

# SY19 - TP01

## Positionnement multidimensionnel

Alice Ngwembou - Antoine Hars

October 15, 2013

### Exercice 1 : Exercice théorique

#### Première partie : ACP

##### Question 1 :

Afin de calculer les axes factoriels de l'ACP du nuage de points définis, nous centrons, dans un premier temps, la matrice donnée, ce qui nous permet de prendre comme nouvelle origine le centre de gravité dans l'espace des individus.

Nous calculons ensuite la matrice de covariance  $S$  pour appliquer la fonction `eigen()` sur cette dernière, ce qui nous donne les résultats suivants :

	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	$\lambda_6$	$\lambda_7$	$\lambda_8$
Valeurs propres	24.26	0.006	3.85e-16	3.65e-17	1.52e-19	-1.84e-19	-7.17e-16	-1.50e-15
Inerties expliquées	98.45%	100%	100%	100%	100%	100%	100%	100%

Tableau des valeurs propres et obtenues sur le nuage de points définis.

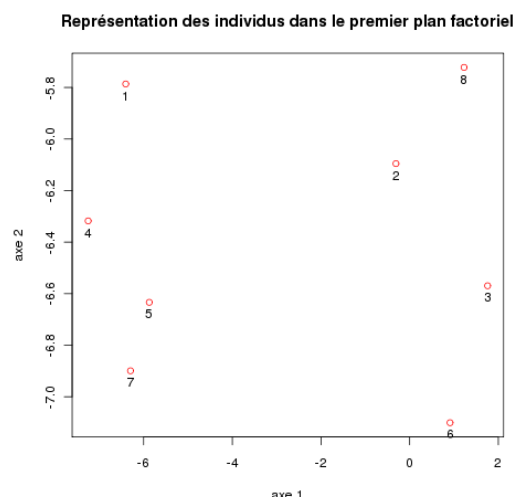
D'après les valeurs obtenues, nous pouvons dire que seules les 2 premières valeurs de valeur propres sont intéressantes car les valeurs suivantes sont excessivement proches de zéro (positives ou négatives) mais nous veillons à calculer les inerties expliquées en nous basant sur la somme des valeurs absolues des valeurs propres.

Nous pouvons donc remarquer que les 2 premiers axes factoriels, donc le plan factoriel défini par ces 2 axes, cumulent environ 100% de l'information.

##### Question 2 :

La matrice de composantes principales nous permettent de représenter les 8 individus dans le premier plan factoriel :

$$C = X * M * U = \begin{pmatrix} -3.62 & 0.60 \\ 2.47 & 0.29 \\ 4.54 & -0.18 \\ -4.47 & 0.07 \\ -3.09 & -0.24 \\ 3.69 & -0.71 \\ -3.52 & -0.51 \\ 4.01 & 0.67 \end{pmatrix}$$



L'ACP applique un changement de base au jeu de données car on a cherché à obtenir une représentation fidèle du nuage de points en le projetant sur un espace de plus faible dimension (dans notre cas, de dimension 2) sans perdre trop d'informations. Les variables obtenues par la projection sont donc des combinaisons linéaires des variables du nuage initial.

**Question 3 :**

$\Sigma_{\alpha=1}^k c_{\alpha} u'_{\alpha}$  pour  $k = 1$  et  $k = 2$  nous permet de retrouver notre matrice  $X$  de départ, à savoir la matrice contenant le jeu de données :  $X = \Sigma_{\alpha=1}^k c_{\alpha} u'_{\alpha}$ .

**Deuxième partie : MDS****Question 1 :**

Le calcul de  $D^2$  nous donne le tableau suivant (utilisation de la méthode R `dist(X, method = "euclidian")`):

$$D^2 = \begin{vmatrix} 37.25 & & & & & & \\ 67.25 & 4.50 & & & & & \\ 1.00 & 48.25 & 81.25 & & & & \\ 1.00 & 31.25 & 58.25 & 2.00 & & & \\ 55.25 & 2.50 & 1.00 & 67.25 & 46.25 & & \\ 1.25 & 36.50 & 65.00 & 1.25 & 0.25 & 52.00 & \\ 58.25 & 2.50 & 1.00 & 72.25 & 51.25 & 2.00 & 58.00 \end{vmatrix}$$

**Question 2 :**

Pour calculer la matrice des produits scalaires  $W$ , 2 méthodes sont possibles :

- En utilisant la matrice centrée des données de départ  $X$  :  $W = X * X^t$
- En partant du tableau des distances euclidiennes  $D^2$  :  $W = -\frac{1}{2} * Q_n * D^2 * Q_n$  avec  $n = 8$

**Question 3 :**

Pour déterminer si  $W$  ou  $\frac{1}{n}W$  est semi-définie positive, nous devons observer si les valeurs propres de la matrice  $\frac{1}{n}W$  sont positives :

Le calcul des valeurs propres nous donne le même résultat que celui trouvé à partir de la matrice de covariance  $S$  dans la première partie. Nous pouvons donc observer que nous avons 5 valeurs positives (dont 3 valeurs comprises entre  $10^{-16}$  et  $10^{-19}$ ) et 3 valeurs négatives (comprises entre  $10^{-19}$  et  $10^{-15}$ ), nous choisissons donc de les considérer comme nulles vu qu'elles sont minimales (vu que les 2 premières valeurs propres regroupent environ 100% de l'information).

Donc vu que nous disposons de valeurs positives et nulles, la matrice  $W$  est semi-définie positive.

**Question 4 :**

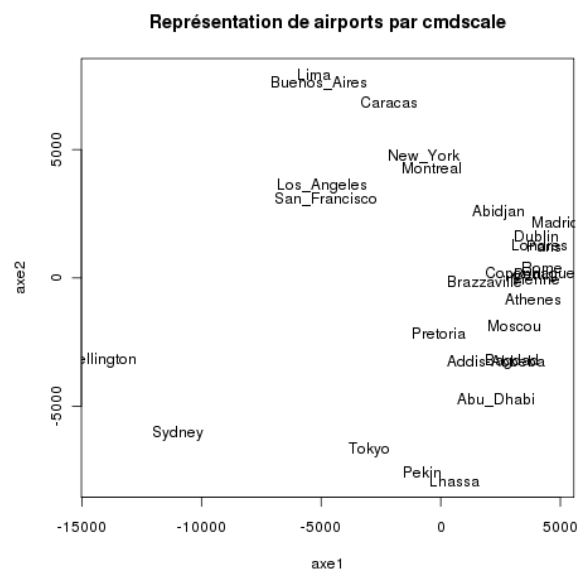
La matrice diagonale des valeurs propres et obtenue

**Question 5 :****Question 6 :****Question 7 :****Exercice 2 : Les données de mutations****Question 1 :****Question 2 :****Question 3 :****Exercice 3 : Les données de distances entre aéroports**

Dans cet exercice, nous avons travaillé sur les données contenues dans le fichier `airport2.txt` qui représentent les distances de vol entre des aéroports situés dans le monde. L'objectif de cet exercice était d'appliquer les différentes méthodes de positionnement multidimensionnel à ces données et de comparer les résultats obtenus.

**Question 1 :**

L'exécution de la fonction `cmdscale` sur le tableau de distance des données nous a donné le graphique suivant :



Question 2 :

Question 3 :