

Chapitre 3:

Analyse en Composantes Principales (ACP)

Souhir BOUAZIZ AFFES

souhir.bouaziz@isims.usf.tn

Amal ABBES

amal.abbes@isims.usf.tn

Plan



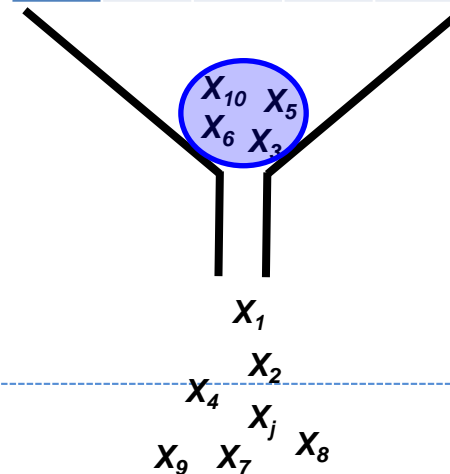
- ▶ Principe de l'ACP
 - Motivation, Applications, Situation, Objectifs, Notion de projection
- ▶ ACP et la covariance
- ▶ ACP et la corrélation
- ▶ Procédure de l'ACP à partir d'un exemple
- ▶ Procédure de l'ACP
- ▶ Combien de composantes principales?
- ▶ Typologies des individus et des variables
- ▶ Exercices

Principe de l'ACP : Motivation

- ▶ Le but de l'ACP est de **réduire la dimensionnalité** d'un ensemble de données (échantillon) en trouvant un **nouvel ensemble de variables, plus petit** que l'ensemble de variables initiales, qui conserve néanmoins la plupart des informations de l'échantillon :
 - Simplification de la réalité
 - Concentration d'une information de départ diluée
 - Description du maximum de **variabilité** dans un espace réduit



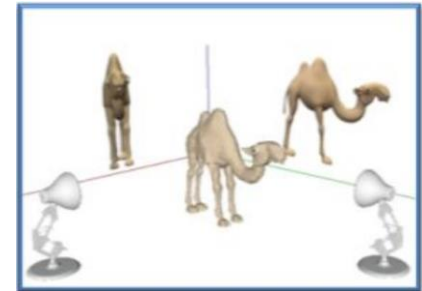
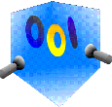
	X_1	X_2	X_3	...
ind 1	33	12	55	...
ind 2	25	11	50	...
ind 3	29	11	43	...
...



Principe de l'ACP : Motivation

► Dessinez c'est gagné?

➤ Simplifier un objet 3D (réalité) par une représentation en 2D (plan)



Comment choisir le bon espace de projection ?

Principe de l'ACP : Intérêt

- ▶ Les données du monde réel sont constituées de nombreuses caractéristiques (variables), qui peuvent être redondantes. Lorsque nous travaillons avec ces caractéristiques, nous pouvons rencontrer plusieurs problèmes :
 - Sur-ajustement (ou *sur-apprentissage*) du modèle
 - Consommation du temps
- ▶ Pour résoudre ces problèmes, nous pouvons appliquer la réduction de la dimensionnalité aux données. Il existe deux types de réduction de la dimensionnalité :
 - Élimination de caractéristiques
 - Extraction de caractéristiques – **ACP**



L'ACP est souvent utilisée pour réduire la dimension des données et faciliter leur exploration et leur visualisation.

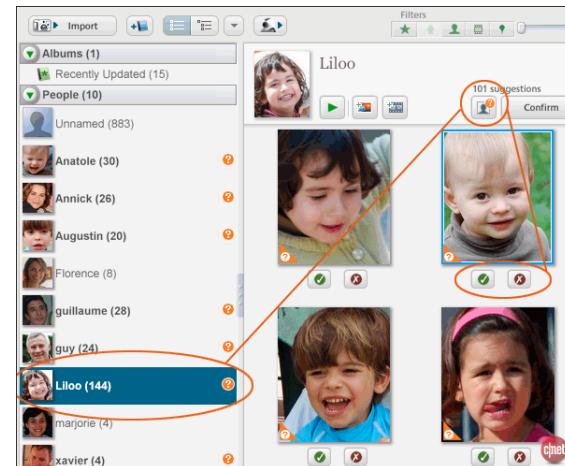
Principe de l'ACP : Applications

► Analyses explicatives ou prédictives:

- Réduction du nombre de variables explicatives ($X_{(j)}$) avant modélisation
- Obtention de nouvelles variables explicatives ($CP_{(j)}$) non corrélées

► Imagerie :

- Compression d'image
- Reconnaissance faciale

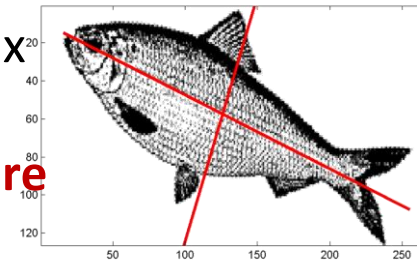


Principe de l'ACP : Situation

- ▶ p variables quantitatives ont été mesurées sur n individus :

	X_1	X_2	...	X_j	...	X_p
ind 1	x_{11}	x_{11}	...	x_{1j}	...	x_{1p}
ind 2	x_{12}	x_{22}	...	x_{2j}	...	x_{2p}
...
ind i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{ip}
...
ind n	x_{n1}	x_{n2}	...	x_{nj}	...	x_{np}

- ▶ **Question** : Peut-on « simplifier », « concentrer » ou « compresser » l'essentiel de l'information contenue dans ce tableau?
- ▶ **Exemple** : cas d'une image, les pixels sont représentés dans un plan et considérés comme une variable aléatoire à deux dimensions.
 - ▶ L'ACP va déterminer les deux axes qui expliquent le mieux la **dispersion** de l'objet. Elle va aussi les ordonner par **quantité d'information**, le second axe étant **perpendiculaire au premier**.

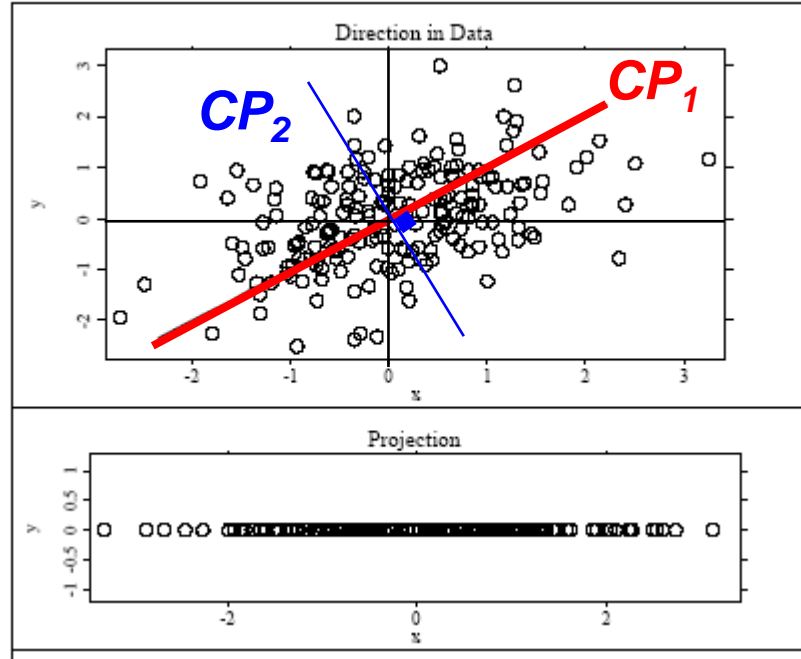


Principe de l'ACP : Objectifs

- ▶ Résumer le tableau de façon à identifier les variables ou combinaisons de variables selon lesquelles les n individus se différencient le plus
 - Transformer p variables quantitatives initiales inter-corrélées en p nouvelles variables « **Composantes Principales** » (CP)
 - Ces **Composantes Principales** sont non corrélées, et sont ordonnées par la part de l'information totale que chacun d'eux retient.
- ▶ Examiner la position des n individus le long de ces « composantes principales »
 - Typologie des individus
- ▶ Etudier les relations des p variables le long de ces « composantes principales »
 - Typologie des variables

Principe de l'ACP : Notion de projection

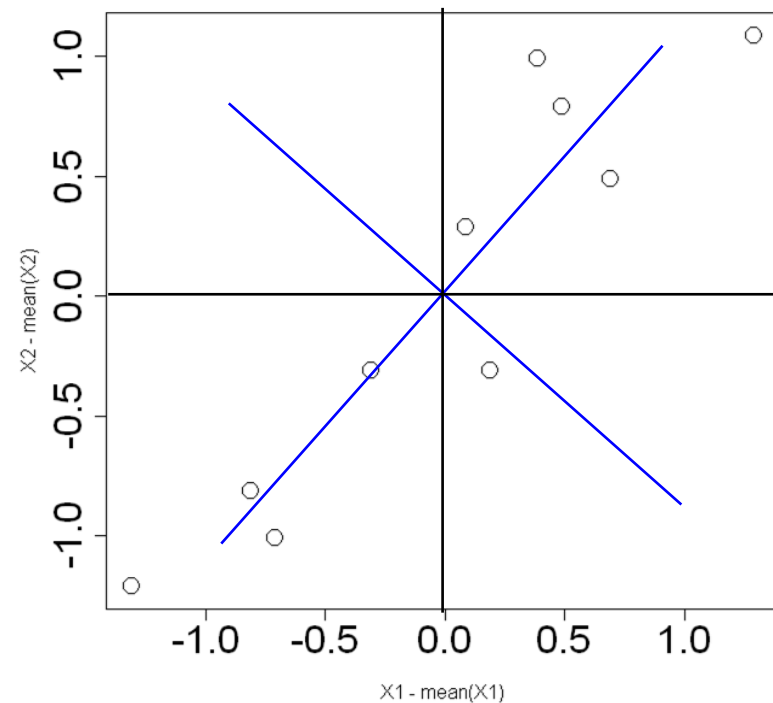
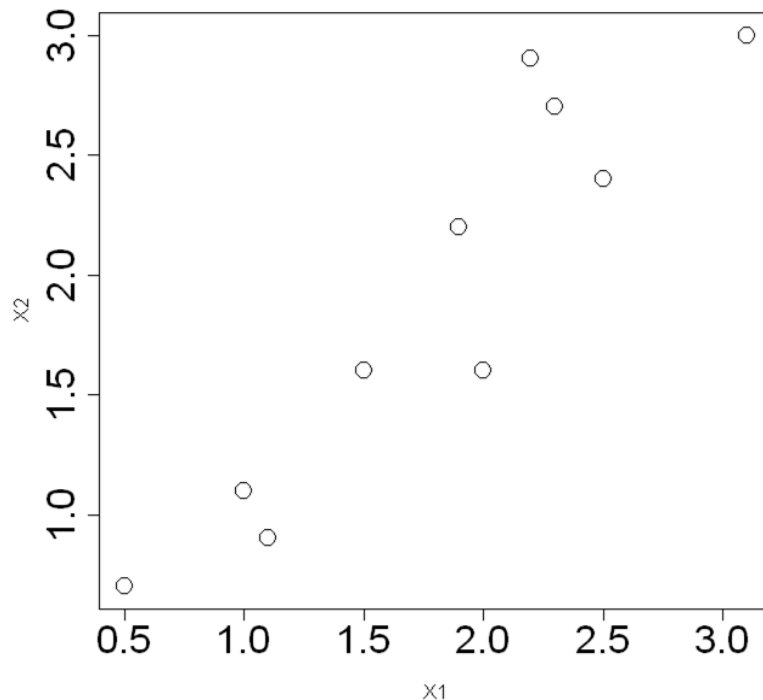
- ▶ Cas simple d'un tableau à 2 variables ($p = 2, X_{(1)}$ et $X_{(2)}$) et n individus :
 - On pourrait résumer ce tableau par une « composante principale » CP_1
 - **Projection orthogonale** des individus le long de CP_1



- ▶ Orthogonalité de CP_2 par rapport à CP_1

Principe de l'ACP : Notion de projection

- ▶ Il est recommandé de toujours **centrer** les valeurs associées à chaque variable $X_{(j)}$ pour éviter le problème de la translation au cours du changement de repère
 - Utilisez la matrice de données $(X - \bar{X})$ au lieu de la matrice X



- ▶ Comment allez-vous positionner CP_1 puis CP_2 ?

ACP et la covariance

- ▶ **ACP : calculs des CPs basées sur la covariance entre variables**
- ▶ Qu'est ce que la covariance entre deux variables $X_{(1)}$ et $X_{(2)}$?
 - Indique si à un écart positif de $X_{(1)}$ pour un individu i par rapport à la moyenne sur $X_{(1)}$ correspond un écart positif ou négatif de $X_{(2)}$ pour ce même individu i par rapport à la moyenne sur $X_{(2)}$

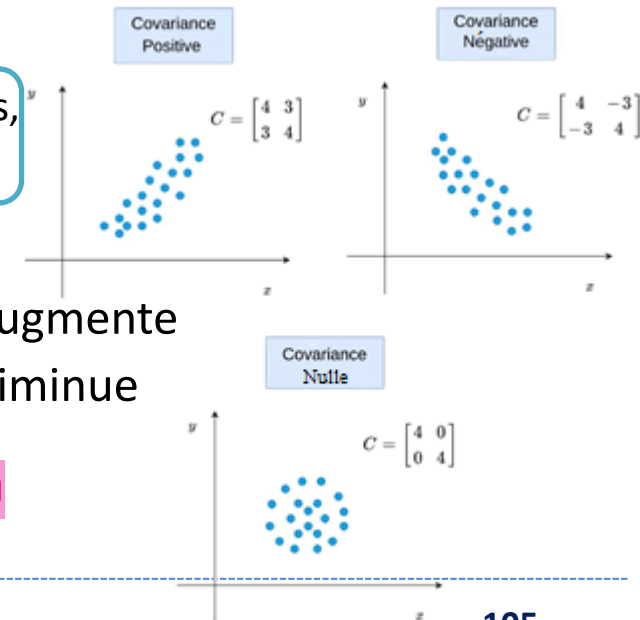
$$\begin{aligned} \text{Cov}(X_{(1)}, X_{(2)}) &= \frac{1}{n} < (X - \bar{X})_{(1)}, (X - \bar{X})_{(2)} > = \frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{X}_{(1)})(x_{i2} - \bar{X}_{(2)}) \\ &= \left(\frac{1}{n} \sum_{i=1}^n x_{i1} x_{i2} \right) - \bar{X}_{(1)} \bar{X}_{(2)} \end{aligned}$$

$$\text{Cov}(X_{(1)}, X_{(1)}) = \text{Var}(X_{(i)})$$

Si $X_{(1)}$ et $X_{(2)}$ sont centrées,
alors $\bar{X}_{(1)} = \bar{X}_{(2)} = 0$

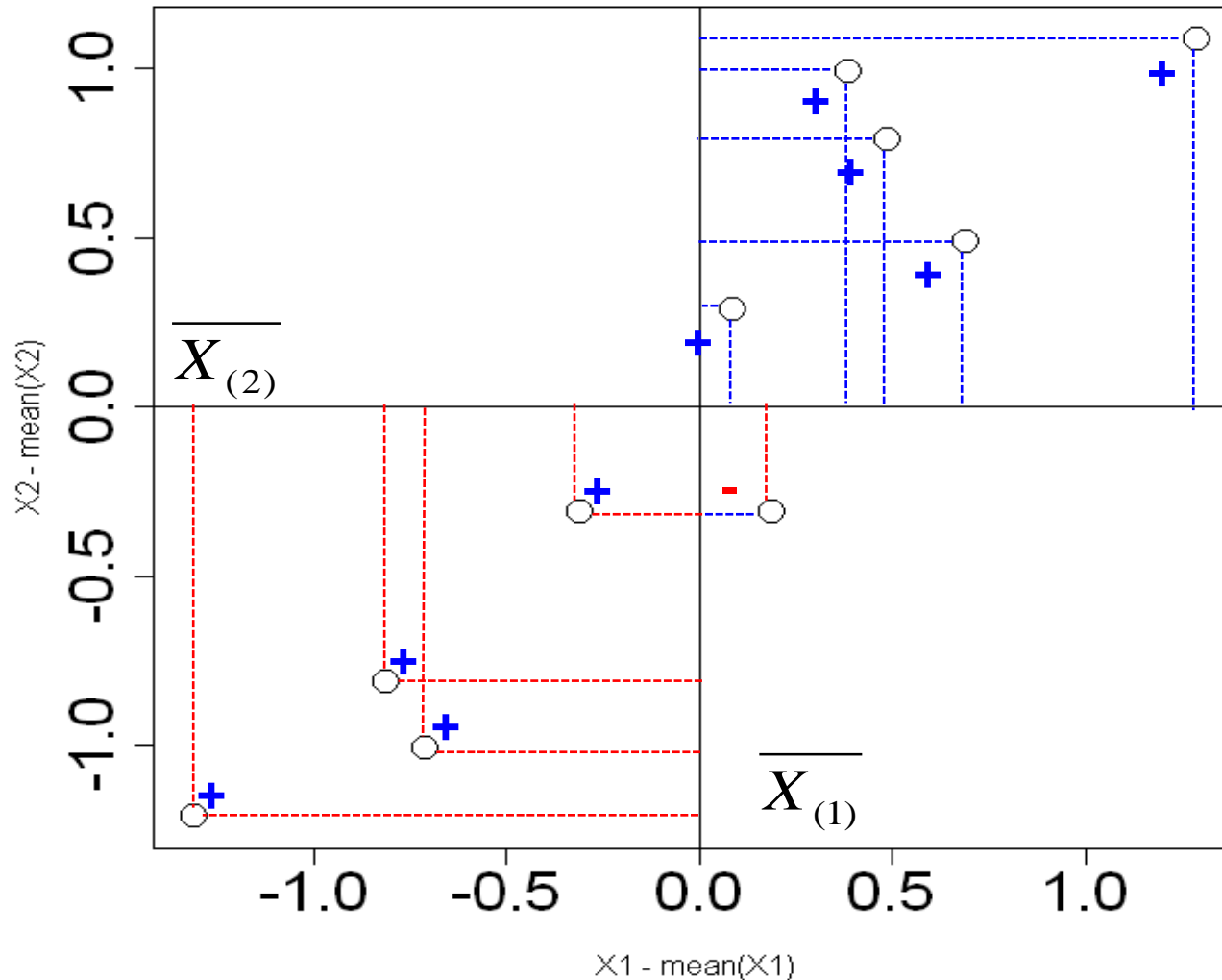
- ▶ C'est le signe de la covariance qui importe :
 - $\text{Cov}(X_{(1)}, X_{(2)}) > 0$: $X_{(1)}$ augmente quand $X_{(2)}$ augmente
 - $\text{Cov}(X_{(1)}, X_{(2)}) < 0$: $X_{(1)}$ augmente quand $X_{(2)}$ diminue

- ▶ La covariance indique **la direction de la relation linéaire** entre les variables.



ACP et la covariance

- Visualisation graphique de la covariance sur variables centrées :



$$\text{Cov}(X_{(1)}, X_{(2)}) = 0,5539 > 0$$

ACP et la covariance

- Pour $p > 2$, on calcule la covariance pour toutes les paires de variables possibles : Matrice $C(X)$ de covariances

$$C(X) = \begin{pmatrix} \text{var}(X_{(1)}) & \text{cov}(X_{(1)}, X_{(2)}) & \dots & \text{cov}(X_{(1)}, X_{(j)}) & \dots & \text{cov}(X_{(1)}, X_{(p)}) \\ \text{cov}(X_{(2)}, X_{(1)}) & \text{var}(X_{(2)}) & \dots & \text{cov}(X_{(2)}, X_{(j)}) & \dots & \text{cov}(X_{(2)}, X_{(p)}) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \text{cov}(X_{(j)}, X_{(1)}) & \text{cov}(X_{(j)}, X_{(2)}) & \dots & \text{var}(X_{(j)}) & \dots & \text{cov}(X_{(j)}, X_{(p)}) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \text{cov}(X_{(p)}, X_{(1)}) & \text{cov}(X_{(p)}, X_{(2)}) & \dots & \text{cov}(X_{(p)}, X_{(j)}) & \dots & \text{var}(X_{(p)}) \end{pmatrix}$$

- **Propriétés :**

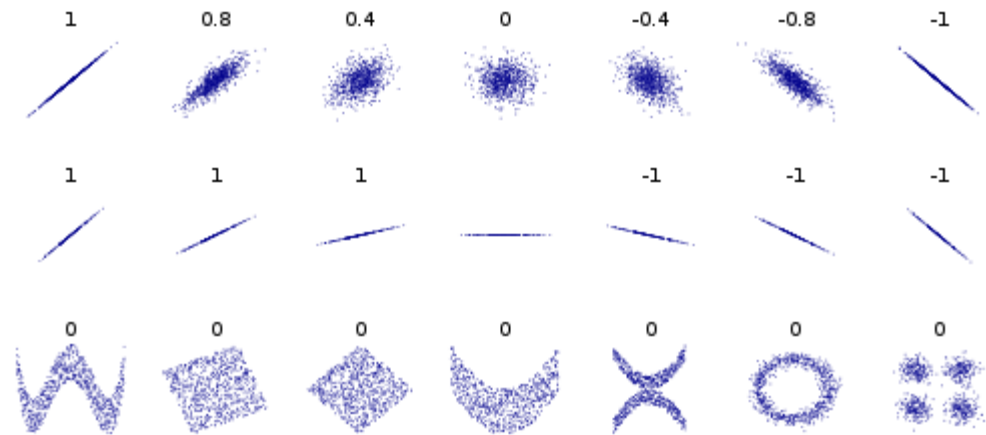
- $C(X)$ est une matrice carré de taille $p \times p$
- $C(X)$ est une matrice symétrique

ACP et la corrélation

- **Corrélation = covariance « standardisée » : réduction**
- Comprise entre -1 et 1, la corrélation mesure à la fois l'**intensité** et la **direction** de la **liaison linéaire** entre deux variables $X_{(1)}$ et $X_{(2)}$

$$r(X_{(1)}, X_{(2)}) = \frac{\text{Cov}(X_{(1)}, X_{(2)})}{\sigma(X_{(1)}) \sigma(X_{(2)})}$$

$$\begin{aligned} r(X_{(1)}, X_{(1)}) &= \frac{\text{Cov}(X_{(1)}, X_{(1)})}{\sigma(X_{(1)}) \sigma(X_{(1)})} \\ &= \frac{\text{Var}(X_{(1)})}{\text{Var}(X_{(1)})} = 1 \end{aligned}$$



ACP et la corrélation

- ▶ Si l'ACP (non-normée) est basée sur la matrice de covariances, l'**ACP normée** est basée sur la matrice de corrélations :

$$R(X) = \begin{pmatrix} 1 & r(X_{(1)}, X_{(2)}) & \dots & r(X_{(1)}, X_{(j)}) & \dots & r(X_{(1)}, X_{(p)}) \\ r(X_{(2)}, X_{(1)}) & 1 & \dots & r(X_{(2)}, X_{(j)}) & \dots & r(X_{(2)}, X_{(p)}) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r(X_{(j)}, X_{(1)}) & r(X_{(j)}, X_{(2)}) & \dots & 1 & \dots & r(X_{(j)}, X_{(p)}) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r(X_{(p)}, X_{(1)}) & r(X_{(p)}, X_{(2)}) & \dots & r(X_{(p)}, X_{(j)}) & \dots & 1 \end{pmatrix}$$

- ▶ **Propriétés :**

- $R(X)$ est une matrice carré de taille $p \times p$
- $R(X)$ est une matrice symétrique
- $R(X)$ possède une diagonale de 1

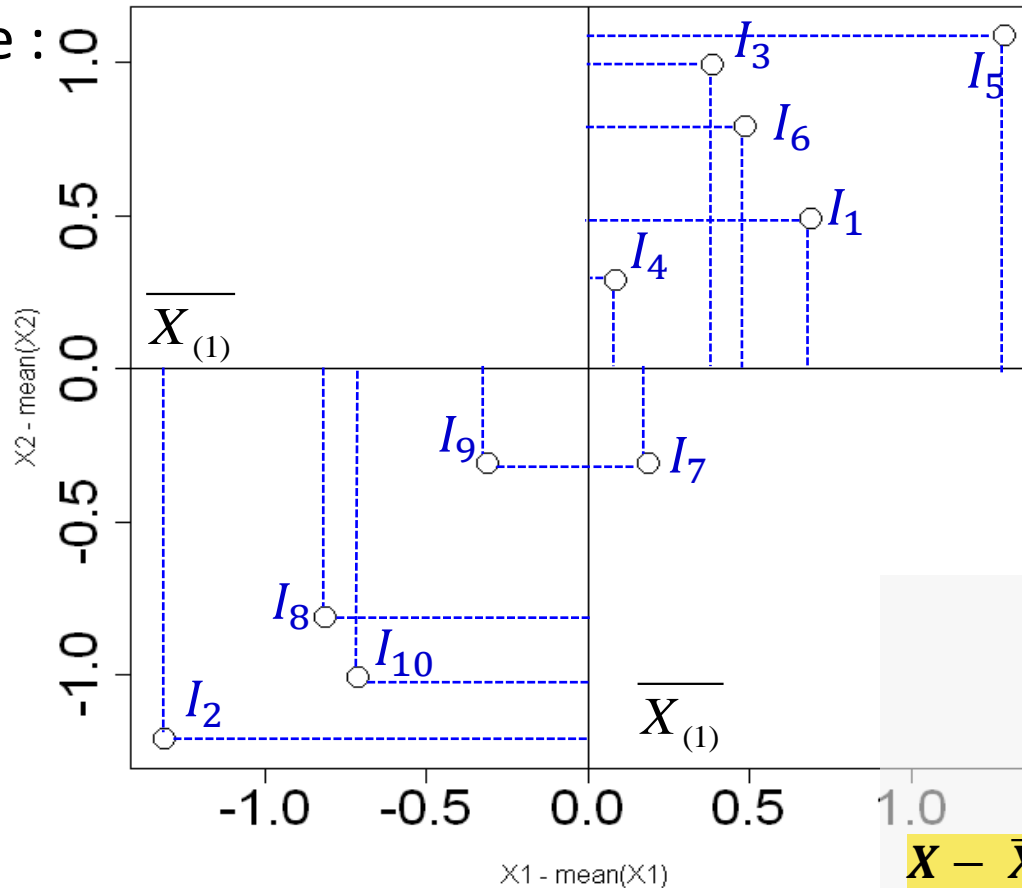
ACP et la corrélation

- ▶ S'il est recommandé de toujours « **centrer** » ses données en **ACP**, la question de les « **réduire** » (**ACP normée**) dépend de vos données :
 - Si vos données sont toutes dans la même unité de mesure et varient dans des gammes de valeurs identiques : l'ACP non-normée suffit
 - Si vos données sont dans des unités de mesure différentes et varient dans des gammes de valeurs différentes : l'ACP normée est recommandée

Procédure de l'ACP à partir d'un exemple

► Soit l'exemple :

	$X_{(1)}$	$X_{(2)}$
I_1	2,5	2,4
	0,5	0,7
	2,2	2,9
	1,9	2,2
	3,1	3
	2,3	2,7
	2	1,6
	1	1,1
	1,5	1,6
I_{10}	1,1	0,9



$X - \bar{X} =$

0,69	0,49
-1,31	-1,21
0,39	0,99
0,09	0,29
1,29	1,09
0,49	0,79
0,19	-0,31
-0,81	-0,81
-0,31	-0,31
-0,71	-1,01

► $\bar{X}_{(1)} = 1.81$
 ► $\bar{X}_{(2)} = 1.91$

Procédure de l'ACP à partir d'un exemple

▶ ACP non-normée → matrice de covariances

$$C(X) = \frac{1}{n} (X - \bar{X})^t (X - \bar{X})$$

$$C(X) = \begin{pmatrix} 0.5549 & 0.5539 \\ 0.5539 & 0.6449 \end{pmatrix}$$

$$\begin{aligned} &\text{> } Var(X_{(1)}) = 0,5549 \\ &\text{> } Var(X_{(2)}) = 0,6449 \\ &\text{> } Cov(X_{(1)}, X_{(2)}) = 0.5539 \end{aligned}$$

- ▶ Utilisation de la matrice de covariances pour changer de repère:
 - Calcul du vecteur directeur de CP_1
 - Calcul du vecteur directeur de CP_2

Procédure de l'ACP à partir d'un exemple

- ▶ Une histoire d'algèbre linéaire et de calculs matriciels :
 - Détermination des p valeurs propres λ_j
 - Détermination des p vecteurs propres V_j
- ▶ Soit la matrice de covariances C de taille $p \times p$, elle admet p valeurs propres et p vecteurs propres associés, tels que :

$$CV_j = \lambda_j V_j$$

- ▶ Dans notre exemple à deux variables, C admet 2 valeurs propres et 2 vecteurs propres tels que soit vérifié les égalités suivantes :

$$\begin{pmatrix} 0.5549 & 0.5539 \\ 0.5539 & 0.6449 \end{pmatrix} \begin{pmatrix} v_{1,1} \\ v_{1,2} \end{pmatrix} = \lambda_1 \begin{pmatrix} v_{1,1} \\ v_{1,2} \end{pmatrix}$$

$$\begin{pmatrix} 0.5549 & 0.5539 \\ 0.5539 & 0.6449 \end{pmatrix} \begin{pmatrix} v_{2,1} \\ v_{2,2} \end{pmatrix} = \lambda_2 \begin{pmatrix} v_{2,1} \\ v_{2,2} \end{pmatrix}$$

Procédure de l'ACP à partir d'un exemple

► Détermination des 2 valeurs propres λ_1 et λ_2 :

➤ Calcul du déterminant de $C - \lambda I$

➤ Résolution de l'équation $\det(C - \lambda I) = 0$

$$C - \lambda I = \begin{pmatrix} 0.5549 & 0.5539 \\ 0.5539 & 0.6449 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$C - \lambda I = \begin{pmatrix} 0.5549 - \lambda & 0.5539 \\ 0.5539 & 0.6449 - \lambda \end{pmatrix}$$

$$\det(C - \lambda I) = (0.5549 - \lambda)(0.6449 - \lambda) - 0.5539^2$$

$$\lambda^2 - 1.1998\lambda + 0.0510498 = 0$$

$$\Delta = b^2 - 4ac = 1.23532084$$

$$\lambda_1 = \frac{-b + \sqrt{\Delta}}{2a} = 1.155625 \quad \lambda_2 = \frac{-b - \sqrt{\Delta}}{2a} = 0.044175$$

Procédure de l'ACP à partir d'un exemple

► Détermination des 2 valeurs propres λ_1 et λ_2 ...

- Chaque **valeur propre** représente la **variance des données** autour d'un nouvel axe **CP** ou « **composante principale** » qui est une combinaison linéaire des variables de départ

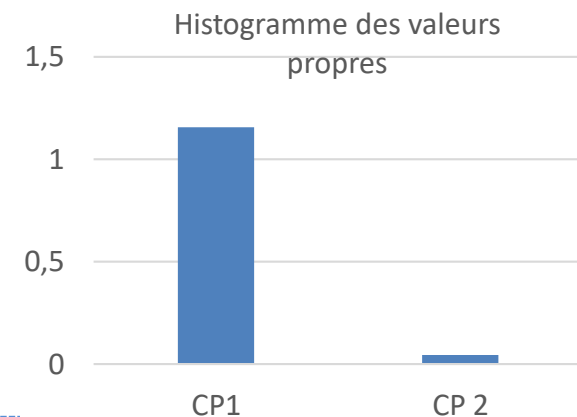
$$\lambda_1 + \lambda_2 = \text{Var}(X_{(1)}) + \text{Var}(X_{(2)})$$

$$1.155625 + 0.044175 = 0.6449 + 0.5549 = 1.1998$$

- La première composante principale ou CP_1 associée à λ_1 porte $\approx 96\%$ de la variance totale
- La deuxième composante principale ou CP_2 associée à λ_2 porte $\approx 4\%$ de la variance totale

	Valeur propre	Pourcentage	Pourcentage cumulé
CP_1	1,155625	96,3181	96,3181
CP_2	0,044175	3,6819	100

➡ A partir d'une seule dimension (CP_1), il est possible ici de résumer $\approx 96\%$ de l'information de départ contenue dans deux dimensions ($X_{(1)}, X_{(2)}$)



Procédure de l'ACP à partir d'un exemple

► Détermination des 2 vecteurs propres V_1 et V_2 :

► Résolution des 2 systèmes d'équations $CV_j - \lambda_j V_j = 0$

$$\lambda_1 \begin{pmatrix} 0.5549 & 0.5539 \\ 0.5539 & 0.6449 \end{pmatrix} \begin{pmatrix} v_{1,1} \\ v_{1,2} \end{pmatrix} = 1.155625 \times \begin{pmatrix} v_{1,1} \\ v_{1,2} \end{pmatrix}$$
$$\begin{cases} 0.5549 \times v_{1,1} + 0.5539 \times v_{1,2} - 1.155625 \times v_{1,1} = 0 \\ 0.5539 \times v_{1,1} + 0.6449 \times v_{1,2} - 1.155625 \times v_{1,2} = 0 \end{cases}$$

$$\lambda_2 \begin{pmatrix} 0.5549 & 0.5539 \\ 0.5539 & 0.6449 \end{pmatrix} \begin{pmatrix} v_{2,1} \\ v_{2,2} \end{pmatrix} = 0.044175 \times \begin{pmatrix} v_{2,1} \\ v_{2,2} \end{pmatrix}$$
$$\begin{cases} 0.5549 \times v_{2,1} + 0.5539 \times v_{2,2} - 0.044175 \times v_{2,1} = 0 \\ 0.5539 \times v_{2,1} + 0.6449 \times v_{2,2} - 0.044175 \times v_{2,2} = 0 \end{cases}$$

► Une solution possible : les vecteurs unitaires :

$$v_{11}^2 + v_{12}^2 = 1 \quad \text{Et} \quad v_{21}^2 + v_{22}^2 = 1$$

Procédure de l'ACP à partir d'un exemple

► Détermination des 2 vecteurs propres V_1 et V_2 ...

► Prenons cette équation (1^{ère} équation du 1^{er} système)

$$0.5549 \times v_{1,1} + 0.5539 \times v_{1,2} - 1.155625 \times v_{1,1} = 0$$

$$\longrightarrow 0.5539 \times v_{1,2} - 0.600725 \times v_{1,1} = 0 \quad \longrightarrow v_{1,1} = 0.9220525 \times v_{1,2}$$

et on a : $v_{1,1}^2 + v_{1,2}^2 = 1$ (car vecteur unitaire)

$$\longrightarrow (0.9220525 \times v_{1,2})^2 + v_{1,2}^2 = 1 \quad \longrightarrow v_{1,2}^2 = \frac{1}{1.8501808}$$

$$\longrightarrow \text{Une solution possible est : } v_{1,2} = 0.7351787$$

$$\text{Et on a : } v_{1,1} = 0.9220525 \times v_{1,2}$$

$$\longrightarrow v_{1,1} = 0.6778734$$

Procédure de l'ACP à partir d'un exemple

► Détermination des 2 vecteurs propres V_1 et V_2 ...

► Prenons cette équation (1^{ère} équation du 2^{ème} système)

$$0.5549 \times v_{2,1} + 0.5539 \times v_{2,2} - 0.044175 \times v_{2,1} = 0$$

$$\longrightarrow 0.5539 \times v_{2,2} + 0.510725 \times v_{2,1} = 0 \quad \longrightarrow v_{2,1} = -1.0845367 \times v_{2,2}$$

et on a : $v_{2,1}^2 + v_{2,2}^2 = 1$ (car vecteur unitaire)

$$\longrightarrow (-1.0845367 \times v_{2,2})^2 + v_{2,2}^2 = 1 \quad \longrightarrow v_{2,2}^2 = \frac{1}{2.1762199}$$

$$\longrightarrow \text{Une solution possible est : } v_{2,2} = 0.6778734$$

$$\text{Et on a : } v_{2,1} = -1.0845367 \times v_{2,2}$$

$$\longrightarrow v_{2,1} = -0.7351786$$

Procédure de l'ACP à partir d'un exemple

- Une solution possible (cf. sous la contrainte que V_1 et V_2 soient tout 2 des **vecteurs unitaires de norme égale à 1**)

$$V_1 = \begin{pmatrix} 0.6778734 \\ 0.7351787 \end{pmatrix} \quad \text{NB: } v_{1,1}^2 + v_{1,2}^2 = 1$$

$$V_2 = \begin{pmatrix} -0.7351786 \\ 0.6778734 \end{pmatrix} \quad \text{NB: } v_{2,1}^2 + v_{2,2}^2 = 1$$

CP_1

CP_2

$$Q = \begin{matrix} X_{(1)} \\ X_{(2)} \end{matrix} \begin{pmatrix} 0.6778734 & -0.7351786 \\ 0.7351787 & 0.6778734 \end{pmatrix} \quad \text{Et } D = C(Y) = \begin{pmatrix} 1.155625 & 0 \\ 0 & 0.044175 \end{pmatrix}$$

$$\rightarrow C(X) = QDQ^t \quad \text{Et} \quad Y = XQ$$

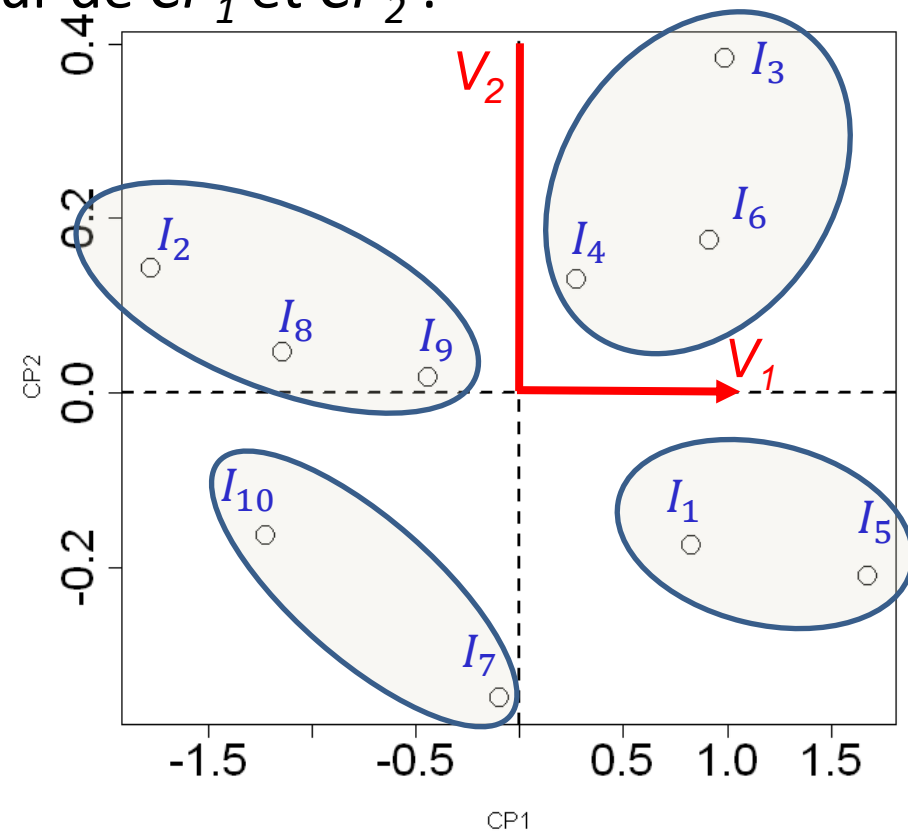
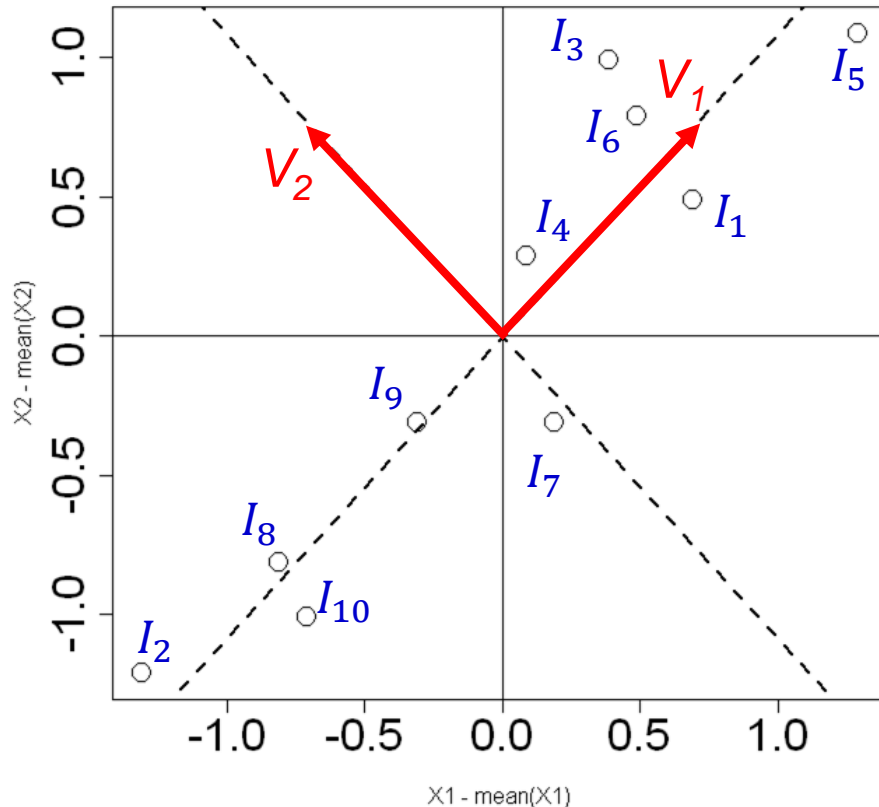
Procédure de l'ACP à partir d'un exemple

$$Y = (X - \bar{X}) Q = \begin{matrix} & \begin{matrix} X_{(1)} & X_{(2)} \end{matrix} \\ \begin{matrix} I_1 \\ \\ \\ \\ \\ \\ \\ \\ I_{10} \end{matrix} & \begin{bmatrix} 0,69 & 0,49 \\ -1,31 & -1,21 \\ 0,39 & 0,99 \\ 0,09 & 0,29 \\ 1,29 & 1,09 \\ 0,49 & 0,79 \\ 0,19 & -0,31 \\ -0,81 & -0,81 \\ -0,31 & -0,31 \\ -0,71 & -1,01 \end{bmatrix} \end{matrix} \times \begin{matrix} \begin{matrix} CP_1 & CP_2 \end{matrix} \\ \begin{matrix} X_{(1)} \\ X_{(2)} \end{matrix} \end{matrix} \begin{bmatrix} 0,6778734 & -0,735186 \\ 0,7351787 & 0,6778734 \end{bmatrix} = \begin{matrix} \begin{matrix} I_1 \\ \\ \\ \\ \\ \\ \\ \\ I_{10} \end{matrix} & \begin{bmatrix} 0,828 & -0,175 \\ -1,778 & 0,143 \\ 0,992 & 0,384 \\ 0,274 & 0,13 \\ 1,676 & -0,209 \\ 0,913 & 0,175 \\ -0,099 & -0,35 \\ -1,145 & 0,046 \\ -0,438 & 0,018 \\ -1,224 & -0,163 \end{bmatrix} \end{matrix}$$

Y : matrice des **scores** des individus : matrice des valeurs des composantes principales sur les individus

Procédure de l'ACP à partir d'un exemple

- V_1 et V_2 sont les vecteurs directeur de CP_1 et CP_2 :



- CP_1 porte 96% de l'inertie totale du nuage de point
- NB : $r(X_1, X_2) = 0.93$ mais $r(CP_1, CP_2) = 0$

Procédure de l'ACP à partir d'un exemple

- ▶ L'information (variance) portée par CP_1 est tellement importante que l'on peut se passer de CP_2 :
 - Cela revient à compresser l'information originale portée par deux dimensions sur une seule dimension avec une perte ici de 4% de l'information d'origine
 - Par analogie, une fois que l'on a vu le chameau de profil, le voir de face n'apporte pas beaucoup plus d'information...
- ▶ **N.B.** L'ACP non-normée est une application rare et en général, on travail avec la matrice des corrélations (ACP normée)

Procédure de l'ACP à partir d'un exemple

► Cas de l'ACP normée sur le même jeu de données :

$$R(X) = \frac{1}{n} Z^t Z = \begin{pmatrix} 1 & 0.93 \\ 0.93 & 1 \end{pmatrix}$$

$$R - \lambda I = \begin{pmatrix} 1 - \lambda & 0.93 \\ 0.93 & 1 - \lambda \end{pmatrix}$$

$$\det(R - \lambda I) = (1 - \lambda)(1 - \lambda) - 0.93^2$$

$$\lambda^2 - 2\lambda + 0.1351 = 0$$

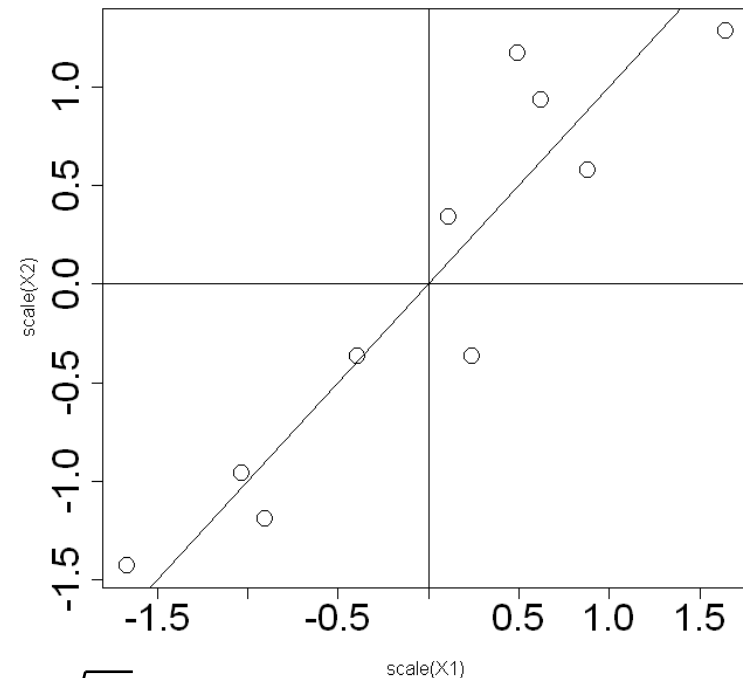
$$\Delta = b^2 - 4ac$$

$$\lambda_1 = \frac{-b + \sqrt{\Delta}}{2a} = 1.93$$

$$\lambda_1 = 1 + r$$

$$\lambda_2 = \frac{-b - \sqrt{\Delta}}{2a} = 0.07$$

$$\lambda_2 = 1 - r$$



Procédure de l'ACP à partir d'un exemple

► Détermination des 2 vecteurs propres V_1 et V_2 :

► Résolution des 2 systèmes d'équations $CV_j - \lambda_j V_j = 0$

$$\lambda_1 \quad \begin{pmatrix} 1 & 0.93 \\ 0.93 & 1 \end{pmatrix} \begin{pmatrix} v_{1,1} \\ v_{1,2} \end{pmatrix} = 1.93 \times \begin{pmatrix} v_{1,1} \\ v_{1,2} \end{pmatrix}$$

$$\begin{cases} v_{1,1} + 0.93 \times v_{1,2} - 1.93 \times v_{1,1} = 0 \\ 0.93 \times v_{1,1} + v_{1,2} - 1.93 \times v_{1,2} = 0 \end{cases}$$

$$\lambda_2 \quad \begin{pmatrix} 1 & 0.93 \\ 0.93 & 1 \end{pmatrix} \begin{pmatrix} v_{2,1} \\ v_{2,2} \end{pmatrix} = 0.07 \times \begin{pmatrix} v_{2,1} \\ v_{2,2} \end{pmatrix}$$

$$\begin{cases} v_{2,1} + 0.93 \times v_{2,2} - 0.07 \times v_{2,1} = 0 \\ 0.93 \times v_{2,1} + v_{2,2} - 0.07 \times v_{2,2} = 0 \end{cases}$$

► Une solution possible : les vecteurs unitaires :

$$v_{11}^2 + v_{12}^2 = 1 \quad \text{Et} \quad v_{21}^2 + v_{22}^2 = 1$$

Procédure de l'ACP à partir d'un exemple

► Détermination des 2 vecteurs propres V_1 et V_2 ...

➤ Prenons cette équation (1^{ère} équation du 1^{er} système)

$$v_{1,1} + 0.93 \times v_{1,2} - 1.93 \times v_{1,1} = 0$$

$$\longrightarrow 0.93 \times v_{1,2} - 0.93 \times v_{1,1} = 0 \quad \longrightarrow \quad v_{1,1} = v_{1,2}$$

et on a : $v_{1,1}^2 + v_{1,2}^2 = 1$ (car vecteur unitaire)

$$\longrightarrow v_{1,2}^2 + v_{1,2}^2 = 1 \quad \longrightarrow \quad v_{1,2}^2 = \frac{1}{2}$$

$$\longrightarrow \text{Une solution possible est:} \quad v_{1,2} = \frac{1}{\sqrt{2}}$$

$$\longrightarrow \quad v_{1,1} = \frac{1}{\sqrt{2}}$$

Procédure de l'ACP à partir d'un exemple

► Détermination des 2 vecteurs propres V_1 et V_2 ...

► Prenons cette équation (1^{ère} équation du 2^{ème} système)

$$v_{2,1} + 0.93 \times v_{2,2} - 0.07 \times v_{2,1} = 0$$

$$\longrightarrow 0.93 \times v_{2,2} + 0.93 \times v_{2,1} = 0 \qquad \longrightarrow V_{2,1} = -V_{2,2}$$

et on a : $v_{2,1}^2 + v_{2,2}^2 = 1$ (car vecteur unitaire)

$$\longrightarrow (-v_{2,2})^2 + v_{2,2}^2 = 1 \qquad \longrightarrow v_{2,2}^2 = \frac{1}{2}$$

$$\longrightarrow \text{Une solution possible est: } v_{2,2} = \frac{1}{\sqrt{2}}$$

$$\longrightarrow v_{1,2} = -\frac{1}{\sqrt{2}}$$

Procédure de l'ACP à partir d'un exemple

- Une solution possible (cf. sous la contrainte que V_1 et V_2 soient tout 2 des **vecteurs unitaires de norme 1**)

$$V_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad V_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad \rightarrow \quad Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \quad D = \begin{pmatrix} 1+r & 0 \\ 0 & 1-r \end{pmatrix}$$

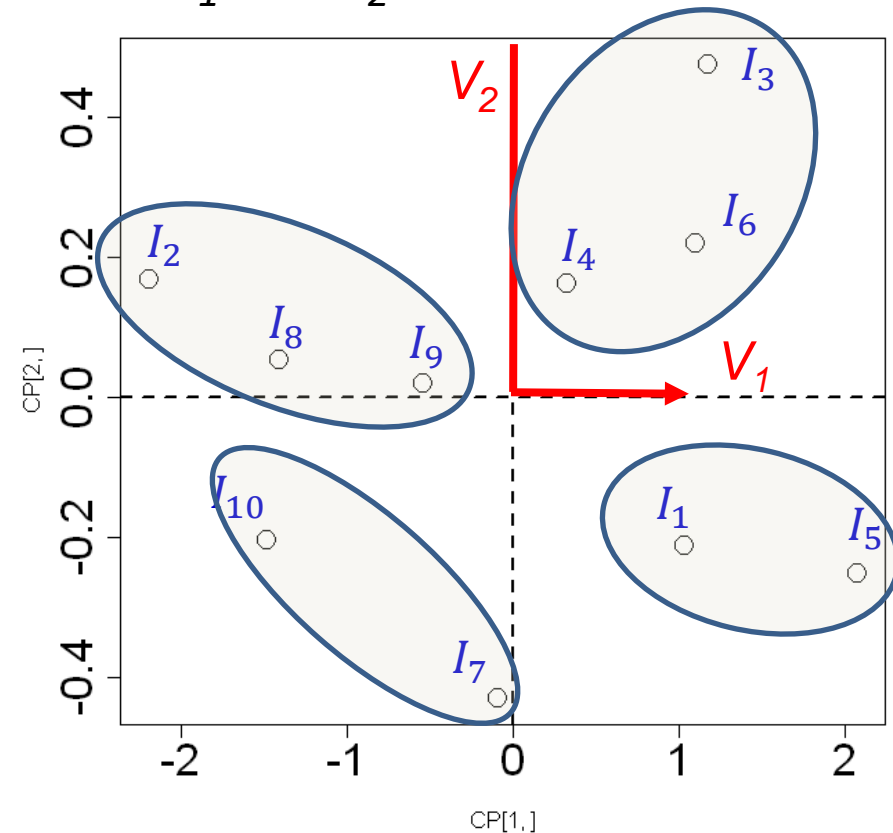
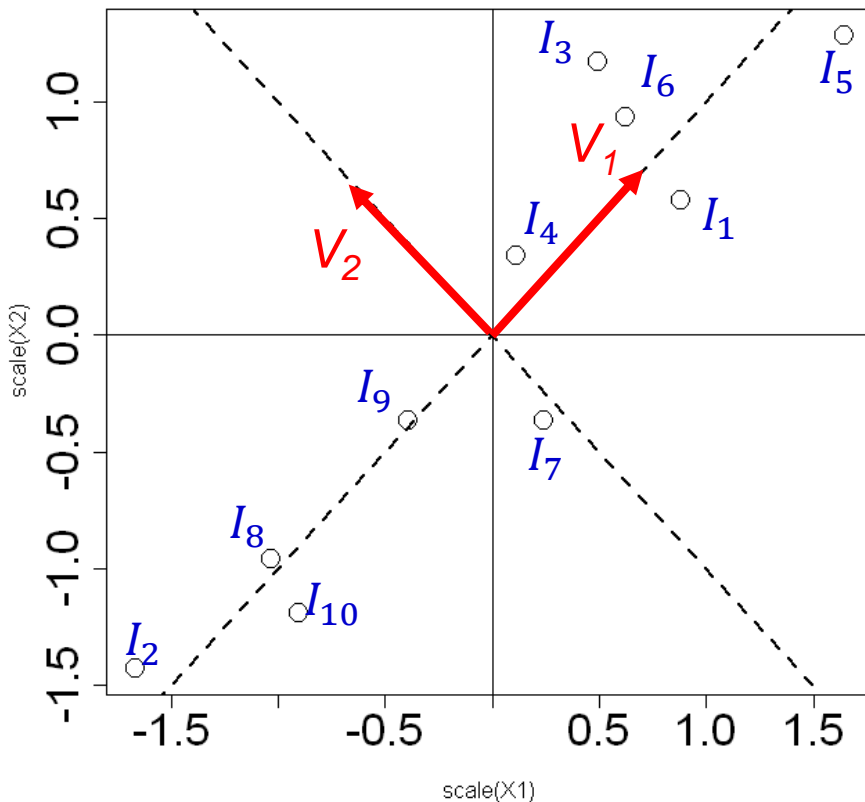
$v_{1,1}^2 + v_{1,2}^2 = 1$ $v_{2,1}^2 + v_{2,2}^2 = 1$

Y: matrice des **scores** des individus

$$Y = Z Q = \begin{bmatrix} 0,926 & 0,61 \\ -1,759 & -1,506 \\ 0,524 & 1,233 \\ 0,121 & 0,361 \\ 1,732 & 1,357 \\ 0,658 & 0,984 \\ 0,255 & -0,386 \\ -1,087 & -1,009 \\ -0,416 & -0,386 \\ -0,953 & -1,258 \end{bmatrix} \times \begin{bmatrix} 0,707 & -0,707 \\ 0,707 & 0,707 \end{bmatrix} = \begin{matrix} I_1 \\ I_{10} \end{matrix} \begin{bmatrix} 1,086 & -0,223 \\ -2,309 & 0,178 \\ 1,242 & 0,501 \\ 0,341 & 0,17 \\ 2,184 & -0,264 \\ 1,161 & 0,23 \\ -0,093 & -0,453 \\ -1,482 & 0,056 \\ -0,567 & 0,021 \\ -1,563 & -0,215 \end{bmatrix}$$

Procédure de l'ACP à partir d'un exemple

- V_1 et V_2 sont les vecteurs directeur de CP_1 et CP_2 :



Procédure de l'ACP

► Étape 1: Changement de repère

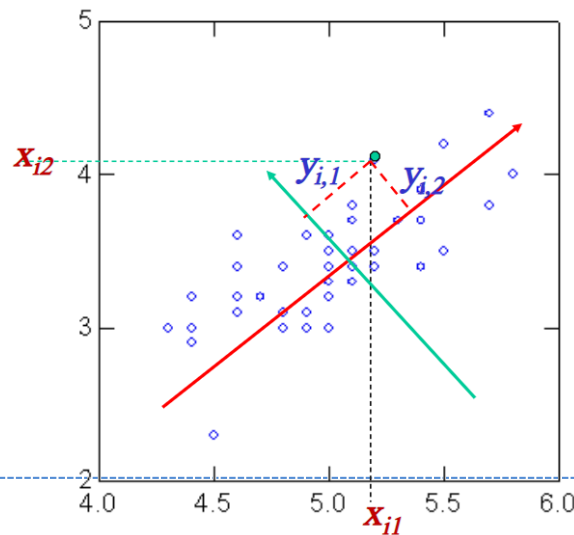
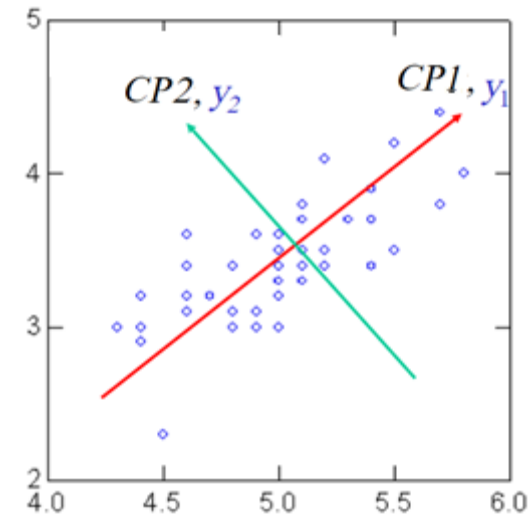
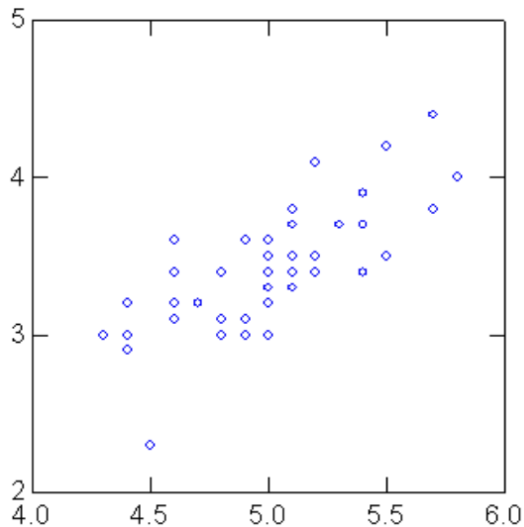
- Utiliser les **données centrées** (la matrice $(X - \bar{X})$ au lieu de la matrice X) ou les **données centrées-réduites** (la matrice Z au lieu de la matrice X) selon vos données
- **Ancien repère:**
 - chaque individu est représenté par un point de \mathbb{R}^p
 - chacun des axes de \mathbb{R}^p représente une des p variables
- **Nouveau repère:**
 - nouveaux axes (droites) passant par l'origine et p **vecteurs unitaires** et **deux à deux orthogonaux** (CP_1, \dots, CP_p)
 - chacun des axes représente une **nouvelle variable**, qui est **combinaison linéaire** des anciennes variables.
 - chaque individu est représenté sur ce nouveau repère, par sa **projection orthogonale** sur ces axes

$$Y = (X - \bar{X})Q \quad \text{ou} \quad Y = ZQ$$

Procédure de l'ACP

► Étape 1: Changement de repère ...

Plan d'origine



Procédure de l'ACP

► Étape 2: Choix du nouveau repère

- **But:** trouver le nouveau repère (CP_1, \dots, CP_p) tel que la quantité d'information expliquée par CP_1 soit maximale, puis celle expliquée par CP_2 , etc.
- **Mesure de la quantité d'information :**
 - Utiliser la **covariance** pour les *données centrées* ou la **corrélation** pour les *données centrées-réduites*
 - Plus la variance d'une variable est grande, plus les données de cette variable sont dispersées, et plus la quantité d'information apportée est importante
 - Quantité d'information = **variabilité totale** des données

ACP non-normé

$$\begin{aligned}\text{Tr}(C(X - \bar{X})) &= \text{Tr}(C(X)) = \sum_{j=1}^p \text{Var}(X_{(j)}) \\ &= \text{Tr}(C(Y))\end{aligned}$$

Ou

ACP normé

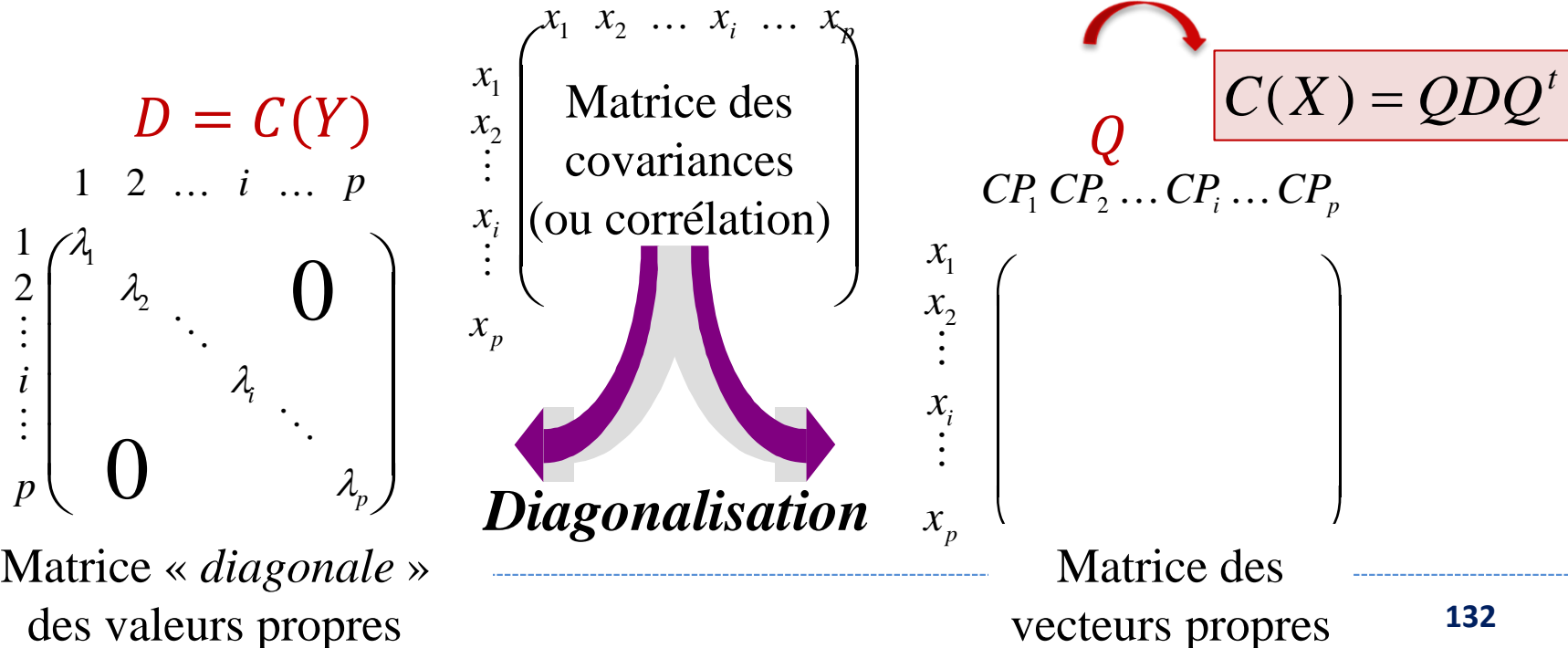
$$\begin{aligned}\text{Tr}(C(Z)) &= \sum_{j=1}^p \text{Var}(Z_{(j)}) = p \times 1 \\ &= \text{Tr}(C(Y))\end{aligned}$$

Procédure de l'ACP

► Étape 2: Choix du nouveau repère ...

► Choix du nouveau repère :

- Déterminer Q de sorte que la part de la variabilité totale expliquée par les données $Y_{(1)}$ de la nouvelle variable CP_1 soit maximale, puis celle expliquée par les données $Y_{(2)}$ de la nouvelle variable CP_2 , etc.
- En appliquant le **Théorème spectral** pour les matrices symétriques :



Procédure de l'ACP

► Étape 2: Choix du nouveau repère ...

➤ Choix du nouveau repère ...

- On a donc :
 - $Var(Y_{(j)}) = \lambda_j$
 - $Cov(Y_{(i)}, Y_{(j)}) = 0$ quand $i \neq j$
 - $Var(Y_{(1)}) \geq \dots \geq Var(Y_{(p)})$
- Les colonnes CP_1, \dots, CP_p de la matrice Q décrivent les nouvelles variables, appelées les **composantes principales**.
- Dans le cas des données centrées réduites (Matrice de données Z), la matrice de covariance est la matrice de corrélation $R(X)$

$$Cov(Z_{(i)}, Z_{(j)}) = Cov\left(\frac{X_{(i)} - \overline{X_{(i)}}}{\sigma(X_{(i)})}, \frac{X_{(j)} - \overline{X_{(j)}}}{\sigma(X_{(j)})}\right) = r(X_{(i)}, X_{(j)})$$

Procédure de l'ACP

► Étape 2: Choix du nouveau repère ...

➤ Projection des individus sur le nouveau repère:

- Y : matrice des **scores** des individus : matrice des valeurs des composantes principales sur les individus

$$Y = (X - \bar{X})Q \text{ ou } Y = Z Q$$

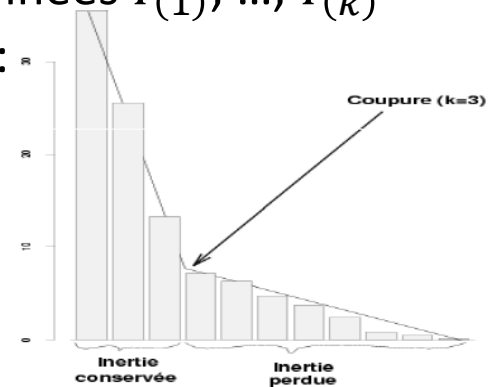
Procédure de l'ACP

► Étape 3: Interprétation des résultats

► Parts de la variabilité totale

- La part de la variabilité totale expliquée par les données $Y_{(1)}, \dots, Y_{(k)}$ des k premières nouvelles variables ($k \leq p$), est :

$$\frac{\text{Var}(Y_{(1)}) + \dots + \text{Var}(Y_{(k)})}{\text{Var}(Y_{(1)}) + \dots + \text{Var}(Y_{(p)})} = \frac{\lambda_1 + \dots + \lambda_k}{\lambda_1 + \dots + \lambda_p}$$



- Dans la pratique, on calcule cette quantité pour $k = 2$ ou 3 , en utilisant le **test du coude**
- En multipliant par 100, ceci donne **le pourcentage de la variabilité totale expliquée par les données** des 2 ou 3 premières nouvelles variables.
- Si ce pourcentage est raisonnable, on choisira de se restreindre aux 2 ou 3 premiers axes.

Procédure de l'ACP

► Étape 3: Interprétation des résultats ...

➤ Saturation des variables : Corrélation entre les données des anciennes et des nouvelles variables

- Etant donné que les nouvelles variables sont dans un sens “artificielles”, on souhaite comprendre la **corrélation** entre les données $(X - \bar{X})_{(j)}$ de la $j^{\text{ème}}$ ancienne variable et celle $Y_{(k)}$ de la $k^{\text{ème}}$ nouvelle variable.
- La matrice de covariance $C(X, Y)$ de $(X - \bar{X})$ et Y est donnée par :

$$C(X, Y) = QD$$

Ainsi:

$$\begin{aligned} \text{Cov}(X_{(j)}, Y_{(k)}) &= C(X, Y)_{jk} = q_{jk}\lambda_k \\ r(X_{(j)}, Y_{(k)}) &= \frac{\text{Cov}(X_{(j)}, Y_{(k)})}{\sigma(X_{(j)}) \sigma(Y_{(k)})} = \frac{q_{jk}\lambda_k}{\sqrt{c_{jj}\lambda_k}} = \frac{q_{jk}\sqrt{\lambda_k}}{\sqrt{c_{jj}}} \end{aligned}$$

Avec: $c_{jj} = \text{Var}(X_{(j)})$

Procédure de l'ACP

► Étape 3: Interprétation des résultats ...

➤ Saturation des variables ...

- Dans le cas des données centrées réduites (Matrice de données Z), la matrice de covariance est la matrice de corrélation $R(X)$

- La corrélation entre $Z_{(j)}$ et $Y_{(k)}$ est :

$$r(Z_{(j)}, Y_{(k)}) = \sqrt{\lambda_k} q_{jk}$$

Car les coefficients diagonaux de la matrice de covariance (qui est la matrice de corrélation) sont égaux à 1

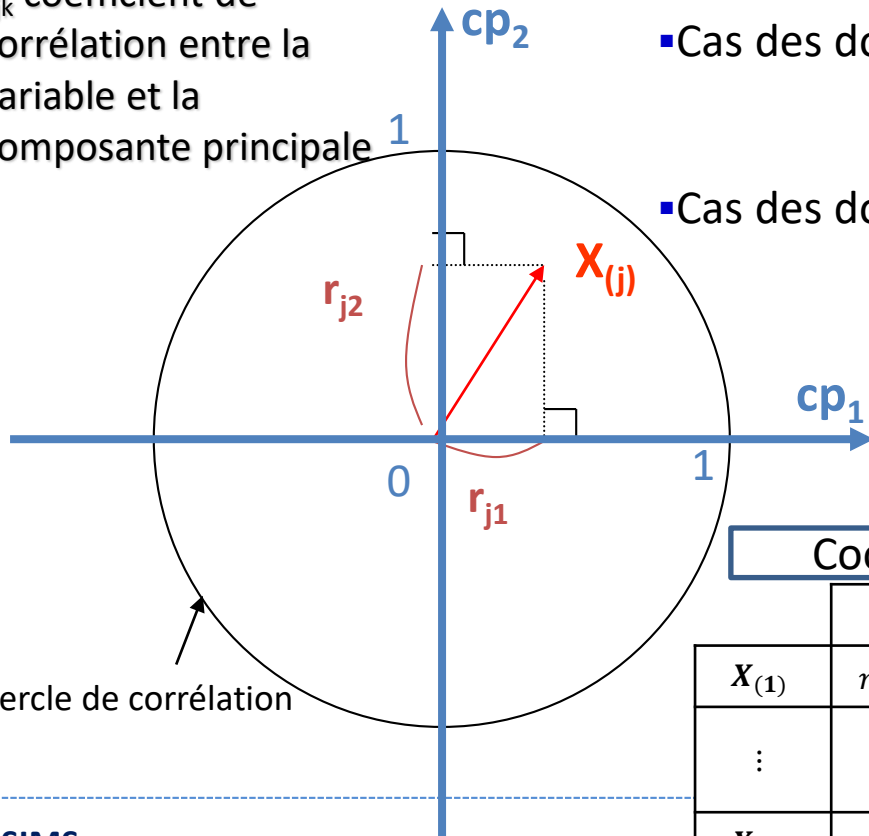
Procédure de l'ACP

► Étape 3: Interprétation des résultats ...

➤ Saturation des variables ...

■ Cercle de corrélation

r_{jk} coefficient de corrélation entre la variable et la composante principale

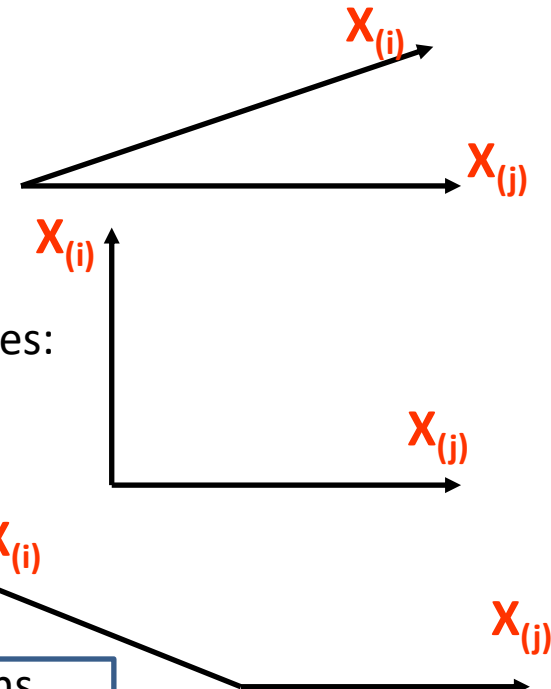


■ Cas des données centrées:

$$X_{(j)} = (X - \bar{X})_{(j)}$$

■ Cas des données centrées-réduites:

$$X_{(j)} = Z_{(j)}$$



Coordonnées - Corrélations

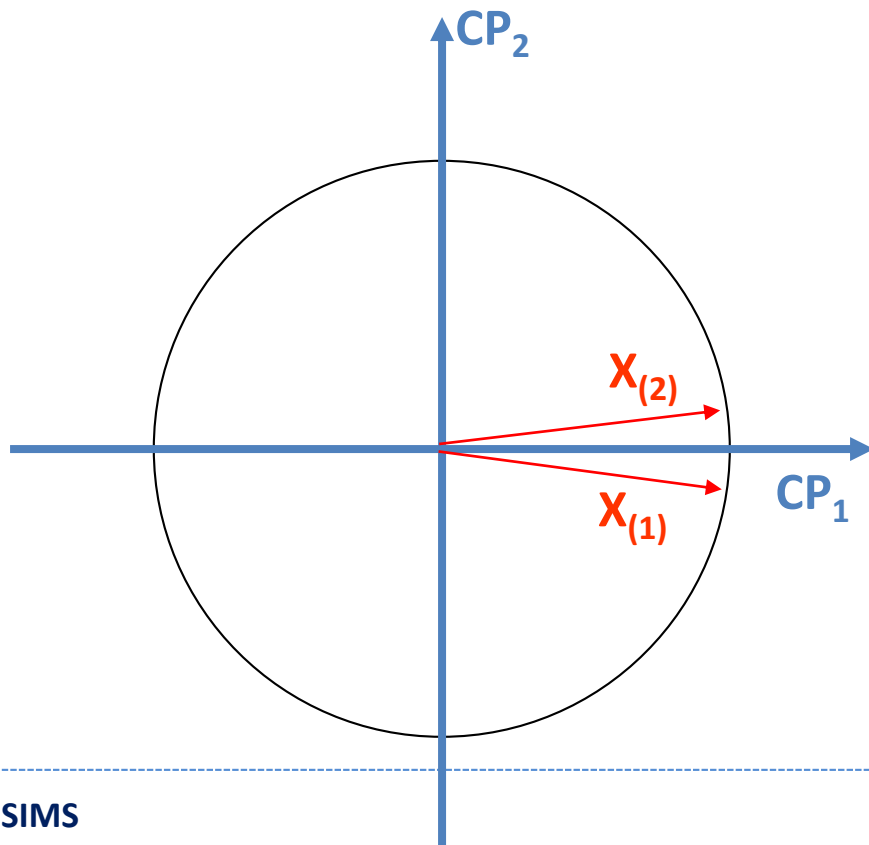
	$Y_{(1)}$...	$Y_{(p)}$
$X_{(1)}$	$r(X_{(1)}, Y_{(1)})$...	$r(X_{(1)}, Y_{(p)})$
\vdots	...	$r(X_{(j)}, Y_{(k)})$	\vdots
$X_{(p)}$	$r(X_{(p)}, Y_{(1)})$...	$r(X_{(p)}, Y_{(p)})$

Procédure de l'ACP

► Étape 3: Interprétation des résultats ...

➤ Saturation des variables ...

■ Cercle de corrélation : Notre exemple



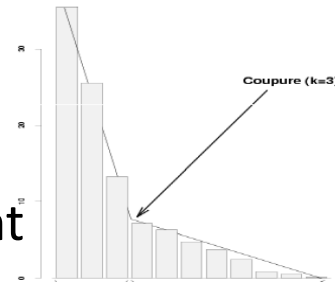
Coordonnées - Corrélations

	$Y_{(1)}$	$Y_{(2)}$
$X_{(1)}$	0,97825	-0,20743
$X_{(2)}$	0,98414	0,17742

	$Y_{(1)}$	$Y_{(2)}$
$Z_{(1)}$	0,98234	-0,18708
$Z_{(2)}$	0,98234	0,18708

Combien de composantes principales?

- ▶ L'ACP d'une matrice de données à p variables et n individus admet p valeurs propres, p vecteurs propres, et p composantes principales :
 - Conservez au moins 70-80% de la variance en cumulé
 - **Critère de Kaiser:** Conservez toute les composantes principales dont $\lambda > 1$
 - **Critère du coude:** Utilisez l'histogramme des valeurs propres (**scree plot**):
 - sur l'évolution des valeurs propres, on observe un décrochement (coude) suivi d'une décroissance régulière. On sélectionne les axes avant le décrochement
- ▶ Attention :
 - L'ACP sur une matrice de données tel que $p > n$ est impossible



Typologies des individus et des variables

► Typologie des individus : *projection des individus sur le plan factoriel*

- **Courbe de score de l'ACP:** La lecture graphique de la position des individus le long des CPs permet de dresser une typologie
- Les individus proches le long d'une CP sont des individus qui partagent les mêmes *caractéristiques* vis-à-vis des variables quantitatives étudiées

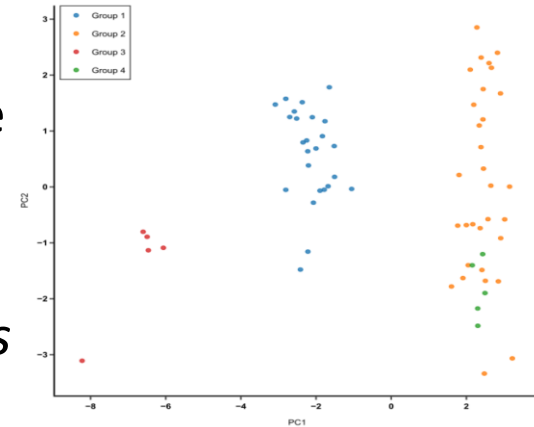
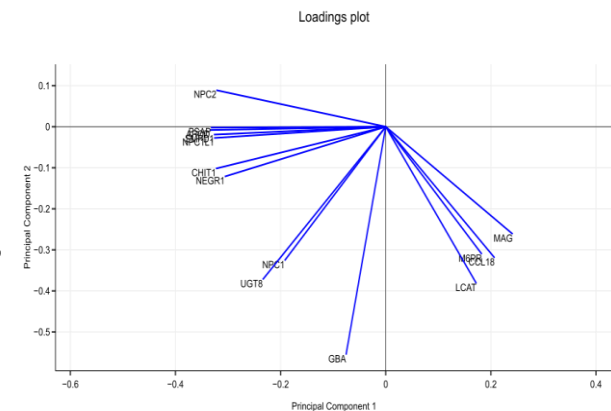


Figure 1. PCA plot.

► Typologie des variables : *projection des variables sur le plan factoriel*

- **Courbe de chargement de l'ACP:** montre l'influence de chaque variable initiale sur une composante principale
- La lecture graphique du cercle des corrélations permet de juger du poids des différentes variables de départ sur chacune des CPs



Exemple

Données initiales

$$X =$$

X1	X2	X3
10	20	10
2	5	2
8	17	7
9	20	10
12	22	11

$$\overbrace{X - \bar{X}}^M =$$

Données centrées

1.8	3.2	2
-8.2	-11.8	-8
-0.2	0.2	-1
0.8	3.2	2
3.8	5.2	3

$$\text{Cov}(X) = \frac{1}{n} (X - \bar{X})^t (X - \bar{X}) =$$

$$\text{Cov}(X) = \frac{1}{n} M^t M$$

	X1	X2	X3
X1	14.2	25.3	13.5
X2	25.3	46.7	24.75
X3	13.5	24.75	13.5

Exemple

	X1	X2	X3
X1	14.2	25.3	13.5
X2	25.3	46.7	24.75
X3	13.5	24.75	13.5

Matrice de covariance

$$\lambda_1 = 73.718$$

$$\lambda_2 = 0.384$$

$$\lambda_3 = 0.298$$

Valeurs propres

Remarque: $\lambda_1 + \lambda_2 + \lambda_3 = 74.4$

= trace de C

(Somme des variances dans la diagonale)

Vecteurs propres

$Q =$

q1	q2	q3
0.434	0.900	-0.044
0.795	-0.406	-0.451
0.424	-0.161	0.891

	Valeur propre	Pourcentage	Pourcentage cumulé
CP_1	73,718	99,083	99,083
CP_2	0,384	0,516	99,599
CP_3	0,298	0,401	100

Exemple

$$Y = (X - \bar{X})Q = M Q =$$

1.8	3.2	2
-8.2	-11.8	-8
-0.2	0.2	-1
0.8	3.2	2
3.8	5.2	3

x

0.434	0.900	-0.044
0.795	-0.406	-0.451
0.424	-0.161	0.891

=

<i>y1</i>	<i>y2</i>	<i>y3</i>
4.173	0.000	-0.259
-14.615	-0.172	-0.252
-0.352	0.100	0.973
3.739	0.900	-0.303
7.055	-0.828	-0.159

Prenant uniquement la 1^{ère} composante principale:

<i>y1</i>
4.173
-14.615
-0.352
3.739
7.055

Exercice 1

► Soit la matrice de données X suivante : **/40**

	Statistiques	Math	Cpta	G° Fi
Individu n° 1	19	14	8	18
Individu n° 2	20	12	4	4
Individu n° 3	10	10	32	38
Individu n° 4	13	17	4	4
Individu n° 5	6	8	26	24
Individu n° 6	6	3	28	32
Individu n° 7	19	16	8	20
Individu n° 8	15	18	6	6
Individu n° 9	9	2	32	30
Individu n° 10	8	7	20	20
Moyenne	12,5	10,7	16,8	19,6
Écart type	5,2010	5,3861	11,3208	11,3772

Exercice 1

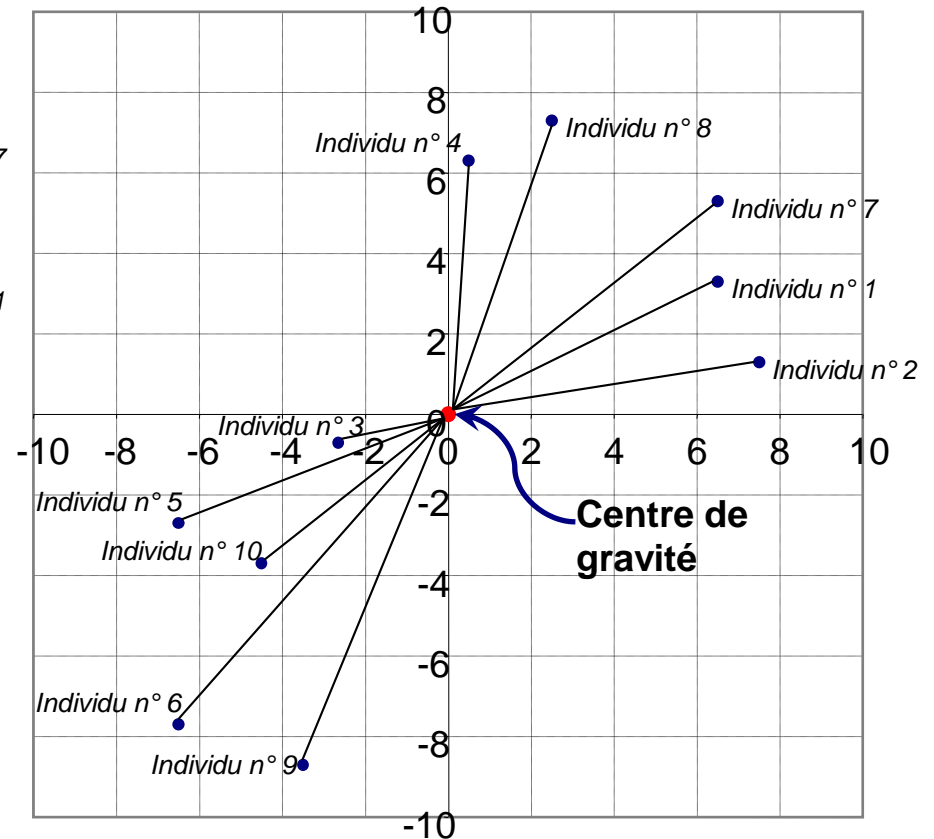
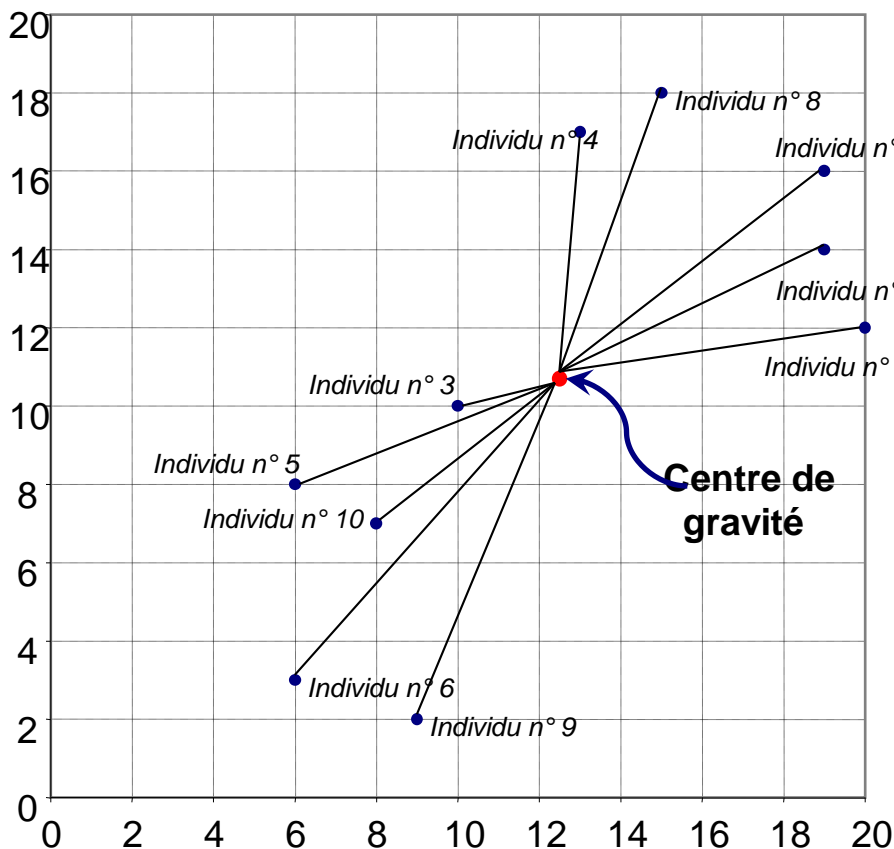
- Déterminer la matrice centrée $(X - \bar{X})$ de X :

	Statistiques	Math	Cpta	G° Fi
Individu n° 1	6,50	3,30	-8,80	-1,60
Individu n° 2	7,50	1,30	-12,80	-15,60
Individu n° 3	-2,50	-0,70	15,20	18,40
Individu n° 4	0,50	6,30	-12,80	-15,60
Individu n° 5	-6,50	-2,70	9,20	4,40
Individu n° 6	-6,50	-7,70	11,20	12,40
Individu n° 7	6,50	5,30	-8,80	0,40
Individu n° 8	2,50	7,30	-10,80	-13,60
Individu n° 9	-3,50	-8,70	15,20	10,40
Individu n° 10	-4,50	-3,70	3,20	0,40

Moyenne	0,00	0,00	0,00	0,00
Écart type	5,20	5,39	11,32	11,38

Exercice 1

- Nuage de points de la matrice X et celui de la matrice $(X - \bar{X})$



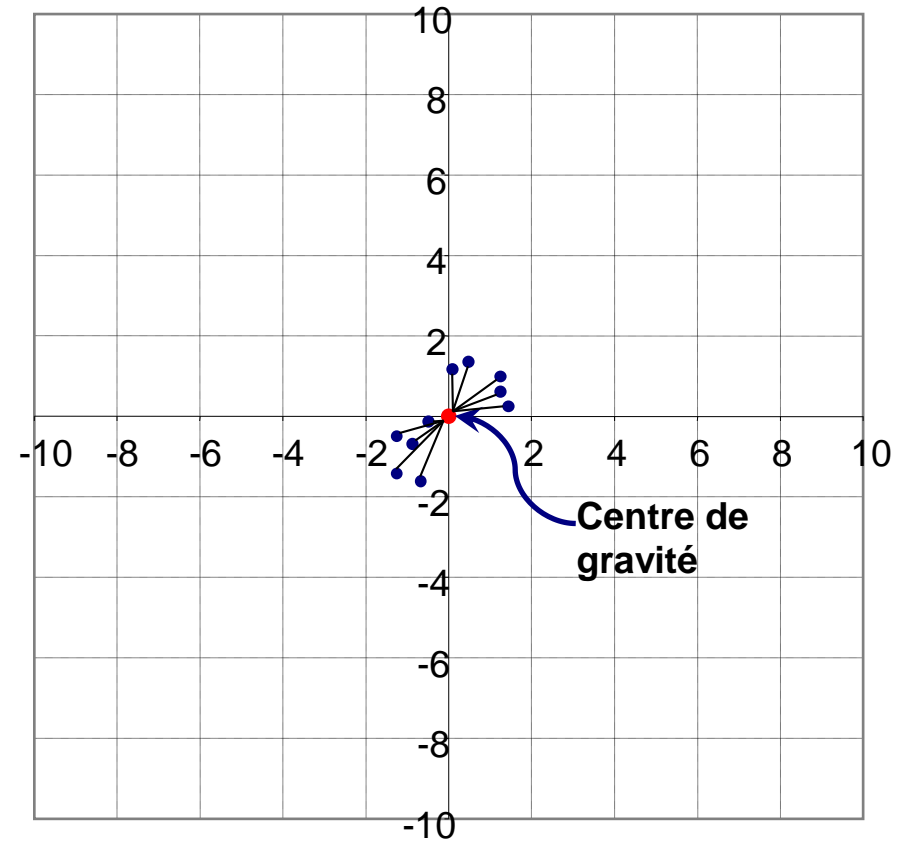
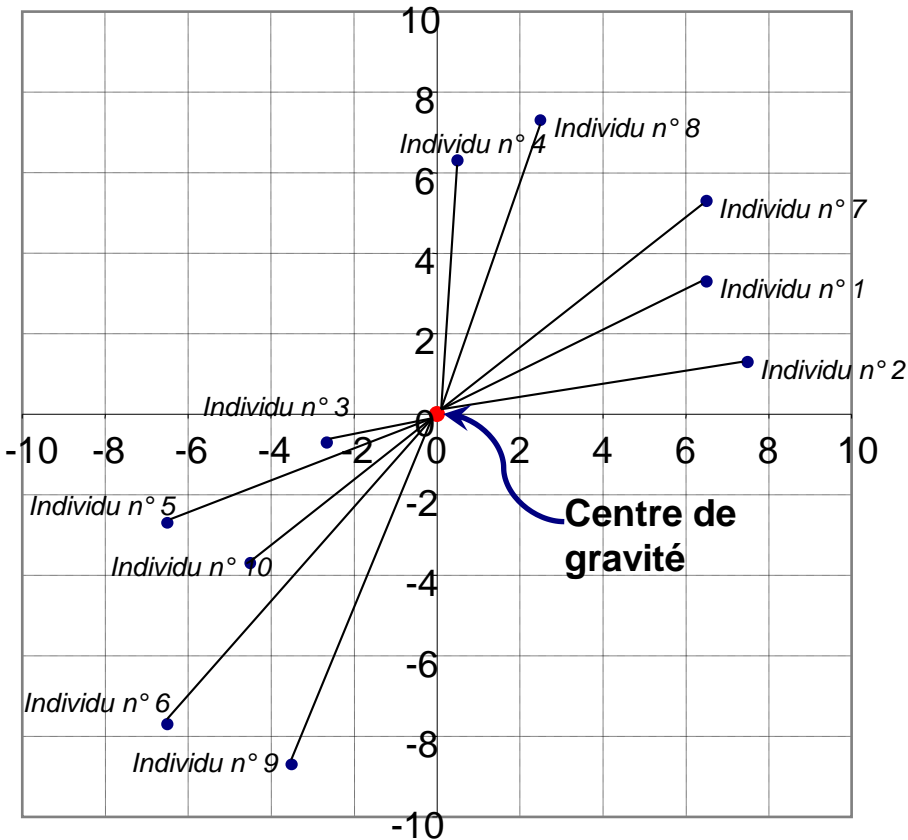
Exercice 1

- Déterminer la matrice centrée réduite (Z) de X :

	Statistiques	Math	Cpta	G° Fi
Individu n° 1	1,2498	0,6127	-0,7773	-0,1406
Individu n° 2	1,4420	0,2414	-1,1307	-1,3712
Individu n° 3	-0,4807	-0,1300	1,3427	1,6173
Individu n° 4	0,0961	1,1697	-1,1307	-1,3712
Individu n° 5	-1,2498	-0,5013	0,8127	0,3867
Individu n° 6	-1,2498	-1,4296	0,9893	1,0899
Individu n° 7	1,2498	0,9840	-0,7773	0,0352
Individu n° 8	0,4807	1,3553	-0,9540	-1,1954
Individu n° 9	-0,6730	-1,6153	1,3427	0,9141
Individu n° 10	-0,8652	-0,6870	0,2827	0,0352
Moyenne	0,00	0,00	0,00	0,00
Écart type	1	1	1	1

Exercice 1

- Nuage de points de la matrice $(X - \bar{X})$ et celui de la matrice Z



Exercice 1

- Déterminer la matrice de corrélation de Bravais-Pearson $R(X)$:

$$R(X) = \frac{1}{n} Z^t Z$$

	Statistiques	Math	Cpta	G° Fi
Statistiques	1,0000	0,7265	-0,8186	-0,6084
Math	0,7265	1,0000	-0,8489	-0,7069
Cpta	-0,8186	-0,8489	1,0000	0,9124
G° Fi	-0,6084	-0,7069	0,9124	1,0000

- Résoudre: $\det(R - \lambda I) = 0$

$$R - \lambda I = \begin{bmatrix} 1 - \lambda & 0,7265 & -0,8186 & -0,6084 \\ 0,7265 & 1 - \lambda & -0,8489 & -0,7069 \\ -0,8186 & -0,8489 & 1 - \lambda & 0,9124 \\ -0,6084 & -0,7069 & 0,9124 & 1 - \lambda \end{bmatrix}$$



Résoudre: $\det(R - \lambda I) = 0$



$$\begin{cases} \lambda_1 = 3.3189 \\ \lambda_2 = 0.4035 \\ \lambda_3 = 0.2508 \\ \lambda_4 = 0.0268 \end{cases}$$

Exercice 1

► Interprétation des valeurs propres

A.C.P. sur données
centrées-réduites

∀ variable → variance = 1

Σ variance
=
nombre de variables

Exemple : $\frac{\lambda_1}{\sum \lambda_i} = \frac{3.3189}{4}$
 $= 0.829725 = 82.97\%$

Part de l'information
initiale restituée par l'axe i

$$\begin{cases} \lambda_1 = 3.3189 \\ \lambda_2 = 0.4035 \\ \lambda_3 = 0.2508 \\ \lambda_4 = 0.0268 \end{cases}$$

$$\sum \lambda_i = 4$$

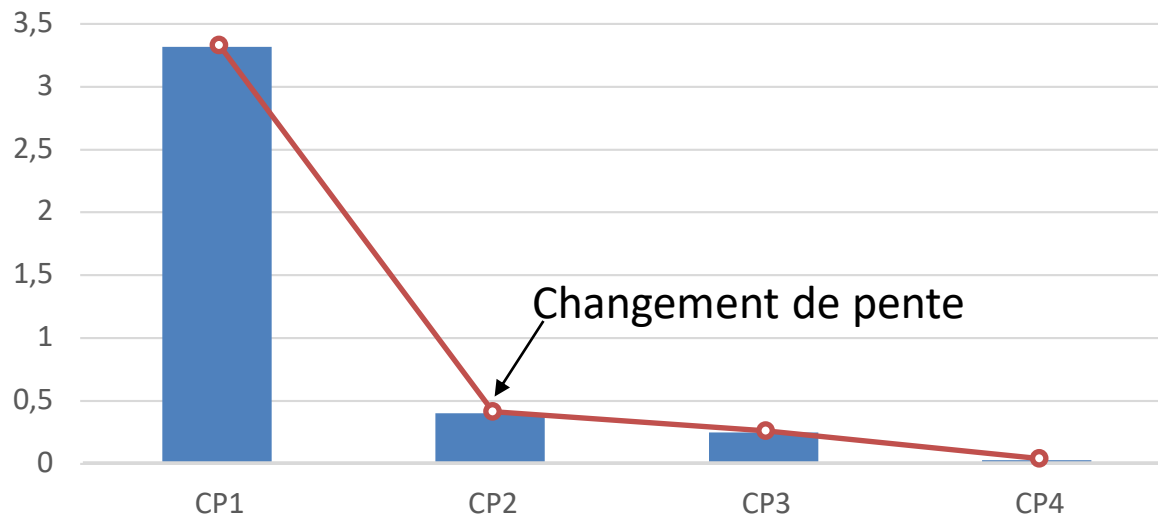
$$\frac{\lambda_i}{\sum \lambda_i}$$

Exercice 1

► Interprétation des valeurs propres

	Valeur propre	Pourcentage	Pourcentage cumulé
1	3.3189	82.9700	82.9700
2	0.4035	10.0900	93.0600
3	0.2508	6.2700	99.3300
4	0.0268	0.6700	100.0000

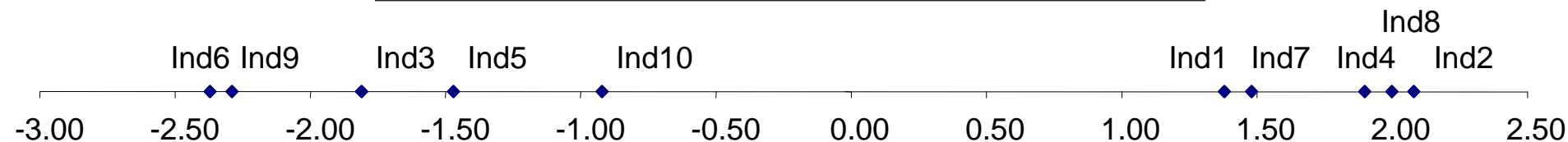
Histogramme des valeurs propres



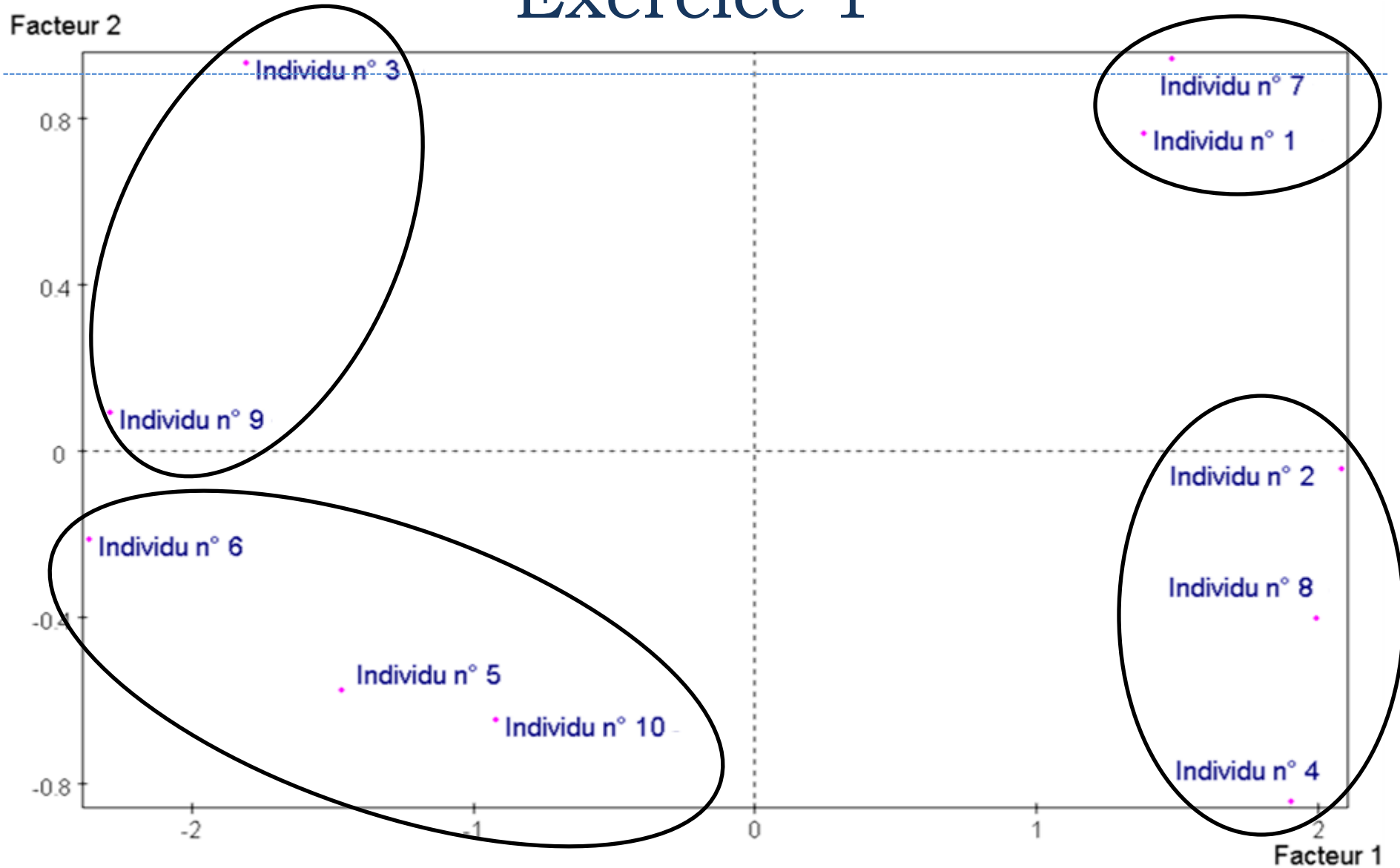
Exercice 1

► Les nouvelles coordonnées des individus

	Coordonnées			
	1	2	3	4
Individu n° 1	1.38	0.77	0.20	0.14
Individu n° 2	2.08	-0.04	0.97	-0.15
Individu n° 3	-1.81	0.93	-0.71	-0.17
Individu n° 4	1.90	-0.85	-0.43	0.04
Individu n° 5	-1.47	-0.58	-0.36	-0.07
Individu n° 6	-2.37	-0.22	0.21	0.28
Individu n° 7	1.48	0.94	-0.16	0.16
Individu n° 8	2.00	-0.40	-0.48	-0.15
Individu n° 9	-2.29	0.09	0.63	-0.21
Individu n° 10	-0.92	-0.65	0.13	0.13

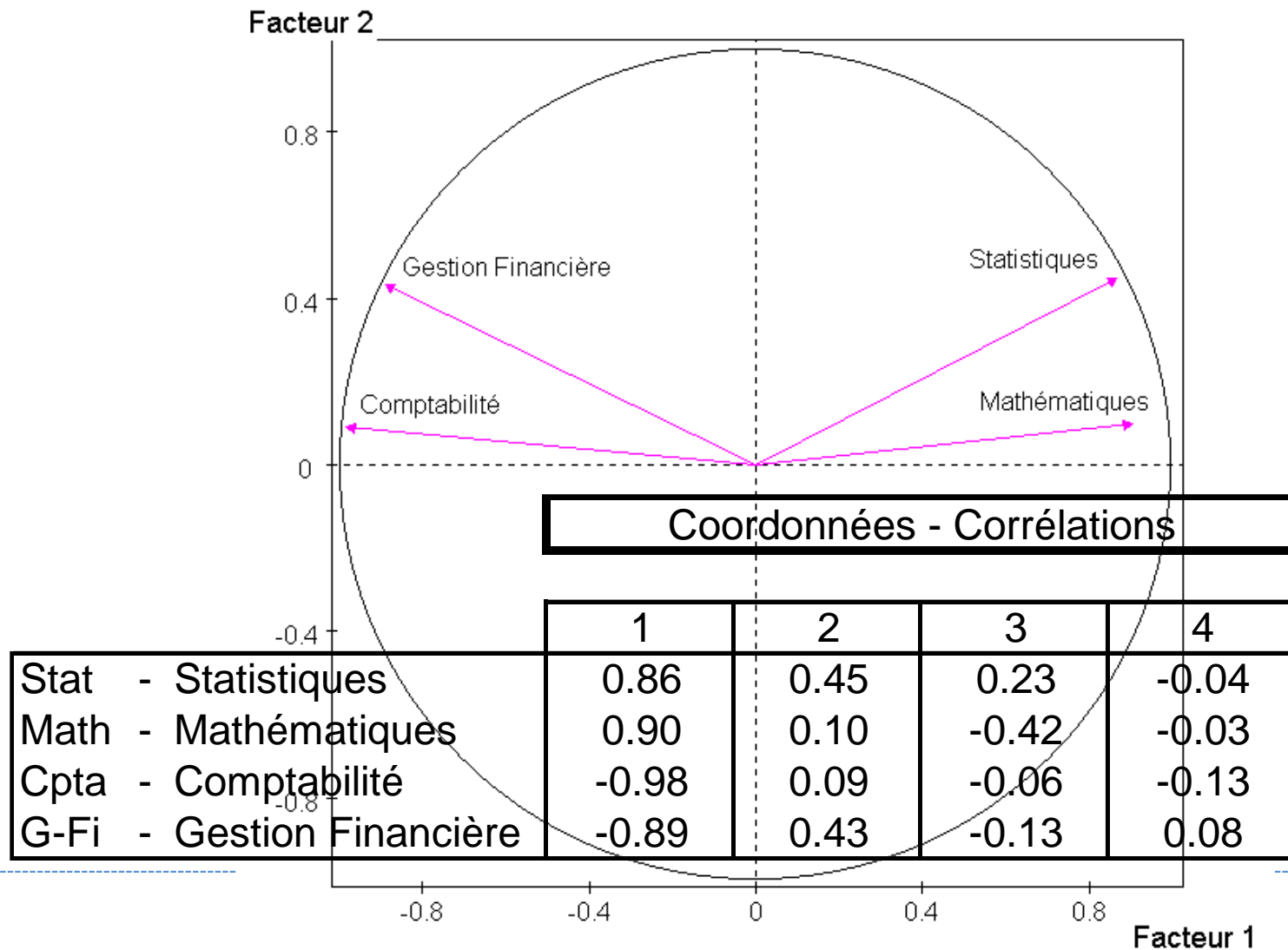


Exercise 1



Exercice 1

► Cercle de corrélation



Exercice 2

- Soit le tableau de données suivant :

	Informatique	Gestion
Noam	4	5
Jean	6	7
Li	8	0

- Où les lignes représentent les individus (noms de quelques étudiants) et les colonnes les variables (notes en informatique et gestion).
- Ce tableau de données peut être représenté par la matrice X de données brutes :

$$X = \begin{pmatrix} 4 & 5 \\ 6 & 7 \\ 8 & 0 \end{pmatrix}$$

- 1) Calculer la matrice $M = X - \bar{X}$ des données centrées et la matrice Z de données centrées et réduites.

Exercice 2

- 2) Calculer la matrice $C(X)$ des covariances de X et la matrice $R(X)$ des corrélations de X .
- 3) Vérifier que $\lambda_1 = 10,152$ et $\lambda_2 = 1,188$ sont deux valeurs propres associées à $C(X)$. Puis vérifier que les vecteurs unitaires :

$$V_1 = \begin{pmatrix} -0.407 \\ 0.914 \end{pmatrix} \text{ et } V_2 = \begin{pmatrix} 0.914 \\ 0.407 \end{pmatrix}$$

sont les vecteurs propres de $C(X)$ associés à λ_1 et λ_2 respectivement.

- 4) À partir des résultats précédents, déterminer les deux composantes principales de l'ACP du nuage des individus associé au tableau X . Pour chacun de ces axes, préciser le pourcentage de la quantité d'information projetée sur l'axe considéré, et le pourcentage cumulé.
- 5) Déterminer la matrice des scores des individus.
- 6) Calculer la saturation des variables et interpréter le cercle de corrélation.

Exercice 3

- Une étude sur des fournisseurs de matériel informatique a conduit à apprécier le service, la qualité et le prix de quatre fournisseurs. Pour cela un expert a noté ces entreprises avec des notes allant de -3 à 3. Les résultats sont consignés ci-dessous :

Ent	Service	Qualité	Prix
E1	-2	3	-1
E2	-1	1	0
E3	2	-1	-1
E4	1	-3	2

- 1) Calculer le vecteur moyen des individus. Qu'en conclure?
- 2) Calculer la matrice $C(X)$ des covariances
- 3) Dans le but d'appliquer une ACP non normé, nous pouvons vérifier facilement que C admet une valeur propre nulle (λ_3). On donne aussi $\lambda_1 = \frac{61}{8}$. En déduire λ_2 .

Exercice 3

-
- 4) Préciser le pourcentage de la quantité d'information projetée sur chaque composante principale, et le pourcentage cumulé. Interpréter ces résultats.
- 5) On donne :

$$V_1 = \begin{pmatrix} 0,5 \\ -0,8 \\ 0,3 \end{pmatrix} \text{ et } V_2 = \begin{pmatrix} 0,65 \\ 0,11 \\ -0,75 \end{pmatrix}$$

les vecteurs propres de $C(X)$ associés à λ_1 et λ_2 respectivement.
Déterminer la matrice des scores des individus.

- 6) Calculer la saturation des variables et interpréter le cercle de corrélation dans le plan factoriel (CP_1 , CP_2).