

Théorème spectral et analyse en composantes principales

Cours 9 — Algèbre linéaire et analyse de données

Christophe Ambroise

L3 GBI — Université d'Évry

Matrices symétriques

Le cas favorable

$$S^{\top} = S.$$

Le cas favorable

$$S^{\top} = S.$$

Les échecs du cours 8 — vecteurs propres en nombre insuffisant, racines complexes — n'y surviennent **jamais**.

Théorème spectral

Ce que change l'orthogonalité

Théorème Toute matrice symétrique réelle est diagonalisable dans une **base orthonormée** de vecteurs propres, et toutes ses valeurs propres sont **réelles**.

$$S = PDP^{\top}, \quad P^{\top}P = I_n$$

P est orthogonale : $P^{-1} = P^{\top}$.

Ce que change l'orthogonalité

P est orthogonale : $P^{-1} = P^{\top}$.

$$S = PDP^{-1} = PDP^{\top}$$

Ce que change l'orthogonalité

P est orthogonale : $P^{-1} = P^{\top}$.

$$S = PDP^{-1} = PDP^{\top}$$

Aucun inverse à calculer. Une **rotation** du repère, au sens du cours 6.

Orthogonalité des vecteurs propres

$$S\mathbf{v}_1 = \lambda_1 \mathbf{v}_1, \quad S\mathbf{v}_2 = \lambda_2 \mathbf{v}_2, \quad \lambda_1 \neq \lambda_2.$$

$$\lambda_1 \langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle S\mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, S\mathbf{v}_2 \rangle = \lambda_2 \langle \mathbf{v}_1, \mathbf{v}_2 \rangle$$

Orthogonalité des vecteurs propres

$$S\mathbf{v}_1 = \lambda_1 \mathbf{v}_1, \quad S\mathbf{v}_2 = \lambda_2 \mathbf{v}_2, \quad \lambda_1 \neq \lambda_2.$$

$$\lambda_1 \langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle S\mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, S\mathbf{v}_2 \rangle = \lambda_2 \langle \mathbf{v}_1, \mathbf{v}_2 \rangle$$

$$(\lambda_1 - \lambda_2) \langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0, \text{ donc } \mathbf{v}_1 \perp \mathbf{v}_2.$$

Orthogonalité des vecteurs propres

$$S\mathbf{v}_1 = \lambda_1 \mathbf{v}_1, \quad S\mathbf{v}_2 = \lambda_2 \mathbf{v}_2, \quad \lambda_1 \neq \lambda_2.$$

$$\lambda_1 \langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle S\mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, S\mathbf{v}_2 \rangle = \lambda_2 \langle \mathbf{v}_1, \mathbf{v}_2 \rangle$$

$$(\lambda_1 - \lambda_2) \langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0, \text{ donc } \mathbf{v}_1 \perp \mathbf{v}_2.$$

La symétrie **produit** l'orthogonalité.

Un exemple

$$S = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}, \quad \lambda_1 = 4, \quad \lambda_2 = 2$$

Un exemple

$$S = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}, \quad \lambda_1 = 4, \lambda_2 = 2$$

Vecteurs propres $(1, 1)$ et $(-1, 1)$: orthogonaux, comme annoncé.

$$P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad P^\top P = I_2$$

Forme quadratique associée

$$q(\mathbf{u}) = \mathbf{u}^\top S \mathbf{u} = \sum_{i=1}^n \lambda_i \langle \mathbf{u}, \mathbf{v}_i \rangle^2$$

Forme quadratique associée

$$q(\mathbf{u}) = \mathbf{u}^\top S \mathbf{u} = \sum_{i=1}^n \lambda_i \langle \mathbf{u}, \mathbf{v}_i \rangle^2$$

Dans la base propre, une simple somme pondérée de carrés.

Forme quadratique associée

$$q(\mathbf{u}) = \mathbf{u}^\top S \mathbf{u} = \sum_{i=1}^n \lambda_i \langle \mathbf{u}, \mathbf{v}_i \rangle^2$$

Dans la base propre, une simple somme pondérée de carrés.

Toute l'ACP repose sur cette écriture.

Encadrement

$\|\mathbf{u}\| = 1$: les $\langle \mathbf{u}, \mathbf{v}_i \rangle^2$ somment à 1.

$$\lambda_{\min} \leq \mathbf{u}^\top S \mathbf{u} \leq \lambda_{\max}$$

Démonstration de l'encadrement

$\mathbf{u} = \sum_i \alpha_i \mathbf{v}_i$ dans la base propre, avec $\sum_i \alpha_i^2 = 1$.

$$\mathbf{u}^\top S \mathbf{u} = \sum_i \lambda_i \alpha_i^2$$

Encadrement

$\|\mathbf{u}\| = 1$: les $\langle \mathbf{u}, \mathbf{v}_i \rangle^2$ somment à 1.

$$\lambda_{\min} \leq \mathbf{u}^\top S \mathbf{u} \leq \lambda_{\max}$$

Le maximum est atteint $\mathbf{u} = \mathbf{v}_{\max}$ donne exactement $\lambda_{\max}$.

Démonstration de l'encadrement

$\mathbf{u} = \sum_i \alpha_i \mathbf{v}_i$ dans la base propre, avec $\sum_i \alpha_i^2 = 1$.

$$\mathbf{u}^\top S \mathbf{u} = \sum_i \lambda_i \alpha_i^2$$

Des poids positifs, de somme 1 : une **moyenne pondérée** des valeurs propres.

Démonstration de l'encadrement

$\mathbf{u} = \sum_i \alpha_i \mathbf{v}_i$ dans la base propre, avec $\sum_i \alpha_i^2 = 1$.

$$\mathbf{u}^\top S \mathbf{u} = \sum_i \lambda_i \alpha_i^2$$

Des poids positifs, de somme 1 : une **moyenne pondérée** des valeurs propres.

L'argument Une moyenne ne sort jamais de l'intervalle de ses termes.

Démonstration de l'encadrement

$\mathbf{u} = \sum_i \alpha_i \mathbf{v}_i$ dans la base propre, avec $\sum_i \alpha_i^2 = 1$.

$$\mathbf{u}^\top S \mathbf{u} = \sum_i \lambda_i \alpha_i^2$$

Des poids positifs, de somme 1 : une **moyenne pondérée** des valeurs propres.

L'argument Une moyenne ne sort jamais de l'intervalle de ses termes.

$\mathbf{u} = \mathbf{v}_{\max}$ met tout le poids sur un seul terme :

$$\mathbf{v}_{\max}^\top S \mathbf{v}_{\max} = \lambda_{\max} \|\mathbf{v}_{\max}\|^2 = \lambda_{\max}$$

Le tableau de données

Matrice de covariance

X : n individus en lignes, p variables en colonnes.

Le tableau de données

Définition

X : n individus en lignes, p variables en colonnes.

Colonnes **centrées** — projetées sur l'orthogonal de $\mathbf{1}$, au sens du cours 7.

$$S = \frac{1}{n} X^\top X \in \mathcal{M}_{p,p}(\mathbb{R})$$

Définition

$$S = \frac{1}{n} X^\top X \in \mathcal{M}_{p,p}(\mathbb{R})$$

$$S_{jk} = \frac{1}{n} \langle \mathbf{x}_j, \mathbf{x}_k \rangle = \text{cov}(\mathbf{x}_j, \mathbf{x}_k)$$

Définition

$$S = \frac{1}{n} X^\top X \in \mathcal{M}_{p,p}(\mathbb{R})$$

$$S_{jk} = \frac{1}{n} \langle \mathbf{x}_j, \mathbf{x}_k \rangle = \text{cov}(\mathbf{x}_j, \mathbf{x}_k)$$

Diagonale : les **variances**. Hors diagonale : les **covariances**.

$$S^\top = \left(\frac{1}{n}X^\top X\right)^\top = \frac{1}{n}X^\top X = S$$

Valeurs propres positives

$$\mathbf{u}^\top S \mathbf{u} = \frac{1}{n} \mathbf{u}^\top X^\top X \mathbf{u} = \frac{1}{n} \|X \mathbf{u}\|^2 \geq 0$$

$$S^\top = \left(\frac{1}{n}X^\top X\right)^\top = \frac{1}{n}X^\top X = S$$

Conséquence Le théorème spectral s'applique. **Sans condition.**

Valeurs propres positives

$$\mathbf{u}^\top S \mathbf{u} = \frac{1}{n} \mathbf{u}^\top X^\top X \mathbf{u} = \frac{1}{n} \|X \mathbf{u}\|^2 \geq 0$$

Appliqué à un vecteur propre : $\lambda \|\mathbf{v}\|^2 \geq 0$.

Valeurs propres positives

$$\mathbf{u}^\top S \mathbf{u} = \frac{1}{n} \mathbf{u}^\top X^\top X \mathbf{u} = \frac{1}{n} \|X \mathbf{u}\|^2 \geq 0$$

Appliqué à un vecteur propre : $\lambda \|\mathbf{v}\|^2 \geq 0$.

$$\forall i, \quad \lambda_i \geq 0$$

Variables réduites

Variables d'unités différentes : centimètres, grammes, milliers.

Chaque colonne divisée par son écart-type :

$$S \longrightarrow R, \quad R_{jk} = \text{cor}(\mathbf{x}_j, \mathbf{x}_k), \quad R_{jj} = 1$$

Variables réduites

Variables d'unités différentes : centimètres, grammes, milliers.

Variables réduites

Variables d'unités différentes : centimètres, grammes, milliers.

Chaque colonne divisée par son écart-type :

$$S \longrightarrow R, \quad R_{jk} = \text{cor}(\mathbf{x}_j, \mathbf{x}_k), \quad R_{jj} = 1$$

La matrice de corrélation. Ses coefficients sont les **cosinus** du cours 6.

Covariance ou corrélation ?

S — les variables gardent leurs unités. Celle de plus grande variance domine.

Covariance ou corrélation ?

S — les variables gardent leurs unités. Celle de plus grande variance domine.

R — chaque variable pèse pareil.

Covariance ou corrélation ?

S — les variables gardent leurs unités. Celle de plus grande variance domine.

R — chaque variable pèse pareil.

Décision de méthode Unités hétérogènes : **réduire**. Le choix change les axes, donc les conclusions.

Inertie totale

$$I = \sum_{j=1}^p \text{var}(\mathbf{x}_j) = \text{tr}(S) = \sum_{i=1}^p \lambda_i$$

Inertie totale

$$I = \sum_{j=1}^p \text{var}(\mathbf{x}_j) = \text{tr}(S) = \sum_{i=1}^p \lambda_i$$

La dispersion totale du nuage, répartie sur les valeurs propres.

Inertie totale

$$I = \sum_{j=1}^p \text{var}(\mathbf{x}_j) = \text{tr}(S) = \sum_{i=1}^p \lambda_i$$

La dispersion totale du nuage, répartie sur les valeurs propres.

Sur données réduites : $I = p$.

Le problème posé

L'analyse en composantes principales

Un axe, porté par $\mathbf{u}$ unitaire.

Le problème posé

Un axe, porté par $\mathbf{u}$ unitaire.

Projeter les n individus dessus : les coordonnées forment $X\mathbf{u}$.

$$\text{var}(X\mathbf{u}) = \frac{1}{n} \|X\mathbf{u}\|^2 = \mathbf{u}^\top S \mathbf{u}$$

Le critère

Objectif Trouver la direction $\mathbf{u}$ qui **maximise la variance** de la projection.

$$\max_{\|\mathbf{u}\|=1} \mathbf{u}^\top S \mathbf{u}$$

Le critère

Objectif Trouver la direction $\mathbf{u}$ qui **maximise la variance** de la projection.

$$\max_{\|\mathbf{u}\|=1} \mathbf{u}^\top S \mathbf{u}$$

La question a déjà sa réponse.

La solution

Théorème Le maximum vaut λ_1 , la plus grande valeur propre de S , et il est atteint pour $\mathbf{u} = \mathbf{v}_1$, vecteur propre associé.

La solution

Théorème Le maximum vaut λ_1 , la plus grande valeur propre de S , et il est atteint pour $\mathbf{u} = \mathbf{v}_1$, vecteur propre associé.

$$\mathbf{u}^\top S \mathbf{u} = \sum_i \lambda_i \alpha_i^2 \leq \lambda_1 \sum_i \alpha_i^2 = \lambda_1$$

Axes principaux

$\mathbf{v}_2$: la meilleure direction **orthogonale** à $\mathbf{v}_1$. Variance λ_2 .

La solution

Théorème Le maximum vaut λ_1 , la plus grande valeur propre de S , et il est atteint pour $\mathbf{u} = \mathbf{v}_1$, vecteur propre associé.

$$\mathbf{u}^\top S \mathbf{u} = \sum_i \lambda_i \alpha_i^2 \leq \lambda_1 \sum_i \alpha_i^2 = \lambda_1$$

L'encadrement de la section 1, appliqué à S .

Axes principaux

$\mathbf{v}_2$: la meilleure direction **orthogonale** à $\mathbf{v}_1$. Variance λ_2 .

Et ainsi de suite.

Résultat Les axes principaux sont les **vecteurs propres** de S , ordonnés par valeurs propres décroissantes.

Composantes principales

$$\mathbf{c}_k = X\mathbf{v}_k, \quad \text{var}(\mathbf{c}_k) = \lambda_k$$

Composantes principales

$$\mathbf{c}_k = X\mathbf{v}_k, \quad \text{var}(\mathbf{c}_k) = \lambda_k$$

n nouvelles coordonnées par axe. Des combinaisons linéaires des variables initiales.

Composantes principales

$$\mathbf{c}_k = X\mathbf{v}_k, \quad \text{var}(\mathbf{c}_k) = \lambda_k$$

n nouvelles coordonnées par axe. Des combinaisons linéaires des variables initiales.

$$\text{cov}(\mathbf{c}_k, \mathbf{c}_l) = 0 \quad (k \neq l)$$

Des variables décorrélées

Les composantes principales sont **deux à deux orthogonales**.

Les composantes principales sont **deux à deux orthogonales**.

Ce qu'a fait l'ACP Remplacer p variables corrélées par p variables décorrélées, ordonnées par la variance qu'elles portent.

Les composantes principales sont **deux à deux orthogonales**.

Ce qu'a fait l'ACP Remplacer p variables corrélées par p variables décorrélées, ordonnées par la variance qu'elles portent.

Un changement de base orthonormée. Rien de plus.

Le fil, refermé

Cours	Étape
5	coordonnées d'un vecteur dans une base
6	matrice de passage, cas orthonormé
8	$A = PDP^{-1}$, changement de base
9	l'ACP est le changement de base qui diagonalise S

[Lire les résultats](#)

Part d'inertie

$$\tau_k = \frac{\lambda_k}{\lambda_1 + \dots + \lambda_p} = \frac{\lambda_k}{I}$$

Part d'inertie

$$\tau_k = \frac{\lambda_k}{\lambda_1 + \dots + \lambda_p} = \frac{\lambda_k}{I}$$

La fraction de la dispersion totale portée par l'axe k .

Part d'inertie

$$\tau_k = \frac{\lambda_k}{\lambda_1 + \dots + \lambda_p} = \frac{\lambda_k}{I}$$

La fraction de la dispersion totale portée par l'axe k .

Leur somme vaut 1.

Éboulis des valeurs propres

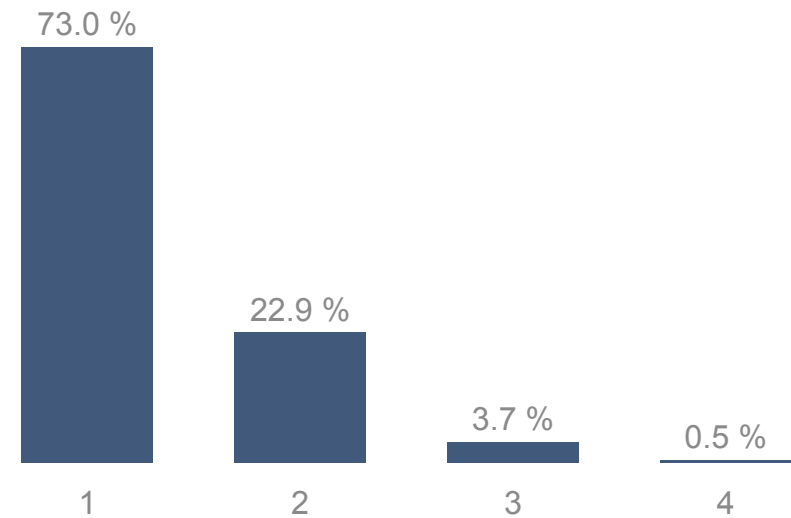


Figure 1

Quatre variables, mais deux axes suffisent

Choisir le nombre d'axes

Le décrochement de l'éboulis.

Choisir le nombre d'axes

Le décrochement de l'éboulis.

Un seuil cumulé — 80 %, 90 %.

Choisir le nombre d'axes

Le décrochement de l'éboulis.

Un seuil cumulé — 80 %, 90 %.

Aucune règle absolue Le choix est **interprétatif**. Les mathématiques fournissent les λ_k , pas la décision.

Le plan principal

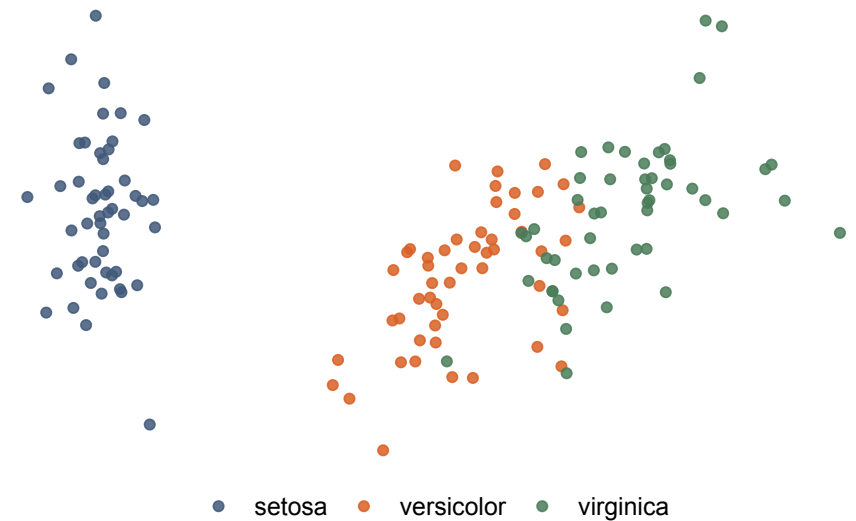


Figure 2

Quatre mesures florales, réduites à deux coordonnées.

Ce que montre le graphique

Aucune information d'espèce n'a servi au calcul.

Cercle des corrélations

$$\text{cor}(\mathbf{x}_j, \mathbf{c}_k) = \sqrt{\lambda_k} v_{jk}$$

Ce que montre le graphique

Aucune information d'espèce n'a servi au calcul.

Le même phénomène qu'au cours 1 Une structure biologique apparaît, sans avoir été fournie.

Cercle des corrélations

$$\text{cor}(\mathbf{x}_j, \mathbf{c}_k) = \sqrt{\lambda_k} v_{jk}$$

Chaque variable devient un point du plan, dans le disque unité.

Cercle des corrélations

$$\text{cor}(\mathbf{x}_j, \mathbf{c}_k) = \sqrt{\lambda_k} v_{jk}$$

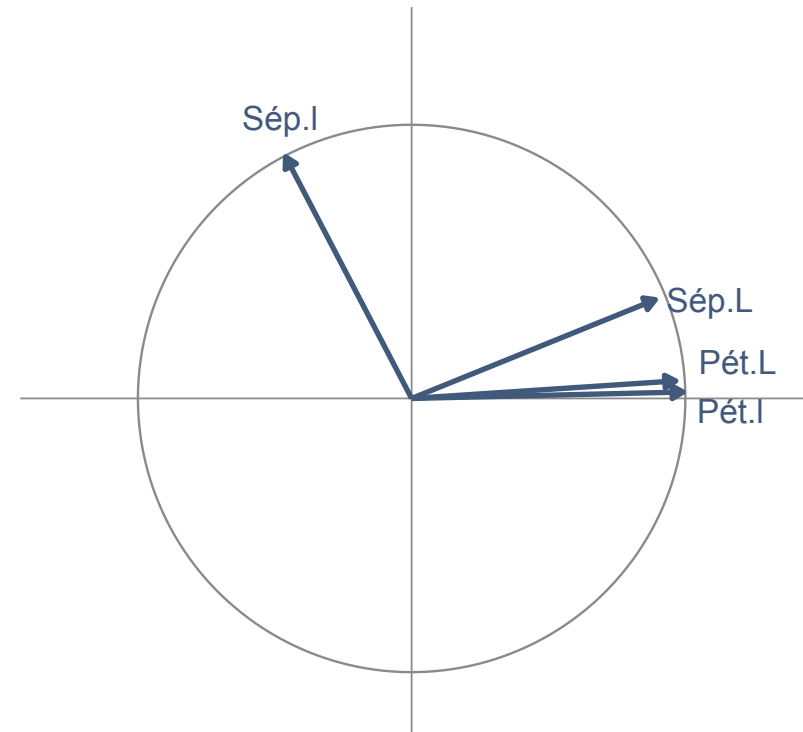
Chaque variable devient un point du plan, dans le disque unité.

Sa distance au bord mesure sa **qualité de représentation**.

Lecture du cercle

Deux variables proches et près du bord : fortement **corrélées**.

Le cercle d'iris



Lecture du cercle

Deux variables proches et près du bord : fortement **corrélées**.

Deux variables à angle droit : **décorrélées**.

Lecture du cercle

Deux variables proches et près du bord : fortement **corrélées**.

Deux variables à angle droit : **décorrélées**.

Variables opposées : corrélation négative.

Lecture du cercle

Deux variables proches et près du bord : fortement **corrélées**.

Deux variables à angle droit : **décorrélées**.

Variables opposées : corrélation négative.

Toujours les cosinus du cours 6.

Une mise en garde

Une variable proche du centre est **mal représentée**.

Une mise en garde

Une variable proche du centre est **mal représentée**.

Piège classique Un angle droit près du centre ne signifie **rien**.
L'interprétation exige d'abord une bonne représentation.

Garder q axes revient à **projeter** le nuage sur un sous-espace de dimension q .

$$X \approx \sum_{k=1}^q \mathbf{c}_k \mathbf{v}_k^\top$$

Garder q axes revient à **projeter** le nuage sur un sous-espace de dimension q .

$$X \approx \sum_{k=1}^q \mathbf{c}_k \mathbf{v}_k^\top$$

La formule de projection du cours 7, tronquée.

Ce qui est perdu

Ce qui est perdu

$$\underbrace{\sum_{k=1}^p \lambda_k}_{\text{inertie totale}} = \underbrace{\sum_{k=1}^q \lambda_k}_{\text{retenue}} + \underbrace{\sum_{k=q+1}^p \lambda_k}_{\text{perdue}}$$

$$\underbrace{\sum_{k=1}^p \lambda_k}_{\text{inertie totale}} = \underbrace{\sum_{k=1}^q \lambda_k}_{\text{retenue}} + \underbrace{\sum_{k=q+1}^p \lambda_k}_{\text{perdue}}$$

Pythagore, une dernière fois La décomposition du cours 7 : part gardée, part perdue. Les valeurs propres en donnent le décompte exact.

Ce que l'ACP ne fait pas

Elle ne détecte que des liaisons **linéaires**.

Ce que l'ACP ne fait pas

Elle ne détecte que des liaisons **linéaires**.

Elle dépend du choix de centrer et réduire.

Ce que l'ACP ne fait pas

Elle ne détecte que des liaisons **linéaires**.

Elle dépend du choix de centrer et réduire.

Elle n'établit **aucune causalité**.

Ce que l'ACP ne fait pas

Elle ne détecte que des liaisons **linéaires**.

Elle dépend du choix de centrer et réduire.

Elle n'établit **aucune causalité**.

Un outil descriptif L'ACP résume. Elle ne teste rien, ne prouve rien.

[Retour à Novembre 2008](#)

[Le semestre, en une page](#)

3 000 individus, 500 000 SNP.

[Retour à Novembre 2008](#)

[Retour à Novembre 2008](#)

3 000 individus, 500 000 SNP.

Matrice de covariance, théorème spectral, deux premiers axes.

3 000 individus, 500 000 SNP.

Matrice de covariance, théorème spectral, deux premiers axes.

La carte de l'Europe Le calcul du cours 1 est celui que vous venez de faire.

Étape	Objet	Cours
Centrer, réduire	X	6, 7
Covariance	$S = \frac{1}{n} X^\top X$, symétrique	3, 9
Diagonaliser	$S = PDP^\top$	8, 9
Axes principaux	vecteurs propres $\mathbf{v}_k$	9
Composantes	$\mathbf{c}_k = X\mathbf{v}_k$, variance λ_k	9
Inertie	$\tau_k = \lambda_k / \text{tr}(S)$	9

Résoudre — le pivot de Gauss, du système à la base.

Résoudre — le pivot de Gauss, du système à la base.
Mesurer — le produit scalaire, la corrélation comme cosinus.

Résoudre — le pivot de Gauss, du système à la base.
Mesurer — le produit scalaire, la corrélation comme cosinus.
Synthétiser — le théorème spectral, l'ACP.

Les trois problématiques

Résoudre — le pivot de Gauss, du système à la base.

Mesurer — le produit scalaire, la corrélation comme cosinus.

Synthétiser — le théorème spectral, l'ACP.

Une seule opération Les trois se ramènent à des **combinaisons linéaires**.