

ACM

DJEBALI Wissam

October 12, 2018

Rappel

Il existe plusieurs techniques d'analyse factorielle dont les plus courantes sont :

- l'analyse en composante principale (ACP) portant sur des variables quantitatives,
- l'analyse factorielle des correspondances (AFC) portant sur deux variables qualitatives
- l'analyse des correspondances multiples (ACM) portant sur plusieurs variables qualitatives (il s'agit d'une extension de l'AFC).

Pour combiner des variables à la fois quantitatives et qualitatives, on pourra avoir recours à l'analyse mixte de Hill et Smith.

Bien que ces techniques soient disponibles dans les extensions standards de R, il est souvent préférable d'avoir recours à deux autres extensions plus complètes, `ade4` et `FactoMineR`, chacune ayant ses avantages et des possibilités différentes.

Voici les fonctions les plus fréquentes :

Analyse	Variables	Fonction standard	Fonction <code>ade4</code>	Fonctions <code>FactoMineR</code>
ACP	plusieurs variables quantitatives	<code>princomp</code>	<code>dudi.pca</code>	PCA
AFC	deux variables qualitatives	<code>corresp</code>	<code>dudi.coa</code>	CA
ACM	plusieurs variables qualitatives	<code>mca</code>	<code>dudi.acm</code>	MCA
Analyse mixte de Hill et Smith	plusieurs variables quantitatives et/ou qualitatives	—	<code>dudi.mix</code>	—

Principes générales

L'analyse des correspondances multiples est une technique descriptive visant à résumer l'information contenu dans un grand nombre de variables afin de faciliter l'interprétation des corrélations existantes entre ces différentes variables. On cherche à savoir quelles sont les modalités corrélées entre elles.

L'ACM est une extension de l'AFC. L'ensemble des individus peut être représenté dans un espace à plusieurs dimensions où chaque axe représente les différentes variables utilisées pour décrire chaque individu. Plus précisément, pour chaque variable qualitative, il y a autant d'axes que de modalités moins un. Ainsi il faut trois axes pour décrire une variable à quatre modalités. Un tel nuage de points est aussi difficile à interpréter que de lire directement le fichier de données. On ne voit pas les corrélations qu'il peut y avoir entre modalités, par exemple qu'aller au cinéma est plus fréquent chez les personnes habitant en milieu urbain. Afin de mieux représenter ce nuage de points, on va procéder à un changement de systèmes de coordonnées. Les individus seront dès lors projetés et représentés sur un nouveau système d'axe. Ce nouveau système d'axes est choisis de telle manière que la majorité des variations soit concentrées sur les premiers axes. Les deux-trois premiers axes permettront d'expliquer la majorité des différences observées dans l'échantillon, les autres axes n'apportant qu'une faible part additionnelle d'information. Dès lors, l'analyse pourra se concentrer sur ses premiers axes qui constitueront un bon résumé des variations observables dans l'échantillon.