



Département de mathématique
Licence d'excellence
Machine Learning and artificial intelligence



Unsupervised Learning

Chapitre 4 :

Analyse en Composantes Principales ACP : 2 Variables

Pr. Harchli Fidaé

Table of Contents

1 Introduction

- Histoire
- Objectifs du cours
- Pré-requis au cours

2 Quelques Notions d'Algèbre Linéaire

- Produit scalaire
- Norme d'un vecteur
- Orthogonalité

Table of Contents

3 Méthodes Géométriques en Statistique

- Moyenne
- Variance
- Covariance et Corrélation
- Corrélation

4 Analyse en Composantes Principales

- Définition des Espaces
- Régression Orthogonale
- Notion d'Inertie
- Valeurs et Vecteurs propres
- Coordonnées Factorielles

5 Exemple d'Application

- Méthode Analytique
- Résolution par Logiciel

Introduction

Pourquoi ce Cours ?

Introduction

L'analyse factorielle est une technique statistique utilisée surtout pour dépouiller des enquêtes : elle permet, quand on dispose d'une population d'individus pour lesquelles on possède de nombreux renseignements concernant les opinions, les pratiques et le statut (sexe, âge, etc.), d'en donner une représentation géométrique, c'est-à-dire en utilisant un graphique qui permet de voir les rapprochements et les oppositions entre les caractéristiques des individus.

Introduction

Cette technique est déjà centenaire : elle a été créée en 1904 par le psychologue anglais **Charles Spearman**, dans le but de mesurer l'intelligence. Sa technique porte le nom aujourd'hui **d'analyse factorielle des psychologues**.

D'autres techniques d'analyse factorielle ont été développées par la suite. Créée dans les années 1960 par **Jean-Paul Benzécri**, l'analyse en composantes principales (**ACP**) est l'une des variétés les plus importantes de l'analyse factorielle des correspondances (**AFC**).

Introduction

Ils ont dit :

G. Morlat

Les services rendus montrent bien que l'Analyse des Données constitue aujourd'hui, et de loin, la partie la plus immédiatement rentable de la statistique.

Objectifs du cours

- Présenter de nouvelles techniques de l'analyse de tas de données et exposer les outils de base de l'analyse de données.
- Former l'étudiant pour qu'il soit capable de :
 - Extraire, traiter et décrire l'information contenue dans des amas de données ;
 - Lire correctement les résultats des logiciels ;
 - Résoudre des problèmes avec données réelles.

Pré-requis du cours

- **Algèbre linéaire et géométrie dans l'espace**
- **Analyse statistique unidimensionnelle et bidimensionnelle**
- **Inférence statistique (de préférence)**

1 Introduction

- Histoire
- Objectifs du cours
- Pré-requis au cours

2 Quelques Notions d'Algèbre Linéaire

- Produit scalaire
- Norme d'un vecteur
- Orthogonalité

Produit scalaire canonique

Soit $f(X, Y)$ la forme bilinéaire symétrique ($f(x, y) = f(y, x)$) définie par :

$$f(X, Y) = X'AY$$

avec A est une matrice carrée symétrique (n lignes et n colonnes).

On obtient le produit scalaire sur \mathbb{R}^n lorsque $A = I_n$:

$$((x_1, \dots, x_n), (y_1, \dots, y_n)) \longrightarrow \langle X | Y \rangle = X' I_n Y$$

Produit scalaire canonique

$$\langle X|Y \rangle = (x_1 \dots x_2 \dots x_n) \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^n x_i y_i$$

Produit scalaire défini par une matrice des poids

Soit P une **matrice diagonale** dont les éléments diagonaux sont *des nombres réels strictement positifs* :

$$\langle X|Y \rangle_P = X'PY = \sum_{i=1}^n \delta_i x_i y_i$$

avec δ_i est dit le $i^{\text{ème}}$ poids.

Remarque 1

Le produit scalaire **canonique** correspond au cas où la matrice P est la matrice unité I_n .

Remarque 2

Si $P = \frac{1}{n}I_n$, tous les poids sont égaux à $\frac{1}{n}$ et la somme des poids vaut 1.

Produit scalaire défini par une matrice des poids

Nous avons alors :

$$\langle X|Y \rangle_P = X'PY = \frac{1}{n}X'I_nY = \frac{1}{n}\sum_{i=1}^n x_i y_i$$

Norme d'un vecteur

Si f est un produit scalaire sur \mathbb{R}^n , le nombre réel positif

$$||X||_f = \sqrt{f(X, X)} = \sqrt{\langle X|X \rangle_f}$$

est la **norme de X** , ou encore **la longueur de X** .

On dit qu'un vecteur est **normé** pour f (ou unitaire) si sa longueur $||X||_f$ est égale à **1**.

Par exemple, dans \mathbb{R}^2 muni du produit scalaire canonique, la longueur de $X = (x_1, x_2) = (1, 0)$ est :

$$||X||_f = \sqrt{x_1^2 + x_2^2} = \sqrt{1^2 + 0^2} = 1$$

Norme d'un vecteur

Remarque

Si le vecteur x est **non-normée** alors le vecteur transformé $\frac{x}{\|x\|}$ est normé pour n'importe quel vecteur x

La **distance** entre deux points $X = (x_1, \dots, x_n)$ et $Y = (y_1, \dots, y_n)$ est :

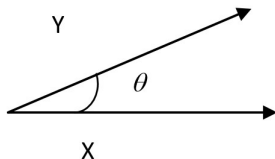
$$d(X, Y) = \|X - Y\|_{l_n} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Norme d'un vecteur

Soit X et Y deux vecteurs de \mathbb{R}^n , et soit θ l'angle que forme les deux vecteurs entre eux.

Alors :

$$\langle X|Y \rangle_f = \|X\|_f \|Y\|_f \cos(\theta)$$



Orthogonalité

Deux vecteurs X et Y de \mathbb{R}^n sont **orthogonaux** si, et seulement si, leur produit scalaire est nul :

$$\langle X|Y \rangle_f = 0$$

- Le vecteur nul est orthogonal à tout vecteur de \mathbb{R}^n
- L'angle entre deux vecteurs non nuls orthogonaux est $\frac{\pi}{2}$
- La base **canonique** de \mathbb{R}^n muni du produit scalaire canonique est formée de vecteurs **normés orthogonaux** deux à deux : on parle alors de la **Base orthonormée**

3 Méthodes Géométriques en Statistique

- Moyenne
- Variance
- Covariance et Corrélation
- Corrélation

4 Analyse en Composantes Principales

- Définition des Espaces
- Régression Orthogonale
- Notion d'Inertie
- Valeurs et Vecteurs propres
- Coordonnées Factorielles

5 Exemple d'Application

- Méthode Analytique
- Résolution par Logiciel

Moyenne

La moyenne \bar{x} de X n'est autre que le produit scalaire de X par

le vecteur unité $1_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$:

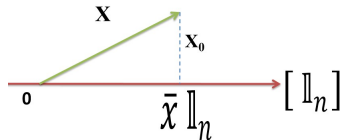
$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1, \dots, x_n) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = X' P 1_n = \langle X | 1_n \rangle_P$$

Remarque : le vecteur unité 1_n est unitaire pour ce produit scalaire :

$$\|1_n\| = \sqrt{\frac{1}{n}(1^2 + \dots + 1^2)} = 1$$

Notons $X_0 = X - \bar{x}1_n$ la variable **centrée**. On vérifie qu'elle est orthogonale à 1_n .

$$X_0 = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} - \begin{pmatrix} \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} - \bar{x} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = X - \bar{x}1_n$$



Variance

La variance de la variable X est le carré de la norme de la variable centrée X_0 :

$$V(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \langle X_0 | X_0 \rangle_{I_n}$$

$$V(X) = \frac{1}{n} X_0' X_0 = \|X_0\|_p^2$$

Covariance

La covariance est le produit scalaire des variables centrées :

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^n X_{0i} Y_{0i}$$

$$\text{Cov}(X, Y) = \frac{1}{n} X_0' Y_0 = X_0' \frac{1}{n} I_n Y_0 = \langle X_0 | Y_0 \rangle_p$$

Corrélation

Le coefficient de corrélation linéaire est le cosinus de l'angle entre les variables centrées. En effet :

$$r_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{\langle X_0 | Y_0 \rangle_P}{\|X_0\|_P \|Y_0\|_P}$$

$$r_{XY} = \frac{\|X_0\|_P \|Y_0\|_P \cos(\theta)}{\|X_0\|_P \|Y_0\|_P}$$

$$r_{XY} = \cos(\theta)$$

3 Méthodes Géométriques en Statistique

- Moyenne
- Variance
- Covariance et Corrélacion
- Corrélacion

4 Analyse en Composantes Principales

- Définition des Espaces
- Régression Orthogonale
- Notion d'Inertie
- Valeurs et Vecteurs propres
- Coordonnées Factorielles

5 Exemple d'Application

- Méthode Analytique
- Résolution par Logiciel

ACP à deux variables

On se donne un couple de deux variables (X, Y) qu'on appelle :
la variable Statistique **quantitative** bidimensionnelle à valeurs dans \mathbb{R}^2 .

Dans cette section, notre objectif est d'étudier une Population de n individus selon la Variable Statistique Bidimensionnelle (X, Y) .

Définition des Espaces

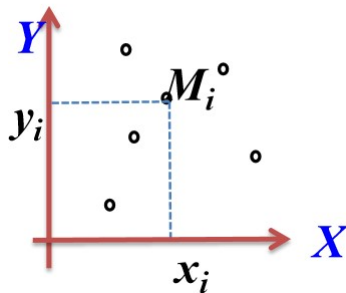
Espace des individus :

C'est l'espace vectoriel dans lequel évoluent les n individus de la population et où chaque axe est représenté par une variable statistique. Sa dimension est donc égale aux nombres de variables statistiques mises en jeu.

Ici, puisqu'on a que deux variables ($p = 2$), l'espace des individus est $\mathbb{R}^p = \mathbb{R}^2$

On muni l'espace des individus, \mathbb{R}^2 , du produit scalaire canonique grâce à la matrice identité $Q = I_2$ et de la base canonique $\{e_1, e_2\}$ orthonormée pour ce produit scalaire.

Espace des individus



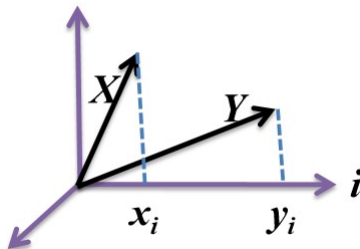
Définition des Espaces

Espace des variables :

- Les n valeurs $(x_1, \dots, x_i, \dots, x_n)$ de X pour les n individus peuvent être considérées comme les coordonnées d'un vecteur de \mathbb{R}^n .
- Même chose pour les n valeurs $(y_1, \dots, y_i, \dots, y_n)$ de Y .
- \mathbb{R}^n est appelé **l'espace des variables** de dimension n .

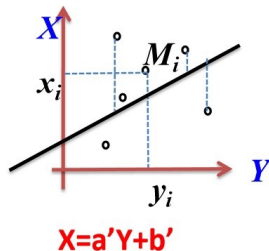
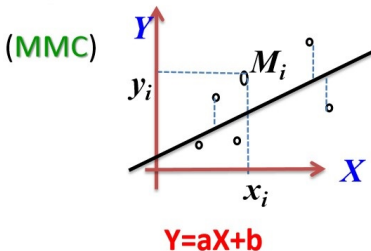
On muni l'espace des variables, \mathbb{R}^n , du produit scalaire défini par la matrice des poids uniformes $P = \frac{1}{n}I_n$.

Espace des variables



Régression Orthogonale

Dans la régression, lorsqu'on veut expliquer Y par X (i.e ajuster le nuage des "points individus" par une droite), on minimise la somme des carrés des écarts entre les individus et leurs projetés sur la droite parallèlement à l'axe Y . Et si on veut expliquer X par Y , on refait ceci mais parallèlement à X .



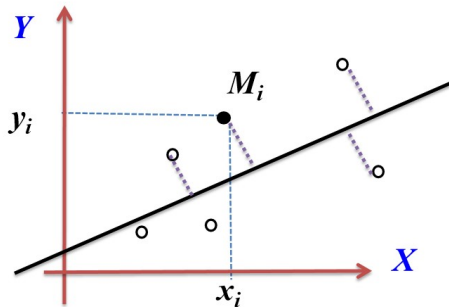
Régression Orthogonale

Par contre, si on ne veut expliquer aucune des deux variables statistiques par l'autre, nous serons alors conduits à une autre technique, **celle de la réduction des données**.

C'est la Régression orthogonale

Droite de la régression orthogonale

Nous cherchons alors dans \mathbb{R}^2 une droite (D) qui minimise la somme des carrés des écarts entre les individus et leurs **Projetés Orthogonaux** sur la droite.



Présentation des données

On présente les données sous forme d'un tableau à double entrée (**Observation** / **Variable**), soit par une matrice $M_{(n \times 2)}$ à n lignes et deux colonnes.

$$\begin{array}{c} M_1 \\ \vdots \\ M_n \end{array} \begin{array}{c} \xrightarrow{\quad} \\ \\ \xrightarrow{\quad} \end{array} M = \begin{array}{cc} \overset{X}{x_1} & \overset{Y}{y_1} \\ \vdots & \vdots \\ x_n & y_n \end{array}$$

la ligne i représente la valeur prise par le $i^{\text{ème}}$ individu (M_i) pour les deux variables quantitatives X et Y .

Présentation des données

Dans l'espace des Individus, chaque individu est représenté par un point M_i de coordonnées (X_i, Y_i) . Soit $G(\bar{x}, \bar{y})$ le barycentre du nuage.

G sera pris comme nouvelle origine de l'espace, par conséquent le point M_i aura pour coordonnées dans le nouveau repère (X_{0i}, Y_{0i}) et on notera $\overrightarrow{GM_i} \begin{pmatrix} X_{0i} \\ Y_{0i} \end{pmatrix}$

$$Z = \begin{pmatrix} X_{01} & Y_{01} \\ \vdots & \vdots \\ X_{0n} & Y_{0n} \end{pmatrix}$$

Avec Z est la matrice des données centrées

Position du problème

Soit (D) une droite passant par G et considérons u_{\perp} le vecteur unitaire de \mathbb{R}^2 orthogonal à la droite (D) :

$$u_{\perp} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

Et

$$\|u_{\perp}\|_2^2 = \alpha^2 + \beta^2 = 1$$

Le vecteur unitaire u porté par la droite (D) est $\begin{pmatrix} \beta \\ -\alpha \end{pmatrix}$

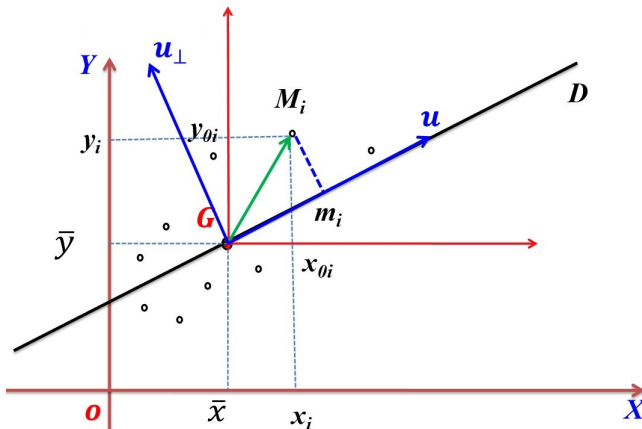
La droite (D) est l'ensemble des points $m = (x, y)$ vérifiant :

$$\langle u_{\perp} | \overrightarrow{Gm} \rangle = 0$$

$$\Rightarrow \alpha x_0 + \beta y_0 = 0$$

Position du problème

Soit m_i la projection orthogonale du point M_i sur la droite D , le vecteur $\overrightarrow{Gm_i}$ est le projeté orthogonal de $\overrightarrow{GM_i}$ sur le vecteur u :



On cherche la droite qui minimise la moyenne des carrés des écarts entre m_i et M_i , c-à-d :

$$\left\{ \begin{array}{l} \min_{\alpha, \beta} \quad S^2 \\ S.C. \\ \alpha^2 + \beta^2 = 1 \end{array} \right. \quad / \quad S^2 = \frac{1}{n} \sum_{i=1}^n ||\overrightarrow{m_i M_i}||^2$$

Cherchons alors le vecteur $\overrightarrow{m_i M_i}$.

On a $\overrightarrow{Gm_i}$ est la projection de $\overrightarrow{GM_i}$ sur le vecteur unitaire u alors :

$$\overrightarrow{Gm_i} = \frac{\langle \overrightarrow{GM_i} | u \rangle}{\langle u | u \rangle} u$$

$$\overrightarrow{Gm_i} = \langle \overrightarrow{GM_i} | u \rangle u = u' \overrightarrow{GM_i} u = (\beta x_{0i} - \alpha y_{0i}) \begin{pmatrix} \beta \\ -\alpha \end{pmatrix}$$

$$\begin{aligned} \overrightarrow{m_i M_i} &= \overrightarrow{GM_i} - \overrightarrow{Gm_i} = \begin{pmatrix} x_{0i} \\ y_{0i} \end{pmatrix} - (\beta x_{0i} - \alpha y_{0i}) \begin{pmatrix} \beta \\ -\alpha \end{pmatrix} \\ &= \begin{pmatrix} (1 - \beta^2)x_{0i} + \alpha\beta y_{0i} \\ (1 - \alpha^2)y_{0i} + \alpha\beta x_{0i} \end{pmatrix} = \begin{pmatrix} \alpha^2 x_{0i} + \alpha\beta y_{0i} \\ \beta^2 y_{0i} + \alpha\beta x_{0i} \end{pmatrix} \\ &= (\alpha x_{0i} + \beta y_{0i}) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \\ &= (\alpha x_{0i} + \beta y_{0i}) u_{\perp} \end{aligned}$$

$$||\overrightarrow{m_i M_i}||^2 = (\alpha x_{0i} + \beta y_{0i})^2 ||u_{\perp}||^2 = (\alpha x_{0i} + \beta y_{0i})^2 \times 1$$

D'où :

$$\begin{aligned} S^2 &= \frac{1}{n} \sum_{i=1}^n ||\overrightarrow{m_i M_i}||^2 \\ &= \frac{1}{n} \sum_{i=1}^n (\alpha x_{0i} + \beta y_{0i})^2 \\ &= ||(\alpha X_0 + \beta Y_0)||_{\mathbf{P}}^2 \end{aligned}$$

Donc chercher la droite (D) de régression orthogonale, revient à chercher, dans l'espace des variables \mathbb{R}^n , un vecteur $\alpha X_0 + \beta Y_0$ (combinaison linéaire fictive des deux variables X_0 et Y_0) minimal, sous la contrainte $\alpha^2 + \beta^2 = 1$ et où \mathbf{P} est la matrice des poids uniformes.

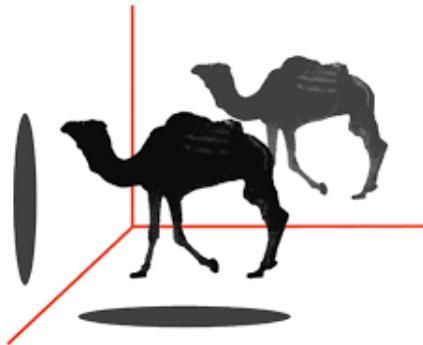
C'est l'objet de **l'analyse en composantes principales**

Notion d'Inertie

- Après l'approche géométrique du problème, nous allons nous pencher vers une approche analytique (notion d'Inertie).
- Rappelons qu'on se place dans \mathbb{R}^2 , l'espace des Individus, qu'on muni du produit scalaire canonique (i.e. défini par la matrice Q).
- La distance d'un point M_i au centre G est :

$$d(M_i, G) = \|\overrightarrow{GM_i}\|$$

Notion d'Inertie



- Meilleure approximation par projection
- Meilleure représentation de la diversité, de la variabilité

Inertie total d'un nuage

La variance est une mesure **unidimensionnelle** de la dispersion. Alors que l'**Inertie** en est une mesure **multidimensionnelle**.

Définition :

L'inertie du nuage des points est la moyenne des carrés des distances de G à tous les points :

$$I_T = \frac{1}{n} \sum_{i=1}^n \|\overrightarrow{GM_i}\|^2 = \frac{1}{n} \sum_{i=1}^n (x_{0i}^2 + y_{0i}^2) = V(X) + V(Y)$$

Inertie total d'un nuage par rapport à une droite

Définition :

Soit D_u la droite de vecteur directeur u , traversant le nuage et passant par son centre de gravité G , alors L'inertie du nuage par rapport à D_u est :

$$I(D_u) = \frac{1}{n} \sum_{i=1}^n d(M_i, D_u)^2 = \frac{1}{n} \sum_{i=1}^n ||\overrightarrow{M_i m_i}||^2 = S^2$$

Puisque m_i est la projeté orthogonal de M_i sur D_u .

Inertie total d'un nuage par rapport à une droite

Le problème maintenant revient à minimiser $I(D_u)$ (puisque égale à S^2). Or comme on a selon le théorème de Pythagore :

$$||\overrightarrow{GM_i}||^2 = ||\overrightarrow{Gm_i}||^2 + ||\overrightarrow{m_iM_i}||^2$$

Ce qui implique :

$$I_T = I(D_{u_\perp}) + I(D_u)$$

$$I(D_{u_\perp}) = I_T - I(D_u)$$

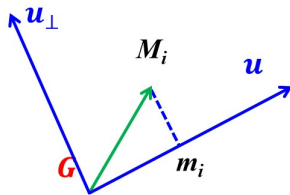
où $I(D_{u_\perp})$ est la droite orthogonale à D_u

Inertie total d'un nuage par rapport à une droite

Minimiser $I(D_u)$ revient donc à maximiser $I(D_{u_\perp})$

$$I(D_{u_\perp}) = \frac{1}{n} \sum_{i=1}^n \|\overrightarrow{Gm_i}\|^2$$

- $I(D_u)$ est appelée l'inertie résiduelle.
- $I(D_{u_\perp})$ est appelée l'inertie expliquée par D_u .



Liaison avec la variance-covariance

Nous avons déjà vu que l'inertie expliquée par D_u s'écrit :

$$\begin{aligned} I(D_{u_{\perp}}) &= \frac{1}{n} \sum_{i=1}^n ||\overrightarrow{Gm_i}|^2 \\ &= \frac{1}{n} \sum_{i=1}^n (\beta x_{0i} - \alpha y_{0i})^2 \\ &= \alpha^2 V(Y) + \beta^2 V(X) - 2\beta\alpha \text{Cov}(X, Y) \\ &= (\beta \quad -\alpha) \begin{pmatrix} V(X) & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & V(Y) \end{pmatrix} \begin{pmatrix} \beta \\ -\alpha \end{pmatrix} \\ &= \mathbf{u}' \mathbf{A} \mathbf{u} \end{aligned}$$

\mathbf{A} est la matrice des **variances-covariances**

Liaison avec la variance-covariance

En introduisant la matrice Z des variables centrées, la matrice des variances-covariances s'écrit sous les formes :

$$\begin{aligned} \mathbf{A} &= \begin{pmatrix} V(X) & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & V(Y) \end{pmatrix} \\ &= \frac{1}{n} \begin{pmatrix} x_{01} & \dots & x_{0n} \\ y_{01} & \dots & y_{0n} \end{pmatrix} \begin{pmatrix} x_{01} & y_{01} \\ \vdots & \vdots \\ x_{0n} & y_{0n} \end{pmatrix} \end{aligned}$$

$$\mathbf{A} = \frac{1}{n} \mathbf{Z}' \mathbf{I}_n \mathbf{Z}$$

L'inertie totale est la trace de la matrice \mathbf{A} (somme des éléments diagonaux $V(X)$ et $V(Y)$) : $I_T = \text{Tr}(\mathbf{A}) = V(X) + V(Y)$

Valeurs et Vecteurs propres

Rappel

La matrice des variances-covariances \mathbf{A} est, comme on le voit, symétrique réelle.

Une valeur propre de \mathbf{A} est un nombre réel λ tel qu'il existe un vecteur $\mathbf{u} \neq \mathbf{0}$, appelé vecteur propre vérifiant :

$$\mathbf{A} \mathbf{u} = \lambda \mathbf{u}$$

Méthode des multiplicateurs de LAGRANGE

Revenons au problème de maximisation de l'inertie expliquée.

On doit maximiser $u' A u$ sous la contrainte :

$$||u||_Q^2 = u' u = 1$$

En utilisant la méthode des multiplicateurs de Lagrange, on cherche u tel que :

$$\begin{cases} \frac{\partial(u' A u - \lambda(u' u - 1))}{\partial u} = 0; & \lambda \in \mathbb{R} \\ u' u = 1 \end{cases}$$

Méthode des multiplicateurs de LAGRANGE

D'où :

$$2Au - 2\lambda u = 0$$

$$Au = \lambda u$$

Ce qui montre que u est un vecteur propre de A associé à la valeur propre λ .

D'où :

$$l(D_{u^\perp}) = u' Au = u'(\lambda u) = \lambda u' u = \lambda$$

1^{er} Axe Factoriel

Donc λ doit être la plus grande des valeurs propres de A et par conséquent le vecteur cherché est le vecteur propre normé u de A associé à la plus grande valeur propre λ .

- Notons ce vecteur u_1 et la plus grande valeur propre λ_1 .
- u_1 définit l'axe D_{u_1} expliquant la plus forte inertie.
- Autrement, u_1 définit la direction du plus grand allongement du nuage.

Cet axe est appelé **premier axe factoriel** ou **1^{ère} composante principale**

2^{ème} Axe Factoriel

- Réutilisons de nouveau la méthode des multiplicateurs de LAGRANGE.
- Cherchons un deuxième axe D_{u_2} qui soit orthogonal à D_{u_1} et que $I(D_{u_1})$ soit maximale
- On maximise $u_2' A u_2$ sous les contraintes :

$$||u_2||_Q^2 = u_2' u_2 = 1$$

Et

$$u_1' u_2 = 0$$

Méthode des multiplicateurs de LAGRANGE

On cherche u_2 tel que :

$$\begin{cases} \frac{\partial(u_2' A u_2 - \lambda(u_2' u_2 - 1) - \mu(u_1' u_2 - 0))}{\partial u_2} = 0; & \lambda \in \mathbb{R} \\ u_2' u_2 = 1 \\ u_1' u_2 = 0 \end{cases}$$

D'où :

$$2A u_2 - 2\lambda u_2 - \mu u_1 = 0$$

Multiplions cette équation par u_1' :

$$2u_1' A u_2 - 2\lambda u_1' u_2 - \mu u_1' u_1 = 0$$

$$2(A u_1)' u_2 - 2\lambda u_1' u_2 - \mu u_1' u_1 = 0$$

$$2(\lambda_1 u_1)' u_2 - 2\lambda u_1' u_2 - \mu u_1' u_1 = 0$$

d'où :

$$2\lambda_1 u'_1 u_2 - 2\lambda u'_1 u_2 - \mu u'_1 u_1 = 0$$

Or, nous avons :

$$u'_1 u_2 = 0$$

Et

$$u'_1 u_1 = 1$$

D'où :

$$\mu = 0$$

De ce fait, on revient à la première équation :

$$2Au_2 - 2\lambda u_2 = 0$$

$$Au_2 = \lambda u_2$$

Ce qui montre que u_2 est un vecteur propre de A associé à la valeur propre λ_2 .

D'où :

$$I(D_{u_2}) = u_2' A u_2 = u_2' (\lambda_2 u_2) = \lambda_2 u_2' u_2 = \lambda_2$$

On prend alors λ_2 la deuxième plus grande valeur propre de A associée à u_2 vecteur propre normé et orthogonal à u_1 .

Inertie

Remarque :

λ_1 (respectivement λ_2) n'est autre que l'inertie expliquée par le **premier** (respectivement **deuxième**) axe principal.

$$\textit{Inertie} = \textit{Variance} = \textit{Information}$$

$$\begin{aligned} I_T &= I(D_{u_1}) + I(D_{u_2}) \\ &= \lambda_1 + \lambda_2 \\ &= V(X) + V(Y) \\ &= \textit{Tr}(A) \end{aligned}$$

Valeurs propres de A

Soient λ_1 et λ_2 les valeurs propres de la matrice des variances-covariances, rangées par ordre décroissant : $\lambda_1 > \lambda_2 > 0$

Pour déterminer les deux valeurs propres, on résout l'équation suivante :

$$\det(\mathbf{A} - \lambda \mathbf{I}_2) = \det \begin{pmatrix} V(X) - \lambda & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & V(Y) - \lambda \end{pmatrix} = 0$$

Après le calcul, nous trouvons les expressions des deux solutions de l'équation :

$$\lambda_1 = \frac{1}{2} \left(V(X) + V(Y) + \sqrt{[V(X) - V(Y)]^2 + 4(\text{Cov}(X, Y))^2} \right)$$

$$\lambda_2 = \frac{1}{2} \left(V(X) + V(Y) - \sqrt{[V(X) - V(Y)]^2 + 4(\text{Cov}(X, Y))^2} \right)$$

Vecteurs propres de A

Pour déterminer le vecteur propre associé à la première valeur propre, on résout le système suivant (E) :

$$\begin{cases} A.u_{1_{NN}} = \lambda_1.u_{1_{NN}} \\ \begin{pmatrix} V(X) & Cov(X, Y) \\ Cov(X, Y) & V(Y) \end{pmatrix} \begin{pmatrix} u_{1_{NN}}^1 \\ u_{1_{NN}}^2 \end{pmatrix} = \begin{pmatrix} \lambda_1.u_{1_{NN}}^1 \\ \lambda_1.u_{1_{NN}}^2 \end{pmatrix} \end{cases}$$

$$\begin{cases} V(X).u_{1_{NN}}^1 + Cov(X, Y).u_{1_{NN}}^2 - \lambda_1 u_{1_{NN}}^1 = 0 \\ Cov(X, Y).u_{1_{NN}}^1 + V(Y).u_{1_{NN}}^2 - \lambda_1 u_{1_{NN}}^2 = 0 \end{cases}$$

Vecteurs propres de A

On trouve le vecteur propre associé à λ_1 sera :

$$u_{1_{NN}} = \begin{pmatrix} V(Y) - \lambda_1 \\ -Cov(X, Y) \end{pmatrix}$$

De la même manière, on cherche le vecteur propre associée à λ_2 et on aura :

$$u_{2_{NN}} = \begin{pmatrix} V(Y) - \lambda_2 \\ -Cov(X, Y) \end{pmatrix}$$

Vecteurs propres normés de A

Les deux vecteurs propres **non-normés** associés à λ_1 et λ_2 forment une base de \mathbb{R}^2 . Donc les deux vecteurs **normés** u_1 , u_2 forment donc une base orthonormée de \mathbb{R}^2 où :

$$u_1 = \frac{u_{1NN}}{\|u_{1NN}\|} = \frac{1}{\sqrt{(V(Y) - \lambda_1)^2 + Cov(X, Y)^2}} \begin{pmatrix} V(Y) - \lambda_1 \\ -Cov(X, Y) \end{pmatrix}$$

$$u_2 = \frac{u_{2NN}}{\|u_{2NN}\|} = \frac{1}{\sqrt{(V(Y) - \lambda_2)^2 + Cov(X, Y)^2}} \begin{pmatrix} V(Y) - \lambda_2 \\ -Cov(X, Y) \end{pmatrix}$$

Matrice des vecteurs propres

Désormais, on note V la matrice des vecteurs propres :

$$\begin{array}{cc}
 u_1 & u_2 \\
 \downarrow & \downarrow \\
 V = \begin{pmatrix} \frac{V(Y) - \lambda_1}{\sqrt{(V(Y) - \lambda_1)^2 + \text{Cov}(X, Y)^2}} & \frac{V(Y) - \lambda_2}{\sqrt{(V(Y) - \lambda_2)^2 + \text{Cov}(X, Y)^2}} \\ \frac{-\text{Cov}(X, Y)}{\sqrt{(V(Y) - \lambda_1)^2 + \text{Cov}(X, Y)^2}} & \frac{-\text{Cov}(X, Y)}{\sqrt{(V(Y) - \lambda_2)^2 + \text{Cov}(X, Y)^2}} \end{pmatrix}
 \end{array}$$

Taux d'Inertie

Le taux d'inertie totale expliquée par **le premier axe factoriel** est le rapport :

$$\tau_1 = \frac{I(D_{u_1})}{I_T} = \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{\lambda_1}{V(X) + V(Y)}$$

Le taux d'inertie totale expliquée par **le deuxième axe factoriel** est le rapport :

$$\tau_2 = \frac{I(D_{u_2})}{I_T} = \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{\lambda_2}{V(X) + V(Y)}$$

Coordonnées factorielles

On dispose maintenant, dans \mathbb{R}^2 , de deux bases : l'une est la base orthonormée Canonique $\{e_1, e_2\}$ et l'autre est la base propre orthonormée $\{u_1, u_2\}$.

Le vecteur $\overrightarrow{GM_i}$ a pour coordonnées, dans $\{e_1, e_2\}$, (x_0, y_0) et dans la base $\{u_1, u_2\}$, $(\langle \overrightarrow{GM_i} | u_1 \rangle, \langle \overrightarrow{GM_i} | u_2 \rangle)$. Avec :

$$\begin{aligned}(\langle \overrightarrow{GM_i} | u_1 \rangle, \langle \overrightarrow{GM_i} | u_2 \rangle) &= ((x_{0i} \ y_{0i})u_1, (x_{0i} \ y_{0i})u_2) \\ &= (x_{0i} \ y_{0i})V\end{aligned}$$

Coordonnées factorielles

D'où les n coordonnées du nuage centré sur G , dans la base $\{u_1, u_2\}$:

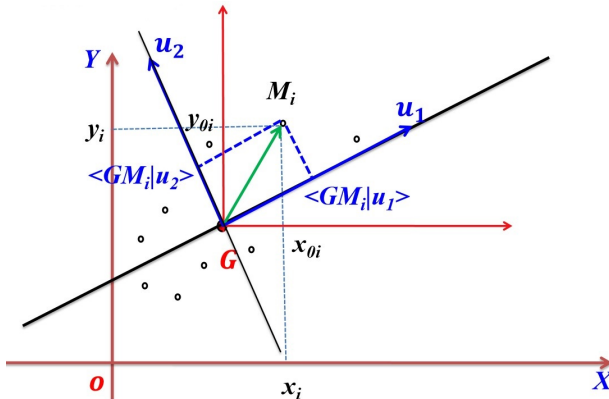
$$F = \begin{pmatrix} \langle \overrightarrow{GM_1} | u_1 \rangle & \langle \overrightarrow{GM_1} | u_2 \rangle \\ \vdots & \vdots \\ \langle \overrightarrow{GM_n} | u_1 \rangle & \langle \overrightarrow{GM_n} | u_2 \rangle \end{pmatrix} = \begin{pmatrix} x_{01} & y_{01} \\ \vdots & \vdots \\ x_{0n} & y_{0n} \end{pmatrix} V$$

$$F = Z V$$

F est donc la matrice, à n lignes et 2 colonnes, dont les lignes sont **les coordonnées factorielles** du nuage de points dans \mathbb{R}^2 muni de la base $\{u_1, u_2\}$.

Coordonnées factorielles

Les deux colonnes de la matrice F constituent les deux composantes principales de la variable statistique (X_0, Y_0) .



La première colonne de la matrice $F = ZV$ est donc le vecteur :

$$F^{(1)} = Z u_1$$

De même, la deuxième colonne de la matrice $F = ZV$ est le vecteur :

$$F^{(2)} = Z u_2$$

Remarque :

$F^{(1)}$ et $F^{(2)}$ peut s'écrire aussi sous la forme suivante :

$$\begin{cases} F^{(1)} &= u_{1_1} X_0 + u_{1_2} Y_0 \\ F^{(2)} &= u_{2_1} X_0 + u_{2_2} Y_0 \end{cases}$$

Propriétés des composantes principales

Propriété 1

Les composantes principales sont centrées.

$$\overline{F^{(1)}} = \langle F^{(1)} | 1_n \rangle_P = 0; \overline{F^{(2)}} = \langle F^{(2)} | 1_n \rangle_P = 0$$

Propriété 2

La variance d'une composante principale est la valeur propre correspondante.

$$V(F^{(1)}) = \|F^{(1)}\|_P^2 = \langle F^{(1)} | F^{(1)} \rangle_P = \lambda_1$$

$$V(F^{(2)}) = \|F^{(2)}\|_P^2 = \langle F^{(2)} | F^{(2)} \rangle_P = \lambda_2$$

Propriétés des composantes principales

Propriété 3

Les composantes principales sont non corrélées.

$$\begin{aligned} \text{Cov}(F^{(1)}, F^{(2)}) &= \langle F^{(1)} | F^{(2)} \rangle_P \\ &= \frac{1}{n} u_1' Z' Z u_2 \\ &= u_1' A u_2 \\ &= \lambda_2 u_1' u_2 \\ &= 0 \end{aligned}$$

3 Méthodes Géométriques en Statistique

- Moyenne
- Variance
- Covariance et Corrélation
- Corrélation

4 Analyse en Composantes Principales

- Définition des Espaces
- Régression Orthogonale
- Notion d'Inertie
- Valeurs et Vecteurs propres
- Coordonnées Factorielles

5 Exemple d'Application

- Méthode Analytique
- Résolution par Logiciel

Exemple d'application

Pour mieux comprendre la méthode d'ACP à deux variables, on présente l'exemple suivant :

Application :

On dispose de Deux systèmes d'exploitation (WIN et MAC) qui sont évalués par cinq individus (A, B, C, D, E) par une note de 0 à 20 exprimant leur appréciation.

INDIVIDUS	WIN	MAC
A	10	2
B	20	4
C	16	6
D	15	8
E	5	18

Étape 1 :

1- La présentation des données sous forme de matrice :

$$M = \begin{pmatrix} 10 & 2 \\ 20 & 4 \\ 16 & 6 \\ 15 & 8 \\ 5 & 18 \end{pmatrix}$$

Les éléments de cette ACP :

$$n = 5 \quad p = 2 \quad P = \frac{1}{5} I_5$$

n : Nombre d'observations

p : Nombre de variables

P : Matrice des poids (espace des individus)

Étape 2 :

2- Centrage des données

La moyenne des variables **WIN** et **MAC** :

$$\overline{WIN} = \frac{1}{n} \sum WIN_i = \frac{1}{5}(10 + 20 + 16 + 15 + 5) = 13.2$$

$$\overline{MAC} = \frac{1}{n} \sum MAC_i = \frac{1}{5}(2 + 4 + 6 + 8 + 18) = 7.6$$

La matrice centrée est donc :

$$Z = \begin{pmatrix} 10 - 13.2 & 2 - 7.6 \\ 20 - 13.2 & 4 - 7.6 \\ 16 - 13.2 & 6 - 7.6 \\ 15 - 13.2 & 8 - 7.6 \\ 5 - 13.2 & 18 - 7.6 \end{pmatrix} = \begin{pmatrix} -3.2 & -5.6 \\ 6.8 & -3.6 \\ 2.8 & -1.6 \\ 1.8 & 0.4 \\ -8.2 & 10.4 \end{pmatrix}$$

Étape 3 :

3- Calcul de la matrice de Var-Cov

$$A = Z'PZ = \frac{1}{n}Z'I_5Z = \frac{1}{5}Z'Z$$

$$A = \frac{1}{5} \begin{pmatrix} -3.2 & 6.8 & 2.8 & 1.8 & -8.2 \\ -5.6 & -3.6 & -1.6 & 0.4 & 10.4 \end{pmatrix} \begin{pmatrix} -3.2 & -5.6 \\ 6.8 & -3.6 \\ 2.8 & -1.6 \\ 1.8 & 0.4 \\ -8.2 & 10.4 \end{pmatrix}$$

$$A = \begin{pmatrix} 26.96 & -19.12 \\ -19.12 & 31.04 \end{pmatrix}$$

Étape 4 :

4- Calcul des valeurs propres de A

Nous avons $p = 2$, ce qui implique que nous aurons deux valeurs propres de A (soit λ_1 et λ_2).

$$\lambda_1 = \frac{1}{2} \left(V(X) + V(Y) + \sqrt{[V(X) - V(Y)]^2 + 4(\text{Cov}(X, Y))^2} \right)$$

$$\lambda_1 = \frac{\text{Tr}(A) + \sqrt{\text{Tr}(A)^2 - 4\det(A)}}{2}$$

$$\lambda_1 = \frac{58 + \sqrt{58^2 - 4 \times 471.264}}{2}$$

$$\lambda_1 = 48.228$$

Nous avons :

$$Tr(A) = V(WIN) + V(MAC) = \lambda_1 + \lambda_2$$

d'où :

$$\lambda_2 = Tr(A) - \lambda_1$$

$$\lambda_2 = 58 - 48.228$$

$$\lambda_2 = 9.772$$

Étape 5 :

5- Calcul des vecteurs propres de **A**

Le vecteur **non-normé** $u_{1_{NN}}$:

$$u_{1_{NN}} = \begin{pmatrix} V(\text{MAC}) - \lambda_1 \\ -\text{Cov}(\text{WIN}, \text{MAC}) \end{pmatrix} = \begin{pmatrix} 31.04 - 48.228 \\ 19.12 \end{pmatrix} = \begin{pmatrix} -17.188 \\ 19.12 \end{pmatrix}$$

La norme de $u_{1_{NN}}$:

$$\|u_{1_{NN}}\| = \sqrt{(-17.188)^2 + (19.12)^2} = 25.71$$

Le vecteur **normé** u_1 :

$$u_1 = \frac{u_{1_{NN}}}{\|u_{1_{NN}}\|} = \begin{pmatrix} \frac{-17.188}{25.71} \\ \frac{19.12}{25.71} \end{pmatrix} = \begin{pmatrix} -0.668 \\ 0.744 \end{pmatrix}$$

Le vecteur **non-normé** $u_{2_{NN}}$:

$$u_{2_{NN}} = \begin{pmatrix} V(\text{MAC}) - \lambda_2 \\ -\text{Cov}(\text{WIN}, \text{MAC}) \end{pmatrix} = \begin{pmatrix} 31.04 - 9.772 \\ 19.12 \end{pmatrix} = \begin{pmatrix} 21.268 \\ 19.12 \end{pmatrix}$$

La norme de $u_{2_{NN}}$:

$$\|u_{2_{NN}}\| = \sqrt{(21.268)^2 + (19.12)^2} = 28.6$$

Le vecteur **normé** u_2 :

$$u_2 = \frac{u_{2_{NN}}}{\|u_{2_{NN}}\|} = \begin{pmatrix} \frac{21.268}{28.6} \\ \frac{19.12}{28.6} \end{pmatrix} = \begin{pmatrix} 0.744 \\ 0.668 \end{pmatrix}$$

Vérification qu'ils sont orthonormés :

$$u_1' u_2 = (-0.668 \quad 0.744) \begin{pmatrix} 0.744 \\ 0.668 \end{pmatrix}$$

$$u_1' u_2 = (-0.668 \times 0.744) + (0.744 \times 0.668)$$

$$u_1' u_2 = 0$$

Étape 6 :

6- Calcul de l'inertie totale et des taux d'inertie

L'inertie totale I_T :

$$I_T = \lambda_1 + \lambda_2 = V(\text{WIN}) + V(\text{MAC}) = \text{Tr}(A) = 58$$

Le taux d'inertie du **premier axe** :

$$\tau_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{48.228}{58} = \mathbf{83.15\%}$$

Le taux d'inertie du **deuxième axe** :

$$\tau_2 = \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{9.772}{58} = \mathbf{16.85\%}$$

Étape 7 :

7- La matrice des vecteurs propres V

$$\begin{array}{cc} u_1 & u_2 \\ \downarrow & \downarrow \end{array}$$

$$V = \begin{pmatrix} -0.668 & 0.744 \\ 0.744 & 0.668 \end{pmatrix}$$

Étape 8 :

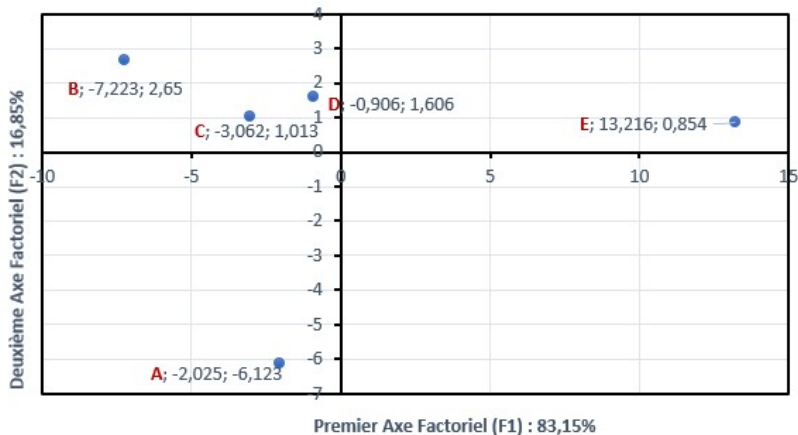
8- Les coordonnées des individus dans le plan factoriel F

$$F = ZV$$

$$F = \begin{pmatrix} -3.2 & -5.6 \\ 6.8 & -3.6 \\ 2.8 & -1.6 \\ 1.8 & 0.4 \\ -8.2 & 10.4 \end{pmatrix} \begin{pmatrix} -0.668 & 0.744 \\ 0.744 & 0.668 \end{pmatrix}$$

$$F = \begin{pmatrix} -2.025 & -6.123 \\ -7.223 & 2.650 \\ -3.062 & 1.013 \\ -0.906 & 1.606 \\ 13.216 & 0.854 \end{pmatrix}$$

Individus (axes F1 et F2 : 100,00 %)



Variables (axes F1 et F2 : 100,00 %)

