Formules de transition

But on cherche une relation entre les vecteurs $\mathbf{a}_k$ et $\mathbf{b}_k$ pour éviter de faire deux diagonalisation de matrice. Par exemple, si $m_1 < m_2$, on diagonalisera la matrice $\mathbf{D}_1^{-1}\mathbf{N}\mathbf{D}_2^{-1}\mathbf{N}'$.

Formules un calcul simple donne les formules suivantes

$$\mathbf{b}_k = \frac{1}{\sqrt{\lambda_k}}\mathbf{D}_2^{-1}\mathbf{N}'\mathbf{a}_k, \text{ soit } b_{jk} = \frac{1}{\sqrt{\lambda_k}}\sum_{i=1}^{m_1}\frac{n_{ij}}{n_{\cdot j}}a_{ik},$$

$$\mathbf{a}_k = \frac{1}{\sqrt{\lambda_k}}\mathbf{D}_1^{-1}\mathbf{N}\mathbf{b}_k, \text{ soit } a_{ik} = \frac{1}{\sqrt{\lambda_k}}\sum_{j=1}^{m_2}\frac{n_{ij}}{n_{i\cdot}}b_{jk}.$$

Méthode comme $\mathbf{a}_k$ est (à une normalisation près) le facteur principal associé à $\mathbf{b}_k$, on sait que $\mathbf{b}_k = \alpha\mathbf{D}_2^{-1}\mathbf{N}'\mathbf{a}_k$. Pour déterminer α, il suffit d'écrire que $\mathbf{b}_k'\frac{\mathbf{D}_2}{n}\mathbf{b}_k = \lambda_k$.

Le χ^2 d'écart à l'indépendance

Utilité Il permet d'évaluer la dépendance entre les variables.

Définition c'est la grandeur suivante (parfois aussi notée χ^2 ou X^2)

$$d^2 = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\frac{\left(n_{ij} - \frac{n_{i\cdot}n_{\cdot j}}{n}\right)^2}{\frac{n_{i\cdot}n_{\cdot j}}{n}} = n\left[\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}} - 1\right].$$

$d^2 = 0 \iff$ les variables sont indépendantes.

Contribution au χ^2 c'est le terme

$$\frac{\left(n_{ij} - \frac{n_{i\cdot}n_{\cdot j}}{n}\right)^2}{\frac{n_{i\cdot}n_{\cdot j}}{n}}$$

qui permet de mettre en évidence les associations significatives entre modalités de deux variables.

Borne supérieure comme $n_{ij} \leq n_{i\cdot}$, on a

$$\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}} \leq \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\frac{n_{ij}}{n_{\cdot j}} = \sum_{j=1}^{m_2}\frac{\sum_{i=1}^{m_1}n_{ij}}{n_{\cdot j}} = \sum_{j=1}^{m_2}\frac{n_{\cdot j}}{n_{\cdot j}} = m_2,$$

et donc $d^2 \leq n(m_2 - 1)$. On fait de même pour m_1 et

$$\varphi^2 = \frac{d^2}{n} \leq \min(m_1 - 1, m_2 - 1).$$

Dépendance fonctionnelle si $\varphi^2 = m_2 - 1$, alors pour chaque i soit $n_{ij} = n_{i\cdot}$, soit $n_{ij} = 0$: il existe une unique case non nulle par ligne. $\mathcal{X}_2$ est donc fonctionnellement liée à $\mathcal{X}_1$.

Dépendance inverse cette relation ne signifie pas que $\mathcal{X}_1$ est fonctionnellement liée à $\mathcal{X}_2$, sauf si $m_1 = m_2$. On peut alors représenter le tableau comme une matrice diagonale.

Caractère significatif du χ^2

Problème à partir de quelle valeur de d^2 doit-on considérer que les variables $\mathcal{X}_1$ et $\mathcal{X}_2$ sont dépendantes ?

Méthode on suppose que $\mathcal{X}_1$ et $\mathcal{X}_2$ sont issus de tirages de deux variables aléatoires indépendantes. On peut alors montrer que d^2 est une réalisation d'une variable aléatoire D^2 qui suit une loi $\chi^2_{(m_1-1)(m_2-1)}$.

Définition Loi du khi-deux à ℓ degrés de libertés χ^2_ℓ est la loi de la variable $\sum_{i=1}^{\ell}U_i^2$, où les U_i sont des variables gaussiennes réduites indépendantes.

Le test du χ^2 Ingrédients :

- on se fixe un risque d'erreur α (0.01 ou 0.05 en général)
- on calcule la valeur d_c^2 telle que $\mathrm{P}\left(\chi^2_{(m_1-1)(m_2-1)} > d_c^2\right) = \alpha$.
- Si $d^2 > d_c^2$ on considère que l'événement est trop improbable et que donc que l'hypothèse originale d'indépendance doit être rejetée.

$d^2 > d_c^2 \implies$ variables liées $\quad d^2 < d_c^2 \implies$ pas de conclusion

Mode de calcul du χ^2

Calcul par table du χ^2 Traditionnellement, on trouvait ces valeurs dans une table précalculée pour $\ell \leq 30$.

- la ligne indique le nombre de degrés de liberté ℓ ;
- la colonne indique la probabilité cumulative $\mathrm{P}\left(\chi^2_\ell > d_c^2\right)$;
- la case de la table donne la valeur de d_c^2.

Quand $\ell > 30$, on considère que $\sqrt{2\chi^2_\ell} - \sqrt{2\ell - 1}$ est distribué comme une variable gaussienne centrée réduite $N(0, 1)$.

Utilisation de la p-value L'utilisation d'un logiciel de statistique permet en général de calculer directement la p–value $\mathrm{P}\left(\chi^2_{(m_1-1)(m_2-1)} > d^2\right)$ et rend inutile de se fixer un seuil d'erreur préalable.

Décomposition de l'inertie

Nombre de valeurs propres Comme $\mathbf{D}_1^{-1}\mathbf{N}\mathbf{D}_2^{-1}\mathbf{N}'$ est carrée de dimension m_1 et que $\mathbf{D}_2^{-1}\mathbf{N}'\mathbf{D}_1^{-1}\mathbf{N}$ est de dimension m_2, le nombre de valeurs propres non nulles est $\min(m_1, m_2)$. Mais comme une des valeurs propres est 1 (associée à $\mathbf{g}$) et n'est pas intéressante :

Il y a au plus $\min(m_1 - 1, m_2 - 1)$ *valeurs propres non nulles*

φ^2 et valeurs propres l'inertie totale (et donc la somme des valeurs propres) est égale à φ^2. Donc si si $m_1 < m_2$, on obtient $\varphi^2 = \sum_{k=1}^{m_1-1}\lambda_k$.

Choix du nombre de valeurs propres On se contente souvent de regarder le premier plan principal car

- la règle de Kaiser $\lambda_k > \varphi^2/(m_1 - 1)$ s'applique mal ;
- la règle du coude reste valide, mais est subjective ;
- il existe un test sur de la part d'inertie non expliquée, mais il est un peu compliqué.

Partie V. Analyse des correspondances multiples

Analyse des correspondances multiples

But on veut étendre l'AFC au cas de $p \geq 2$ variables $\mathcal{X}_1, \mathcal{X}_2 \ldots, \mathcal{X}_p$ à $m_1, m_2, \ldots, m_p$ modalités. Ceci est particulièrement utile pour l'exploration d'enquêtes où les questions sont à réponses multiples.

Problème l'analyse des correspondances utilise une table de contingence qui nécessite $p = 2$.

Méthode on cherche un moyen différent d'analyser $p > 2$ variables et on vérifie que les résultats sont comparables à l'AFC pour $p = 2$.

Les données

Données brutes chaque individu est décrit par les numéros des modalités qu'il possède pour chacune des p variables. Il n'est pas possible de faire des calculs sur ce tableau, où les valeurs sont arbitraires.

Tableau disjonctif on remplace la v-ième colonne par m_v colonnes d'indicatrices : on met un zéro dans chaque colonne, sauf celle correspondant à la modalité de l'individu i qui reçoit 1.

Exemple On interroge 6 personnes sur la couleur de leurs cheveux (`CB`, `CC` et `CR` pour blond, châtain et roux), la couleur de leurs yeux (`YB`, `YV` et `YM` pour bleu, vert et marron) et leur sexe (`H`/`F`). On a donc trois variables (avec respectivement 3, 3 et 2 modalités) mesurées sur 6 individus. Les tableaux brut (ci-dessous à gauche) sont équivalents aux tableaux disjonctifs (à droite).

$$\begin{pmatrix} \texttt{CB} \\ \texttt{CB} \\ \texttt{CC} \\ \texttt{CC} \\ \texttt{CR} \\ \texttt{CB} \end{pmatrix} \begin{pmatrix} \texttt{YB} \\ \texttt{YV} \\ \texttt{YB} \\ \texttt{YM} \\ \texttt{YV} \\ \texttt{YB} \end{pmatrix} \begin{pmatrix} \texttt{H} \\ \texttt{H} \\ \texttt{F} \\ \texttt{H} \\ \texttt{F} \\ \texttt{F} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

Tableau disjonctif et tableau de contingence

Tableau disjonctif à la variable $\mathcal{X}_v$ on associe le tableau disjonctif $\mathbf{X}_v$ à n lignes et m_v colonnes.

Tableau de contingence on vérifie facilement que le tableau de contingence des variables $\mathcal{X}_v$ et $\mathcal{X}_w$ est donné par

$$\mathbf{N}_{vw} = \mathbf{X}_v' \mathbf{X}_w.$$

Effectifs marginaux la matrice diagonale des effectifs marginaux de la variable $\mathcal{X}_v$ est donnée par

$$\mathbf{D}_v = \mathbf{X}_v' \mathbf{X}_v.$$

Exemple (suite) Table de contingence Cheveux/Yeux et matrice d'effectif marginaux de la couleur de cheveux

$$\mathbf{N}_{12} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \qquad \mathbf{D}_1 = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Tableau disjonctif joint

Définition c'est la matrice $\mathbf{X} = (\mathbf{X}_1 | \mathbf{X}_2 | \cdots | \mathbf{X}_p)$, qui possède n lignes et $m_1 + \cdots + m_p$ colonnes. Chaque colonne représente une *catégorie*, c'est-à-dire une modalité d'une variable.

Exemple pour l'exemple de variables précédentes, on a le tableau disjonctif joint suivant

$$\mathbf{X} = \left(\begin{array}{ccc|ccc|cc} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right)$$

Chaque ligne somme à 3. Les sommes de colonnes sont

$$\left(\begin{array}{ccc|ccc|cc} 3 & 2 & 1 & 3 & 2 & 1 & 3 & 3 \end{array}\right)$$

Le tableau de Burt

Définition c'est un super-tableau de contingence des variables $\mathcal{X}_1, \ldots, \mathcal{X}_p$, formé de tableaux de contingence et de matrices d'effectifs marginaux. :

$$\mathbf{B} = \mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 & \cdots & \mathbf{X}_1'\mathbf{X}_p \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 & & \\ \vdots & & \ddots & \vdots \\ \mathbf{X}_p'\mathbf{X}_1 & \cdots & & \mathbf{X}_p'\mathbf{X}_p \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{N}_{12} & \cdots & \mathbf{N}_{1p} \\ \mathbf{N}_{21} & \mathbf{D}_2 & & \\ \vdots & & \ddots & \vdots \\ \mathbf{N}_{p1} & \cdots & & \mathbf{D}_p \end{bmatrix}$$

Exemple Toujours pour les mêmes variables

$$\mathbf{B} = \left(\begin{array}{ccc|ccc|cc} 3 & 0 & 0 & 2 & 1 & 0 & 2 & 1 \\ 0 & 2 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ \hline 2 & 1 & 0 & 3 & 0 & 0 & 1 & 2 \\ 1 & 0 & 1 & 0 & 2 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ \hline 2 & 1 & 0 & 1 & 1 & 1 & 3 & 0 \\ 1 & 1 & 1 & 2 & 1 & 0 & 0 & 3 \end{array}\right)$$

Partie VI. L'ACM : une AFC sur tableau disjonctif

Comment utiliser l'AFC pour analyser p variables

But on cherche à faire une représentation des $m_1 + \cdots + m_p$ catégories comme points d'un espace de faible dimension.

Méthode on fait une AFC sur le tableau disjonctif joint $\mathbf{X} = (\mathbf{X}_1 | \mathbf{X}_2 | \cdots | \mathbf{X}_p)$.

Les lignes la somme des éléments de chaque ligne de $\mathbf{X}$ est égale à p. Le tableau des profils-lignes est donc $\frac{1}{p}\mathbf{X}$.

Les colonnes la somme des éléments de chaque colonne de $\mathbf{X}$ est égale à l'effectif marginal de la catégorie correspondante. Le tableau des profils colonnes est donc $\mathbf{X}\mathbf{D}^{-1}$, où $\mathbf{D}$ est la matrice diagonale par blocs

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{D}_p \end{pmatrix}$$

Les coordonnées factorielles des catégories

Notation On note $\mathbf{a}_k = (\mathbf{a}_{1k}, \ldots, \mathbf{a}_{pk})'$ le vecteur à $m_1 + \cdots + m_p$ composantes des coordonnées factorielles des catégories sur l'axe k.

Calcul de l'AFC sur $\mathbf{X}$ comme la matrice des profils lignes est $\frac{1}{p}\mathbf{X}$ et celle des profils colonnes $\mathbf{X}\mathbf{D}^{-1}$, $\mathbf{a}_k$ est vecteur propre de

$$(\mathbf{X}\mathbf{D}^{-1})'\frac{1}{p}\mathbf{X} = \frac{1}{p}\mathbf{D}^{-1}\mathbf{X}'\mathbf{X} = \frac{1}{p}\mathbf{D}^{-1}\mathbf{B}$$

et donc l'équation des coordonnées des catégories est

$$\frac{1}{p}\mathbf{D}^{-1}\mathbf{B}\mathbf{a}_k = \mu_k \mathbf{a}_k$$

avec la convention de normalisation $\frac{1}{np}\mathbf{a}_k'\mathbf{D}\mathbf{a}_k = \mu_k$.

Propriétés des valeurs propres

Valeur propres triviales La valeur propre 1 est associée (comme en AFC) à la composante $\mathbf{z}^0 = (1, \ldots, 1)$ dans l'espace des individus. Les autres vecteurs propres lui sont orthogonaux, et donc de moyenne nulle.

Autres valeurs propres Si $n > \sum_{v=1}^{p} m_v$, le rang de $\mathbf{X}$ est $\sum_{v=1}^{p} m_v - p + 1$ et le nombre de valeurs propres non trivialement égales à 0 ou 1 est

$$q = \sum_{v=1}^{p} m_v - p.$$

Somme La somme des valeurs propres non triviales est

$$\sum_{k=1}^{q} \mu_k = \mathrm{Tr}\left(\frac{1}{p}\mathbf{D}^{-1}\mathbf{B}\right) - 1 = \frac{1}{p}\sum_{v=1}^{p} m_v - 1 = \frac{q}{p}$$

et leur moyenne vaut donc $1/p$.

Résolution dans le cas $p = 2$

On note $\mathbf{a}_k$ (resp. $\mathbf{b}_k$) les m_1 premières (resp. m_2 dernières) coordonnées de la composante principale k et μ_k la valeur propre correspondante :

$$\frac{1}{2}\mathbf{D}^{-1}\mathbf{B}\begin{bmatrix} \mathbf{a}_k \\ \mathbf{b}_k \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \mathbf{I}_{m_1} & \mathbf{D}_1^{-1}\mathbf{N} \\ \mathbf{D}_2^{-1}\mathbf{N}' & \mathbf{I}_{m_2} \end{bmatrix}\begin{bmatrix} \mathbf{a}_k \\ \mathbf{b}_k \end{bmatrix} = \mu_k \begin{bmatrix} \mathbf{a}_k \\ \mathbf{b}_k \end{bmatrix}.$$

On obtient les équations

$$\begin{cases} \mathbf{D}_1^{-1}\mathbf{N}\mathbf{b}_k = (2\mu_k - 1)\mathbf{a}_k \\ \mathbf{D}_2^{-1}\mathbf{N}'\mathbf{a}_k = (2\mu_k - 1)\mathbf{b}_k \end{cases},$$

et donc on retrouve les coordonnées des modalités de lignes et de colonnes dans l'AFC classique (avec $\lambda_k = (2\mu_k - 1)^2$) :

$$\begin{cases} \mathbf{D}_2^{-1}\mathbf{N}'\mathbf{D}_1^{-1}\mathbf{N}\mathbf{b}_k = (2\mu_k - 1)^2\mathbf{b}_k \\ \mathbf{D}_1^{-1}\mathbf{N}\mathbf{D}_2^{-1}\mathbf{N}'\mathbf{a}_k = (2\mu_k - 1)^2\mathbf{a}_k \end{cases}.$$

Différences ACM/AFC pour $p = 2$

Nombre de valeurs propres on a *a priori* $m_1 + m_2 - 2$ valeurs propres non nulles, ce qui est plus important que dans le cas classique. En particulier pour chaque λ_k, on a deux μ_k possibles

$$\begin{cases} \mu_k = \frac{1+\sqrt{\lambda_k}}{2} & \text{associée à } \begin{bmatrix} \mathbf{a}_k \\ \mathbf{b}_k \end{bmatrix} \\ \mu_k' = \frac{1-\sqrt{\lambda_k}}{2} & \text{associée à } \begin{bmatrix} \mathbf{a}_k \\ -\mathbf{b}_k \end{bmatrix} \end{cases}$$

On ne garde donc que les valeurs $\mu_k > \frac{1}{2}$. On peut montrer qu'il y en a $\min(m_1 - 1, m_2 - 1)$.

Inertie L'interprétation de la part d'inertie expliquée par les valeurs propres est maintenant très différente. En particulier les valeurs propres qui étaient très séparées dans l'AFC de $\mathbf{N}$ le sont beaucoup moins dans celle de $\mathbf{X}$.

Partie VII. Aspects pratiques

Formules barycentriques

Les coordonnées des individus Soit $\mathbf{c}_k$ le vecteur à n composantes des coordonnées des n individus sur l'axe factoriel associé à la valeur propre μ_k. D'après les résultats sur l'AFC, on a

$$\mathbf{c}_k = \frac{1}{\sqrt{\mu_k}}\frac{1}{p}\mathbf{X}\mathbf{a}_k \quad \text{et donc} \quad c_{ik} = \frac{1}{\sqrt{\mu_k}}\frac{1}{p}\sum_{j \text{ catégorie de } i} a_{jk}$$

Les seuls termes non nuls dans le calcul de $\mathbf{X}\mathbf{a}_k$ sont les coordonnées de la catégorie de chaque variable possédée par l'individu. Comme on est est dans le cadre de l'AFC, la variance de $\mathbf{c}_k$ est toujours $\mathrm{var}\,\mathbf{c}_k = \frac{1}{n}\mathbf{c}_k'\mathbf{c}_k = \mu_k$.

Barycentre des catégories *À $1/\sqrt{\mu_k}$ près, la coordonnée d'un individu est égale à la moyenne arithmétique simple des coordonnées des catégories auxquelles il appartient.*

Les coordonnées des catégories On a de même la seconde formule

$$\mathbf{a}_k = \frac{1}{\sqrt{\mu_k}}\mathbf{D}^{-1}\mathbf{X}'\mathbf{c}_k \quad \text{c-à-d} \quad a_{jk} = \frac{1}{\sqrt{\mu_k}}\frac{1}{n_j}\sum_{i \text{ de catégorie } j} c_{ik}$$

Les seuls termes non nuls de $\mathbf{X}'\mathbf{c}_k$ sont les coordonnées des individus ayant une catégorie donnée. Là encore, $\mathrm{var}\,\mathbf{a}_k = \mu_k$.

Barycentre des individus *À $1/\sqrt{\mu_k}$ près, la coordonnée d'une catégorie est égale à la moyenne arithmétique des coordonnées des n_j individus de cette catégorie.*

Barycentres et représentation

Représentation commune Les points représentatifs des catégories sont barycentres des groupes d'individus. On peut donc représenter individus et catégories dans un même plan factoriel.

Moyennes Comme $\mathbf{c}_k$ est une variable de moyenne nulle, la formule de barycentre indique que pour chaque variable $\mathcal{X}_i$, les coordonnées de ses catégories (pondérés par les effectifs) sont de moyenne nulle. Aucun centrage n'est donc nécessaire

Échelle pour que les catégories se trouvent visuellement au barycentre des individus qui les représentent on peut remplacer $\mathbf{a}_k$ par

$$\boldsymbol{\alpha}_k = \mathbf{D}^{-1}\mathbf{X}'\mathbf{c}_k = \sqrt{\mu_k}\mathbf{a}_k$$

Sélection de variables et axes

Sélection des variables on décide souvent de ne garder qu'un nombre réduit de variables actives et de garder les autres comme variables supplémentaires.

Sélection des axes

- règle courante : garder les axes tels que $\mu_k > 1/p$ (la moyenne des valeurs propres est $1/p$).
- les axes intéressants sont ceux que l'on peut interpréter, en regardant les contributions des variables actives et les valeurs-tests associées aux variables supplémentaires (définies plus tard).
- En pratique on se contente souvent d'interpréter le premier plan principal.

Inertie expliquée elle est moins intéressante qu'en ACP.

Catégories et axes factoriels

Catégorie Comme $\operatorname{var}\mathbf{a}_k = \sum_j \frac{n_j}{np}(a_{jk})^2 = \mu_k$, la contribution de la catégorie j à l'axe k est

$$\frac{n_j}{np}\frac{(a_{jk})^2}{\mu_k},$$

intéressante si elle est supérieure au poids n_j/np (à un facteur près comme en ACP et AFC).

Variable la contribution totale de la variable $\mathcal{X}_v$ à l'axe factoriel est

$$\frac{1}{\mu_k}\frac{1}{np}\sum_{j \text{ modalité de } \mathcal{X}_v} n_j(a_{jk})^2$$

Qualité de la représentation pour le sous-espace formé par les ℓ premier axes, la qualité de la représentation de la catégorie j est le cosinus carré habituel

$$\frac{\sum_{k=1}^{\ell}(a_{jk})^2}{\sum_{k=1}^{q}(a_{jk})^2}.$$

Individus et axes factoriels

La normalisation de $\mathbf{c}_k$ est $\sum_{i=1}^n (c_{ik})^2 = n\mu_k$, où c_{ik} est la coordonnée de l'individu i sur l'axe factoriel k associé à la valeur propre μ_k.

Contribution d'un individu pour l'individu i, c'est

$$\frac{1}{n}\frac{(c_{ik})^2}{\mu_k}$$

Cette contribution est jugée en la comparant au poids $1/n$ comme en ACP et AFC.

Qualité de la représentation pour le sous-espace formé par les ℓ premier axes, la qualité de la représentation de l'individu i est

$$\frac{\sum_{k=1}^{\ell}(c_{ik})^2}{\sum_{k=1}^{q}(c_{ik})^2}.$$

Contribution à l'inertie totale

Soit $\mathbf{x}^j = (x_i^j)$ le vecteur colonne de $\mathbf{X}$ correspondant à une catégorie j. On rappelle que l'inertie totale vaut

$$I_{\mathbf{g}} = \sum_{j\in\text{catégories}} \frac{n_j}{np} d^2(\mathbf{z}^j, \mathbf{g}) = \frac{1}{p}\sum_{v=1}^{p} m_v - 1,$$

où la distance du profil-colonne j au centre de gravité des profils-colonnes $\mathbf{g} = \mathbf{1}/n$ est

$$\begin{aligned} d^2(\mathbf{z}^j, \mathbf{g}) &= \sum_{i=1}^{n} \frac{np}{p}\left(\frac{x_i^j}{n_j} - \frac{1}{n}\right)^2 = n\sum_{i=1}^{n}\left(\frac{x_i^j}{n_j^2} + \frac{1}{n^2} - \frac{2x_i^j}{nn_j}\right) \\ &= n\left(\frac{n_j}{n_j^2} + \frac{n}{n^2} - \frac{2n_j}{nn_j}\right) = \frac{n}{n_j} - 1 \end{aligned}$$

Contribution d'une catégorie La contribution absolue de la catégorie j à l'inertie est

$$\frac{n_j}{np} d^2(\mathbf{z}^j, \mathbf{g}) = \frac{1}{p}\left(1 - \frac{n_j}{n}\right),$$

qui est une fonction décroissante de l'effectif. Il faut donc éviter les catégories d'effectif trop faible, qui d'ailleurs se retrouveront dans les premiers axes

Contribution d'une variable La contribution de la variable $\mathcal{X}_v$ est

$$\sum_{j \text{ modalité de } \mathcal{X}_v} \frac{1}{p}\left(1 - \frac{n_j}{n}\right) = \frac{m_v - 1}{p}$$

Elle est d'autant plus grande que le nombre de modalités de $\mathcal{X}_i$ est élevé. Il faut donc éviter les disparités trop grandes entre les nombre de modalités (quand on a le choix du découpage...)

Correspondances multiples et ACP non linéaire

Problème l'ACP vise à trouver une combinaison *linéaire* $u_1\mathbf{x}^1 + \cdots + u_p\mathbf{x}^p$ des variables qui soit de variance maximale. Si les relations entre variables ne sont pas linéaires, l'ACP échoue à extraire des données intéressantes.

Extension non-linéaire on cherche des fonctions $\phi^1, \ldots, \phi^p$ qui maximisent la variance

$$\operatorname{var}\left(\phi^1(\mathbf{x}^1) + \cdots + \phi^p(\mathbf{x}^p)\right)$$

Fonctions en escalier On peut prendre des fonctions en escalier : on découpe l'intervalle de variation de $\mathbf{x}^j$ en m_j classes et on se donne un vecteur $\mathbf{a}_j = (a_{j1}, \ldots, a_{jm_j})$ de poids. Alors $\phi^j(x) = a_{j\ell}$ si x est dans la ℓ-ième classe.