



Licence d'Informatique 3, AD Programmeur (C5-160512-INFO) TD – Analyse en Composantes Principales

Carl FRÉLICOT – Dpt Info / Lab MIA

0. Tableau de Données

[\[TD précédent\]](#)

$$X = \begin{bmatrix} 3 & -1 \\ -1 & 2 \\ 4 & 2 \end{bmatrix} \quad X' = \begin{bmatrix} 1 & -2 \\ -3 & 1 \\ 2 & 1 \end{bmatrix} \quad V = \begin{bmatrix} 14/3 & -1 \\ -1 & 2 \end{bmatrix} \quad X'' = \begin{bmatrix} 0.4629 & -1.4142 \\ -1.3887 & 0.7071 \\ 0.9258 & 0.7071 \end{bmatrix} \quad R = \begin{bmatrix} 1 & -0.3273 \\ -0.3273 & 1 \end{bmatrix}$$

1. Composantes Principales – Solution

$$\begin{matrix} v_1 & v_2 \\ \begin{bmatrix} -3 & -1 \\ 1 & -3 \end{bmatrix} \end{matrix}$$

Le tableau des vecteurs propres de V est ci-contre.

- 1-1) Faut-il réaliser une ACP normée ?
- 1-2) Déduisez les valeurs propres de V .
- 1-3) La 1ère composante principale permet-elle de bien représenter le nuage X ?
- 1-4) Calculez le tableau des vecteurs unitaires U définissant les composantes principales.
- HW) Vous vérifierez (par produit matriciel) qu'ils sont bien unitaires.
- HW) Vous vérifierez qu'ils sont V -orthogonaux (par produit matriciel).
- 1-5) Projetez x'_2 sur u_1 , puis généralisez à tous les individus de sorte de définir C'_1 .
- 1-6) Généralisez à toutes les composantes, c'est-à-dire projetez X' sur U , afin d'obtenir le tableau C' .
- HW) Vous calculerez (avec la formule pratique en dim. 1) la variance de chaque composante principale.
- 1-7) Soit $D_n = I_n/n$. Calculez ${}^t C' D_n C'$.
- 1-8) Par quel tableau faut-il diviser C' (élément par élément) pour le réduire ?
- 1-9) Déduisez le tableau C'' des composantes principales normées.
- HW) Vous calculerez ${}^t C'' D_n C''$. Était-il utile de faire ce calcul ?

2. Composantes Principales – Analyse des Variables

- 2-1) Tracez un cercle de rayon 1, centré à l'origine.
- 2-2) Projetez X''_1 le long de C''_2 .
- 2-3) Généralisez (par produit matriciel) de sorte de projeter le nuage X'' dans l'espace des composantes normées. Qu'obtient-on ?
- 2-4) Placez les variables réduites dans le cercle des corrélations (plan 1-2).
- 2-5) Pourquoi sont-elles au bord du cercle ?
- 2-6) Quantifiez la qualité de représentation de X_2 par C'_1 , puis C'_2 , et enfin $\{C'_1, C'_2\}$. Commentez cette dernière.

3. Composantes Principales – Analyse des Individus

- 3-1) Tracez un repère de base $\{u_1, u_2\}$ et placez les individus projetés.
- 3-2) À la question 2.3) du TD-1, vous avez calculé les normes des individus en prenant la racine carrée de la diagonale du produit matriciel $X {}^t X$. Calculez de la même façon celles des individus centrés.
- 3-3) Obtenez-vous le même résultat ? Pourquoi ?
- 3-4) Pourquoi aurait-on pu aussi les obtenir à l'aide de $C' {}^t C'$?
- HW) Vous calculerez $C' {}^t C'$ pour vérifier.
- 3-5) Quantifiez la qualité de représentation de x_2 sur C'_1 . Aurait-on pu se prévoir ce résultat ?
- 3-6) Pouvez-vous proposer une construction (matricielle et/ou tableau) permettant de généraliser à tous les individus et toutes les composantes ?

L'Analyse en Composantes Principales des ces données n'a aucun d'intérêt pour des raisons assez évidentes :

- on ne sait pas ce que sont les variables, ni pour quels individus on les a observées ; il est alors impossible de donner du sens aux composantes
- n est petit et surtout $p = 2$; à quoi bon réduire une si petite dimensionnalité ?

Si on se place du seul point de vue technique, on peut néanmoins se prêter à l'exercice consistant à :

- (HW) choisir combien de composantes retenir
(HW) pour chacune d'entre elle, choisir (et positionner) quel(s) individus et quelle(s) variables utiliser pour l'interpréter.