

Analyse exploratoire

TP numéro 2.

Analyse Factorielle des Correspondances (AFC)

Élections présidentielles 2012

Le jeu de données à analyser représente les résultats du premier tour de l'élection présidentielle 2012. Le tableau comporte en lignes les départements (métropole et outremer) et en colonnes les différents candidats.

Dans chaque case figure le nombre de voix obtenues par un candidat au sein d'un département. Le tableau inclut également le nombre total de voix de chaque candidat et le total de bulletins exprimés dans chaque département.

1. Importer le jeu de données `President12.txt`. Dans la suite, le jeu de données importé sous R est appelé `president12`.

Déclarer le département comme identificateur des individus (*nom des cas* dans R).

2. Représenter par un diagramme en bâtons les nombres de voix obtenues par les différents candidats dans les départements du Bas-Rhin, de Paris, de la Seine Saint-Denis, du Vaucluse.

Exemple de commande pour le département de l'Ain :

```
> barplot(as.matrix(president12[1,2:11]),names.arg=colnames(president12)[2:11],
col="purple3", border="white", main="Ain", horiz=T, las=1,
xlab="nombre de voix", cex.lab=1.2)
```

- ~~3. Modifier l'espace de la marge de gauche afin de visualiser les libellés des candidats intégralement.~~

~~**Indication :** la commande~~

~~`> par("mar")`~~

~~renvoie dans l'ordre les espaces par défaut attribués aux marges en bas, à gauche, en haut, à droite.~~

~~Essayer alors la commande :~~

~~`> par(mar = c(6,10,4,2))`~~

~~Puis soumettre à nouveau la commande `barplot`.~~

4. Construire :

- le **profil-ligne** du Bas-Rhin (tableau des fréquences).
- le graphique représentant le profil-ligne du Bas-Rhin.

5. Réaliser le test du χ^2 permettant d'étudier le lien de dépendance entre les 96 départements de France métropolitaine et les dix candidats.

Indication : effectuer la commande suivante :

```
> resu.chi2 = chisq.test(president12[1:96,2:11])
```

- ~~– Pour savoir quels sont les résultats fournis par ce test, taper la commande :~~
- ~~> summary(resu.chi2)~~
- Extraire des résultats de ce test : la *p-value* puis les données attendues en cas d'indépendance.
- Quel est le profil ligne attendu d'un département quelconque sous l'hypothèse d'indépendance ?
- 6. Effectuer l'analyse factorielle des correspondances (AFC) du tableau avec les choix suivants :
 - lignes actives = les 96 départements de la France métropolitaine
 - colonnes actives = les 10 candidats
 - ~~– colonnes supplémentaires = TotCand~~
 - ~~– lignes supplémentaires = TotDep~~
- 7. Créer un diagramme en bâtons pour étudier la décroissance de l'inertie des axes :


```
> barplot (res$eig[,2], names=paste("Dim",1:length(res$eig[,2])),
main="Inertie (en %) des axes factoriels", col="orange", border="white")
```
- 8. Retrouver l'inertie totale de deux façons :
 - à partir du test du χ^2 ;
 - en calculant la somme des inerties de tous les axes factoriels issus de l'AFC.
 - En déduire que la valeur du *V* de Cramer est égale à 0.063.
- 9. Calculer la valeur propre moyenne. En déduire le nombre d'axes dont l'inertie est supérieure à l'inertie moyenne d'un axe.
- ~~10. Déterminer le nombre d'axes potentiellement intéressants selon le critère du bâton brisé (*broken stick*).~~
- 11. Produire le premier plan factoriel (axes 1 et 2) des seuls individus (départements).
Remarque : pour rendre le graphique plus lisible, il est possible de réduire la taille des libellés des départements en soumettant à nouveau les commandes de l'AFC après avoir ajouté l'argument `cex=0.7` à la fin de la commande `plot.CA`.
- 12. Produire le premier plan factoriel des colonnes actives (candidats) et supplémentaires.
- 13. Produire le premier plan factoriel avec la totalité des lignes et colonnes (actives et supplémentaires).
- ~~14. Pourquoi l'élément supplémentaire TOTALDept ne se positionne-t-il pas à l'origine ?~~
- 15. Quels sont les départements ayant le plus fortement contribué à la construction de l'axe 1 ? de l'axe 2 ? Pour obtenir les contributions triées par ordre décroissant des axes 1 et 2, on peut soumettre les commandes :


```
> sort(res$row$contrib[,1], T)
> sort(res$row$contrib[,2], T)
```
- 16. Faire de même pour les candidats.
- 17. Quels sont les départements et les candidats ayant le plus fortement contribué à l'inertie globale ?
- 18. Calculer la distance de chacun des départements au centre de gravité du nuage.
Comment interpréter la distance la plus courte ?
- 19. Interpréter les axes factoriels 1 et 2.
- 20. Refaire l'AFC avec les départements et territoires d'outremer en éléments illustratifs (supplémentaires). Interpréter la position de ces derniers sur le plan factoriel (valider votre commentaire en construisant les graphiques des profils-lignes de ces nouveaux départements.)