

Analyse en composantes principales

M2R santé publique, option biostatistiques

Introduction

L'Analyse en Composantes Principales (ACP) fait partie des méthodes d'analyses factorielles.

Elle permet de résumer un tableau de données individus \times variables, lorsque les variables sont **quantitatives**.

L'ACP permet d'étudier les ressemblances entre individus du point de vue de l'ensemble des variables et dégage des profils d'individus.

L'ACP permet également de réaliser un bilan des liaisons **linéaires** entre variables.

L'ACP est une méthode descriptive (ce n'est pas une fin en soi). Elle permet de mieux connaître les données sur lesquelles on travaille, de détecter d'éventuelles valeurs suspectes et de formuler des hypothèses qu'il faudra étudier à l'aide de modèles et d'études statistiques inférentielles.

Les représentations graphiques fournies par l'ACP peuvent servir pour illustrer certains résultats.

Introduction

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5.0	3.6	1.4	0.2
6	5.4	3.9	1.7	0.4

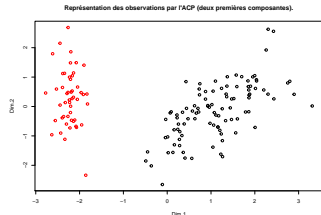
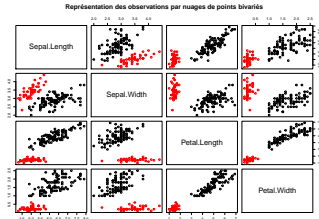


Tableau de données

Les données sont composées de n observations décrites par p variables **quantitatives**. Le tableau des données brutes (à partir duquel on fait l'analyse) est noté \mathbf{X} .

$$\mathbf{X} = \begin{bmatrix} x_{11} & \dots & \dots & \dots & x_{1p} \\ \vdots & \ddots & & & \vdots \\ x_{i1} & & x_{ij} & & x_{ip} \\ \vdots & & & \ddots & \vdots \\ x_{n1} & \dots & \dots & \dots & x_{np} \end{bmatrix}$$

On note $U_i \in \mathbb{R}^p$ le vecteur des p variables décrivant l'observation i .

On note $V_j \in \mathbb{R}^n$ le vecteur des n mesures de la variable p .

$$U_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{bmatrix} \quad \text{et} \quad V_j = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{bmatrix}$$

Tableau de données

Chaque observation peut être représentée dans \mathbb{R}^p . L'ensemble des points qui représentent les observation est appelé **nuage des individus**.

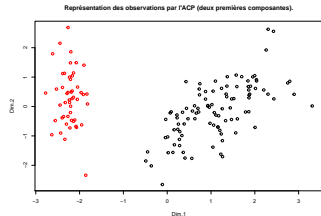
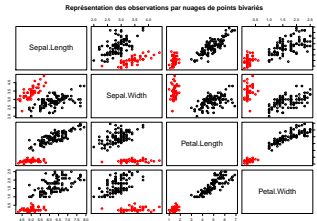
Chaque variable peut être représentée dans \mathbb{R}^n . L'ensemble des points qui représentent les variables est appelé **nuage des variables**.

Ainsi, les espaces considérés sont généralement de dimension supérieure à 3. On ne peut donc généralement pas visualiser les nuages de points.

Tableau de données

L'idée des méthodes factorielles et de trouver un système d'axes et de plans tel que les projections de ces nuages de points sur ces axes et ces plans permettent de reconstituer au mieux les positions des points les uns par rapport aux autres.

Cela revient à trouver une représentation qui conserve au mieux les **distances** entre les points. C'est une représentation **géométrique**.



Choix d'une distance

Pour faire une représentation géométrique, il faut choisir une distance entre deux points de l'espace.

La distance utilisée par l'ACP dans l'espace où sont représentées les observations est la **distance euclidienne classique** (on reviendra sur ce point à la fin du cours). Ainsi la distance entre les observations i et i' est égale à $d(i, i')$ avec

$$d^2(i, i') = \sum_{j=1}^p (x_{ij} - x_{i'j})^2.$$

Avec cette distance, toutes les variables jouent le même rôle. Les variables constituent une base orthogonale, à cette distance on associe un produit scalaire entre deux vecteurs ainsi que la norme d'un vecteur

$$U_i^\top U_{i'} = \sum_{j=1}^p x_{ij} x_{i'j} \text{ et } ||U_i||^2 = \sum_{j=1}^p x_{ij}^2.$$

Choix d'une distance

On peut alors définir l'angle α entre les vecteurs des observations i et i' par

$$\begin{aligned}\cos(\alpha) &= \frac{U_i^\top U_{i'}}{\|U_i\| \|U_{i'}\|} \\ &= \frac{\sum_{j=1}^p x_{ij} x_{i'j}}{\sqrt{\sum_{j=1}^p x_{ij}^2 \sum_{j=1}^p x_{i'j}^2}}\end{aligned}$$

Choix de l'origine

Le point o correspondant au vecteur de coordonnées toutes nulles n'est pas forcément une origine satisfaisante, car si les coordonnées des points du nuage des individus sont grandes, le nuage est éloigné de cette origine.

Il apparaît plus judicieux de choisir une origine liée au nuage lui-même: le **centre de gravité du nuage**.

Pour définir ce centre de gravité, il faut choisir un système de pondération des observations. On notera p_i le poids l'observation i tel que $\sum_{i=1}^n p_i = 1$. Le centre de gravité G est défini par

$$\sum_{i=1}^p p_i \overrightarrow{GU_i}$$

Choix de l'origine

Pour la suite, on considère que $p_i = 1/n$. Ainsi le centre de gravité correspond à la valeurs moyennes des variables

$$G = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{ip} \end{bmatrix} = \begin{bmatrix} \bar{x}_{\bullet 1} \\ \vdots \\ \bar{x}_{\bullet p} \end{bmatrix}$$

Prendre G comme origine revient alors à travailler sur le tableau des données centrées \mathbf{X}_c :

$$\mathbf{X}_c = \begin{bmatrix} x_{11} - \bar{x}_{\bullet 1} & \dots & \dots & \dots & x_{1p} - \bar{x}_{\bullet p} \\ \vdots & \ddots & & & \vdots \\ x_{i1} - \bar{x}_{\bullet 1} & & x_{ij} - \bar{x}_{\bullet j} & & x_{ip} - \bar{x}_{\bullet p} \\ \vdots & & & \ddots & \vdots \\ x_{n1} - \bar{x}_{\bullet 1} & \dots & \dots & \dots & x_{np} - \bar{x}_{\bullet p} \end{bmatrix}$$

Choix de l'origine

On en déduit le vecteur des coordonnées centrées de l'observation i noté U_{ci} ainsi que le vecteur des coordonnées centrées des variables noté V_{cj} .

$$U_{ci} = \begin{bmatrix} x_{i1} - \bar{x}_{\bullet 1} \\ \vdots \\ x_{ip} - \bar{x}_{\bullet p} \end{bmatrix} \text{ et } V_{cj} = \begin{bmatrix} x_{1j} - \bar{x}_{\bullet j} \\ \vdots \\ x_{nj} - \bar{x}_{\bullet j} \end{bmatrix}$$

Moments d'inertie

On note I_G le **moment d'inertie totale** du nuage des individus par rapport au centre de gravité G :

$$I_G = \frac{1}{n} \sum_{i=1}^n d^2(G, i) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_{\bullet j})^2 = \frac{1}{n} \sum_{i=1}^n U_{ci}^\top U_{ci}.$$

Le moment d'inertie totale est intéressant, car c'est une mesure de la dispersion du nuage des individus par rapport à son centre de gravité.

Si I_G est grand, cela signifie que le nuage est très dispersé, tandis que s'il est petit, alors le nuage est très concentré sur son centre de gravité.

Moments d'inertie

On a I_G

$$\begin{aligned} I_G &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_{\bullet j})^2 \\ &= \sum_{j=1}^p \left[\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_{\bullet j})^2 \right] = \sum_{j=1}^p [\text{Var}(V_j)] \end{aligned}$$

où $\text{Var}(V_j)$ est la variance empirique de la variable j .

Ainsi, on constate que l'inertie totale est égale à la trace de la matrice de covariance Σ des p variables.

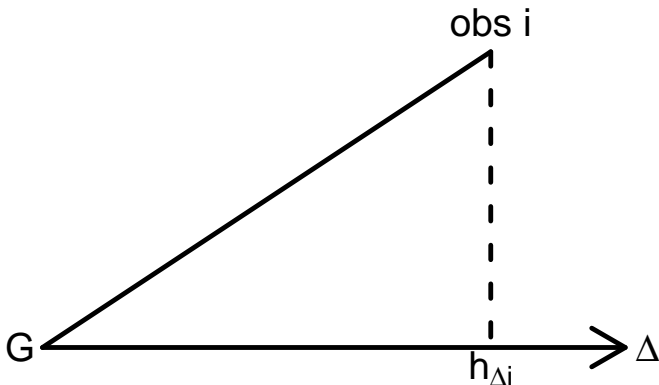
$$I_G = \text{trace}(\Sigma).$$

Moments d'inertie

L'inertie du nuage des individus par rapport à un axe Δ passant par G est définie par

$$I_{\Delta} = \frac{1}{n} \sum_{i=1}^n d^2(h_{\Delta i}, i)$$

où $h_{\Delta i}$ est la projection orthogonale de l'observation i sur l'axe Δ .



Moments d'inertie

De la même façon, l'inertie du nuage des individus par rapport à un sous-espace vectoriel W passant par G est

$$I_W = \frac{1}{n} \sum_{i=1}^n d^2(h_{Wi}, i)$$

où h_{Wi} est la projection orthogonale de l'observation i sur le sous-espace W .

Décomposition de l'inertie totale

Si on note \tilde{W} le complémentaire orthogonal de W dans \mathbb{R}^p et $h_{\tilde{W}i}$ la projection orthogonale de l'observation i sur \tilde{W} , en appliquant le théorème de Pythagore, on peut écrire

$$d^2(h_{Wi}, i) + d^2(h_{\tilde{W}i}, i) = d^2(G, i) = d^2(G, h_{Wi}) + d^2(G, h_{\tilde{W}i}).$$

On en déduit le théorème de Huygens:

$$I_W + I_{\tilde{W}} = I_G$$

Dans le cas particulier où le sous-espace est de dimension 1, $I_{\tilde{W}}$ est une mesure de l'allongement du nuage selon cet axe.

On emploie pour $I_{\tilde{W}}$ les expressions **d'inertie portée par l'axe** ou bien **d'inertie expliquée par l'axe**.

Décomposition de l'inertie totale

En projetant le nuage des individus sur un sous-espace W , on perd l'inertie mesurée par I_W , on ne conserve que celle mesurée par $I_{\tilde{W}}$.

Si on décompose l'espace \mathbb{R}^p comme la somme de sous-espaces de dimension 1 et orthogonaux entre eux :

$$\Delta_1 \oplus \Delta_2 \oplus \dots \oplus \Delta_p$$

on peut écrire

$$I_G = I_{\tilde{\Delta}_1} + I_{\tilde{\Delta}_2} + \dots + I_{\tilde{\Delta}_p}$$

Décomposition de l'inertie totale

En effet, supposons que l'on décompose l'espace \mathbb{R}^p comme la somme de sous-espaces de dimension 1 et orthogonaux entre eux :

$$\Delta_1 \oplus \Delta_2 \oplus \dots \oplus \Delta_p$$

En appliquant le théorème de Pythagore, on a

$$d^2(G, i) = d^2(G, h_{\Delta_2 \oplus \dots \oplus \Delta_p} i) + d^2(i, h_{\Delta_2 \oplus \dots \oplus \Delta_p} i)$$

où $h_{\Delta_2 \oplus \dots \oplus \Delta_p} i$ est la projection orthogonale de l'observation i sur le sous-espace $\tilde{\Delta}_1 = \Delta_2 \oplus \dots \oplus \Delta_p$ complémentaire orthogonal de Δ_1 dans \mathbb{R}^p . Comme tous les axes sont orthogonaux

$$d^2(i, h_{\Delta_2 \oplus \dots \oplus \Delta_p} i) = d^2(G, h_{\Delta_1} i)$$

Et donc

$$d^2(G, i) = d^2(G, h_{\Delta_2 \oplus \dots \oplus \Delta_p} i) + d^2(G, h_{\Delta_1} i)$$

Décomposition de l'inertie totale

On peut de la même façon, décomposer $d^2(G, h_{\Delta_2 \oplus \dots \oplus \Delta_p i})$

$$d^2(G, h_{\Delta_2 \oplus \dots \oplus \Delta_p i}) = d^2(G, h_{\Delta_3 \oplus \dots \oplus \Delta_p i}) + d^2(G, h_{\Delta_2 i})$$

et aboutir de proche en proche à

$$d^2(G, i) = d^2(G, h_{\Delta_1 i}) + d^2(G, h_{\Delta_2 i}) + \dots + d^2(G, h_{\Delta_p i})$$

En utilisant cette expression du carré de la distance de l'observation i au centre de gravité, on retrouve la décomposition de l'inertie totale comme somme des inerties expliquées par les p axes.

Recherche du premier axe

Attention, pour simplifier les notations, on suppose de $G = o$.

Cela revient à centrer le tableau avant l'étude.

Recherche du premier axe

On cherche un axe Δ_1 passant par o d'inertie I_{Δ_1} **minimum**.

En effet, dans ce cas, Δ_1 est l'axe le plus proche de l'ensemble des points du nuage des individus, et donc, si l'on doit projeter ce nuage sur cet axe, c'est lui qui donnera l'image la moins déformée du nuage.

Si on utilise les propriétés des inerties, alors rechercher Δ_1 tel que I_{Δ_1} est minimum, est équivalent à chercher Δ_1 tel que $I_{\tilde{W}_1}$ est maximum

On définit l'axe Δ_1 par son vecteur directeur unitaire a_1 . Il faut donc trouver a_1 tel que $I_{\tilde{W}_1}$ est maximum sous la contrainte que $\|a_1\|^2 = 1$.

Recherche du premier axe

Quelques expressions algébriques de $l_{\tilde{\Delta}}$ et de $\|a_1\|^2$.

$$d^2(0, h_{\Delta_1 i}) = (U_i^\top a_1)^2 = a_1^\top U_i U_i^\top a_1$$

Comme $l_{\tilde{\Delta}_1} = \frac{1}{n} \sum_{i=1}^n d^2(0, h_{\Delta_1 i})$, on en déduit que

$$l_{\tilde{\Delta}_1} = a_1^\top \left(\underbrace{\frac{1}{n} \sum_{i=1}^n U_i U_i^\top}_{\Sigma} \right) a_1$$

Le problème d'optimisation est donc de trouver a_1 tel que $a_1^\top \Sigma a_1$ soit maximum sous la contrainte que $\|a_1\|^2 = 1$.

Recherche du premier axe

$$a_1 = \arg \max a^\top \Sigma a \quad \text{s.c.} \quad \|a\|^2 = 1$$

C'est un problème qui peut être résolu en utilisant les multiplicateurs de Lagrange.

Méthode des multiplicateurs de Lagrange

On cherche les optimum d'une fonction $f(t_1, \dots, t_p)$ de p variables liées par une relation $\ell(t_1, \dots, t_p) = c$.

Alors, on calcule les dérivées partielles de la fonction

$$g(t_1, \dots, t_p) = f(t_1, \dots, t_p) - \lambda(\ell(t_1, \dots, t_p) - c)$$

par rapport à chacune des variables.

En annulant ces p dérivées partielles et celle par rapport à λ , on obtient un système de $p + 1$ équations à $p + 1$ inconnues.

On peut généraliser ce problème aux cas où d contraintes sont considérées, dans ce cas on considère d multiplicateurs de Lagrange et le problème suivant

$$g(t_1, \dots, t_p) = f(t_1, \dots, t_p) - \sum_{h=1}^k \lambda_h(\ell_h(t_1, \dots, t_p) - c_h)$$

Recherche du premier axe

Dans le cas de la recherche de a_1 tel que

$$a_1 = \arg \max a^\top \Sigma a \quad \text{s.c.} \quad \|a\|^2 = 1$$

On cherche donc à maximiser

$$g(a) = a^\top \Sigma a - \lambda_1 (a^\top a - 1)$$

On note $\frac{\partial g(a)}{\partial a}$ le vecteur de \mathbb{R}^p dont les composantes sont les dérivées partielles de $g(a)$ par rapport à chacune des composantes du vecteur $a \in \mathbb{R}^p$.

$$\frac{\partial g(a)}{\partial a} = \begin{bmatrix} \frac{\partial g(a)}{\partial a_1} \\ \vdots \\ \frac{\partial g(a)}{\partial a_j} \\ \vdots \\ \frac{\partial g(a)}{\partial a_p} \end{bmatrix}$$

Recherche du premier axe

On peut montrer que

$$\frac{\partial a^\top \Sigma a}{\partial a} = 2\Sigma a$$

En effet

$$\frac{\partial g(a)}{\partial a} = \begin{bmatrix} \frac{\partial a^\top}{\partial a_1} \Sigma a \\ \vdots \\ \frac{\partial a^\top}{\partial a_j} \Sigma a \\ \vdots \\ \frac{\partial a^\top}{\partial a_p} \Sigma a \end{bmatrix} + \begin{bmatrix} a^\top \Sigma \frac{\partial a}{\partial a_1} \\ \vdots \\ a^\top \Sigma \frac{\partial a}{\partial a_j} \\ \vdots \\ a^\top \Sigma \frac{\partial a}{\partial a_p} \end{bmatrix} = 2 \frac{\partial a^\top}{\partial a} \Sigma a$$

On conclut en remarquant que $\frac{\partial a^\top}{\partial a} = I_p$.

De la même façon, on a $\frac{\partial a^\top a}{\partial a} = 2a$.

Recherche du premier axe

On cherche à maximiser

$$g(a) = a^T \Sigma a - \lambda_1 (a^T a - 1)$$

On obtient donc

$$\begin{cases} 2\Sigma a_1 - 2\lambda_1 a_1 = 0 \\ a_1^T a_1 - 1 = 0 \end{cases}$$

On en déduit que a_1 est vecteur propre de la matrice Σ associé à la valeur propre λ . En multipliant la première ligne à gauche par a_1^T et en utilisant la seconde ligne du système on trouve que

$$a_1^T \Sigma a_1 = \lambda_1$$

On reconnaît que le premier membre de l'équation précédente est égal à l'inertie $I_{\tilde{\Delta}_1}$ qui doit être maximum. Cela signifie que la valeur propre λ_1 est la plus grande valeur propre de la matrice de covariance Σ et que cette valeur propre est égale à l'inertie portée par l'axe Δ_1 .

Recherche des axes suivants

On recherche ensuite une deuxième axe Δ_2 orthogonal au premier et d'inertie minimum.

A nouveau, on définit l'axe Δ_2 par son vecteur unitaire a_2 .

L'inertie du nuage des individus par rapport à son complémentaire orthogonal est égale à

$$I_{\tilde{\Delta}_2} = a_2^\top \Sigma a_2$$

elle doit être maximum avec les deux contraintes suivantes

$$a_2^\top a_2 = 1 \text{ et } a_2^\top a_1 = 0$$

En appliquant les multiplicateurs de Lagrange, on trouve que a_2 est le vecteur propre de Σ correspondant à la deuxième plus grande valeur propre.

Recherche des axes suivants

On peut rechercher les autres axes en suivant la même procédure.

Les nouveaux axes sont tous les vecteurs propres de Σ correspondant aux valeurs propres ordonnées. La matrice Σ étant une matrice symétrique réelle, elle possède p vecteurs propres réels formant une base orthogonale de \mathbb{R}^p

$$\left\{ \begin{array}{cccccc} \Delta_1 & \perp & \Delta_2 & \perp & \dots & \perp & \Delta_p \\ a_1 & \perp & a_2 & \perp & \dots & \perp & a_p \\ \lambda_1 & \geq & \lambda_2 & \geq & \dots & \geq & \lambda_p \\ l_{\tilde{\Delta}_1} & \geq & l_{\tilde{\Delta}_2} & \geq & \dots & \geq & l_{\tilde{\Delta}_p} \end{array} \right.$$

On passera de la base orthogonale initiale des variables centrées à la nouvelle base orthogonale des vecteurs propres de Σ . On appelle les nouveaux axes, **axes principaux**.

Contributions des axes à l'inertie totale

En utilisant le théorème de Huygens, on peut décomposer l'inertie totale du nuage des individus

$$I_G = I_{\tilde{\Delta}_1} + I_{\tilde{\Delta}_2} + \dots + I_{\tilde{\Delta}_p} = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

La contribution absolue de l'axe Δ_k à l'inertie totale du nuage des individus est égale à

$$CTA(\Delta_k) = \lambda_k$$

Sa contribution relative (pourcentage d'inertie expliquée par Δ_k) est égale à

$$CTR(\Delta_k) = \frac{\lambda_k}{I_G}$$

On peut étendre ces définitions à tous les sous-espaces engendrés par les nouveaux axes. Ainsi le pourcentage d'inertie expliqué par le plan engendré par les deux premiers axes Δ_1 et Δ_2 est égal à

$$CTR(\Delta_1 \oplus \Delta_2) = \frac{\lambda_1 + \lambda_2}{I_G}$$

Contributions des axes à l'inertie totale

Ces pourcentages d'inertie sont des indicateurs qui rendent compte de la part de variabilité du nuage des individus expliquée par ces sous-espaces.

Si les dernières valeurs propres ont des valeurs faibles, on pourra négliger la variabilité qu'expliquent les axes correspondant.

On se contente souvent de faire des représentations du nuage des individus dans un sous-espace engendré par les d premiers axes si ce sous-espace explique un pourcentage d'inertie proche de 1.

On peut ainsi réduire l'analyse à un sous-espace de dimension $d < p$.

Représentation des individus dans les nouveaux axes

Pour faire la représentation des individus dans les plans définis par les nouveaux axes, il suffit de calculer les coordonnées des individus dans le nouveaux axes.

Pour obtenir y_{ik} , coordonnée de l'observation i sur l'axe Δ_k , on projette orthogonalement le vecteur de l'observation i sur cet axe. On obtient alors (rappel, on suppose $G = o$)

$$Y_i = A^T U_i$$

où Y_i est le vecteur des coordonnées de l'observation i dans la nouvelle base, et A est la matrice du changement de base

A matrice des vecteurs propres orthogonaux et de norme 1 est une matrice orthogonale, son inverse est égale à sa transposée.

Représentation des individus dans les nouveaux axes

On veut montrer que $A^\top = A^{-1}$ où A est une matrice orthogonale.
Si on écrit cette matrice suivant ses vecteurs colonnes, on obtient

$$A = \begin{bmatrix} C_1 & C_2 & \dots & C_p \end{bmatrix}$$

avec

$$\|C_j\|^2 = 1 \text{ et } C_j^\top C_{j'} = 0, \forall j = 1, \dots, p \text{ et } j' \neq j.$$

Ainsi

$$A^\top A = \begin{bmatrix} C_1^\top C_1 & \dots & C_1^\top C_j & \dots & C_1^\top C_p \\ \vdots & \ddots & \vdots & & \vdots \\ C_j^\top C_1 & \dots & C_j^\top C_j & \dots & C_j^\top C_p \\ \vdots & & \vdots & \ddots & \vdots \\ C_p^\top C_1 & \dots & C_p^\top C_j & \dots & C_p^\top C_p \end{bmatrix}$$

On a alors $A^\top A = I_p$ et donc $A^\top = A^{-1}$.

Représentation des individus dans les nouveaux axes

Remarque: l'orientation des axes est complètement arbitraire et peut différer d'un logiciel à l'autre. Le signe des coordonnées des individus sur un axe n'a donc pas de signification. En revanche, la comparaison des signes peut s'interpréter. Si deux individus i et i' ont sur un axe Δ , le premier une coordonnée positive et le second une coordonnée négative, cela signifie qu'ils s'opposent entre eux.

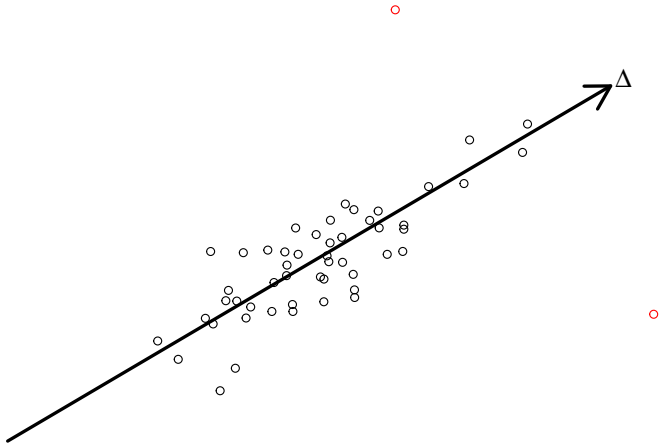
Qualité de la représentation des individus

Lorsque des individus sont éloignés sur un axe (ou sur un plan), on peut assurer que les points représentant ces individus sont éloignés dans l'espace natif.

En revanche, deux individus dont les projections sont proches sur un axe (ou sur un plan) peuvent ne pas être proches dans l'espace.

Pour interpréter correctement la proximité des projections de deux individus sur un plan, il faut donc s'assurer que ces individus sont bien représentés dans le plan.

Qualité de la représentation des individus



Qualité de la représentation des individus

Pour que l'individu i soit bien représenté sur un axe (ou sur un sous-espace), il faut que l'angle entre x_i et l'axe soit petit. On calcule donc le carré du cosinus de cet angle.

En effet, on peut montrer que le carré du cosinus de l'angle d'un vecteur avec un plan engendré par deux vecteurs orthogonaux est égal à la somme des carrés des cosinus des angles du vecteur avec chacun des deux vecteurs qui engendrent le plan.

Ainsi, le carré du cosinus de l'angle α_{ik} entre l'observation i et l'axe Δ_k de vecteur directeur unitaire a_k est égal à

$$\cos^2(\alpha_{ik}) = \frac{\langle U_i, a_k \rangle^2}{\|U_i\|^2 \|a_k\|^2} = \frac{a_k^\top U_i U_i^\top a_k}{U_i^\top U_i} = \frac{(\sum_{j=1}^p x_{ij} a_{kj})^2}{\sum_{j=1}^p x_{ij}^2}$$

De plus, le carré du cosinus de l'angle $\alpha_{ikk'}$ entre l'observation i et le plan engendré par deux axes $\Delta_k \oplus \Delta_{k'}$ est

$$\cos^2(\alpha_{ikk'}) = \cos^2(\alpha_{ik}) + \cos^2(\alpha_{ik'})$$

Interprétation des nouveaux axes en fonction des individus

Lorsqu'on calcule l'inertie $I_{\tilde{\Delta}_k}$ portée par l'axe Δ_k , on peut voir quelle est la part de cette inertie due à un individu en particulier.

En rappelant que $I_{\tilde{\Delta}_k} = \frac{1}{n} d^2(h_{\Delta_{ki}}, o)$, la contribution absolue de l'observation i à cette inertie est égale à

$$CTA(U_i; k) = \frac{1}{n} d^2(h_{\Delta_{ki}}, o)$$

Un individu contribuera d'autant plus à la construction d'un axe, que sa projection sur cet axe sera éloignée du centre de gravité du nuage(ici o). Inversement un individu dont la projection sur un axe sera proche du centre de gravité contribuera faiblement à l'inertie portée par cette axe.

On se sert de ces contributions pour interpréter les nouveaux axes de l'ACP en fonction des individus.

Contribution relative d'un individu à un axe

On peut, pour un individu i , donner sa contribution relative à l'inertie portée par cet axe

$$CTR(U_i; k) = \frac{\frac{1}{n} d^2(h_{\Delta_{ki}}, o)}{I_{\tilde{\Delta}_k}} = \frac{\frac{1}{n} a_k^\top U_i U_i^\top a_k}{\lambda_k}$$

Représentation des variables

On peut envisager le problème de la représentation des variables de façon complètement symétrique de celui des individus. Les raisonnements se font dans \mathbb{R}^n au lieu de \mathbb{R}^p .

Les individus étaient décrits dans l'espace des anciennes variables et on a fait un changement de base dans cet espace. Les nouveaux axes sont des combinaisons linéaires des anciens axes et peuvent donc être considérés comme de nouvelles variables (combinaisons linéaires des anciennes). On appelle généralement ces nouvelles variables "composantes principales".

On note Z_1, \dots, Z_p les composantes principales, Z_k étant la nouvelle variable correspondant à l'axe Δ_k :

$$Z_k = \sum_{j=1}^p a_{kj} V_j = Xa_k$$

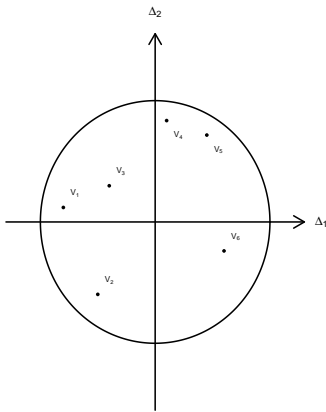
et de façon générale:

$$Z = \begin{bmatrix} Z_1 & Z_2 & \dots & Z_p \end{bmatrix} = XA$$

Représentation des variables

Il est alors intéressant de voir comment les anciennes variables sont liées aux nouvelles. Pour cela, on calcule les corrélations des anciennes variables avec les nouvelles. La représentation des anciennes variables se fera en prenant comme coordonnées des anciennes variables leurs coefficients de corrélation avec les nouvelles variables.

On obtient alors “le cercle des corrélations”:



Représentation des variables

On peut montrer que les variances, covariances et coefficients de corrélations empiriques des composantes principales entre elles ou avec les variables de départ sont

$$\text{Var}(Z_k) = \frac{1}{n} a_k^\top X^\top X a_k = a_k^\top \Sigma a_k = \lambda_k$$

$$\text{Cov}(Z_k, V_j) = \frac{1}{n} a_k^\top X^\top V_j = \frac{1}{n} a_k^\top X^\top X = a_k^\top \Sigma a_k \mathbf{1}_j = \lambda_k a_k^\top \mathbf{1}_j = \lambda_k a_{kj}$$

$$\text{où } \mathbf{1}_j^\top = \begin{bmatrix} 0 & 0 & \dots & 1 & 0 & \dots & 0 \end{bmatrix}$$

$$\text{Cor}(Z_k, V_j) = \frac{\sqrt{\lambda_k} a_{kj}}{\sqrt{\text{Var}(V_j)}}$$

Représentation des variables

De façon générale, la matrice de covariance des composantes principales est égale à

$$\Sigma_Z = \frac{1}{n} A^\top X^\top X A = A^\top \Sigma A = \Lambda$$

où $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ est la matrice diagonale des valeurs propres de Σ .

La matrice des covariances entre les composantes principales et les anciennes variables vaut

$$\text{Cov}(Z, V) = \frac{1}{n} X^\top X A = \Sigma A = A \Lambda$$

Remarquons que la variance empirique d'une variable est égale au carrée de la norme du vecteur qui la représente. De plus, le coefficient de corrélation empirique de deux variables est égal au produit scalaire des deux vecteurs qui les représentent, on pourra donc interpréter les angles des vecteurs comme des corrélations!

Interprétation des axes en fonction des anciennes variables

On peut interpréter les axes principaux en fonction des anciennes variables.

Une ancienne variables V_j expliquera d'autant mieux un axe principal qu'elle sera fortement corrélée avec la composante principale correspondant à cet axe.

Une variable sera d'autant mieux représentée sur un axe que sa corrélation avec la composante principale correspondante est en valeur absolue proche de 1.

Étude des liaisons entre les variables

Sur le graphique du cercle des corrélations, on peut aussi interpréter les positions des anciennes variables les unes par rapport aux autres en termes de corrélations.

Deux variables très proches du cercle des corrélations, donc bien représentées dans le plan, seront très corrélées positivement entre elles.

Si elles sont proches du cercle, mais dans des positions symétriques par rapport à l'origine, elles seront très corrélées négativement.

Deux variables proches du cercle des corrélations et donc les vecteurs qui les joignent à l'origine forment un angle droit ne seront pas corrélées entre elles.

Attention Le coefficient de corrélation est une mesure de liaison linéaire uniquement!

Analyse en composantes principales normée

Nous avons étudié l'ACP simple:

- les individus ont le même poids dans l'analyse
- les variables sont traitées de façon symétrique (on leur fait jouer le même rôle)

Deux critiques:

- si les anciennes variables sont hétérogènes (poids, taille, age...), quels sens peut-on donner aux combinaisons linéaires de ces variables?
- si on change d'unité sur les variables, on peut changer complètement les résultats de l'ACP.

Solution:

- normaliser les variables (par exemple, peut travailler avec les variables centrées et réduites)

Choix du nombre d'axes

Lorsque l'ACP est utilisée pour réduire la dimension des données, le nombre d'axes à conserver doit être inférieur au nombre de variables.

Pour choisir le nombre d'axes à conserver, voici trois critères ad-hoc:

- Règle de la part d'inertie : on décide de garder q tel que la part d'inertie expliquée est supérieure à un seuil fixé a priori

$$r_q = \frac{\sum_{k=1}^q \lambda_k}{\sum_{k=1}^p \lambda_k}$$

- Règle de Kaiser : on ne conserve que les axes dont la valeur propre est supérieure à leur moyenne car ce sont les seules jugées informatives.
- Éboulis des valeurs propres : on sélectionne q en localisant le « coude » dans l'éboulis des valeurs propres.

Effet taille

Lorsque toutes les variables sont corrélées positivement entre elles, la première composante principale définit un « facteur taille ».

Dans ce cas, la première composante n'est pas utilisée pour la visualisation.

La seconde composante principale différencie alors les individus de taille semblable, c'est le facteur de « forme ».

Formule de reconstitution

Analyse en composantes principales normée

Travailler avec les variables centrées et réduites revient à faire l'ACP simple en utilisant une autre métrique. La distance choisie est alors

$$d^2(x_i, x_{i'}) = \sum_{j=1}^p \frac{1}{\sigma_j^2} (x_{ij} - x_{i'j})^2.$$

Si on reprend tous les calculs de l'ACP simple, en utilisant les variables centrées réduites, on voit que c'est maintenant la matrice de corrélation **R** qui intervient pour la recherche des nouveaux axes.

Par conséquent, l'inertie totale du nuage des individus dans \mathbb{R}^p est toujours égale à p pour l'ACP normée. Cette particularité peut donner la règle empirique suivante pour choisir le nombre d'axes: on considère comme significatives les composantes issues des valeurs propres dont la valeur est supérieure à 1.

Individus et variables supplémentaires

Il arrive que l'on veuille faire apparaître dans les représentation graphiques certains individus sans qu'ils interviennent dans la détermination des axes.

Il en est de même pour les variables. On peut, par exemple, vouloir représenter une variables qui dépend de façon synthétique des p variables choisies pour faire l'ACP mais on ne souhaite pas qu'elle intervienne dans la confection des axes car sa liaisons avec les p variables de départ fausseraient la représentation.

Pour représenter un individu supplémentaire, il suffit d'exprimer les coordonnées de cet individu dans la nouvelle base des axes principaux.

Pour une variable supplémentaire, il suffit de calculer ses coefficients de corrélation empiriques avec les composantes principales.