*Analyse des données*

*Chapitre 2*

*Analyse en Composantes Principales*

## 6. Analyse en composantes principales

***Définition 1***

A chaque axe principal (factoriel) ($\Delta u_j$) est associé une **variable** $C^j$ appelée ***composante principale***. Sa dimension est égale à $m$ (nombre d'individus).

Pour tout $j$,

$$(\Delta u_j) \longrightarrow C^j$$

***Définition 2***

Une ***composante principale*** n'est que ***le vecteur constitué des projections des individus*** sur ***l'axe principal*** associé à cette nouvelle variable.

Par exemple,

$$C^1 = (x_1, x_2, \ldots, x_i, \ldots, \ldots, x_m)t.$$

Où,

$x_i$ **Represente la projection de l'individu** $X_i$ sur le premier axe ($\Delta u_1$).

Pour tout : $i = 1, 2, \ldots, m$.

### 6.1 Calcul explicite des composantes principales

La détermination explicite des composantes principales se fait moyennant le produit scalaire.

$$C^1 = u_{11}X^1 + u_{21}X^2 + \cdots + u_{n1}X^n.$$

C'est-à-dire :

$$C^1 = \sum_{j=1}^{n} u_{j1}X^j = Xu_1.$$

De même,

$$C^2 = u_{12}X^1 + u_{22}X^2 + \cdots + u_{n2}X^n.$$

C'est-à-dire :

$$C^2 = \sum_{j=1}^{n} u_{j2}X^j = Xu_2.$$

**Remarque**

**Dans le cas d'une *ACP* centrée réduite c'est-à-dire normée,** $C^i = Zu_i$.

*Analyse des données*

> ***Conclusion***
> **Les composantes *principales sont des combinaisons linéaires des variables initiales.***

### 6.2 Propriétés des composantes principales

Les composantes principales vérifient les propriétés suivantes :

1) La variance d'une composante principale n'est que : $\boldsymbol{Var(C^i) = \lambda_i}$, $i = 1,2,\dots,q$.
Où, $q$ représente la dimension du sous espace ajustant le nuage de points.
2) Les composantes principales sont ***non corrélées deux à deux*** c'est-à-dire pour tout $i,j = 1,2,\dots,q$, $\mathbf{Cov(C^i, C^j) = 0.}$

### 6.3 Représentation des individus

N'oublions pas que notre but est de passer de l'espace $\Re^n$ vers l'espace $\Re^q$ de dimension faible par à rapport à celle d'origine $(q <<< n)$ afin que nous puissions visualiser les données (ici, il s'agit des individus).

Par conséquent représenter les individus revient à les représenter dans l'espace $\Re^q$.

***Exemple***

Pour $q = 2$, il faut représenter les individus dans l'espace $(C^1, C^2)$ avec $C^1, C^2$ sont les ***deux nouvelles variables*** (***composantes principales***).

Pour tout $j = 1, 2$, $C^j$ est constituée des $\boldsymbol{m}$ coordonnées des individus sur la j^ème^ composante c'est-à-dire le j^ème^ axe principal.

Posons :

$$C^j = (c_1^j, c_2^j, \dots, c_m^j)^t \text{ pour tout } j = 1, 2.$$

Alors :

Chaque individu $\boldsymbol{X_i}$ aura $(\boldsymbol{c_i^1, c_i^2})$ comme coordonnées dans le nouveau plan. Voir l'illustration graphique ci-dessous (figure 5).

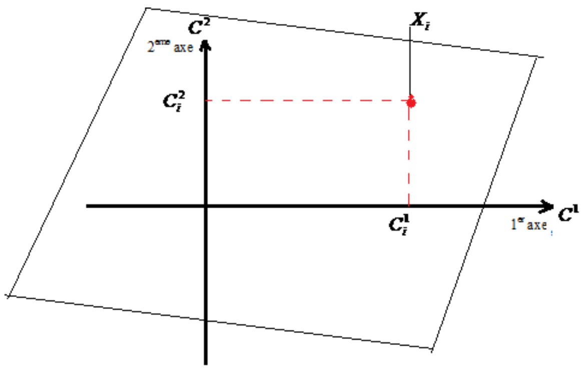



*Figure 5. Projection d'un point sur un plan.*

D'après le paragraphe précédent,

$$C^1 = X \cdot u_1,\ C^2 = X \cdot u_2$$

Avec :

$$c_i^1 = X_i \cdot u_1 \text{ et } c_i^2 = X_i \cdot u_2 \text{ pour tout } i = 1,2,\ldots,m.$$

Mais dans le cas d'une **ACP normée**, bien-sûr nous prenons :

$$C^1 = Z \cdot u_1 \text{ et } C^2 = Z \cdot u_2.$$

**Remarque**

***D'une manière générale le choix du nouveau sous-espace c'est-à-dire le choix de $q$ dépend aussi du problème à résoudre ou à modéliser et la précision à exiger.***

## 7. Nombres d'axes

Le choix du nombre de composantes suffisamment représentatives est lié à la notion de l'inertie exprimée par le sous espace principal.

***Définition 3***

Nous appelons ***pourcentage d'inertie reproduite*** ou ***expliquée*** par le sous espace de $dim = q$, le rapport :

$$\sum_{i=1}^{i=q} \lambda_i \Big/ \sum_{i=1}^{n} \lambda_i.$$

**Cas particulier**

Le rapport défini ci-dessus nous permet donc de mesurer la ***part d'inertie expliquée*** par ***chaque axe***. Pour tout $\boldsymbol{i} = 1,2,\ldots,q$ :

$$\frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}.$$

Sachant que par définition l'**inertie totale** ($I_g$) est la ***somme des variances*** c'est-à-dire :

$$I_g = \sum_{i=1}^{n} var(X^i) = Trace(V).$$

Alors, dans le cas où les ***données sont centrées-réduites***, ***l'inertie totale*** est égale à $\boldsymbol{n}$ car la variance de chaque variable transformée est égale à 1.

Donc

$$\boldsymbol{I_g = n = Trace(R)}.$$

Par conséquent, les rapports précédents deviennent :

$$\sum_{i=1}^{q} \lambda_i/n, \quad \text{et} \quad \lambda_i/n \quad \text{pour tout } i = 1, 2, \ldots q.$$

Par suite, les critères de base utilisés pour le choix du nombre d'axes sont définis par :

# *Analyse des données*

## 7.1 Critères globaux

Parmi ces critères, nous citons :

- ***Taux d'inertie*** souhaité à ***priori*** par l'utilisateur, en général **80** %**.**
- Retenir que les axes principaux associés aux valeurs propres ***supérieures à 1***. C'est le critère de ***Kaiser.***
- ***Histogramme*** des valeurs propres ou bien critère du ***coude*** de ***Cattell***,

Dans le graphe de cette figure, nous choisissons les valeurs propres : $\lambda_3, \lambda_5, \lambda_2$.

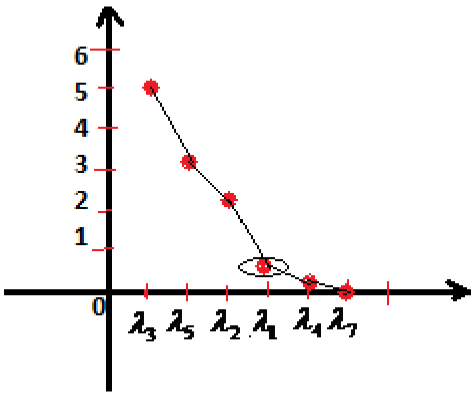


Figure 6. Illustration du critère de Cattell.

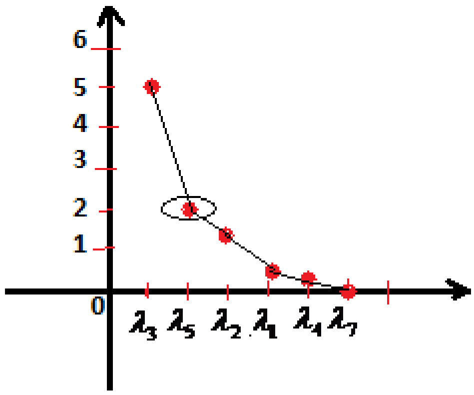


Figure 7. Illustration du critère de Cattell.

## 7.2 Critères de qualité

D'autres critères sont présentés dans la littérature, ils sont basés sur la qualité de la représentation des individus.

- ***Qualité de la représentation d'individu.*** La qualité de la représentation d'un individu $X_i$ quelconque sur le nouveau sous espace de $dim = q$ est donnée par :

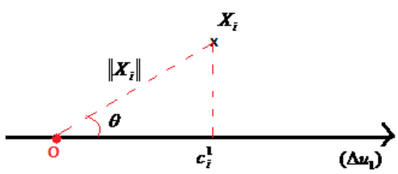
$$\boldsymbol{\varphi_i^q = \sum_{j=1}^{q} (c_i^j)^2 \Big/ \|X_i\|^2.}$$

Ce critère peut être interprété géométriquement. En effet,

Dans le cas où $q = 1$, la projection d'un individu $X_i$ quelconque sur le sous espace de $dim = 1$, $(\Delta u_1)$ a comme coordonnée $\boldsymbol{c_i^1}$.

Donc, si on note l'angle inscrit entre le vecteur $(\overrightarrow{OX_i})$ et l'axe $(\Delta u_1)$ par $\boldsymbol{\theta}$**,** nous avons :

$$(cos(\theta))^2 = \frac{(c_i^1)^2}{\|X_i\|^2}.$$




C'est-à-dire :

$$\boldsymbol{\varphi_1^1 = (cos(\theta))^2.}$$

*Analyse des données*

**Conséquences**

- D'après l'exemple explicatif, nous en déduisons que plus que $(\boldsymbol{cos}(\boldsymbol{\theta}))^{\mathbf{2}}$ est proche de 1 meilleure est la qualité de la représentation de l'individu $X_i$ sur l'axe $(\Delta u_1)$.
- D'après l'exemple explicatif, nous en déduisons que $\boldsymbol{\varphi_i^q}$ n'est que **la somme des carrés des cosinus des angles** inscrits entre $(\overrightarrow{OX_i})$ et ses projections sur les axes de l'espace de projection de dimension $q$.
- D'une manière générale, cette qualité de représentation $\boldsymbol{\varphi_i^q}$ est dite la **contribution relative** de l'individu $X_i$ sur le nouveau sous espace de projection de $dim = q$.
- Ce critère reflète aussi l'importance relative d'un axe dans la variabilité des données.

- ***Contribution de l'individu à l'inertie d'un axe.*** Ce critère noté $\boldsymbol{ctr(i,j)}$ ou $\boldsymbol{\gamma_i^j}$ ou encore $\boldsymbol{ctr_j(i)}$ nous permet de calculer la contribution de chaque individu $X_i$, $i = 1, 2, \dots, m$, à **l'inertie** du $\boldsymbol{j}^{\text{ème}}$ **axe** $(\Delta u_j)$ pour tout $j = 1, 2, \dots, q$. Elle est donnée par le rapport suivant :

$$\boldsymbol{\gamma_i^j = \frac{1}{m}(c_i^j)^2 / \lambda_j = \frac{\frac{1}{m}(c_i^j)^2}{\lambda_j}}.$$

- En effet, pour tout $j = 1, 2, \dots, q$, nous avons :

$$\boldsymbol{var(C^j) = \lambda_j}.$$

Alors :

$$var(C^j) = \lambda_j \Leftrightarrow \frac{1}{m}\sum_{i=1}^{m}(c_i^j)^2 = \lambda_j.$$

C'est-à-dire :

$$\frac{1}{m}(c_1^j)^2 + \frac{1}{m}(c_2^j)^2 + \cdots + \frac{1}{m}(c_i^j)^2 + \cdots + \frac{1}{m}(c_m^j)^2 = \lambda_j.$$

Ce qui est équivalent à :

$$\frac{1}{m\lambda_j}(c_1^j)^2 + \frac{1}{m\lambda_j}(c_2^j)^2 + \cdots + \frac{1}{m\lambda_j}(c_i^j)^2 + \cdots + \frac{1}{m\lambda_j}(c_m^j)^2 = 1.$$

Ou encore :

$$\boldsymbol{\sum_{i=1}^{m}\frac{\frac{1}{m}(c_i^j)^2}{\lambda_j} = 1}.$$

D'autre part, comme pour tout $j = 1, 2, \dots, q$, nous avons :

$$\boldsymbol{I\big((\Delta u_j)\big) = \lambda_j}.$$

Alors en termes de pourcentage, l'inertie du $j^{ème}$ axe vérifie :

$$\boldsymbol{\sum_{i=1}^{m} \frac{\frac{1}{m}(c_i^j)^2}{\lambda_j} = 100\ \%.}$$

Ce qui nous conduit à conclure que pour tout $i = 1, 2, \dots, m$, le facteur :

$$\boldsymbol{\frac{\frac{1}{m}(c_i^j)^2}{\lambda_j}}.$$

N'est que ***la contribution du $i^{ème}$ individu à l'inertie** du **$j^{ème}$ axe.***

**Remarque**

Cette contribution est dite ***contribution absolue*** de l'individu $X_i$ à l'inertie du $j^{ème}$ axe.

**Conséquences**

- Ce critère nous permet d'étudier la contribution apportée par les individus à chaque axe.
- Il nous permet de voir quels sont les individus ayant une forte contribution à la construction des axes.
- Les individus ayant une forte contribution peuvent créer une instabilité.

**Exemple**

**Remarque**

La même procédure suivie pour analyser le nuage des individus sera appliquée sur le nuage des variables.

## 8. Analyse duale

Le but est de construire le meilleur sous espace factoriel du nuage de points mais cette fois ci pour le nuage de points variables.

Donc, il s'agit de définir $q$ nouveaux individus comme axes du repère du nuage de points-variables.

Le principe étant le même, alors l'étude sera déduite de la première tout en considérant la matrice $\boldsymbol{X \cdot X^t}$.

D'autre part, comme les deux nuages individus et variables sont complémentaires et duales. Donc nous pouvons passer d'un nuage à un autre via des formules dites ***formules de transition***.

*Analyse des données*

## *Relations de Transition*

L'espace des individus : $N(I)$ L'espace des variables : $N(J)$

$q$ : Variables (C.P) $q$ : Individus (C.P)

Donc, l'ajustement du nuage des variables par un sous espace de $dim = q$ est obtenu par les $q$ vecteurs propres normés associés aux $q$ plus grandes valeurs propres de la matrice $\boldsymbol{X \cdot X^t}$.

***Proposition***

Les valeurs propres non nulles de la matrice $\boldsymbol{X^t . X}$ **coïncident** avec celles de $\boldsymbol{X \cdot X^t}$ et les **autres** valeurs propres sont **nulles**.

### 8.1 Relations de transition

Le passage du nuage des individus où chaque individu est un point de $\mathfrak{R}^n$ vers le nuage des variables où chaque variable est un point de $\mathfrak{R}^m$ se fait en utilisant les relations suivantes.

***Soient***

$\boldsymbol{u_k}$ un vecteur propre normé de la matrice $\boldsymbol{X^t . X}$ associé à la valeur propre $\boldsymbol{\lambda_k}$, et $\boldsymbol{C^k}$ la composante principale associée à $u_k$,

$\boldsymbol{v_k}$ un vecteur propre normé de la matrice $\boldsymbol{X \cdot X^t}$ associé à la valeur propre $\boldsymbol{\lambda_k}$ et $\boldsymbol{\phi^k}$ la composante principale associée à $v_k$.

C'est-à-dire :

$\boldsymbol{\phi^k}$ représente les ***coordonnées des projections des variables*** sur l'axe $(\boldsymbol{\Delta v_k})$.

Alors, nous avons :

$$\begin{cases} \boldsymbol{v_k} = \dfrac{\boldsymbol{1}}{\sqrt{\boldsymbol{\lambda_k}}} \boldsymbol{X \cdot u_k} \\ \boldsymbol{u_k} = \dfrac{\boldsymbol{1}}{\sqrt{\boldsymbol{\lambda_k}}} \boldsymbol{X^t \cdot v_k} \end{cases}.$$

Comme $\boldsymbol{X \cdot u_k = C^k}$ et $\boldsymbol{X^t \cdot v_k = \phi^k}$, les relations précédentes deviennent :

$$\begin{cases} \boldsymbol{v_k} = \dfrac{\boldsymbol{1}}{\sqrt{\boldsymbol{\lambda_k}}} \boldsymbol{C^k} \\ \boldsymbol{u_k} = \dfrac{\boldsymbol{1}}{\sqrt{\boldsymbol{\lambda_k}}} \boldsymbol{\phi^k} \end{cases}.$$

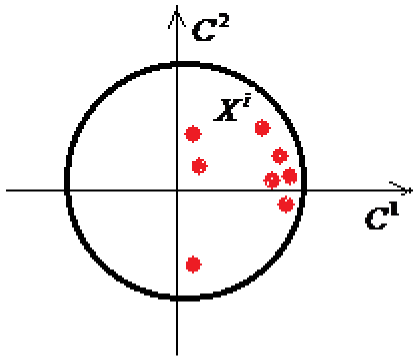
Ce qui donnent pour tout $k = 1, 2, \dots, q$ et $j = 1, 2, \dots, n$ :

$$\boldsymbol{\phi^k = \sqrt{\lambda_k} \cdot u_k} \qquad \text{c'est-à-dire} \qquad \boldsymbol{\phi_j^k = \sqrt{\lambda_k} \cdot u_{jk}}.$$

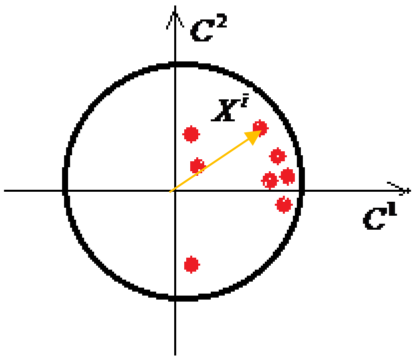
**Conséquences**

- Ces relations nous permettent le passage des coordonnées des projections des individus vers les coordonnées des projections des variables.

- Dans le cas de ***données centrées réduites***, ces relations nous permettent de visualiser les variables initiales dans le nouveau sous espace composé de nouvelles variables.




- D'étudier les corrélations linéaires entre les variables initiales et les axes d'inertie (les nouvelles variables).




### 8.2 Représentation des variables

Les proximités entre les composantes principales et les variables initiales se mesurent par les covariances et surtout les corrélations.
Donc pour des données centrées-réduites,

$$\boldsymbol{\phi_j^k} \text{ n'est que } \boldsymbol{r(X^j, C^k)}.$$

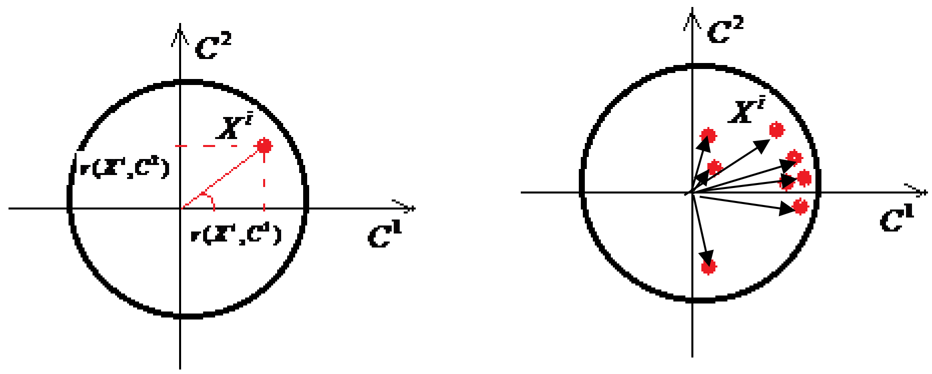
C'est-à-dire :

*Analyse des données*

$r(X^j, C^k)$ est le ***coefficient de corrélation linéaire entre la variable initiale*** $X^j$ ***et l'axe d'inertie*** $C^k$.

Comme la corrélation est dans $[-1, 1]$ , elle represente le cosinus d'angles inscrits entre les variables normées, alors nous pouvons représenter les variables dans un cercle unitaire (cercle de rayon un) dit ***cercle des corrélations.***

Ce cercle n'est que la projection du nuage des variables sur ***le plan*** *des* ***composantes principales*** c'est-à-dire dans le cas où le **sous espace de projection** est de **dimension 2.**




Représentation des variables dans le nouveau plan.