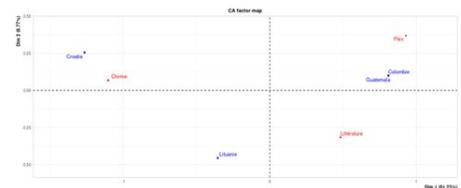


M2_MIAGE_1

1. AFC sur données "Prix Nobel" : 8 prix équitablement répartis dans 4 pays.

- A Les lituaniens obtiennent plutôt des prix en littérature.
- B Les croates obtiennent plutôt des prix en chimie.
- C Il y a tellement peu de données qu'il est hasardeux de tirer des conclusions.

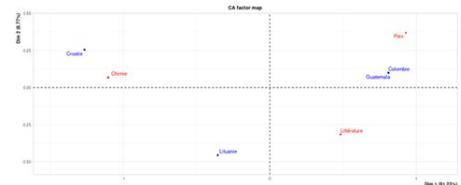


2. AFC sur données "Prix Nobel" : 8 prix équitablement répartis dans 4 pays.

Les points représentant la Colombie et le Guatemala sont confondus.

Quelles en sont les raisons possibles ?

- A Leurs profils sont égaux (même nombre de prix obtenus dans les mêmes catégories).
- B Leurs profils sont proportionnels (par exemple la Colombie a exactement 2 fois plus de prix que le Guatemala, dans les mêmes catégories).
- C On parle espagnol dans ces deux pays.



3. En AFC (Analyse Factorielle des Correspondances), si un individu-ligne se trouve représenté vers le centre du nuage, alors son profil est proche du profil moyen (moyennes sur les colonnes). La réciproque est également vraie.

- T True
- F False

4. On réalise une ACP sur un jeu de données et obtient les inerties cumulées suivantes :

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	4.44033139	49.3370155	49.33702
comp 2	2.04755648	22.7506275	72.08764
comp 3	0.89673574	9.9637304	82.05137
comp 4	0.62468682	6.9409647	88.99234
comp 5	0.54127118	6.0141242	95.00646
comp 6	0.26018411	2.8909345	97.89740
comp 7	0.11251704	1.2501893	99.14759
comp 8	0.05387136	0.5985706	99.74616
comp 9	0.02284589	0.2538432	100.00000

Combien d'axes doit-on garder au minimum pour retenir au moins 85% de l'information ?
(= 85% de l'inertie)

- (A) 2
- (B) 3
- (C) 4
- (D) 5

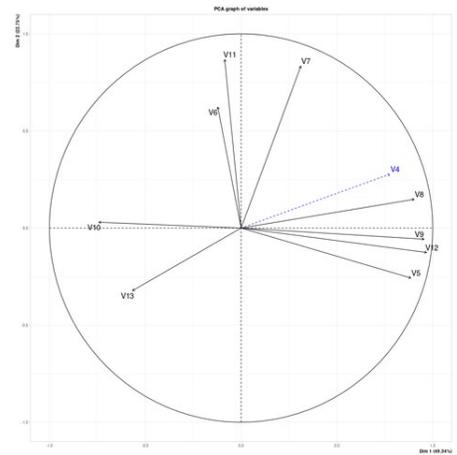
5. Effectuer une ACP (Analyse en Composantes Principales) revient à rechercher un système d'axes deux à deux orthogonaux minimisant l'inertie projetée.

- (T) True
- (F) False

6. ACP sur un jeu de données "pollution à l'ozone" journalières à Los Angeles en 1976, sur un peu plus de 6 mois.

Légende :

- 4 Daily maximum one-hour-average ozone reading
- 5 500 millibar pressure height (m) measured at Vandenberg AFB
- 6 Wind speed (mph) at Los Angeles International Airport (LAX)
- 7 Humidity (%) at LAX
- 8 Temperature (degrees F) measured at Sandburg, CA
- 9 Temperature (degrees F) measured at El Monte, CA
- 10 Inversion base height (feet) at LAX
- 11 Pressure gradient (mm Hg) from LAX to Daggett, CA
- 12 Inversion base temperature (degrees F) at LAX
- 13 Visibility (miles) measured at LAX



Note : "Inversion base height" = altitude de la zone à partir de laquelle la température diminue (certains jours, environ les 2/3 du temps à LA).

- (A) Les températures à Sandburg et El Monte (V8, V9) sont très corrélées.
- (B) La moyenne horaire maximale du niveau d'ozone (V4) semble augmenter quand la visibilité depuis l'aéroport de Los Angeles (V13) augmente.
- (C) Le gradient de pression (V11) semble plutôt décorrélié de la moyenne horaire maximale du niveau d'ozone (V4).
- (D) La vitesse du vent à l'aéroport de LA (V6) est très corrélée au gradient de pression (V11).

7. Quel est l'objectif (global) du clustering ?

- (A) Trouver des groupes d'individus qui se ressemblent (entre eux dans un cluster), tout en étant éloignés des autres groupes.
- (B) Agréger petit à petit n groupes jusqu'à n'en obtenir qu'un.
- (C) Diviser un jeu de données (n lignes) en K parties de même taille (environ n/K donc).

8. Question clustering.

Redéfinir les distances à partir d'un graphe permet dans certains cas d'obtenir une meilleure partition (les distances correspondant mieux à la géométrie des données).

- (T) True
- (F) False

9. Question clustering. On se place dans le cas où les vrais groupes sont inconnus. Un critère "externe" (clusterCrit::extCriteria) permet de mesurer...

- A** la ressemblance entre deux partitions.
- B** la stabilité d'un algorithme de clustering.
- C** la performance (précision) d'un algorithme.
- D** l'adéquation d'un algorithme de clustering avec un autre.

10. On coupe le dendrogramme ci-contre au niveau du trait bleu. Combien de clusters obtient-on, et comment sont-ils constitués ? (Notation R)

- A** 4 groupes : 1:2 - 3:8 - 9:16 - 17:20
- B** 4 groupes : 1:8 - c(9,10,15,16) - 17:20 - 11:14
- C** 3 groupes : 1:8 - 9:16 - 17:20
- D** 3 groupes : 3:8 - 9:16 - c(1,2,17:20)

