Produit Contrainte lignes/colonnes

$$\underset{(n\times p)}{\mathbf{A}} \times \underset{(p\times k)}{\mathbf{B}} \Longrightarrow \underset{(n\times k)}{\mathbf{C}}$$

Nombre de colonnes de la première matrice égal au nombre de lignes de la seconde

$$\mathbf{AB} \neq \mathbf{BA}, \qquad (\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$$
$$\mathbf{I}_n\mathbf{A} = \mathbf{AI}_p = \mathbf{A} \qquad \mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$$

On peut aussi calculer $\alpha\mathbf{A}$: tous les coefficients de $\mathbf{A}$ sont multipliés par le réel α.

Inverse si $\mathbf{A}$ et $\mathbf{B}$ sont carrées de taille n, alors

$$\mathbf{AB} = \mathbf{I}_n \implies \mathbf{BA} = \mathbf{I}_n \qquad \text{On note } \mathbf{B} = \mathbf{A}^{-1} \text{ (inverse de } \mathbf{A})$$

Transposition échange des lignes et des colonnes d'une matrice ; on note $\mathbf{A}'$ la transposée de $\mathbf{A}$.

$$(\mathbf{A}')' = \mathbf{A}, \qquad (\alpha\mathbf{A})' = \alpha\mathbf{A}',$$
$$(\mathbf{A}+\mathbf{B})' = \mathbf{A}'+\mathbf{B}', \qquad (\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$$

Une matrice carrée telle que $\mathbf{A}' = \mathbf{A}$ est dite *symétrique*.

Trace la trace d'une matrice carrée est la somme des termes de sa diagonale

$$\mathrm{Tr}(\alpha\mathbf{A}) = \alpha\mathrm{Tr}(\mathbf{A}), \quad \mathrm{Tr}(\mathbf{A}+\mathbf{B}) = \mathrm{Tr}(\mathbf{A})+\mathrm{Tr}(\mathbf{B}),$$
$$\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA}),$$

$$\mathrm{Tr}(\mathbf{ABC}) = \mathrm{Tr}(\mathbf{CAB}) = \mathrm{Tr}(\mathbf{BCA}) \neq \mathrm{Tr}(\mathbf{CBA})$$

Tableau de données

On note x_i^j la valeur de la *variable* $\mathbf{x}^j$ pour le i^{e} individu. $\mathbf{X} = (\mathbf{x}^1, \ldots, \mathbf{x}^p)$ est une matrice rectangulaire à n lignes et p colonnes.

$$\mathbf{x}^j = \begin{bmatrix} x_1^j \\ x_2^j \\ \\ \vdots \\ \\ x_n^j \end{bmatrix}, \ \mathbf{X} = \begin{bmatrix} x_1^1 & x_1^2 & & & \cdots & x_1^p \\ x_2^1 & x_2^2 & & & & \\ & & & \vdots & & \\ \vdots & & \cdots & x_i^j & & \\ & & & & \ddots & \\ x_n^1 & & & & & x_n^p \end{bmatrix}.$$

Un *individu* est représenté par

$$\mathbf{e}_i' = [x_i^1, \ldots, x_i^j, \ldots, x_i^p]$$

La matrice des poids

Définition on associe aux individus un poids p_i tel que

$$p_1 + \cdots + p_n = 1$$

que l'on représente par la matrice diagonale de taille n

$$\mathbf{D_p} = \begin{bmatrix} p_1 & & & 0 \\ & p_2 & & \\ & & \ddots & \\ 0 & & & p_n \end{bmatrix}.$$

Symétrie La matrice $\mathbf{D_p}$ est diagonale et donc symétrique : $\mathbf{D_p}' = \mathbf{D_p}$.

Cas uniforme tous les individus ont le même poids $p_i = 1/n$ et $\mathbf{D_p} = \frac{1}{n}\mathbf{I}_n$.

Point moyen et tableau centré

Point moyen c'est le vecteur $\mathbf{g}$ des moyennes arithmétiques de chaque variable :

$$\mathbf{g}' = (\bar{x}^1, \ldots, \bar{x}^p) = \sum_{i=1}^{n} p_i\mathbf{e}_i'.$$

On peut écrire sous forme matricielle

$$\mathbf{g} = \mathbf{X}'\mathbf{D_p}\mathbf{1}_n.$$

Tableau centré il est obtenu en centrant les variables autour de leur moyenne

$$y_i^j = x_i^j - \bar{x}^j, \quad \text{c'est-à-dire} \quad \mathbf{y}^j = \mathbf{x}^j - \bar{x}^j\mathbf{1}_n$$

ou, en notation matricielle,

$$\mathbf{Y} = \mathbf{X} - \mathbf{1}_n\mathbf{g}' = (\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n'\mathbf{D_p})\mathbf{X}$$

Matrice de variance-covariance

Définition c'est une matrice *carrée* de dimension p

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & & & \\ \vdots & & \ddots & \\ \sigma_{p1} & & & \sigma_p^2 \end{bmatrix},$$

où $\sigma_{j\ell}$ est la covariance des variables $\mathbf{x}^j$ et $\mathbf{x}^\ell$ et σ_j^2 est la variance de la variable $\mathbf{x}^j$

Symétrie Comme $\sigma_{j\ell} = \sigma_{\ell j}$, la matrice $\mathbf{V}$ est symétrique : $\mathbf{V}' = \mathbf{V}$.

Formule matricielle

$$\mathbf{V} = \mathbf{X}'\mathbf{D_p}\mathbf{X} - \mathbf{gg}' = \mathbf{Y}'\mathbf{D_p}\mathbf{Y}.$$

Matrice de corrélation

Définition Si l'on note $r_{j\ell} = \sigma_{j\ell}/\sigma_j\sigma_\ell$, c'est la matrice $p \times p$

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & & \\ \vdots & & \ddots & \\ r_{p1} & & & 1 \end{bmatrix},$$

Symétrie Comme $r_{j\ell} = r_{\ell j}$, la matrice $\mathbf{R}$ est symétrique : $\mathbf{R}' = \mathbf{R}$.

Formule matricielle $\mathbf{R} = \mathbf{D}_{1/\sigma}\mathbf{V}\mathbf{D}_{1/\sigma}$, où

$$\mathbf{D}_{1/\sigma} = \begin{bmatrix} \frac{1}{\sigma_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_p} \end{bmatrix}$$

Les données centrées réduites

Définition c'est le tableau $\mathbf{Z}$ contenant les données

$$z_i^j = \frac{y_i^j}{\sigma_j} = \frac{x_i^j - \bar{x}^j}{\sigma_j}, \quad \text{c'est-à-dire} \quad \mathbf{z}^j = \frac{\mathbf{y}^j}{\sigma_j}$$

qui se calcule matriciellement comme $\mathbf{Z} = \mathbf{Y}\mathbf{D}_{1/\sigma}$

Pourquoi réduites ?

— pour que les distances soient indépendantes des unités de mesure,

— pour ne pas privilégier les variables dispersées.

Covariances comme $\bar{z}^j = \bar{y}^j = 0$, les covariances des $\mathbf{z}^j$ sont des corrélations :

$$\operatorname{cov}(\mathbf{z}^k, \mathbf{z}^\ell) = \sum_{i=0}^n p_i z_i^k z_i^\ell = \frac{1}{\sigma_k \sigma_\ell} \sum_{i=0}^n p_i y_i^k y_i^\ell = \operatorname{cor}(\mathbf{x}^k, \mathbf{x}^\ell).$$

La matrice de variance-covariance des variables centrées-réduites est donc la matrice de corrélation $\mathbf{R}$.

Partie III. Données : vision géométrique

L'analyse de composantes principales (ACP)

Contexte chaque individu est considéré comme un point d'un espace vectoriel F de dimension p. Ses coordonnées dans F sont

$$(x_i^1, x_i^2, \ldots, x_i^p).$$

L'ensemble des individus est un *nuage de points* dans F et $\mathbf{g}$ est son *centre de gravité*.

Principe on cherche à réduire le nombre p de variables tout en préservant au maximum la structure du problème.

Pour cela on projette le nuage de points sur un sous-espace de dimension inférieure.

Distance entre individus

Motivation afin de pouvoir considérer la structure du nuage des individus, il faut définir une distance, qui induira une géométrie.

Distance euclidienne classique la distance la plus simple entre deux points de $\mathbb{R}^p$ est définie par

$$d^2(\mathbf{u}, \mathbf{v}) = \sum_{j=1}^p (u_j - v_j)^2 = \|\mathbf{u} - \mathbf{v}\|^2$$

Généralisation simple on donne un poids $m_j > 0$ à la variable j

$$d^2(\mathbf{u}, \mathbf{v}) = \sum_{j=1}^p m_j (u_j - v_j)^2$$

Cela revient à multiplier la coordonnée j par $\sqrt{m_j}$

Métrique

Définition soit $\mathbf{M} = \operatorname{diag}(m_j)$, où $m_1, \ldots, m_p$ sont des réels strictement positifs. On pose

$$\|\mathbf{u}\|_{\mathbf{M}}^2 = \sum_{j=1}^p m_j u_j^2 = \mathbf{u}'\mathbf{M}\mathbf{u},$$
$$d_{\mathbf{M}}^2(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|_{\mathbf{M}}^2.$$

Espace métrique il est défini par le produit scalaire

$$\langle \mathbf{u}, \mathbf{v} \rangle_{\mathbf{M}} = \sum_{j=1}^p m_j u_j v_j = \mathbf{u}'\mathbf{M}\mathbf{v}.$$

On notera que $\|\mathbf{u}\|_{\mathbf{M}}^2 = \langle \mathbf{u}, \mathbf{u} \rangle_{\mathbf{M}}$.

Orthogonalité on dit que $\mathbf{u}$ et $\mathbf{v}$ sont $\mathbf{M}$-orthogonaux si $\langle \mathbf{u}, \mathbf{v} \rangle_{\mathbf{M}} = 0$.

Propriétés du produit scalaire

Le produit scalaire est commutatif

$$\langle \mathbf{u}, \mathbf{v} \rangle_{\mathbf{M}} = \langle \mathbf{v}, \mathbf{u} \rangle_{\mathbf{M}}$$

Le produit scalaire est linéaire

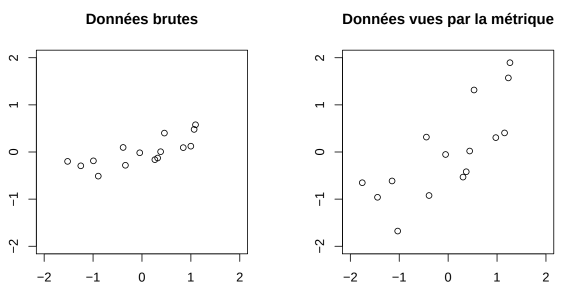
$$\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle_{\mathbf{M}} = \langle \mathbf{u}, \mathbf{v} \rangle_{\mathbf{M}} + \langle \mathbf{u}, \mathbf{w} \rangle_{\mathbf{M}},$$
$$\langle \mathbf{u}, \lambda\mathbf{v} \rangle_{\mathbf{M}} = \lambda \langle \mathbf{u}, \mathbf{v} \rangle_{\mathbf{M}} \quad \text{pour tout } \lambda \in \mathbb{R}.$$

Identité remarquable

$$\|\mathbf{u} + \mathbf{v}\|_{\mathbf{M}}^2 = \|\mathbf{u}\|_{\mathbf{M}}^2 + \|\mathbf{v}\|_{\mathbf{M}}^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle_{\mathbf{M}}$$

Utilisation des métriques

Utiliser une métrique est donc équivalent à « tordre » les données, par exemple pour les rendre comparables



Cas particuliers

Métrique usuelle Si $m_1, \ldots, m_p = 1$, alors $\mathbf{M} = \mathbf{I}_p$ et on note $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle_{\mathbf{I}}$.

Métrique réduite diviser les variables par σ_j est équivalent à prendre $m_j = 1/\sigma_j^2$. On a $\mathbf{D}_{1/\sigma^2} = \mathbf{D}_{1/\sigma}\mathbf{D}_{1/\sigma}$ et donc

$$\langle \mathbf{D}_{1/\sigma}\mathbf{u}, \mathbf{D}_{1/\sigma}\mathbf{v} \rangle = \mathbf{u}'\mathbf{D}_{1/\sigma}\mathbf{D}_{1/\sigma}\mathbf{v} = \mathbf{u}'\mathbf{D}_{1/\sigma^2}\mathbf{v} = \langle \mathbf{u}, \mathbf{v} \rangle_{\mathbf{D}_{1/\sigma^2}}.$$

Travailler avec la métrique $\mathbf{D}_{1/\sigma^2}$, c'est comme utiliser la métrique $\mathbf{I}$ sur des variables réduites.

La plupart du temps en ACP, on fait l'analyse avec la métrique usuelle sur les données centrées-réduites.