Analyse des correspondances multiples

L'analyse des correspondances multiples (ACM) peut être présentée comme un prolongement de l'AFC. Elle permet la representation graphique de tableaux de fréquences contnant plus de deux variables. Un exemple classique d'un tableau de fréquences avec plus de deux variables qualitatives est un tableau présentant les réponses d'individus à un questionnaire contenant Q questions à choix multiples. L'ACM est donc très utile pour visualiser les résultats d'une étude par questionnaire.

L'ACM peut aussi être vue comme une version de l'ACP quand les variables sont mixtes, i.e. comprenant à la fois des variables quantitatives et des variables qualitatives. Le traitement conjoint de ces deux types de données repose sur leur transformation préalable appelée **codage disjonctif complet**.

0.1 Notation

Notons n le nombre d'individus (ou d'observations) et Q le nombre de variables (ou de questions dans le cas d'un questionnaire). Chaque variable possède J_q modalités et le nombre total de modalités est égal à J.

Définition

Le tableau binaire, i.e. ne contenant que des 0 et des 1, à n lignes et J colonnes est appelé **tableau de codage disjonctif complet**. On le note Z.

Ainsi, une variable n'est pas traitée telle quelle mais à travers ses modalités. Elle est découpée en modalités et tout individus est alors codé 1 pour la modalité qu'il possède et 0 dans les autres (i.e. qu'il ne possède pas, les modalités étant exclusives). Ce codage est immédiat pour des variables qualitatives. Cependant, pour une variable qualitative, on procède en découpant au préalable la variable en classes. Ainsi, chaque individu n'appartient qu'à une seule classe. Ce processus de transformation de l'information est appelé **codage disjonctif complet**. Il s'agit bien d'un codage, car l'information initiale est transformée, disjonctif, car tout individu possède au plus une modalité, et complet, car tout individu a au moins une modalité.

Exemple de transformation de tableau

Remarque

Lorsque l'on veut transformer des variables quantitatives en tableau de codage disjonctif complet, on perd de l'information. En effet, comme on doit découper les variables qualitatives en classes, l'appartenance à une classe est moins informatif qu'une valeur précise d'une variable.

En général, pour un questionnaire contenant Q questions, on a un tableau de la forme suivante :

$$Z = \left[Z_1 \mid \cdots \mid Z_Q\right].$$

Notation :

— Q : nombre de questions
— n : nombre d'individus répondant au questionnaire
— p_q : nombre de modalités (choix de réponses) de la question q.
— $p = p_1 + p_Q$

Potentiel problème : plus le nombre de questions est grand, plus il y aura de cellules vides. C'est aussi le cas si le nombre de réponses aux questions est important. La proportion de cellules non vides est

$$\frac{nQ}{np} = \frac{Q}{p}.$$

.

Si toutes les questions ont le même nombre de choix de réponses, alors $p_1 = \cdots = p_Q = \frac{p}{Q}$, de sorte que

$$\frac{Q}{p} = \frac{1}{p_1} \longrightarrow 0, \quad \text{quand } p_1 \to \infty.$$

Le tableau résumé est un tableau de taille $n \times Q$. Il contient le numéro de la modalité associée à la réponse de chaque individu pour chacune des questions.

La tableau de Burt est une autre façon de présenter un tableau de fréquences contenant plus de deux variables. Étant donné un tableau logique $Z = \left[Z_1 \mid \cdots \mid Z_Q\right]$, le tableau de Burt associé est la matrice carrée $p \times p$ définie comme étant

$$B = \begin{pmatrix} Z_1^\top Z_1 & \cdots & Z_1^\top Z_Q \\ \vdots & \ddots & \vdots \\ Z_Q^\top Z_1 & \cdots & Z_Q^\top Z_Q \end{pmatrix}.$$

Propriétés de $Z_q^\top Z_q$

1. $Z_q^\top Z_q$ est une matrice diagonale $p_q \times p_q$ présentant les réponses à la qe question.
2. L'élément (j, j) de la matrice $Z_q^\top Z_q$ est égal au nombre d'individus d_{jj} qui appar-

tiennent à la je catégorie de la qe question.

3. $Z_q^\top Z_r$ est le tableau de fréquences présentant les répones au x qe et re questions.
4. L'élément (j, k) de la matrice $Z_q^\top Z_r$ est égal au nombre d'individus d_{jk} qui appartiennent à la je catégorie de la qe question et à la ke catégorie de la re question.

D'un point de vue mathématique, l'ACM est une AFC effectuée sur la matrice logique Z ou sur le tableau de Burt B. On peut démontrer que l'on obtient les mêmes facteurs, et ce, peu importe la matrice utilisé pour l'analyse.

Note

On peut créé un graphique comme l'AFC. Cependant, en ACM, la distance entre les points de même couleur et la géométrie globale du graphique ne peuvent pas s'interpréter comme en AFC. En fait, on s'intéresse aux points qui sont dans une même direction par rapport à l'origine.