Forme alternative de l'inertie totale : comme $I_{\mathbf{v}} = I_{\mathbf{g}} + \|\mathbf{v}-\mathbf{g}\|^2_{\mathbf{M}}$, on a

$$I_{\mathbf{g}} = I_{\mathbf{0}} - \|\mathbf{g}\|^2_{\chi^2} = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}} - 1.$$

ACP des deux nuages de profils

Comment ? deux possibilités en dualité exacte

	données	métrique	poids
Profils-lignes	$\mathbf{X} = \mathbf{D}_1^{-1}\mathbf{N}$	$\mathbf{M} = n\mathbf{D}_2^{-1}$	$\mathbf{D} = \frac{\mathbf{D}_1}{n}$
Profils-colonnes	$\mathbf{X} = \mathbf{D}_2^{-1}\mathbf{N}'$	$\mathbf{M} = n\mathbf{D}_1^{-1}$	$\mathbf{D} = \frac{\mathbf{D}_2}{n}$

Autres données

- Centre de gravité $\mathbf{g} = \mathbf{X}'\mathbf{D1}$ (comme ACP), avec $\mathbf{Mg} = \mathbf{1}$ et $\mathbf{g}'\mathbf{Mg} = 1$
- Matrice de variance-covariance

$$\mathbf{V} = \mathbf{X}'\mathbf{DX} - \mathbf{gg}' = (\mathbf{X} - \mathbf{1g}')'\mathbf{D}(\mathbf{X} - \mathbf{1g}')$$

Non-nécessité du centrage

Propriété $\mathbf{VM}$ et $\mathbf{X}'\mathbf{DXM}$ ont les mêmes vecteurs propres :

- d'une part $\mathbf{g}$ (associé aux valeurs respectives 0 et 1)
- d'autre part des $\mathbf{u}$ tels que $\mathbf{X}'\mathbf{DXMu} = \mathbf{VMu} = \lambda\mathbf{u}$.

Preuve $\mathbf{g}$ satisfait $\mathbf{VMg} = \mathbf{0}$ car $\mathbf{Mg} = \mathbf{1}$:

$$\mathbf{VMg} = (\mathbf{X} - \mathbf{1g}')'\mathbf{D}(\mathbf{X} - \mathbf{1g}')\mathbf{1} = \mathbf{0}.$$

De même

$$\mathbf{X}'\mathbf{DXMg} = \mathbf{VMg} + \mathbf{gg}'\mathbf{Mg} = \mathbf{0} + \mathbf{g} = \mathbf{g}.$$

Les autres vecteurs propres de $\mathbf{VM}$ sont orthogonaux à $\mathbf{g}$ ($\mathbf{g}'\mathbf{Mu} = 0$) et

$$\mathbf{X}'\mathbf{DXMu} = \mathbf{VMu} + \mathbf{gg}'\mathbf{Mu} = \mathbf{VMu} = \lambda\mathbf{u}.$$

Approche On effectue donc une ACP non centrée et on élimine la valeur propre 1 associée à l'axe principal $\mathbf{g}$

Calcul de l'ACP (profils-lignes)

Facteurs principaux ils sont vecteurs propres de

$$\mathbf{MX}'\mathbf{DX} = (n\mathbf{D}_2^{-1})(\mathbf{D}_1^{-1}\mathbf{N})'\frac{\mathbf{D}_1}{n}(\mathbf{D}_1^{-1}\mathbf{N}) = \mathbf{D}_2^{-1}\mathbf{N}'\mathbf{D}_1^{-1}\mathbf{N}.$$

et on a donc pour chaque axe principal k

$$\mathbf{D}_2^{-1}\mathbf{N}'\mathbf{D}_1^{-1}\mathbf{N}\mathbf{u}_k = \lambda_k\mathbf{u}_k$$

Composantes principales la composante principale associée au facteur $\mathbf{u}_k$ est $\mathbf{a}_k = \mathbf{X}\mathbf{u}_k = \mathbf{D}_1^{-1}\mathbf{N}\mathbf{u}_k$; elle est vecteur propre de la matrice $\mathbf{D}_1^{-1}\mathbf{N}\mathbf{D}_2^{-1}\mathbf{N}'$ car

$$\begin{aligned}\mathbf{D}_1^{-1}\mathbf{N}\mathbf{D}_2^{-1}\mathbf{N}'\mathbf{a}_k &= \mathbf{D}_1^{-1}\mathbf{N}\mathbf{D}_2^{-1}\mathbf{N}'\mathbf{D}_1^{-1}\mathbf{N}\mathbf{u}_k \\ &= \lambda_k\mathbf{D}_1^{-1}\mathbf{N}\mathbf{u}_k = \lambda_k\mathbf{a_k}\end{aligned}$$

Analyse des profils-colonnes on échange les indices 1 et 2 et on transpose $\mathbf{N}$.

Comparaison lignes-colonnes

	ACP profils-lignes	ACP profils-colonnes
Facteurs principaux	Vecteurs propres de $\mathbf{D}_2^{-1}\mathbf{N}'\mathbf{D}_1^{-1}\mathbf{N}$	Vecteurs propres de $\mathbf{D}_1^{-1}\mathbf{N}\mathbf{D}_2^{-1}\mathbf{N}'$
Composantes principales	Vecteurs propres de $\mathbf{D}_1^{-1}\mathbf{N}\mathbf{D}_2^{-1}\mathbf{N}'$ normalisés par $\operatorname{var}\mathbf{a}_k = \mathbf{a}_k'\frac{\mathbf{D}_1}{n}\mathbf{a}_k = \lambda_k$	Vecteurs propres de $\mathbf{D}_2^{-1}\mathbf{N}'\mathbf{D}_1^{-1}\mathbf{N}$ normalisés par $\operatorname{var}\mathbf{b}_k = \mathbf{b}_k'\frac{\mathbf{D}_2}{n}\mathbf{b}_k = \lambda_k$

Comparaison les deux analyses conduisent aux mêmes valeurs propres et les facteurs principaux de l'une sont les composantes principales de l'autre (à un facteur près).

Partie IV. Aspects pratiques

Interprétation des résultats

Coordonnées des points Les coordonnées des points-lignes et points-colonnes s'obtiennent en cherchant les vecteurs propres des produits des deux tableaux de profils. Ce sont les grandeurs principales à obtenir.

Projection des nuages il est possible de projeter les deux nuages de points sur la même représentation. On justifiera plus tard le sens de cette représentation et son interprétation.

Cercle des corrélations il n'a aucun intérêt ici, puisque les véritables variables sont qualitatives.

(non) effet de taille comme les composantes variables sont centrées ($\sum_{i=1}^{m_1} n_{i\cdot}a_{ik} = \sum_{j=1}^{m_2} n_{\cdot j}b_{jk} = 0$), ont sait que les coordonnées des $\mathbf{a}_k$ et $\mathbf{b}_k$ ne peuvent être toutes de même signe ; il n'y a donc jamais d'effet de « taille ».

Contributions à l'inertie

Contribution des profils-lignes On sait que $\lambda_k = \sum_{i=1}^{m_1}\frac{n_{i\cdot}}{n}(a_{ik})^2$, où a_{ik} est la coordonnée du profil-ligne i sur la k-ième composante principale de l'ACP sur les profils-lignes. On définit donc la contribution de la modalité i à l'axe principal k comme

$$\frac{n_{i\cdot}}{n}\cdot\frac{(a_{ik})^2}{\lambda_k}.$$

On considérera les modalités ayant l'influence la plus importante (typiquement $> \alpha n_{i\cdot}/n$, $\alpha = 2$ ou 3) comme constitutives des axes ; on regardera aussi le signe de la coordonnée.

Il n'y a pas ici de modalités sur-représentées, puisqu'on ne peut pas les retirer.

Contribution des profils-colonnes pour les mêmes raisons, la contribution du de la modalité j de $\mathcal{X}_2$ à l'axe k est

$$\frac{n_{\cdot j}}{n}\cdot\frac{(b_{jk})^2}{\lambda_k}.$$

Qualité de la représentation

Profils-lignes l'AFC est une ACP, et on peut donc mesurer la qualité de la représentation de la modalité i (son profil-ligne) par un sous-espace factoriel. La qualité (le $\cos^2$ de l'angle entre le point et sa projection) s'écrit encore, pour le plan formé des k^* premiers axes :

$$\frac{\sum_{k=1}^{k^*}(a_{ik})^2}{\sum_{k=1}^{m_2}(a_{ik})^2}.$$

Comme pour l'ACP, > 0.8 signifie « bien représenté » et < 0.5 veut dire « mal représenté ». Les valeurs sont souvent données en pourcents.

Profils-colonne Le principe est le même, mais la formule devient, pour la modalité j :

$$\frac{\sum_{k=1}^{k^*}(b_{jk})^2}{\sum_{k=1}^{m_1}(b_{jk})^2}.$$

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.