

Analyses de données

École des mines de Saint-Étienne

Thomas Galtier :
thomas.galtier@emse.fr, Bureau 532

Liseth Pasaguayo :
liseth.pasaguayo@emse.fr, Bureau 529

Département GMI - Institut Fayol

2025

Dans ce cours : analyses factorielles

Méthodes abordées dans ce cours :

- ① ACP : Analyse en composantes principales
(anglais : PCA — *Principal Component Analysis*)
- ② AFC : Analyse factorielle des correspondances

6h ACP ; 6h AFC

- ① 1 petit TD/TP 6
- ② 1 gros TP 6

- 1 Statistiques descriptives — à quoi servent les analyses factorielles ?
- 2 Introduction sur l'ACP
- 3 Identifier les composantes principales
- 4 Réduction de dimension par ACP

- 1 Statistiques descriptives — à quoi servent les analyses factorielles ?
- 2 Introduction sur l'ACP
- 3 Identifier les composantes principales
- 4 Réduction de dimension par ACP

Qu'est-ce que les statistiques ?

On dispose d'une **grande quantité d'information**.

→ Généralement via **un jeu de données**

Données quantitatives ou qualitatives (catégorielles / ordinales)

	x_1	x_2	x_3	x_4
1	12.1	0.1	A	2
2	7.2	0.3	A	2
3	8.3	0.2	C	1
4	9.3	0.4	B	3
\vdots	\vdots	\vdots	\vdots	\vdots

But :

- **Résumer l'information** via des indicateurs
- Évaluer la **fiabilité des indicateurs**

Qu'est-ce que les statistiques ?

On dispose d'une **grande quantité d'information**.

→ Généralement via **un jeu de données**

Données quantitatives ou qualitatives (catégorielles / ordinales)

	x_1	x_2	x_3	x_4
1	12.1	0.1	A	2
2	7.2	0.3	A	2
3	8.3	0.2	C	1
4	9.3	0.4	B	3
\vdots	\vdots	\vdots	\vdots	\vdots

But :

- **Résumer l'information** via des indicateurs → statistiques descriptives
- Évaluer la **fiabilité des indicateurs** → statistiques inférentielles

À quoi servent les statistiques ?


Décrire des comportements collectifs
 \neq comportements individuels.

Les indicateurs statistiques cherchent souvent à identifier :

- Les grandes tendances
- Les liens entre variables

Deux types de liens possibles :

- liens de corrélation ou concomitances (constater)
- liens de causalité (expliquer)

Attention : corrélation ou concomitances \neq causalité (cf. ce site )

Une corrélation ne fait qu'indiquer un possible lien de causalité.

Comprendre les liens entre variables permet de
prédire / classifier / \simeq expliquer / constater

Statistiques descriptives : un préalable à toute analyse

Savoir quoi analyser, comment, et avec quels outils

Avant d'utiliser des outils d'apprentissage ou statistiques :

On s'informe sur la structure des données !

Via des stat descriptives :)

- ① Comment les données ont été collectées ?
Données manquantes, données aberrantes ? risque de biais ?
- ② Nombre d'observations, effectifs par classes croisées ?
Suffisants pour la ou les méthodes envisagées ?
- ③ Avoir un 1er aperçu des données.
 - Les individus groupés ou dispersés
 - Les variables mesurent des informations similaires ou complémentaires
 - Si peu de variables : histogrammes, graphes, tableaux croisés, etc.
 - Si beaucoup de variables → analyses factorielles **ACP, AFC**

Statistiques descriptives : un préalable à toute analyse

Savoir quoi analyser, comment, et avec quels outils

Avant d'utiliser des outils d'apprentissage ou statistiques :

On s'informe sur la structure des données !!!

Via des **stat descriptives**

- ① Comment les données ont été collectées ?
Données manquantes, données aberrantes ? risque de biais ?
- ② Nombre d'observations, effectifs par classes croisées ?
Suffisants pour la ou les méthodes envisagées ?
- ③ Avoir un 1er aperçu des données.
 - Les individus groupés ou dispersés
 - Les variables mesurent des informations similaires ou complémentaires
 - Si peu de variables : histogrammes, graphes, tableaux croisés, etc.
 - Si beaucoup de variables → analyses factorielles **ACP, AFC**

→ Vérifier que les hypothèses des modèles envisagés soient respectées.

→ Choisir le ou les modèles adaptés au jeu de données.

Statistiques descriptives

Moins d'outils quand le nombre de variables augmente :(

- Univariées :
 - Quali : proportion dans chaque classe
 - quantitatives : moyenne, variance, quantiles, histogramme
- Bivariées :
 - Quali + Quali : tableau de contingence
 - quantitatives + Quali : boxplot
 - quantitatives + quantitatives : corrélation, graphe / histogramme 2D
- Multivariées : Représentations bien plus difficiles...
 - Peu de variables : analyses des corrélations par tableaux croisés
 - Beaucoup de variables : **analyses factorielles :**
 - *analyse groupée des corrélations, des correspondances, et possibilité de réduire la dimension pour y voir plus clair*

Rappels : covariance (empirique) et tableaux croisés

Deux variables : x_1, x_2

n observations : $x_1 = (x_{11}, \dots, x_{n1}), x_2 = (x_{12}, \dots, x_{n2})$

Moyenne :

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

Covariance :

$$\text{Cov}(x_1, x_2) = \frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)$$

$$\text{Cov}(x_1, x_1) = \text{Var}(x_1)$$

Corrélation :

$$\text{Corr}(x_1, x_2) = \frac{\text{Cov}(x_1, x_2)}{\sqrt{\text{Var}(x_1)}\sqrt{\text{Var}(x_2)}} \in [-1; 1]$$

Rappels : Corrélation

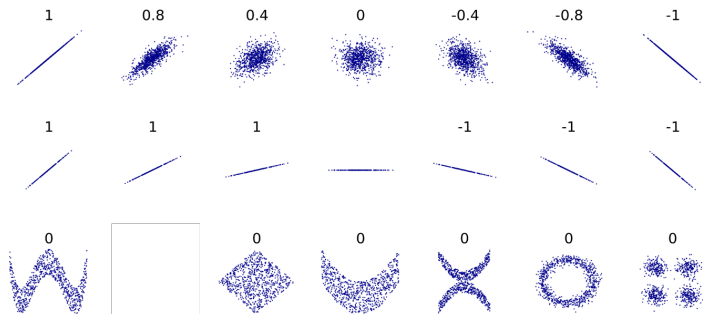


Figure – Exemple de nuage de points et la corrélation associée

→ La corrélation indique le caractère linéaire de la relation entre 2 variables

Rappels : Corrélation

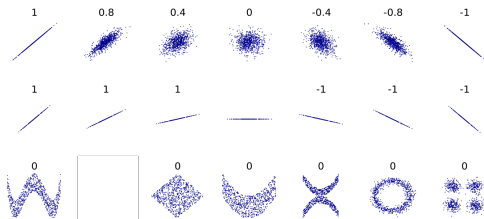


Figure – Exemple de nuage de points et la corrélation associée

- La corrélation n'indique pas :
 - Les "courbures" dans la relation entre les variables
 - Les symétries verticale ou horizontale (passant par la moyenne)
 - L'indépendance des variables
 - Les causalités

Covariance = corrélation \times écart-type des 2 variables

- La covariance indique :
 - Le lien linéaire entre deux variables
 - L'ampleur des variations
- Comme la corrélation, la covariance n'indique pas :
 - les "courbures" de la relation entre deux variables
 - Les symétries verticale ou horizontale (passant par la moyenne)
 - L'indépendance des variables
 - Les causalités

Rappels : Tableaux croisés (pairplots en anglais)

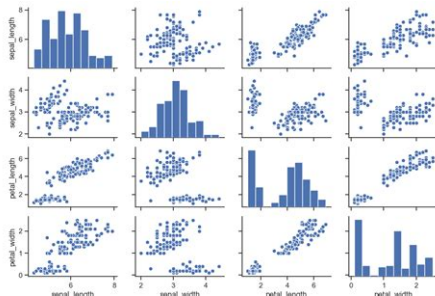


Figure – Exemple de tableau croisé

- Analyse des relations entre variables (deux à deux)
- Offre une visualisation plus complète
 - Va au delà du caractère "linéaire"
 - Permet d'analyser les dépendances
 - Les anomalies
 - Les variables redondantes

Rappels : Tableaux croisés (pairplots en anglais)

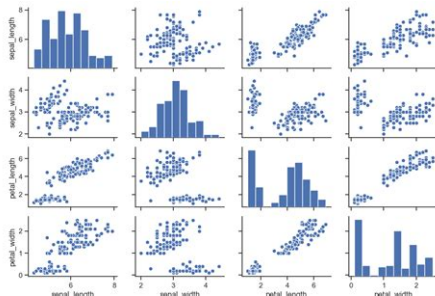


Figure – Exemple de tableau croisé

- Ici on voit que les deux dernières variables sont très similaires. (cf graphe croisé proche de la diagonale, et histogrammes similaires)
→ Est-ce utile de conserver cette redondance dans le jeu de données ?
- Elles divisent les données en deux groupes. → On envisage de séparer ces groupes dans l'analyse (améliore précisions/fiabilité). → Cela pourrait aussi s'expliquer par des données non-récoltées (attention au biais du survivant).

- 1 Statistiques descriptives — à quoi servent les analyses factorielles ?
- 2 Introduction sur l'ACP
- 3 Identifier les composantes principales
- 4 Réduction de dimension par ACP

ACP : Analyse en composantes principales

Identifier et d'analyser l'information utile d'un jeu de données

- Séparer les différents types d'informations contenues dans le jeu de données
- Réduire la dimension tout en conservant le plus d'informations utiles

Remarque :

- pas d'explication des comportements (pas de lien de causalité identifié)
- pas de prédiction
- ⇒ pas de variable Y à expliquer ou prédire comme en régression

Quelle information est analysée dans l'ACP ?

En statistique/probabilité il y a plusieurs façon de mesurer l'information.

Dans l'ACP :
information = variabilité

- Une variable contient de l'information si elle varie beaucoup.
Par opposition à une variable constante qui ne donne pas d'information.
- 2 variables contiennent la même information, si elles varient de la même façon. (lien linéaire)

parti-pris de l'ACP

≠ info = dépendances : parti pris de la méthode ICA

≠ info considère des relations quadratiques ou autre : parti pris des ACP non-linéaires

L'ACP se fait sur **des données quantitatives**

	x_1	x_2	x_3	x_4	x_5
1	12.1	0.1	A	2	-7.0
2	7.2	0.3	A	2	6.2
3	8.3	0.2	C	1	-4.7
4	9.3	0.4	B	3	5.2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

Attention aux variables qualitatives encodées sous forme de nombres.

Ici, si la variable x_4 est suspecte. Si x_3 et x_4 indiquent des classes alors on les enlève avant de faire l'ACP.

	x_1	x_2	x_5
1	12.1	0.1	-7.0
2	7.2	0.3	6.2
3	8.3	0.2	-4.7
4	9.3	0.4	5.2
\vdots	\vdots	\vdots	\vdots

\Rightarrow

Notation matricielle

du jeu de données quantitatives

\mathbb{X} : Matrice des données

x_{ij} : Valeur de la j -ème variable pour l'individu i

x_j : Vecteur colonne des valeurs de la j -ème variable, $x_j = (x_{ij})_{i=1..n}$

\bar{x}_j : Moyenne associée à la j -ème variable

$$\mathbb{X} = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & \dots & x_{2j} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix} \in \mathcal{M}_{n,p}(\mathbb{R})$$

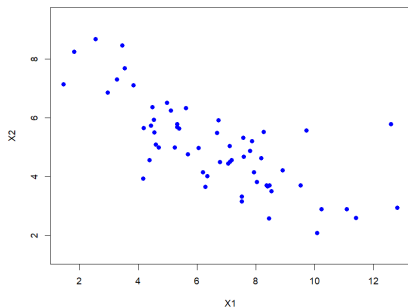
Le jeu de données peut se voir comme un nuage de points

n : nombre d'observations

p : nombre de variables

dimension du nuage de points
= nombre de variables
= p

	x_1	x_2	\dots	x_p
1	x_{11}	x_{12}	\dots	x_{1p}
2	x_{21}	x_{22}	\dots	x_{2p}
\vdots	\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots	\vdots
n	x_{n1}	x_{n2}	\dots	x_{np}



Quelles tendances voit-on dans ce nuage de points ?

Intuitivement

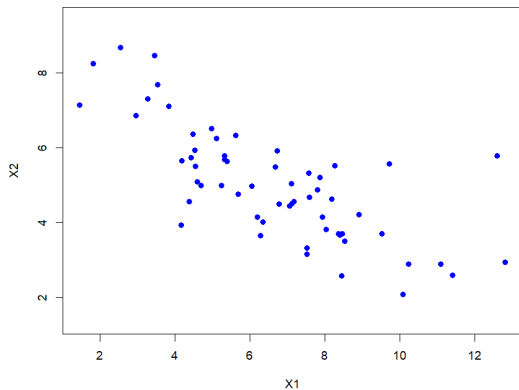


Figure – Nuage de points en 2 dimensions

Quelles tendances voit-on dans ce nuage de points ?

Intuitivement

1) Moyenne (valeur centrale)

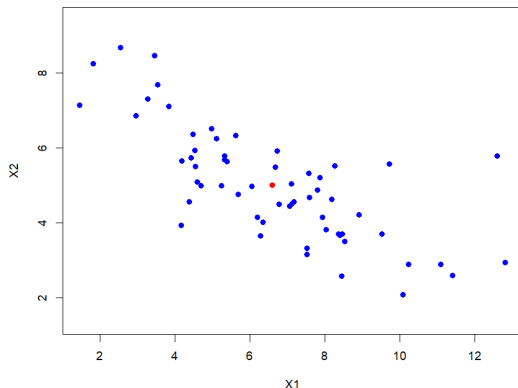


Figure – Nuage de points en 2 dimensions

Tendance de dispersion ?

Quelles tendances voit-on dans ce nuage de points ?

Intuitivement

- 1) Moyenne. 2) dispersion selon une première direction

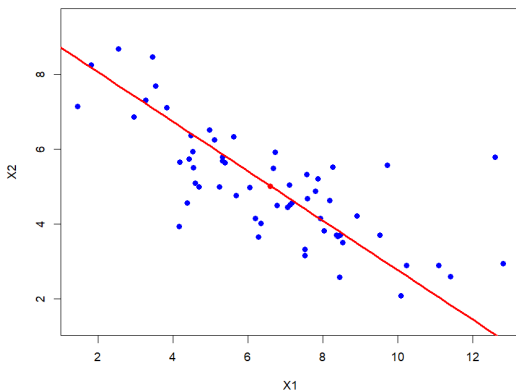


Figure – Nuage de points en 2 dimensions

Quelles tendances voit-on dans ce nuage de points ?

Intuitivement

- 1) Moyenne. 2) dispersion selon un 1^{er} axe. 3) dispersion sur un 2^e axe

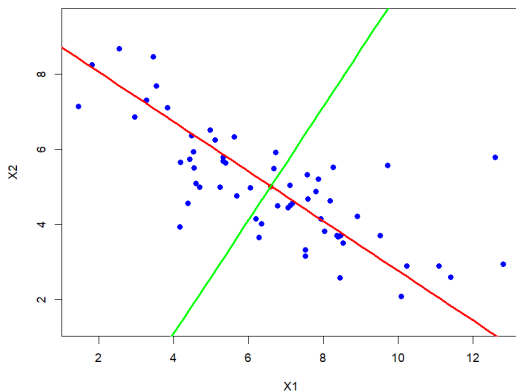


Figure – Nuage de points en 2 dimensions

Première étape de l'ACP

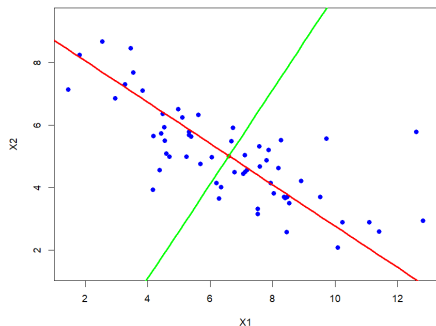


Figure – Nuage de points en 2D
+ directions des dispersions principales



Mission 1 de l'ACP :

Déterminer les axes qui expliquent la dispersion.

Ces axes sont appelés
« axes principaux » ou
« **composantes principales** ».

L'ACP permet un changement de repère

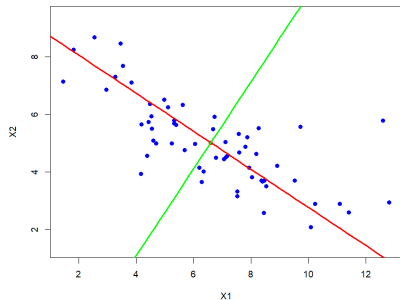


Figure – Données initiales

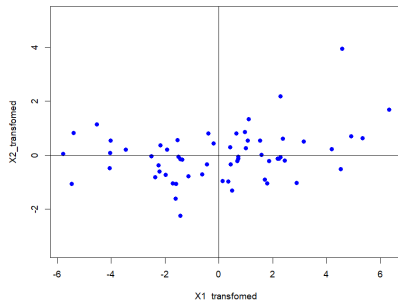


Figure – Données dans le repère ACP

Intérêt :

- Les données transformées sont décorrélées.
Modèle prédictif (ex : régression) avec données décorrélées = meilleures prédictions.
- Détermination des observations opposées plus facile
(avec les coordonnées transformées).

Exemple avec 3 variables

nuage gaussien

Démo via R (package plotly) pour visualiser en 3D

Exemple avec 3 variables

nuage gaussien

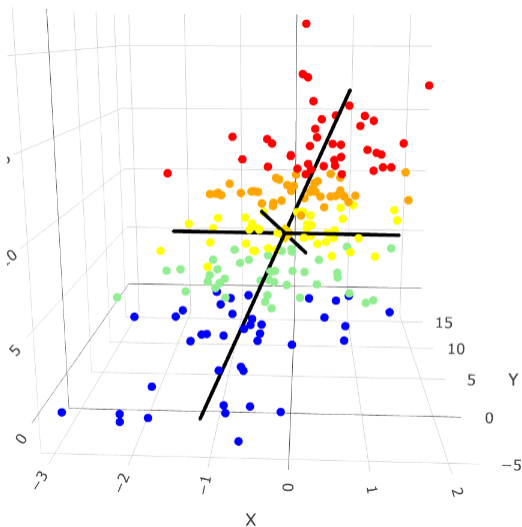


Figure – Nuage de points gaussien + axes de dispersions principales

Exemple avec 3 variables

nuage boomerang

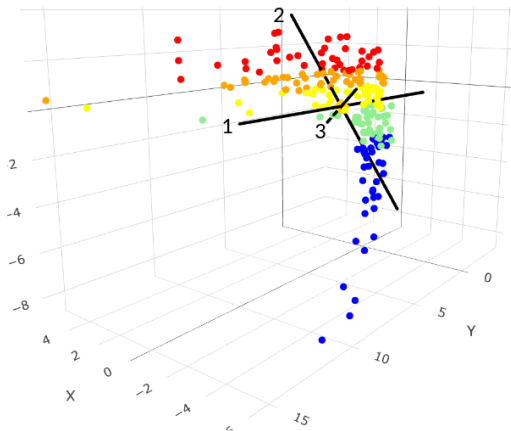


Figure – Nuage de points boomerang

Exemple avec 3 variables

nuage boomerang

Figure – Nuage de points boomerang + axes de dispersions principales

Réduire la dimension

exemple 2 variables \rightarrow 1 variables

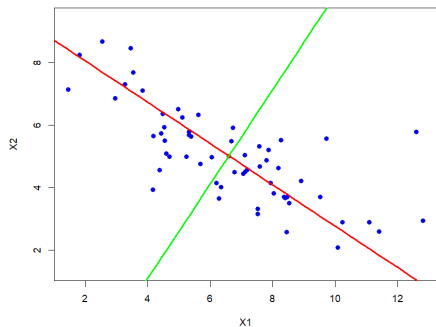


Figure – Nuage de points en 2D
+ directions des dispersions principales

L'axe vert capte peu de variabilité.

Mission 2 de l'ACP :

\rightarrow Négliger les axes associés à une faible quantité de dispersion.

Réduire la dimension

exemple 2 variables \rightarrow 1 variables

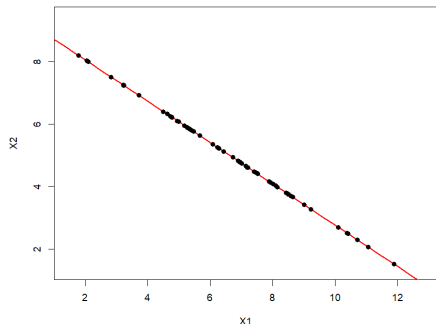


Figure – Nuage de points compressé

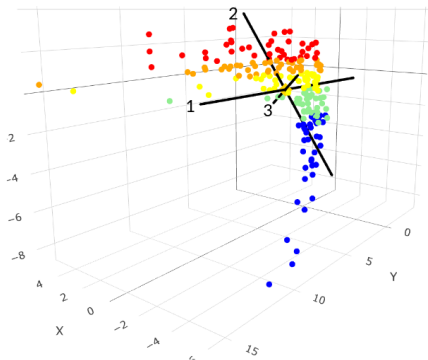
On compresse sur l'axe rouge.

Mission 2 de l'ACP :

\rightarrow Négliger les axes associés à une faible quantité de dispersion.

Réduire la dimension

exemple 3 variables \rightarrow 2 variables



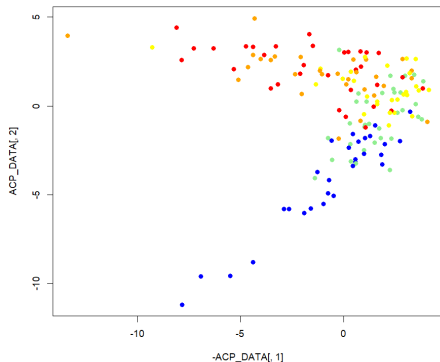
Mission 2 de l'ACP :

\rightarrow Négliger les axes associés à une faible quantité de dispersion.

Figure – Nuage de points en 3D
+ directions des dispersions principales

Réduire la dimension

exemple 3 variables \rightarrow 2 variables



Mission 2 de l'ACP :

Négliger les axes associés à une faible quantité de dispersion.

Figure – Nuage de points compressé en 2D

Réduction de dimension

Initialement p variables \longrightarrow Réduire à k variables transformées ($k < p$).

Exemple avec $p = 4$ et $k = 2$:

	x_1	x_2	x_3	x_4
1	x_{11}	x_{12}	x_{13}	x_{14}
2	x_{21}	x_{22}	x_{23}	x_{24}
\vdots	\vdots	\vdots	\vdots	\vdots
n	x_{n1}	x_{n2}	x_{n3}	x_{n4}

Réduction de dim \longrightarrow

	z_1	z_2
1	z_{11}	z_{12}
2	z_{21}	z_{22}
\vdots	\vdots	\vdots
n	z_{n1}	z_{n2}

Question :

Quelles variables transformées choisir pour garder le plus d'information ?

Avec une réduction de dimension :

- Plus facile de comprendre et visualiser les liens de corrélation principaux en petite dimension.
- Les faibles bruits sont filtrés.

Avec 26 variables, difficile de se représenter les données en dimension 26...
C'est plus facile en 2 ou 3 dimensions, non ?

L'ACP s'articule donc autour de deux points

- ① L'identification des composantes principales.
- ② La réduction de dimension.

- 1 Statistiques descriptives — à quoi servent les analyses factorielles ?
- 2 Introduction sur l'ACP
- 3 Identifier les composantes principales
 - L'inertie comme mesure de dispersion
 - Inertie orientée selon un axe
 - Décomposition en composantes principales
 - Changement de repère
- 4 Réduction de dimension par ACP

- 1 Statistiques descriptives — à quoi servent les analyses factorielles ?
- 2 Introduction sur l'ACP
- 3 Identifier les composantes principales
 - L'inertie comme mesure de dispersion
 - Inertie orientée selon un axe
 - Décomposition en composantes principales
 - Changement de repère
- 4 Réduction de dimension par ACP

L'inertie : une mesure de la dispersion du nuage de points

M_i : coordonnées du i -ème point $\longrightarrow M_i = (x_{i1}, x_{i2}, \dots, x_{ip})$

G : coordonnées du barycentre $\longrightarrow G = \frac{1}{n} \sum_{i=1}^n M_i = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)$

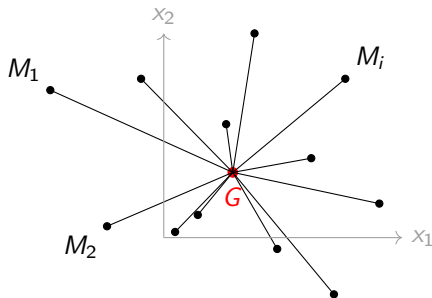


Figure – Exemple pour $p = 2$

Inertie totale :

$$\begin{aligned} I_{tot} &= \sum_{i=1}^n \frac{1}{n} \|M_i - G\|^2 \\ &= \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2 \end{aligned}$$



$$I_{tot} = \sum_{j=1}^p \text{Var}(x_j)$$

(Somme des variances)

Centrer les données

Inertie totale : $I_{tot} = \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2$

$$x_{ij} \leftarrow x_{ij} - \bar{x}_j$$

Moyennes et barycentre deviennent nuls :

$$\forall j, \bar{x}_j = 0 \quad \text{et} \quad G = (0, \dots, 0).$$

Inertie totale reste égale à la somme des variances :

$$I_{tot} = \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^p x_{ij}^2 = \sum_{j=1}^p \text{Var}(x_j)$$

Intérêt :

- Centre l'analyse autour de la tendance principale : la moyenne.
- Facilite les calculs (permet d'utiliser l'algèbre linéaire classique).

Centrage = changement de l'origine du repère

M_i : coordonnées du i -ème point $\longrightarrow M_i = (x_{i1}, x_{i2}, \dots, x_{ip})$

G : barycentre devient l'origine $\longrightarrow G = (0, \dots, 0)$

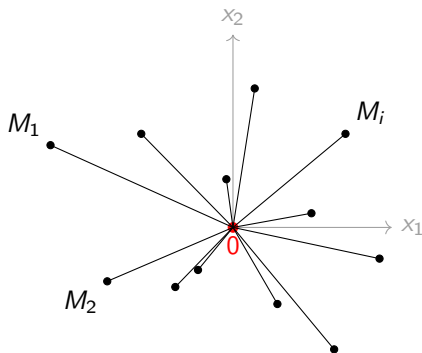


Figure – Exemple pour $p = 2$

\longrightarrow l'origine se cale sur le centre du nuage

Inertie totale :

$$\begin{aligned} l_{tot} &= \sum_{i=1}^n \frac{1}{n} \|M_i\|^2 \\ &= \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^p (x_{ij})^2 \end{aligned}$$



$$l_{tot} = \sum_{j=1}^p \text{Var}(x_j)$$

(Somme des variances)

- 1 Statistiques descriptives — à quoi servent les analyses factorielles ?
- 2 Introduction sur l'ACP
- 3 Identifier les composantes principales
 - L'inertie comme mesure de dispersion
 - Inertie orientée selon un axe
 - Décomposition en composantes principales
 - Changement de repère
- 4 Réduction de dimension par ACP

Comment mesurer la dispersion associée à un axe ?

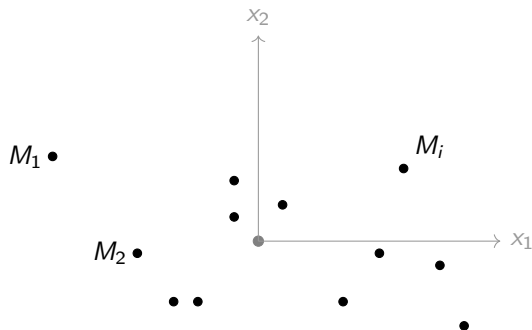


Figure – Un nuage de points

Comment mesurer la dispersion associée à un axe ?

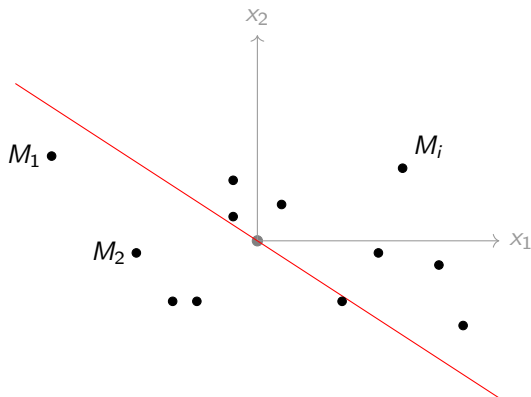


Figure – Un nuage de points et un axe

Comment mesurer la dispersion associée à un axe ?

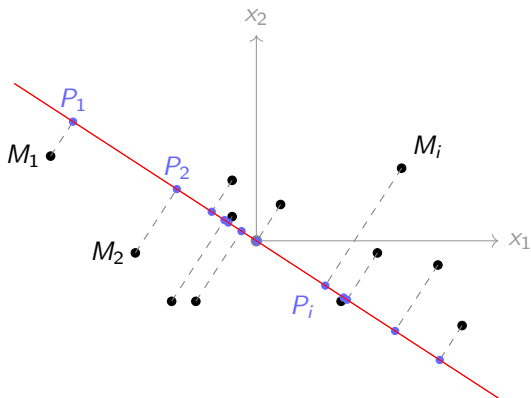


Figure – Projections orthogonales sur l'axe

Comment mesurer la dispersion associée à un axe ?

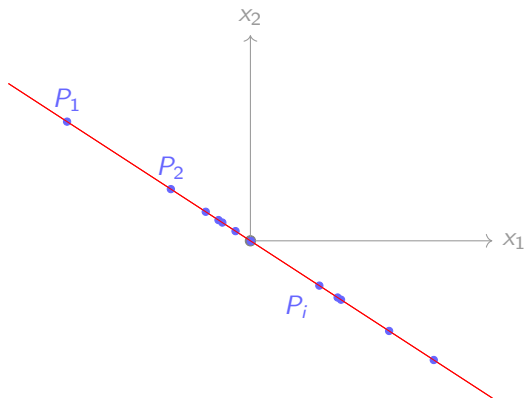


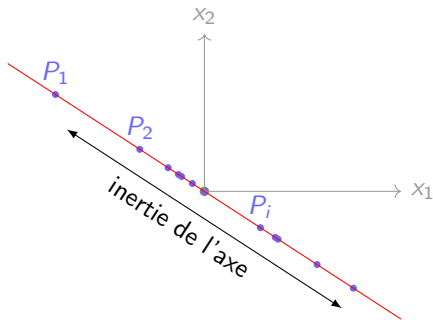
Figure – projetés sur l'axe

Mesurer la quantité de dispersion sur un axe

v : vecteur directeur de l'axe, $\|v\| = 1$.

L'inertie des projetés mesure la dispersion associée à l'axe :

$$I_v = \sum_{i=1}^n \frac{1}{n} \|P_i\|^2 = \text{Var}(\|P_i\|).$$



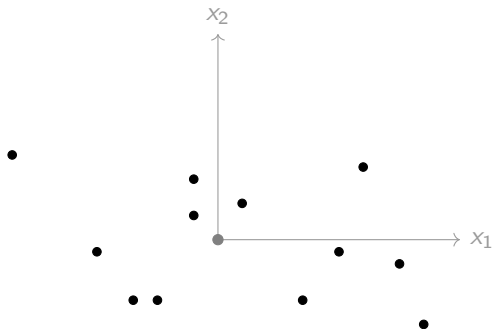


Figure – Le nuage + une droite passant par l'origine

$$I_v = \text{Var}(\|P_i\|)$$

Exemples

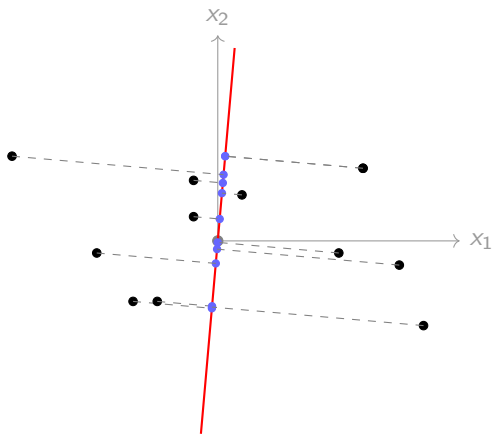


Figure – Inertie axiale faible

$$I_v = \text{Var}(\|P_i\|)$$

Exemples

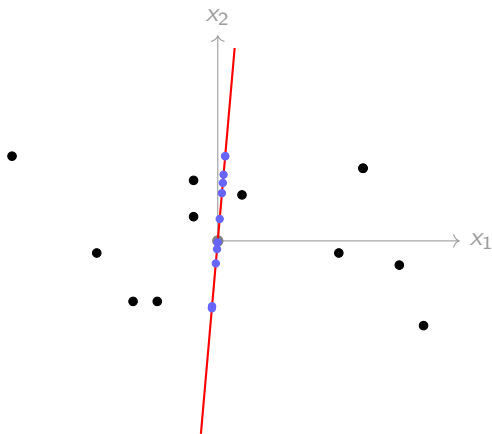


Figure – Inertie axiale faible

$$I_v = \text{Var}(\|P_i\|)$$

Exemples

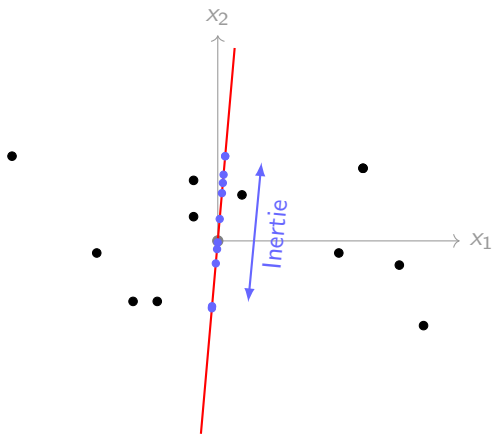


Figure – Inertie axiale faible

$$I_v = \text{Var}(\|P_i\|)$$

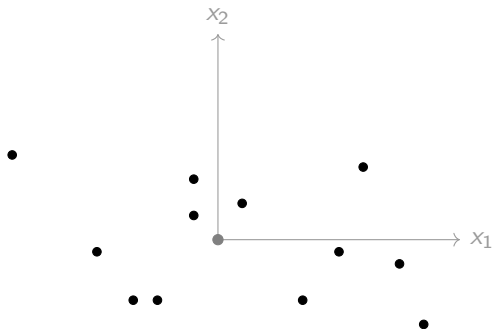


Figure – Le nuage + une droite passant par l'origine

$$I_v = \text{Var}(\|P_i\|)$$

Exemples

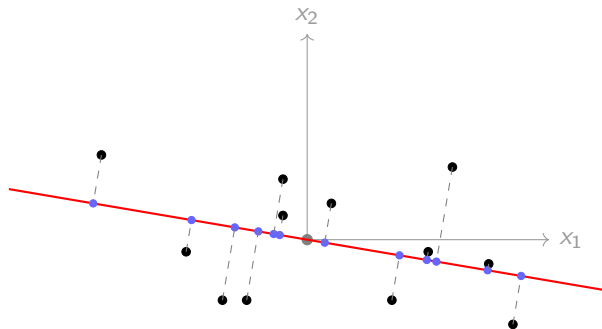


Figure – Inertie axiale maximale

$$I_v = \text{Var}(\|P_i\|)$$

Exemples

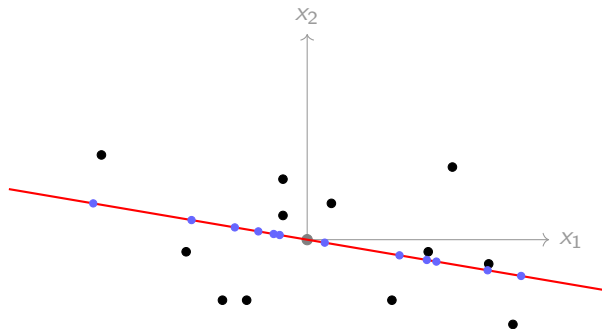


Figure – Inertie axiale maximale

$$I_v = \text{Var}(\|P_i\|)$$

Exemples

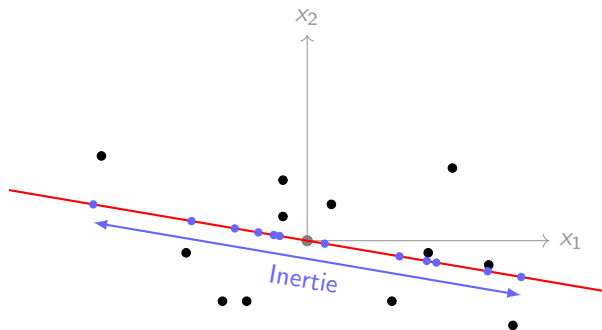


Figure – Inertie axiale maximale

$$I_v = \text{Var}(\|P_i\|)$$

Décomposition de l'inertie totale par l'inertie d'un axe

$$\begin{aligned}I_{tot} &= \frac{1}{n} \sum_{i=1}^n \|M_i\|^2 \\&= \frac{1}{n} \sum_{i=1}^n \|P_i + (M_i - P_i)\|^2 \\&= \frac{1}{n} \sum_{i=1}^n \|P_i\|^2 + \|M_i - P_i\|^2 \quad (\text{par orthogonalité}) \\&= \frac{1}{n} \sum_{i=1}^n \|P_i\|^2 \quad + \quad \frac{1}{n} \sum_{i=1}^n \|M_i - P_i\|^2\end{aligned}$$

I_{tot}	=	I_v	+	$\frac{1}{n} \sum_{i=1}^n \ M_i - P_i\ ^2$
$\underbrace{\hspace{2cm}}$		$\underbrace{\hspace{2cm}}$		$\underbrace{\hspace{2cm}}$
Inertie totale		Inertie de l'axe		Inertie non-orientée selon l'axe

Décomposition de l'inertie totale par l'inertie d'un hyperplan

$M_i - P_i$: projection du point M_i sur l'hyperplan orthogonal à v .

$$\forall i, \quad M_i - P_i \in v^\perp$$

$$I_{tot} = I_{v_1} + \underbrace{\frac{1}{n} \sum_{i=1}^n \|M_i - P_i\|^2}_{\text{Inertie restante dans l'hyperplan orthogonal}}$$

En projetant tous les points sur l'hyperplan v^\perp , on obtient un nouveau nuage de point, qui ne prend pas en compte la dispersion selon v_1 .

L'inertie résiduelle

Exemple en 2D

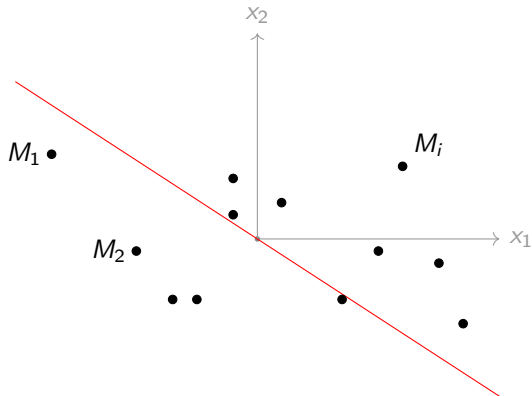


Figure – Un nuage de points

L'inertie résiduelle

Exemple en 2D

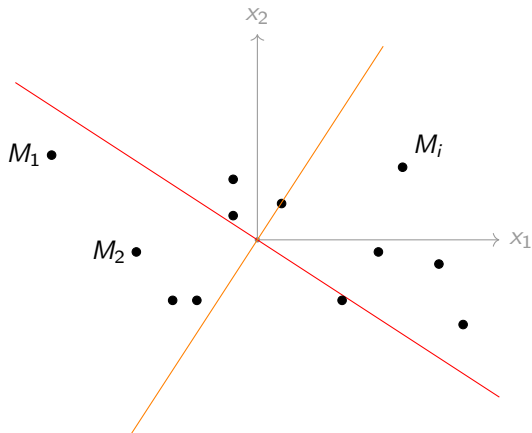


Figure – Un nuage de points et un hyperplan (ici la droite orange)

L'inertie résiduelle

Exemple en 2D

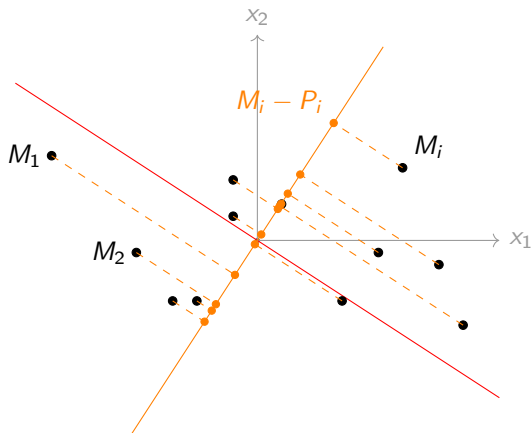


Figure – Projections orthogonales sur l'hyperplan

L'inertie résiduelle

Exemple en 2D

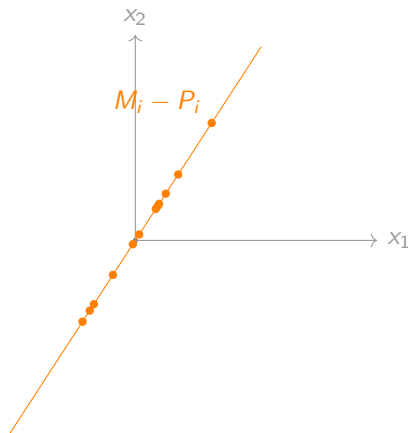


Figure – Nouveau nuage de points

L'inertie résiduelle

Exemple en 2D

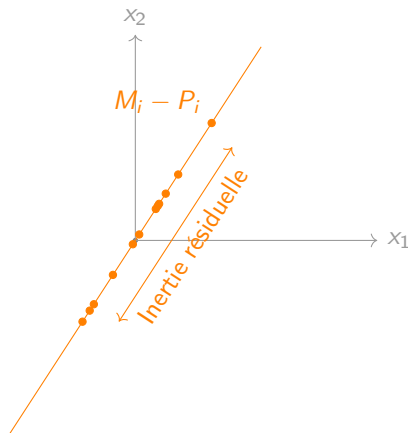


Figure – Nouveau nuage de points

Notation matricielle

\mathbb{X} : Matrice des **données centrées**

x_{ij} : valeur de la j -ème variable pour l'individu i

x_j : vecteur colonne des valeurs de la j -ème variable, $x_j = (x_{ij})_{i=1..n}$

\bar{x}_j : moyenne associée à la j -ème variable, $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} = 0$

$$\mathbb{X} = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & \dots & x_{2j} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix} \in \mathcal{M}_{n,p}(\mathbb{R})$$

dimension du nuage de points = nombre de variable p

Produit scalaire : $u, v \in \mathbb{R}^n$, $(u|v) = \sum_{i=1}^n u_i v_i$

Inertie associée à un axe (notation matricielle)

v : vecteur directeur de l'axe tel que $\|v\| = 1$

P_i : projeté orthogonal de M_i sur l'axe, donc $P_i = (M_i | v) v$

Grâce au centrage ($G = 0_p$), on a :

$$I_v = \sum_{i=1}^n \frac{1}{n} \|P_i\|^2 = \sum_{i=1}^n \frac{1}{n} (M_i | v)^2 = \frac{1}{n} \begin{pmatrix} (M_1 | v) \\ \vdots \\ (M_n | v) \end{pmatrix}^* \begin{pmatrix} (M_1 | v) \\ \vdots \\ (M_n | v) \end{pmatrix}$$

$$\text{or } \forall i, (M_i | v) = \sum_{j=1}^p x_{ij} v_j \implies \begin{pmatrix} (M_1 | v) \\ \vdots \\ (M_n | v) \end{pmatrix} = \mathbb{X} v$$

$$\implies I_v = \frac{1}{n} (\mathbb{X} v)^* (\mathbb{X} v),$$

$$I_v = \frac{1}{n} v^* (\mathbb{X}^* \mathbb{X}) v$$

Liens entre la matrice de covariance et les inerties

C : Matrice de covariance. Coefficients : $C_{ij} = \mathbb{C}ov(x_i, x_j)$

$$C = \begin{pmatrix} \mathbb{V}ar(x_1) & \mathbb{C}ov(x_1, x_2) & \dots & \mathbb{C}ov(x_1, x_p) \\ \mathbb{C}ov(x_2, x_1) & \mathbb{V}ar(x_2) & & \mathbb{C}ov(x_2, x_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{C}ov(x_p, x_1) & \mathbb{C}ov(x_p, x_2) & \dots & \mathbb{V}ar(x_p) \end{pmatrix} \in \mathcal{M}_{pp}(\mathbb{R})$$

$$\mathbb{C}ov(x_i, x_j) = \frac{1}{n} \sum_{k=1}^n x_{ki} x_{kj}$$

$$\Rightarrow \boxed{C = \frac{1}{n} \mathbb{X}^* \mathbb{X}}$$

$$\text{or } l_v = \frac{1}{n} v^* \mathbb{X}^* \mathbb{X} v$$

$$\Rightarrow \boxed{l_v = v^* C v}$$

$$l_{\text{tot}} = \sum_{j=1}^p \mathbb{V}ar(x_j)$$

$$\Rightarrow \boxed{l_{\text{tot}} = \text{tr}(C)}$$

- 1 Statistiques descriptives — à quoi servent les analyses factorielles ?
- 2 Introduction sur l'ACP
- 3 Identifier les composantes principales
 - L'inertie comme mesure de dispersion
 - Inertie orientée selon un axe
 - Décomposition en composantes principales
 - Changement de repère
- 4 Réduction de dimension par ACP

Le premier axe principal

Le premier axe principal est celui qui explique le plus de dispersion.

v_1 : vecteur directeur normé du premier axe principal

Le premier axe principal est celui qui maximise l'inertie axiale.

$$v_1 = \operatorname{argmax}_{v \in \mathbb{R}^p, \|v\|=1} I_v$$

$$\rightarrow v_1 = \operatorname{argmax}_{v \in \mathbb{R}^p, \|v\|=1} v^* C v.$$

Remarque :

C est une matrice symétrique réelle définie positive, donc diagonalisable dans une base orthonormée à valeur propre positives.

On note λ_i ses valeurs propres ordonnées : $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

Inertie du premier axe principal

Le premier axe principal est celui qui maximise l'inertie axiale.

$$v_1 = \operatorname{argmax}_{v \in \mathbb{R}^p, \|v\|=1} v^* C v.$$

Théorème spectral :

- v_1 est un vecteur propre normé de la matrice C
- v_1 est un vecteur propre associé à plus grande valeur propre λ_1
- $I_{v_1} = v_1^* C v_1 = \lambda_1 = \mathbb{V}ar((M_i | v_1))$
 I_{v_1} = variance des normes des projetés

Première décomposition de l'inertie

Sur l'axe principal

$M_i - P_i$: projection du point M_i sur l'hyperplan orthogonal à v_1 .

$$\forall i, \quad M_i - P_i \in v_1^\perp$$

$$I_{tot} = I_{v_1} + \underbrace{\frac{1}{n} \sum_{i=1}^n \|M_i - P_i\|^2}_{\text{Inertie résiduelle dans l'hyperplan orthogonal}}$$

En projetant tous les points sur l'hyperplan v_1^\perp , on obtient un nouveau nuage de point, qui ne prend pas en compte la dispersion selon v_1 .

Première décomposition de l'inertie

Sur l'axe principal

$M_i - P_i$: projection du point M_i sur l'hyperplan orthogonal à v_1 .

$$\forall i, \quad M_i - P_i \in v_1^\perp$$

$$I_{tot} = I_{v_1} + \underbrace{\frac{1}{n} \sum_{i=1}^n \|M_i - P_i\|^2}_{\text{Inertie résiduelle dans l'hyperplan orthogonal}}$$

En projetant tous les points sur l'hyperplan v_1^\perp , on obtient un nouveau nuage de point, qui ne prend pas en compte la dispersion selon v_1 .

→ on applique la même méthode sur ce nouveau nuage pour trouver le second axe principal.

Décomposition de l'inertie selon 2 axes

On note désormais le projeté de M_i sur v_j P_{ij} .

v_1^\perp : hyperplan orthogonal à v_1

$M_i - P_{i1} = x_i - (x_i|v_1)v_1$: Projection de M_i sur v_1^\perp

Si $v_2 \in v_1^\perp$, l'inertie dans l'hyperplan v_1^\perp se décompose :

$$\begin{aligned} I_{v_1^\perp} &= \frac{1}{n} \sum_{i=1}^n \|M_i - P_{i1}\|^2 = \frac{1}{n} \sum_{i=1}^n \|P_{i2} + M_i - P_{i1} - P_{i2}\|^2 \\ &= \frac{1}{n} \sum_{i=1}^n \|P_{i2}\|^2 + \|M_i - P_{i1} - P_{i2}\|^2 \quad (\text{par orthogonalité}) \\ &= I_{v_2} + \frac{1}{n} \sum_{i=1}^n \|M_i - P_{i1} - P_{i2}\|^2 \end{aligned}$$

$$\Rightarrow \boxed{I_{\text{tot}} = I_{v_1} + I_{v_2} + \frac{1}{n} \sum_{i=1}^n \|M_i - P_{i1} - P_{i2}\|^2} \quad \text{décomposition sur 2 axes}$$

Second axe principal, et inertie associée

Le second axe principal est celui qui maximise l'inertie axiale des projections sur l'hyperplan $\text{Vect}(v_1)^\perp$.

v_2 : vecteur directeur du second axe principal, $v_2 \in \text{Vect}(v_1)^\perp$

$$v_2 = \underset{v \in \mathbb{R}^p, \|v\|=1, v \in \text{Vect}(v_1)^\perp}{\text{argmax}} I_v$$

$$\rightarrow v_2 = \underset{v \in \mathbb{R}^p, \|v\|=1, v \in \text{Vect}(v_1)^\perp}{\text{argmax}} v^* C v$$

Théorème spectral :

- v_2 est un vecteur propre normé de la matrice C
- v_2 est un vecteur propre associé à la plus grande valeur propre λ_2
- $I_{v_2} = \lambda_2 = \mathbb{V}ar((M_i | v_2))$ (variance des normes des projections)

* Axes principaux

Pour trouver les axes principaux suivants on répète le raisonnement.

Le k -ième axe principal est celui qui maximise l'inertie axiale des projections dans l'espace $\text{Vect}(v_1, \dots, v_{k-1})^\perp$

v_k : vecteur directeur du k -ième axe principal, $v_k \in \text{Vect}(v_1, \dots, v_{k-1})^\perp$

$$v_k = \underset{\substack{v \in \mathbb{R}^p, \|v\| = 1, \\ v \in \text{Vect}(v_1, \dots, v_{k-1})^\perp}}{\text{argmax}} \quad v^* C v.$$

Théorème spectral :

- v_k est un vecteur propre normé de la matrice C
- v_k est un vecteur propre associé à la plus grande valeur propre λ_k
- $I_{v_k} = \lambda_k = \mathbb{V}ar((M_i | v_k))$ (variance des normes des projections)

* Axes principaux, décomposition complète de l'inertie

Les axes principaux v_j (appelés composantes principales) sont les vecteurs propres de la matrice de covariance C . Ils indiquent dans quelles directions les points sont dispersés.

Leur inertie associée est maximisée successivement et elle vérifie $I_{v_j} = \lambda_j$.

L'inertie totale se décompose de deux façons :

- 1 Décomposition de l'inertie originelle :

$$I_{tot} = \mathbb{V}ar(x_1) + \mathbb{V}ar(x_1) + \dots + \mathbb{V}ar(x_p)$$

- 2 Décomposition de l'inertie par ACP :

Comme (v_1, v_2, \dots, v_p) forme une base orthonormée, on a :

$$I_{tot} = I_{v_1} + I_{v_2} + \dots + I_{v_p}$$

$$I_{tot} = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

Décomposition de l'inertie : proportions d'inertie expliquée

Taux d'inertie expliquée par le j -ème axe principal

$$\mathcal{Prop}_j = \frac{\lambda_j}{\sum_{j=1}^p \lambda_j}$$

Taux d'inertie expliquée par les k premiers axes principaux,
Aussi appelée proportion d'inertie cumulée :

$$T_k = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^p \lambda_j}$$

- 1 Statistiques descriptives — à quoi servent les analyses factorielles ?
- 2 Introduction sur l'ACP
- 3 Identifier les composantes principales
 - L'inertie comme mesure de dispersion
 - Inertie orientée selon un axe
 - Décomposition en composantes principales
 - Changement de repère
- 4 Réduction de dimension par ACP

Formules de transitions vers le "repère ACP"

$V = (v_1 | \dots | v_p)$ et $\Delta = \text{diag}(\lambda_j)$, $C = V\Delta V^*$.

On peut exprimer les données dans la base (v_1, \dots, v_p)

$(\tilde{x}_{i1}, \dots, \tilde{x}_{ip})$: les coordonnées du point M_i dans cette nouvelle base. La coordonnée du point M_i sur l'axe j :

$$\tilde{x}_{ij} = (M_i | v_j),$$

Le vecteur des valeurs prises par la j -ème variable transformée est donc :

$$\tilde{x}_j = \mathbb{X}v_j$$

Le jeu de données transformé est donc :

$$\tilde{\mathbb{X}} = \mathbb{X}V$$

Les données transformées sont décorrélées

$$\text{cov}(\tilde{x}_j, \tilde{x}_k) = \frac{1}{n} \tilde{x}_j^* \tilde{x}_k = \frac{1}{n} (\tilde{x}_j | \tilde{x}_k) = 0$$

L'ACP permet un changement de repère

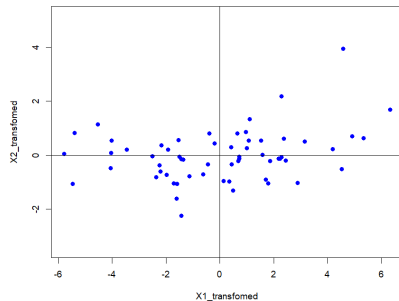
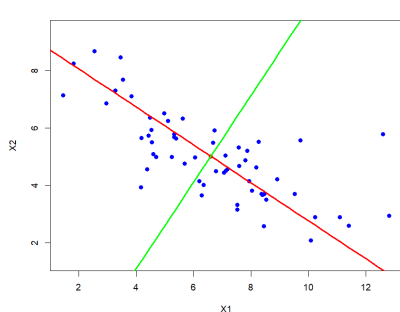


Figure – Données initiales ($p = 2$)

Figure – Données dans le repère ACP ($p = 2$)

Intérêt :

- Les données transformées sont décorrélées.
Modèle prédictif (ex : régression) avec données décorrélées = meilleures prédictions.
- Détermination des observations opposées plus facile.
(Les axes du repère ACP collent au nuage).

Graphes 3D

Vecteurs et valeurs propres

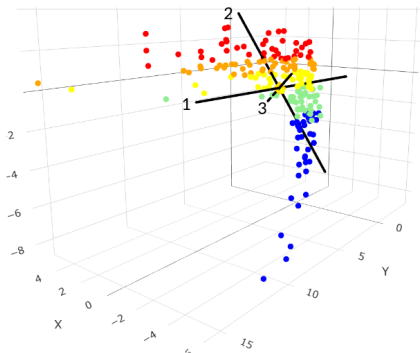


Figure – nuage d'origine

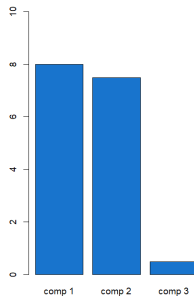


Figure – Valeurs propres

Graphes 3D

changement de repère

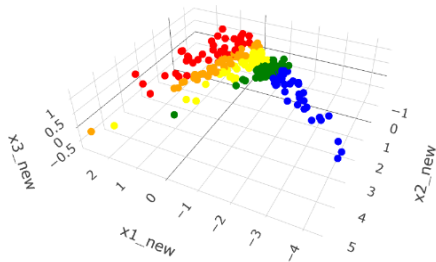


Figure – nuage d'origine

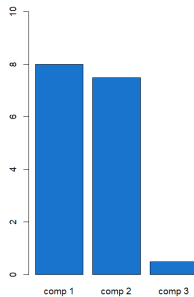


Figure – Valeurs propres

- 1 Statistiques descriptives — à quoi servent les analyses factorielles ?
- 2 Introduction sur l'ACP
- 3 Identifier les composantes principales
- 4 Réduction de dimension par ACP
 - Pourquoi réduire ?
 - Les outils pour choisir le nombre de dimension à conserver

- 1 Statistiques descriptives — à quoi servent les analyses factorielles ?
- 2 Introduction sur l'ACP
- 3 Identifier les composantes principales
- 4 Réduction de dimension par ACP
 - Pourquoi réduire ?
 - Les outils pour choisir le nombre de dimension à conserver

- Compression de données (avec perte).
- Filtrer les faibles bruits.
- Visualiser les données en 2D ou 3D
- Identifier des caractéristiques potentiellement plus fondamentales, et donc potentiellement plus explicatives.
- Prétraitement des données avant utilisation d'un modèle prédictif :
 - Utiliser le jeu de données transformées donne de meilleures prédictions.

Principe de la réduction de dimension par ACP

Initialement p variables \longrightarrow Réduire à k variables **transformées** ($k < p$).

	x_1	x_2	\dots	x_p
1	x_{11}	x_{12}	\dots	x_{1p}
2	x_{21}	x_{22}	\dots	x_{2p}
\vdots	\vdots	\vdots	\vdots	\vdots
n	x_{n1}	x_{n2}	\dots	x_{np}

ACP \longrightarrow

	\tilde{x}_1	\tilde{x}_2	\dots	\tilde{x}_p
1	\tilde{x}_{11}	\tilde{x}_{12}	\dots	\tilde{x}_{1p}
2	\tilde{x}_{21}	\tilde{x}_{22}	\dots	\tilde{x}_{2p}
\vdots	\vdots	\vdots	\vdots	\vdots
n	\tilde{x}_{n1}	\tilde{x}_{n2}	\dots	\tilde{x}_{np}

Principe de la réduction de dimension par ACP

Initialement p variables \longrightarrow Réduire à k variables **transformées** ($k < p$).

	x_1	x_2	\dots	x_p
1	x_{11}	x_{12}	\dots	x_{1p}
2	x_{21}	x_{22}	\dots	x_{2p}
\vdots	\vdots	\vdots	\vdots	\vdots
n	x_{n1}	x_{n2}	\dots	x_{np}

ACP \longrightarrow

	\tilde{x}_1	\tilde{x}_2	\dots	\tilde{x}_p
1	\tilde{x}_{11}	\tilde{x}_{12}	\dots	\tilde{x}_{1p}
2	\tilde{x}_{21}	\tilde{x}_{22}	\dots	\tilde{x}_{2p}
\vdots	\vdots	\vdots	\vdots	\vdots
n	\tilde{x}_{n1}	\tilde{x}_{n2}	\dots	\tilde{x}_{np}

- 1 Trouver un sous-espace linéaire, qui passe au plus proche des données.

Principe de la réduction de dimension par ACP

Initialement p variables \longrightarrow Réduire à k variables **transformées** ($k < p$).

	x_1	x_2	\dots	x_p
1	x_{11}	x_{12}	\dots	x_{1p}
2	x_{21}	x_{22}	\dots	x_{2p}
\vdots	\vdots	\vdots	\vdots	\vdots
n	x_{n1}	x_{n2}	\dots	x_{np}

ACP \longrightarrow

	\tilde{x}_1	\tilde{x}_2	\dots	\tilde{x}_p
1	\tilde{x}_{11}	\tilde{x}_{12}	\dots	\tilde{x}_{1p}
2	\tilde{x}_{21}	\tilde{x}_{22}	\dots	\tilde{x}_{2p}
\vdots	\vdots	\vdots	\vdots	\vdots
n	\tilde{x}_{n1}	\tilde{x}_{n2}	\dots	\tilde{x}_{np}

$k = 2 \downarrow$

- 1 Trouver un sous-espace linéaire, qui passe au plus proche des données.
- 2 Projection orthogonale des données transformées sur un sous-espace linéaire de plus faible dimension.

	\tilde{x}_1	\tilde{x}_2
1	\tilde{x}_{11}	\tilde{x}_{12}
2	\tilde{x}_{21}	\tilde{x}_{22}
\vdots	\vdots	\vdots
n	\tilde{x}_{n1}	\tilde{x}_{n2}

Questions associées à la réduction de dimension

Initialement p variables \longrightarrow Réduire à k variables transformées ($k < p$).

	x_1	x_2	\dots	x_p
1	x_{11}	x_{12}	\dots	x_{1p}
2	x_{21}	x_{22}	\dots	x_{2p}
\vdots	\vdots	\vdots	\vdots	\vdots
n	x_{n1}	x_{n2}	\dots	x_{np}

Réduction de dim \longrightarrow

	\tilde{x}_1	\tilde{x}_2
1	\tilde{x}_{11}	\tilde{x}_{12}
2	\tilde{x}_{21}	\tilde{x}_{22}
\vdots	\vdots	\vdots
n	\tilde{x}_{n1}	\tilde{x}_{n2}

Questions :

- Quelles variables transformées choisir pour garder le plus d'information ?
- Comment choisir k ?

Questions associées à la réduction de dimension

Initialement p variables \longrightarrow Réduire à k variables transformées ($k < p$).

	x_1	x_2	\dots	x_p
1	x_{11}	x_{12}	\dots	x_{1p}
2	x_{21}	x_{22}	\dots	x_{2p}
\vdots	\vdots	\vdots	\vdots	\vdots
n	x_{n1}	x_{n2}	\dots	x_{np}

Réduction de dim \longrightarrow

	\tilde{x}_1	\tilde{x}_2
1	\tilde{x}_{11}	\tilde{x}_{12}
2	\tilde{x}_{21}	\tilde{x}_{22}
\vdots	\vdots	\vdots
n	\tilde{x}_{n1}	\tilde{x}_{n2}

Questions :

- Quelles variables transformées choisir pour garder le plus d'information ? \longrightarrow Choisir les axes expliquant le plus d'inertie
- Comment choisir k ?
(4 critères seront vus ensuite)

Exemple du boomerang

Quels axes choisir ?

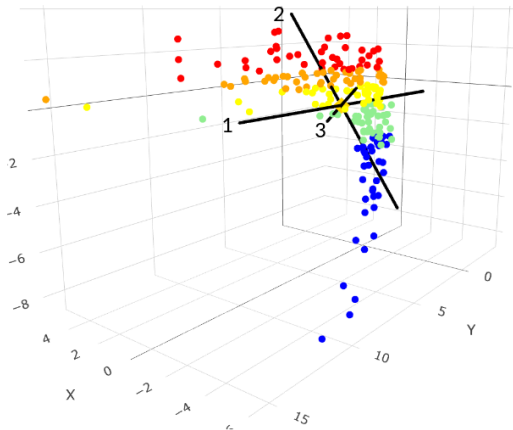
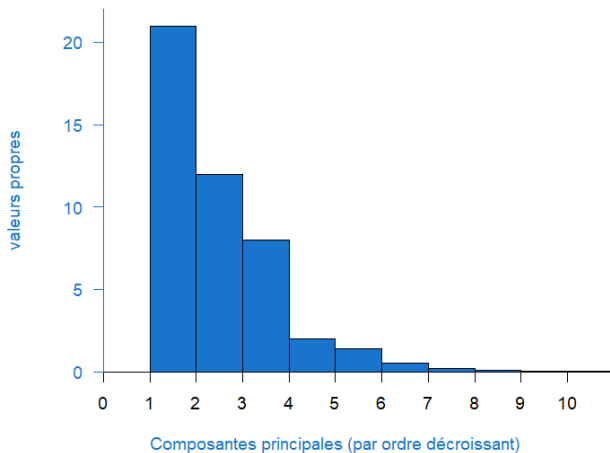


Figure – nuage d'origine

- 1 Statistiques descriptives — à quoi servent les analyses factorielles ?
- 2 Introduction sur l'ACP
- 3 Identifier les composantes principales
- 4 Réduction de dimension par ACP
 - Pourquoi réduire ?
 - Les outils pour choisir le nombre de dimension à conserver

Cascade des valeurs propres

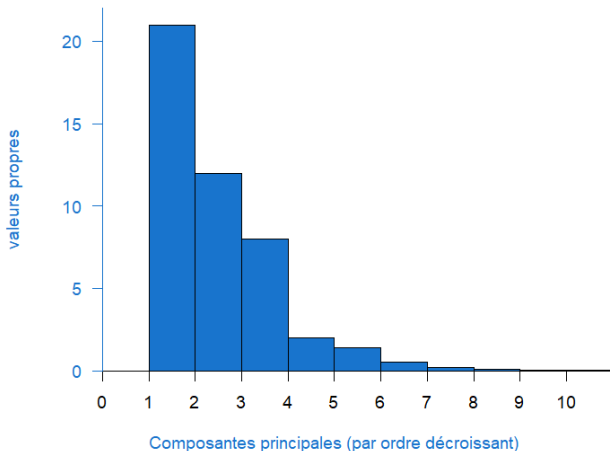
Exemple avec 10 variables



$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_p$$

Cascade des valeurs propres

Exemple avec $p = 10$

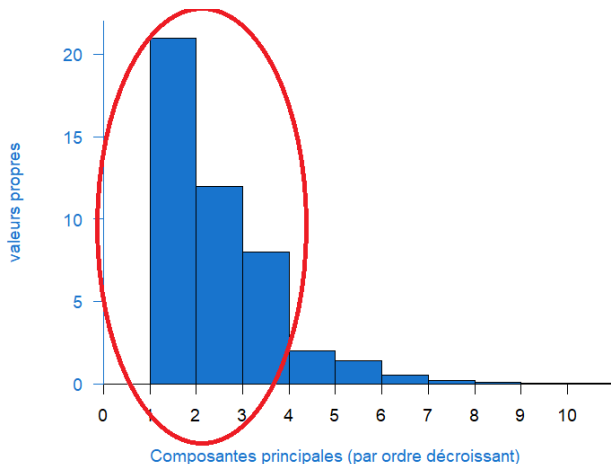


On choisit les composantes principales associés aux k plus grandes valeurs propres.

→ Permet de conserver un maximum d'inertie expliquée / d'information.

Cascade des valeurs propres

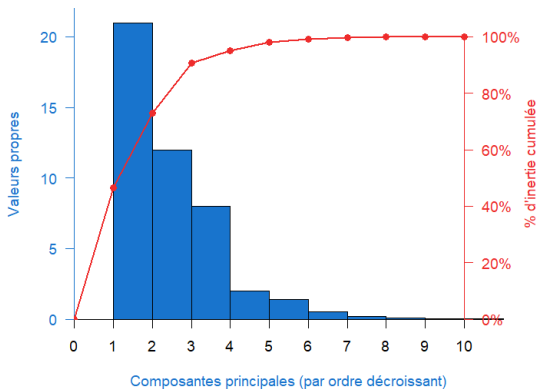
Exemple avec $p = 10$



On choisit les composantes principales associés aux k plus grandes valeurs propres.

→ Permet de conserver un maximum d'inertie expliquée / d'information.

Taux d'inertie cumulée

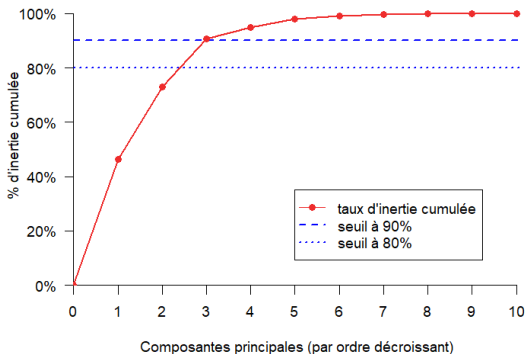


T_k : Taux d'inertie cumulée = proportion d'inertie conservée par les k premières composantes :

$$T_k = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^p \lambda_j}$$

Choisir le nombre de composantes à conserver

Premier critère : Taux d'inertie cumulée suffisant



❶ Seuil à 80%, 90%, ou $\alpha\%$ sur le taux d'inertie cumulée.

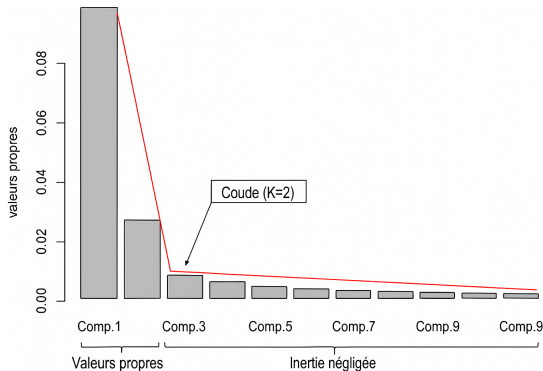
Prendre k tels que $T_k \leq \alpha$ et $T_{k+1} > \alpha$.

+ Facile à réaliser.

– Seuil arbitraire. (Sauf s'il est choisi selon le taux de compression voulu)

Choisir le nombre de composantes à conserver

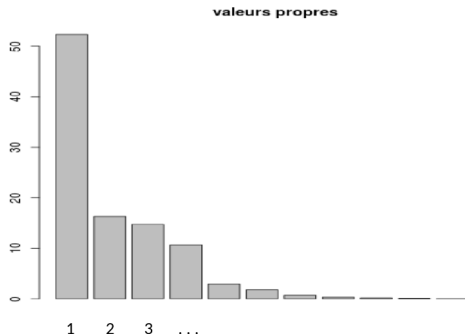
Critère visuel sur la cascade des valeurs propres : Règle de Cattell



2 Identifier visuellement un coude, une cassure.

- + Dépasse l'arbitraire purement numérique.
- Soumise à l'appréciation (critère visuel)
Choix pas évident quand : pas de cassure visible ou plusieurs cassures.

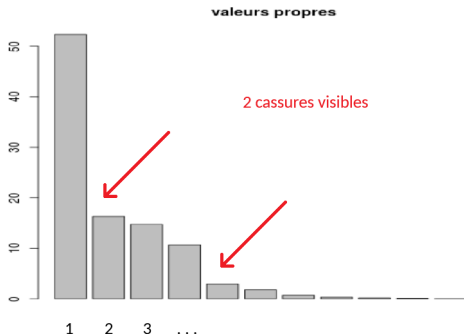
Choisir le nombre de composantes à conserver



- 2 Identifier visuellement un coude, une cassure.
 - + Dépasse l'arbitraire purement numérique.
 - Soumise à l'appréciation (critère visuel)
Choix pas évident quand : pas de cassure visible ou plusieurs cassures.

Choisir le nombre de composantes à conserver

Critère visuel sur la cascade des valeurs propres : Règle de Cattell



② Identifier visuellement un coude, une cassure.

- + Dépasse l'arbitraire purement numérique.
- Soumise à l'appréciation (critère visuel)
Choix pas évident quand : pas de cassure visible ou plusieurs cassures.

Choisir le nombre de composantes à conserver

Autres critères

- ③ Règle de Kaiser - Guttman → que pour ACP normée ($\sum_{j=1}^p \lambda_j = p$)

→ Prendre les axes tels que $\lambda_j \geq 1$.

- + Facile à mettre en place
- Valable uniquement pour ACP normée

- ④ Règle de Karlis - Saporta - Spinaki

→ Prendre les axes tels que $\lambda_j \geq 2\sqrt{\frac{p-1}{n-1}}$.

Seuil = la moyenne des valeurs propres + 2 fois leur écart-type.

- + Facile à mettre en place
- Restrictif (sélectionne peu d'axes)

Choisir le nombre de composantes à conserver

4 critères :

- ① Taux d'inertie expliquée $>$ seuil
- ② Règle de Cattell (critère visuel)
- ③ Règle de Kaiser - Guttman (ACP normée)
- ④ Règle de Karlis - Saporta - Spinaki

Pas de critère parfait :

On regarde tout les critères pour se faire une idée, et on choisit ce le nombre de composante qui parait pertinent.

Réduction

Exemple du nuage boomerang

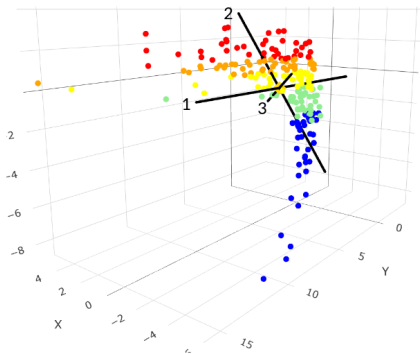


Figure – nuage d'origine

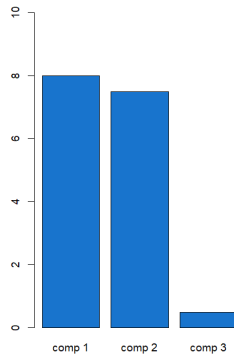


Figure – valeurs propres

→ 2 composantes à conserver.

Plans factoriels

Exemple du nuage boomerang

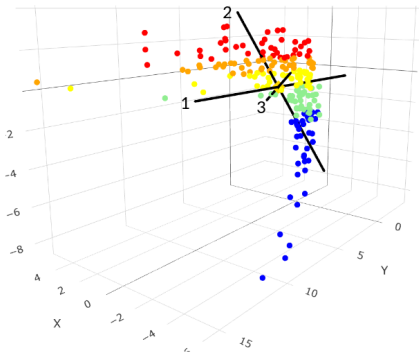


Figure – nuage d'origine

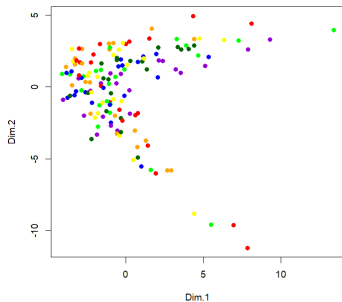


Figure – Plan factoriel axes 1 et 2

Un plan factoriel est un plan associée à 2 composantes principales.
En grande dimension, on y projette les points, se donner une idée de la forme du nuage, et identifier des individus opposés.

Plans factoriels

Exemple du nuage boomerang

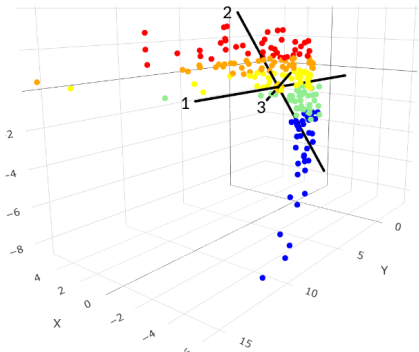


Figure – nuage d'origine

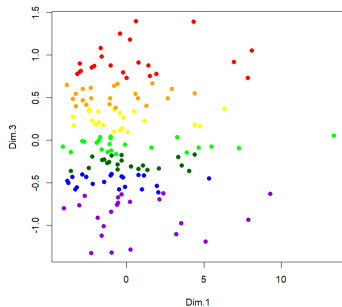


Figure – Plan factoriel axes 1 et 3

Un plan factoriel est un plan associée à 2 composantes principales.
En grande dimension, on y projette les points, se donner une idée de la forme du nuage, et identifier des individus opposés.

Plans factoriels

Exemple du nuage boomerang

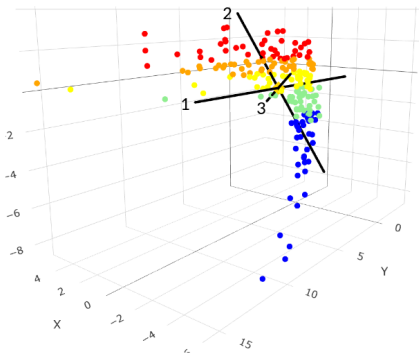


Figure – nuage d'origine

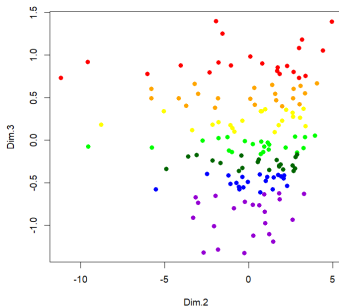
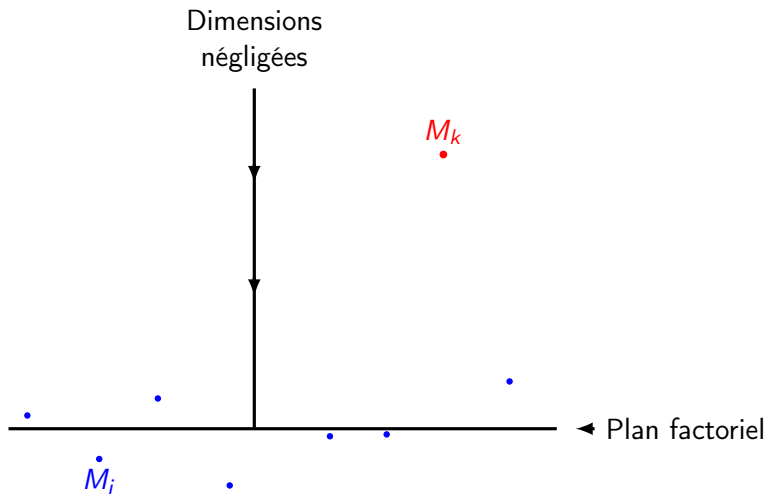


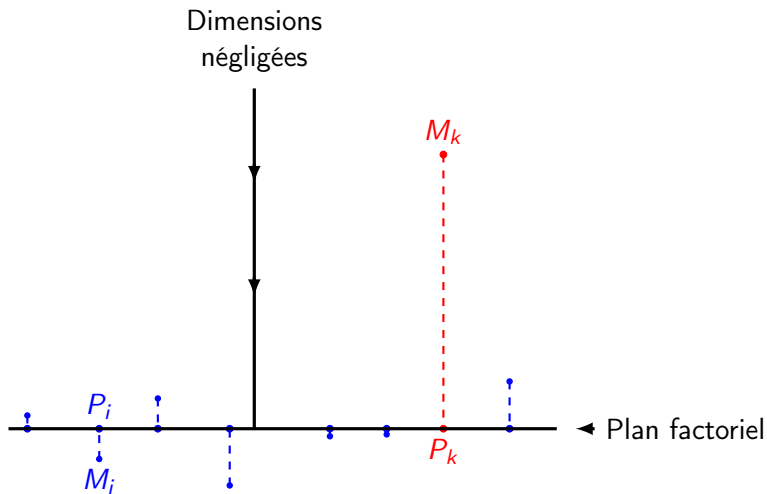
Figure – Plan factoriel axes 2 et 3

Un plan factoriel est un plan associée à 2 composantes principales.
En grande dimension, on y projette les points, se donner une idée de la forme du nuage, et identifier des individus opposés.

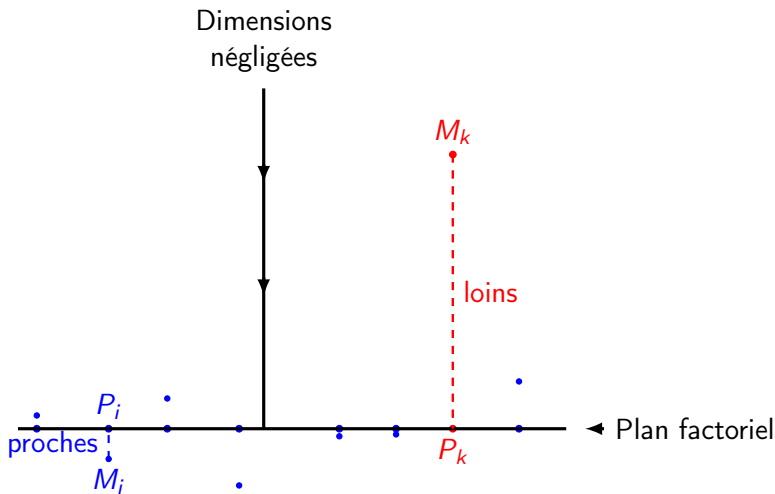
Qualité de représentation par un plan factoriel



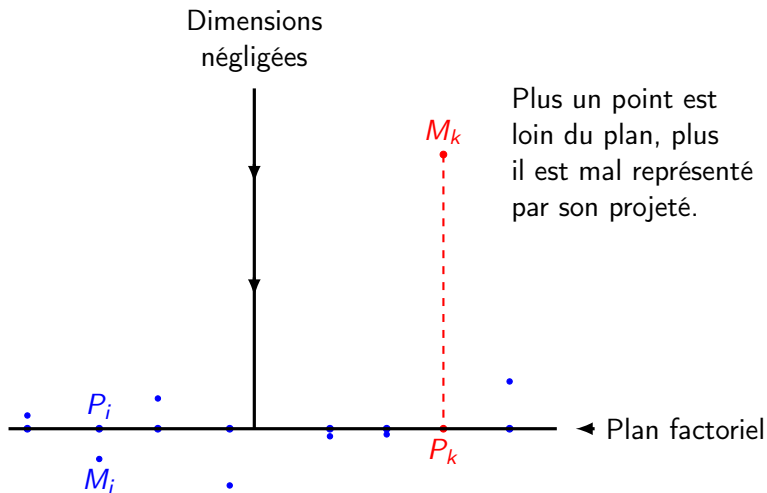
Qualité de représentation par un plan factoriel



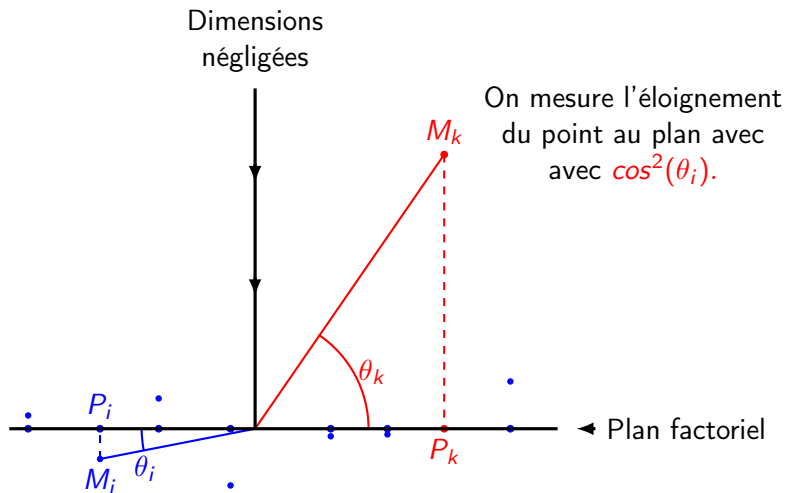
Qualité de représentation par un plan factoriel



Qualité de représentation par un plan factoriel

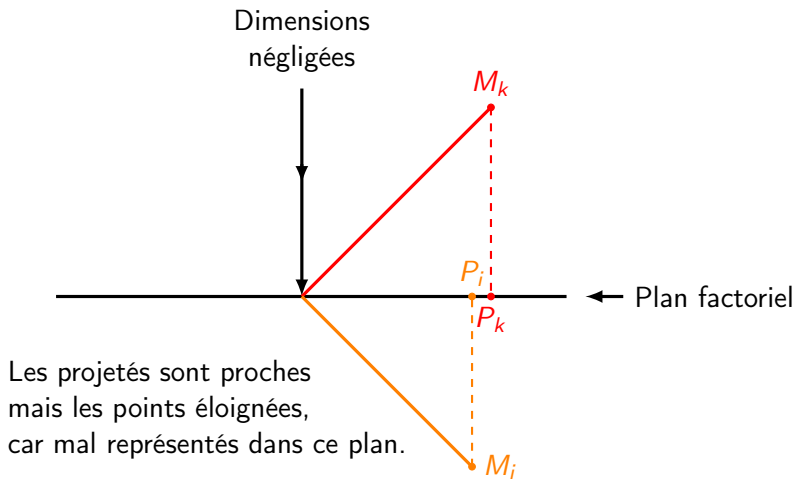


Qualité de représentation par un plan factoriel



Qualité de représentation

Exemple de cas à identifier pour ne pas se tromper dans l'interprétation



Qualité de représentation

F : indices des axes du plan

i : indice d'un individu

Qualité de représentation d'un point sur un plan factoriel :

$$\cos^2(\theta_i) = \frac{\|P_i\|^2}{\|M_i\|^2} = \frac{\sum_{j \in F} \tilde{x}_{ij}^2}{\sum_{j=1}^p \tilde{x}_{ij}^2}$$

Avant d'identifier des points proches via leur projetés sur un plan factoriel, il faut regarder s'ils sont bien représentés par le plan.

Ils peuvent en réalité être très éloignées selon les autres composantes (orthogonales au plan).

Par construction cela arrive peu sur les premiers plans car les dimensions négligées captent peu d'inertie, mais prudence.

Notions à comprendre :

- ① La décomposition en composantes principales :
savoir trouver un repère adapté à la dispersion
- ② La réduction de dimension :
savoir quelles dimensions choisir
- ③ Analyses des résultats :
plans de projections factorielles, et cercle de corrélation des variables