

# Fondements statistiques pour la science des données

## TP n°2. L'Analyse en Composantes Principales (ACP)

### 1 L'ACP des données « température »

1. Importer le jeu de données `temperature`.
2. Choisir la colonne `Ville` comme identificateur des individus : *Données* → *Jeu de données actif* → *nom des cas*.
3. Réaliser l'ACP (*FactoMineR* → *Principal Components Analysis*) du tableau des températures en faisant les choix suivants :
  - variables actives = les 12 mois ;
  - ACP normée : *scaled* ;
  - Outputs : sélectionner l'ensemble des items.
4. Quelle est l'inertie brute associée au premier axe factoriel ? Au second axe ? Quelle est l'inertie totale ? (vérifier votre résultat en exécutant la commande R qui permet de calculer la somme des valeurs propres.) Quel est le pourcentage d'inertie associé au premier plan factoriel ?
5. Représenter à l'aide de deux diagrammes en bâtons : la décroissance de l'inertie des axes ("éboulis des valeurs propres"), puis le pourcentage d'inertie cumulé.
6. Produire les corrélations des variables avec l'axe 1 et les trier de façon décroissante. Idem pour l'axe 2. Quelle est la variable la plus corrélée à l'axe 1 ? à l'axe 2 ?
7. Retrouver par le calcul la contribution de la variable la plus corrélée avec l'axe 1.
8. Quel est l'individu dont la qualité de représentation sur le plan (1,2) est la plus élevée ? la moins élevée ?
9. Quel est l'individu ayant le plus contribué à la formation de l'axe 1 ? le moins contribué ?
10. Quel est l'individu ayant le plus contribué à la formation de l'axe 2 ? le moins contribué ?
11. Trier les individus du plus éloigné au plus proche du centre de gravité du nuage.
12. Retrouver la valeur de la distance au centre de gravité de Strasbourg à partir de sa qualité de représentation et sa coordonnée le long d'un axe.
13. Découper la variable `Latitude` en 3 classes de même effectif avec les libellés *nord*, *centre*, *sud* comme nom de niveaux. Idem pour la variable `Longitude` (avec les libellés *ouest*, *centre*, *est* ). Appeler les nouvelles variables `latiClasse`, `longiClasse`.

14. Réaliser à nouveau l'ACP du tableau des températures en faisant maintenant en plus les choix suivants :
  - variables illustratives quantitatives (supplementary variables) = latitude, longitude, amplitude, moyenne ;
  - variables illustratives qualitatives (supplementary factors) = latiClasse, longiClasse, Coul.Pol, Littoral

On pourra également utiliser les options graphiques consistant à représenter d'une même couleur les individus possédant la même modalité (*coloring for individuals* = *LatiClasse*, par exemple).
15. Interpréter les deux premiers axes factoriels.

## 2 Utilisation du package Factoshiny

1. Installer puis charger le package **Factoshiny**.
2. Appliquer la fonction **PCAshiny** au résultat de l'ACP (voir cours.)

À partir des différents onglets proposés par **Factoshiny**, retrouver les différentes sorties graphiques et numériques suivantes :

3. Le jeu de données ; trier les villes par températures décroissantes en Janvier.
4. Le tableau des inerties et l'éboullis des valeurs propres associé.
5. Les résultats numériques des individus et variables actives (coordonnées, contributions, cos2.)
6. Les résultats des variables illustratives quantitatives puis qualitatives. Repérer l'indicateur **V-test** qui permet de conclure au caractère significatif de la position d'une modalité le long d'un axe.
7. Sur le cercle des corrélations, représenter les variables dont la qualité de représentation sur le plan dépasse 0.95. Faire de même pour le plan des individus.
8. Représenter les villes selon la couleur politique du maire. Représenter des ellipses de confiance autour des modalités. Idem pour la variable Littoral.

## 3 Application « Café »

Un jury d'experts a réalisé une évaluation sensorielle de sept cafés ( $i = 1, \dots, 7$ ) sur la base de 15 attributs sensoriels ( $j = 1, \dots, 15$ ). Dans une case  $(i, j)$  du tableau figure la note moyenne accordée par les experts pour un café et un attribut donné. Par ailleurs, ces cafés ont été dégustés puis notés par 84 consommateurs sur une échelle d'appréciation allant de 1 à 5. L'appréciation moyenne des 84 consommateurs est associée à la variable **Q1t.Glob**. Ces données sont réunies dans le fichier **Café.txt**.

1. Importer le jeu de données **Café.txt**
2. Choisir la colonne **Café** comme identificateur puis supprimer la colonne **Produit** du jeu de données.
3. Réaliser l'ACP normée en choisissant : comme variables actives les 15 attributs sensoriels ; comme variables supplémentaires la qualité globale.

4. Construire le graphe de l'éboulis des valeurs propres.
5. Combien d'axes factoriels vous semblent-ils intéressants à interpréter? Justifier à l'aide du critère de Kaiser.
6. Interpréter les deux premiers axes factoriels.
7. Ajouter enfin comme variables supplémentaires les notes d'appréciation des 84 consommateurs puis interpréter les résultats.

## 4 Application « Appareil photo »

Un magazine a passé au banc d'essai 27 appareils photo de milieu de gamme à objectifs interchangeable : 11 réflex et 16 hybrides. Les caractéristiques principales suivantes ont été relevées pour chacun des appareils :

- Capteur (résolution)
- Focale minimale
- Focale maximale
- Grossissement (zoom optique)
- Écran (résolution)
- Poids (avec batterie)
- Prix
- Écran tactile (oui/non)
- Présence d'un viseur (oui/non)

Suite à une batterie de tests réalisés par des experts, des notes (de 1 à 5) ont été attribuées aux différents appareils concernant les aspects suivants :

- Qualité globale
- Qualité photo
- Qualité vidéo
- Qualité écran
- Définition
- Vignettage
- Distorsion
- Qualité Flash
- Vidéo en lumière jour
- Vidéo en lumière faible
- Son de la vidéo
- Temps au démarrage
- Délai de déclenchement
- Délai entre deux prises de vue

Les données obtenues sont présentes dans le fichier `photo.xlsx`.

À l'aide d'une ACP, on souhaite effectuer une typologie de ces appareils photo du point de vue de leurs caractéristiques principales et on recherchera dans quelle mesure ces dernières sont liées aux notes d'appréciation fournies par les experts.