

# Analyse en composantes principales (ACP)

## TD/TP 1

---

### 1 TD

On considère le nuage de points dans  $\mathbb{R}^2$  constitué des points suivants :

$$M_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad M_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad M_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad M_4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad M_5 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

**Question 1.** Statistiques de base.

1. Donner la matrice de données associées à ce nuage de points
2. En considérant que la première coordonnée des points est associée à une variable  $x_1$  et la seconde à une variable  $x_2$ , donner les moyennes  $\bar{x}_1$ , et  $\bar{x}_2$
3. Donner la matrice de données associée aux données centrées
4. Calculer  $C$  la matrice de covariance des variables.

**Question 2.** Diagonalisation et Inertie.

1. Diagonaliser la matrice de covariance  $C$ .  
(On notera  $v_1$  et  $v_2$  les vecteurs propres associés, et  $\lambda_1$  et  $\lambda_2$  leur valeur propre respective.)
2. On note  $I_{v_i}$  l'inertie associée à un axe dirigé par un vecteur propre  $v_i$ . Vérifier les équations suivantes :

$$I_{tot} = Var(x_1) + Var(x_2) \tag{1}$$

$$= I_{v_1} + I_{v_2} \tag{2}$$

$$= \lambda_1 + \lambda_2 \tag{3}$$

$$I_{v_i} = v_i^* C v_i = \lambda_i \tag{4}$$

**Question 3.** Tracer le nuage de points, ainsi que les axes principaux.

**Question 4.** Calculer le taux d'inertie expliqué par chaque composante

**Question 5.** Projection On note  $P_{21}$  de la projection du point  $M_2$  sur le premier axe principal. Calculer ses coordonnées. Sachant qu'ici  $P_{ij} = (M_i | v_j) v_j$ .

**Question 6.** Qualité de représentation.

Pour savoir si un point  $i$  est loin de l'axe  $j$  on mesure sa "qualité de représentation" par son projeté sur l'axe  $j$  par

$$Q_{ij} = \cos^2(\widehat{M_i G P_{ij}}) = \frac{\|GP_{ij}\|^2}{\|GM_i\|^2}. \tag{5}$$

1. Calculer la qualité de représentation de  $M_2$  sur le premier axe principal.
2. Ce point est-il bien représenté par son projeté ?

**Question 7.** Point supplémentaire.

Soit  $M_6 = \begin{pmatrix} 0 \\ 10 \end{pmatrix}$

1. Calculer la qualité de représentation de  $M_6$  sur le premier axe principal.
2. En comparant  $Q_{21}$  et  $Q_{61}$ , dites si le point  $M_6$  est-il mieux représenté par son projeté sur le premier axe principal que  $M_2$  ?

**Question 8.** Stabilité de l'ACP.

En ajoutant le point  $M_6$  à nos données, tracer (à vu de nez) les axes factoriels associées au nouveau nuage. Sont-ils similaires ou très différents ? En repensant aux formules sur l'inertie et la covariance, pouvez-vous expliquer pourquoi ?

## 2 TP

Dans ce TP vous devrez recoder la décomposition en composantes principales en dimension  $p > 2$  (sans utiliser de package R associé à l'ACP). Puis l'appliquer sur le jeu de données *data\_PDE20.csv* disponible sur campus, en présentant les résultats. Une phase de prétraitement des données sera donc nécessaire (données manquantes ? aberrantes ? type de données ? etc...).

Votre code devra retourner :

- Les vecteurs directeurs des composantes principales
- L'inertie de chaque composante principale
- Les taux d'inertie associés aux  $k$  premières composantes principales
- Les nouvelles coordonnées des points dans le repère ACP