L'analyse des correspondances

Cours SOL6210 Analyse quantitative avancée

© Claire Durand, 2024
Professeur titulaire,
Département de sociologie,
Université de Montréal

1

Benzécri

Énonce cinq principes de sa philosophie

- Le premier :
  - "*Statistique n'est pas probabilité. Sous le nom de statistique mathématique, des auteurs (qui, je vous le dis en français, n'écrivent guère dans notre langue...) ont édifié une pompeuse discipline, riche en hypothèses qui ne sont jamais satisfaites dans la pratique. Ce n'est pas de ces auteurs qu'il faut attendre la solution à nos problèmestypologiques*". Benzécri (1984, t.2 p. 3).
- Le 2ème: Le modèle doit suivre les données et non l'inverse.
- Le 3ème: Il faut traiter simultanément les informations concernant le plus grand nombre possible de dimensions.
- Les 4ème et 5ème ont trait à l'utilisation de l'ordinateur.

2

Ecole française d'analyse des données

Trois périodes de développement (Van Meter et coll., 1994)

- Émergence: 1963-1973
- Isolation: 1973-1981
  - Utilisation de Bourdieu popularise la méthode en France mais elle ne s'étend pas au monde anglo-saxon.
- Adoption réservée: à partir de 1981
  - La méthode fait son apparition dans les grands logiciels polyvalents américains (SPSS, SAS, etc.), éventuellement Stata, R.
  - Excellents logiciels français: Spad, DTM-Vic.

3

Deux types principaux d'analyse des correspondances

- Correspondances simples :
  - Entre DEUX variables seulement, qui peuvent/ doivent avoir plusieurs catégories. Exemple: départements de l'université et nombre d'étudiants de divers types dans chaque département.
- Correspondances multiples :
  - Entre plusieurs variables, de type nominal ou ordinal
  - Qui peuvent être *actives* ou *illustratives*

Les principaux éléments

- Utilisée surtout pour les variables de type nominal et ordinal, rarement pour les variables quantitatives (ou sinon après regroupement des valeurs en un nombre restreint de catégories).
- Les catégories de variables choisies par les mêmes personnes devraient se retrouver proche les unes des autres sur un plan.
- Les individus qui ont un profil de réponse similaire devraient se retrouver proche les uns des autres sur un plan.

Originalité

Il s'agit d'une méthode visuelle

- Les "individus" (cas ou objets en AFC) sont "positionnés" sur le même plan que les variables.
  - En pratique, on utilise cette possibilité seulement lorsqu'il y a peu de cas.
- Les variables peuvent être
  - Actives – et donc contribuer à définir l'espace du plan,
  - Ou illustratives – et donc être simplement projetées sur le plan en fonction de leur relation avec les variables et les cas définissant le plan.
- La procédure peut servir pour l'analyse textuelle et donc, pour l'analyse combinée de données textuelles et de sondage (voir Cibois, 1989)

Le vocabulaire, les termes

- Les notions sont les mêmes qu'en analyse factorielle "classique" mais les termes utilisés sont souvent spécifiques à ce type d'analyse, sans être des "nouveautés".
  - Inertie = variance
  - Axes ≈ facteurs mais aussi axes du plan géométrique
  - Masse: fréquence de la catégorie / fréquence totale x nombre de variables.
  - Modalités = catégories
  - Coordonnées: position des modalités de variables sur les axes
  - Contribution absolue: saturation factorielle, relation entre choix de modalité et axe
  - Contribution relative: proportion de la variance de la modalité expliquée par le facteur (l'axe).

7

Comment procéder?

De multiples logiciels gratuits

- Voir site web du cours pour les multiples logiciels gratuits disponibles -- voir entre autres lebart.org -- pour analyser soit des fichiers de données individuelles, soit des tableaux de contingence.
- DTM-VIC: Logiciel gratuit de Ludovic Lebart (Inclut commandes dans R).
- SPSS : Analyse de base, représentation graphique peu poussée.
- SPAD, logiciel français: Excellent pr graphiques
  - Voir possibilité d'acheter une version étudiante
  - Informations sur l'aide à l'interprétation incluses

8

Quelle information?

Comme en analyse factorielle...

- Le nombre de facteurs (axes) nécessaires pour expliquer la majeure partie de la variance "signifiante" (test du coude de Cattell ou test de l'éboulis)
- Que signifie chacun des axes (interprétation à partir des modalités qui sont reliées aux divers axes).
- Relations entre les modalités en fonction de leur positionnement sur le plan.

9

Quelle information?

Et plus

- La relation entre les modalités de variables illustratives et les modalités de variables actives (équivalent à une régression).
- A quelle modalité ressemble la modalité non répondue (valeur manquante)?
- Le positionnement des cas relativement aux modalités des variables.
- Éventuellement, l'utilisation de certains termes dans les questions ouvertes relativement à des réponses à des questions fermées.