on associe un vecteur

$$\underline{X}^j = \begin{pmatrix} X_1^j \\ . \\ . \\ . \\ X_n^j \end{pmatrix} \in \mathbb{R}^n$$

qui correspont à la$j^{ème}$ colonne du tableau.$\mathbb{R}^n$ est l'espace des variables.

$\underline{X}_i$ le $i^{ème}$ individus et $\underline{X}^j$ la $j^{ème}$variable.

On suppose que chaque individus est muni d'un poids p_i $\left(\sum_{i=1}^n p_i = 1\right)$ et

$$D_p = \begin{pmatrix} p_1 & 0 & 0 \\ & . & \\ 0 & . & 0 \\ & & . \\ 0 & 0 & p_n \end{pmatrix}$$

la metrique du poids.

Elements descriptifs de l'espace des variables

a/ Moyenne $\overline{\underline{X}}^j$ $\quad \overline{\underline{X}}^j =< \underline{X}^j, 1 >_{D_p}$

b/ Variance $Var(\underline{X}^j) = \sigma^2_{\underline{X}^j} = \sum p_i (X_i^j - \overline{\underline{X}}^j)^2 =< \underline{X}^j - 1\overline{\underline{X}}^j \cdot \underline{X}^j - 1\overline{\underline{X}}^j >_{D_p}$.

c/ Ecart-type $\sigma_j = \sqrt{Var(\underline{X}^j)}$.

la Var mesure la dispersion autour de la moyenne.

d/ covariance; $cov(\underline{X}^j, \underline{X}^{j'}) = \sum_{i=1}^{n} p_i(X_i^j - \overline{\underline{X}}^j)(X_i^{j'} - \overline{\underline{X}}^{j'}) =< \underline{X}^j - 1\overline{\underline{X}}^j \cdot \underline{X}^{j'} - 1\overline{\underline{X}}^{j'} >_{D_p}$.

Remarque 5.1.1 $\widetilde{X} = (X_i^j - \overline{\underline{X}}^j;\ i = 1, ..., n\ ;\ j = 1, ..., p)$

Elements descriptifs de l'espace des individus.

a/ Le nuage des points $N = \{(\underline{X}_i, p_i)\ /\ i = 1, ..., n\ ;\ \underline{X}_i \in R^p\}$

b/ Le centre de gravité ou barycentre des ω_i muni des poids p_i (en général, on prend $p_i = \frac{1}{n}$)

$$g = \sum_{i=1}^{n} p_i \underline{X}_i = \frac{1}{n} \sum_{i=1}^{n} \underline{X}_i$$

Utilité de la vairable qualitative

La variable qualitative correspond à la partition apriori $P\ (P_1, ..., P_k)$ de centre de gravité respectivement $G = (g_1, ..., g_k)$ et la matrice de $Var - Covar\ \ (W_1, ..., W_k)\ \ w_l$ designe la dispersion à l'intérieur de la classe P_l.L'ensemble des réalisation de l est donnée par :

$$\Phi(\Omega) = \{y_i^l, i = 1, ..., n; j = 1, ..., k\}$$

Avec

$$y_i^l = \begin{cases} 1\ si \quad w_l \in P_l \\ \quad 0 \text{ sinon} \end{cases}$$

Poids des classes Comme chaque individus est muni d'un poids p_i, alors on attribue à chaque classe P_l un poid :

$$p_l = \sum_{w_i \in P_l} p_i$$

Relation entre g et g_l On a :

$$g = \sum_{l=1}^{k} p_l g_l\ . \qquad Avec \quad g_l = \frac{1}{p_l} \sum_{w_i \in P_l} p_i \underline{X}_i$$

En effet:

$$\begin{aligned}\sum_{l=1}^{k} p_l g_l &= \sum_{l=1}^{k} p_l \frac{1}{p_l} \sum_{\omega_i \in P_l} p_i \underline{X}_i \\ &= \sum_{l=1}^{k} \sum_{\omega_i \in P_l} p_i \underline{X}_i \\ &= \sum_{i=1}^{n} p_i \underline{X}_i = g\end{aligned}$$

Matrice de Variance Covariance

a/ Intérieur des classes: Matrice var-covar intra-classe

On note par $W_{(p,p)}$ (du mot anglais "Within") la matrice var-covar intra-classe

$$W = \sum_{l=1}^{k} p_l w_l \quad \text{Avec} \quad w_l = \frac{1}{p_l} \sum_{w_i \in P_l} p_i (\underline{X}_i - g_l)\ {}^t(\underline{X}_i - g_l)$$

W est la moyenne des matrices w_l ; $l = 1, ..., k$

$$W = \frac{1}{p_l} \sum_{l=1}^{k} \sum_{w_i \in P_l} p_i (\underline{X}_i - g_l)\ {}^t(\underline{X}_i - g_l)$$

b/Entrelesclasses:Matricedevar-covar inter-classe

On note par B (du mot anglais "between") la matrice var-covar inter-classe

$$B = \sum_{l=1}^{k} p_l (g_l - g)\ {}^t(g_l - g)$$

La martice B traduit l'eloignement entre les classes:B mesure la séparation des classes.

Relation entre B et $G = (g_1, ..., g_k)$ On a

$$B = {}^t G D_p G \qquad et \ \ D_p = \begin{pmatrix} p_1 & 0 & 0 \\ & . & \\ 0 & . & 0 \\ & & . \\ 0 & 0 & p_k \end{pmatrix}$$

5.1.1 Matrice de Variance-Covariance totale

Soit $V = (cov(\underline{X}^j, \underline{X}^{j'}))_{\substack{1<j<p \\ 1<j'<p}}$ la matrice de var-covar

$$V = \sum_{i=1}^{n} p_i(\underline{X}_i - g)\ {}^t(\underline{X}_i - g)$$

La matrice de var-covar totale V mesure la dispersion totale de toutes les individus.

Relation entre V et X

Si les variable centrées $V = \sum_{i=1}^{n} p_i \underline{X}_i \ \cdot^t \underline{X}_i$ car $g = 0_{\mathbb{R}^p}$

$V = {}^tXD_pX$ (sinon on pose $\widetilde{X}$ la matrice centrée)

Théorème d'Huygens :

Théorème 5.1.1 *On a*

$$V = W + B$$

Démonstration. (Exercice TD) ■

5.2 Variances et Covariances des Variables discriminantes.

Théorème 5.2.1 *Si les variables discriminantes Z sont données par*

$$Z = \sum_{j=1}^{p} u^j x^j \in \mathbb{R}^n$$

(Combinaisons lineaire des variables initiales), alors

$$Var(Z) = \frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z})^2 =^t uVu,$$

avec $u = \begin{pmatrix} u^1 \\ . \\ . \\ . \\ u^p \end{pmatrix}$ *et* $\overline{Z}$ *est la moyenne de* Z.

La covariance est donnée par

$$cov(Z, Z') =^t uVu' =^t u'Vu$$

Démonstration. (ExerciceTD) ■

Définition 5.2.1 *la vairiance intra-classe de* Z *est donnée par :*

$$Intra(Z) = \sum_{l=1}^{k} \frac{n_l}{n} \sum_{w_i \in P_l} \frac{1}{n_k} (Z_i - \overline{Z})^2 =^t uWu.$$

Définition 5.2.2 *La variance inter-classe de* Z *est donnée par :*

$$Inter(Z) = \sum_{l=1}^{k} \frac{n_l}{n} (\overline{Z} - \overline{Z}_k)^2 =^t uBu.$$

Remarque 5.2.1 *Comme on a* $V = W + B$ *(Variance totale= Variance intra-classe +Variance inter-classe) alors*

$$^t uVu =^t uWu +^t uBu$$

Ce qui signifie que la variance totale de Z *est égale à la vairiance de* Z *intra-classe + variance de* Z *inter-classe .*
La quantité $^t u\ V\ u$ *étant indépendante des groupes, minimiser* $^t u\ W\ u$ *est équivalent à maximiser* $^t u\ B\ u$.

5.3 Choix des variables discriminantes

C'est la deterrmination des axes discriminants.

On cherche les variables discriminantes Z de variance totale $V(Z)$ donnée dont la variance inter-classe B est maximale.(de manière séparer au mieux les classes). Par la relation

$$^{t}uVu =^{t} uWu +^{t} uBu.$$

On aura

$$\max_{u}^{t} uBu \Leftrightarrow \min_{u}^{t} uWu$$

Problème d'optimisation

On peut formuler le problème d'optimisation de plusieurs façons

Problème1

Trouver $u \in R^p$ telque :

$\frac{^{t}uBu}{^{t}uVu}$ max sous la contrainte

$$^{t}uVu = C^{ste}.$$

La discrimination est alors impossible.

La statistique de decision est $\Lambda = \frac{|W|}{|V|}$,

$|.|$ est le determinant

Sous l'hypothèse H_0, Λ suit la loi de wilks de paramètre $(n-1), p, (k-1)$ notée $\Lambda(n-1, p, k-1)$.La région critique du test

$$RC = [\Lambda < \Lambda_{tab}^{\alpha}] \ ,$$

Λ est tabulée au niveau α fixé.

Si H_0 est rejetté, il est possible de projeter les individus dans un nouvel espace factoriel permettant de distinguer au mieux les classes.

Variance expliquée

Le pouvoir discriminant d'un axe factoriel est évalué par le rapport $\frac{\lambda}{\sum_{j=1}^{p-1}\lambda_j}$ inertie ou proportion de variance expliquée par l'axe.

L'inertie ou proportion de variance expliquée par le plan obtenue par les deux premièrs axes factoriels est $\frac{\lambda_1+\lambda_2}{\sum_{j=1}^{p-1}\lambda_j}$.

De manière générale l'inertie expliquée par le s espace factoriels de dim s est

$$\frac{\lambda_1 + \lambda_2 + ... + \lambda_s}{\sum_{j=1}^{p-1} \lambda_j}; \quad s < p - 1.$$

Remarque 5.3.1 *La somme est jusqu'à $k-1$ car il y a au plus $(k-1)$ vps non nulles si $k < p$.*
En effet, la matrice B des variance-covariance intra-classe est la martice associe aux centres de gravité $g_{1,...,}g_k$.
Or on a

$$\sum_{k=1}^{k} p_k g_k = g,$$

d'où B est de rang $\leq k-1 \Rightarrow rang(V^{-1}B) \leq \min(p, k-1)$
Si

$$k < p \Rightarrow k-1 < p \Rightarrow \min(p, k-1) = k-1 \Rightarrow rang(V^{-1}B) \leq (k-1).$$

Bibliographie

[1] J. P. Benzekri. Analyse des données (Tome 1). La taxinomie. Dunod. (1980).

[2] J. P. Benzekri. Analyse des données (Tome 2). L'analyse des correspondances. Dunod. (1980).

[3] G. Celeux, E. Diday, G. Govaert, Y. Lechevallier et H. Ralambondrainy. Classification automatique des données. Dunod (1989).

[4] B. Escoffier et J. Pagées. Analyse factorielle simple et multiples objectifs, Méthodes et interprétations. dunod. (1990)

[5] G. Hebrail et Y. Lechevallier. Analyse des données, chapitre Data Mining et analyse des données. Hermes sciences publications. (2003).

[6] L. Lebart, A. Morineau et M. Piron. Statistique exploratoire multidimensionnelle. Dunod. (1995).

[7] A. Martin. L'analyse de données. Polycopier de cours ESNIETA- ref 1463. (2003).