REPRÉSENTATION DES *p* VARIABLES DANS ($\mathbb{R}^n$, D)

On considère *p* variables quantitatives observées sur *n* individus. Le tableau initial des données (appelé tableau *individus* x *variables*) est présenté dans une matrice $X = (x_{ij})$ de dimension $n \times p$; x_{ij} est la valeur de la variable x_j sur l'individu *i*. La variable x_j est identifiée au vecteur $(x_{1j}, ..., x_{nj})$ de $\mathbb{R}^n$.

L'espace vectoriel $\mathbb{R}^n$ est muni du produit scalaire associé à la matrice $D = \frac{1}{n} I_n$ (produit scalaire classique au facteur $\frac{1}{n}$ près), de la distance et de la norme associées (espace euclidien ($\mathbb{R}^n$, D)).

Le produit scalaire des variables x_j et x_k est alors :

$$\langle \boldsymbol{x_j}, \boldsymbol{x_k} \rangle = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{x_{ij}}\ \boldsymbol{x_{ik}}.$$

Soit $\mathbb{1}$ = (1, 1, ..., 1), le vecteur de $\mathbb{R}^n$ dont les éléments sont tous égaux à 1.
Alors le vecteur $y_j = x_j - \bar{x}_j \mathbb{1}$ représente la variable centrée associée à x_j et le vecteur $z_j = \frac{1}{s_j}(x_i - \bar{x}_j \mathbb{1})$ avec $s_j = \sqrt{\mathrm{var}(x_j)}$ représente la variable centrée et réduite associée à x_j.

Interprétation géométrique des indices statistiques

Moyenne	$\langle x_j, 1 \rangle = \frac{1}{n} \sum_{i=1}^{n} x_{ij} = \bar{x}_j$
Covariance	$\langle y_j, y_k \rangle = \frac{1}{n} \sum_{i=1}^{n} y_{ij}\ y_{ik} = \mathrm{cov}(x_j, x_k)$
Variance	$\lVert y_j \rVert^2 = \langle y_j, y_j \rangle = \frac{1}{n} \sum_{i=1}^{n} (y_{ij})^2 = \mathrm{var}(x_j)$
Écart-type	$\lVert y_j \rVert = s_j$
Corrélation	$\cos(y_i, y_k) = \frac{\langle y_j, y_k \rangle}{\lVert y_j \rVert\ \lVert y_k \rVert} = \frac{\mathrm{cov}(x_j, x_k)}{s_j\ s_k} = r(x_j, x_k)$

$\mathbf{1}^{\perp}$ est le sous-espace de $\mathbb{R}^n$ de dimension $n-1$ (c'est-à-dire un hyperespace), orthogonal au vecteur $\mathbb{1}$; S est la sphère de $\mathbf{1}^{\perp}$ de centre 0 et de rayon 1.

On obtient la moyenne en projetant la variable sur l'axe porté par le vecteur 1.

On obtient la variable centrée en projetant la variable sur le sous-espace orthogonal à 1. (Dans $\mathbb{R}^n$, centrer les variables revient à les projeter sur l'espace orthogonal à 1).

L'extrémité de la variable centrée réduite se trouve sur la sphère unité.

La corrélation des variables est le cosinus de l'angle que forment les variables centrées.

Les variables sont représentées par des *vecteurs* de $\mathbb{R}^n$.

Le cosinus de l'angle que forment ces vecteurs dans $\mathbf{1}^{\perp}$ est égal au coefficient de corrélation de ces variables.