

Fondements statistiques pour la science des données

TP n°3. Analyse factorielle des correspondances simples (AFC) et multiples (ACM)

1 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 outre-mer) 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 département de France métropolitaine et les dix candidat. Indication : effectuer la commande suivante :

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

 - Pour savoir quels sont les résultats fournis par ce test, tapez 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. Retrouver l'inertie totale de deux façons :
 - à partir du test du χ^2 ;
 - à partir des résultats issus de l'AFC.
- 8. Calculer la valeur propre moyenne. En déduire le nombre d'axes dont l'inertie est supérieure à l'inertie moyenne d'un axe.
- 9. Consulter les résultats de l'AFC à l'aide de la fonction `CAshtiny` du package `Factoashiny` (ne pas oublier de charger le package !)
- 10. À l'aide de l'option *invisible elements*, construire les plans factoriels suivants :
 - le premier plan factoriel (axes 1 et 2) des lignes actives et supplémentaires. Pourquoi l'élément supplémentaire `TOTALDept` ne se positionne-t-il pas à l'origine ?
 - Le premier plan factoriel des colonnes actives (candidats) et supplémentaires.
 - Le premier plan factoriel avec la totalité des lignes et colonnes (actives et supplémentaires).
- 11. Quels sont les départements ayant le plus fortement contribué à la construction de l'axe 1 ? de l'axe 2 ?
- 12. Faire de même pour les candidats.
- 13. Interpréter les axes factoriels 1 et 2.
- 14. 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.)

2 Exemple “Enquête OGM”

Une enquête a été réalisée auprès de 135 personnes (adultes françaises) sur le thème des organismes génétiquement modifiés (OGM). Deux ensembles de questions ont été posées :

(I) Un ensemble de 16 questions portant sur la position des enquêtés vis-à-vis des OGM :

- *Vous sentez-vous concernés par la polémique sur les OGM (beaucoup, moyennement, un peu, pas du tout) ?*
- *Quelle est votre position quant à la culture d'OGM en France (très favorable, favorable, plutôt défavorable, pas favorable du tout) ?*

- Quelle est votre position quant à l'incorporation de matières premières OGM dans les produits alimentaires destinés à l'alimentation humaine (très favorable, favorable, plutôt défavorable, pas favorable du tout) ?
- Quelle est votre position quant à l'incorporation de matières premières OGM dans les produits alimentaires destinés à l'alimentation animale (très favorable, favorable, plutôt défavorable, pas favorable du tout) ?
- Avez-vous déjà participé à une manifestation contre les OGM (oui, non) ?
- Estimez-vous que les médias communiquent suffisamment sur le sujet (oui, non) ?
- Faites-vous vous-même la démarche de vous informer sur le sujet (oui, non) ?
- Pensez-vous que les OGM puissent permettre la réduction d'usage de fongicides (oui, non) ?
- Pensez-vous que les OGM puissent permettre la réduction des problème de famine dans le monde (oui, non) ?
- Pensez-vous que les OGM puissent permettre l'amélioration des conditions de vie des agriculteurs (oui, non) ?
- Pensez-vous que les OGM puissent permettre de futurs progrès scientifiques (oui, non) ?
- Pensez-vous que les OGM représentent un éventuel danger pour notre santé (oui, non) ?
- Pensez-vous que les OGM représentent une menace pour l'environnement (oui, non) ?
- Pensez-vous que les OGM représentent un risque économique pour les agriculteurs (oui, non) ?
- Pensez-vous que les OGM représentent un procédé scientifique inutile (oui, non) ?
- Pensez-vous que nos grands-parents avaient une alimentation plus saine (oui, non) ?

(II) Un second groupe de 5 questions portant sur le profil socio-démographique des enquêtés :

- Sexe (masculin, féminin)
- Catégorie socio-professionnelle (agriculteur, étudiant, ouvrier, cadre, fonction publique, libéral, technicien, commerçant, autre actif, non actif, retraité)
- Âge (-25 ans, 25-40, 40-60, +60)
- Exercez-vous des études ou un métier en rapport avec l'agriculture, l'industrie agroalimentaire ou la pharmaceutique (oui, non) ?
- À quel parti politique vous identifiez-vous le plus (extrême gauche, verts, PS, centre, UMP, Front National) ?

Questions

1. Importer le jeu de données `OGM.txt`
2. Fournir le résumé du jeu de données actif. Détecter les deux variables pour lesquelles les modalités *Très favorable* ont été très peu choisies les enquêtés (effectif inférieur ou égal à trois).
3. À l'aide du menu *Recoder les variables*, regrouper ces modalités avec la modalité *Favorable* en suivant la directive de codage suivante :
`"Très Favorable"="Favorable"`
4. Effectuer l'analyse des correspondances multiples avec les choix suivants :
Variables actives : les 16 variables liées à la position des enquêtés sur les OGM
Facteurs supplémentaires : les 5 variables socio-démographiques.
Outputs : tout cocher.

5. Consulter l'éboulis des valeurs propres en utilisant la fonction `MCashiny` du package `Factoshiny`.
6. Construire les graphiques suivants :
 - le plan (1,2) des seuls individus.
 - Le plan (1,2) des modalités actives. Comment s'appelle le phénomène observé sur le premier plan factoriel ?
 - La représentation des variables pour les axes 1 et 2.
7. Retrouver la valeur de l'inertie de l'axe 1 en calculant la moyenne des rapports de corrélation entre le premier facteur et les variables actives.
8. Construire le graphe des individus en coloriant ceux-ci en fonction de leur réponse à la variable *Position.Culture*. Faire de même pour d'autres variables, par exemple : *Parti.Politique*, *Sexe*, *Grds.Parents*.
9. Fournir les contributions triées des modalités actives.
10. Interpréter brièvement les axes 1 et 2.
11. Dessiner des ellipses de confiance autour des modalités des variables *Position.Culture*, *Parti.Politique*, *Sexe*, *Grds.Parents* :


```
plotellipses(res,keepvar=c("Position.Culture", "Parti.Politique", "Sexe",
"Grds.Parents"))
```
12. Éditer le jeu de données. Se positionner au niveau d'un nouvel individu n° 136. Répondre au questionnaire en saisissant vos réponses pour ce nouvel individu. Refaire l'ACM et sélectionner l'individu 136 comme individu supplémentaire. Observer alors sa position sur le plan factoriel.