Regression linéaire.

La variance : $V(x) = \frac{1}{n} \sum_{i=0}^{n} x_i^2 - \bar{x}^2$

Ecart type : $\sqrt{V(x)} = \sigma x = \sqrt{\left(\frac{1}{n} \sum_{i=1} x_i^2\right) - \bar{x}^2}$

La moyenne : $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

Coefficient de variation : $CV = \frac{\sigma x}{\bar{x}}$

⚠ plus la valeur de CV est ↑ plus la dispertion autour la moyenne est grande

La covariance : $cov(x,y) = \left(\frac{1}{n} \sum_{i=1}^{n} x_i y_i\right) - \bar{x}\,\bar{y}$

⚠ $Cov(x,y) > 0 \Rightarrow$ Les variables x et y varient ds le même sens
$Cov(x,y) < 0 \Rightarrow$ " " " " le sens contraire

Le coefficient de corrélation linéaire :

$r = \frac{cov(x,y)}{\sigma x \times \sigma y}$

⚠ $r = -1$: La correlation entre x et y est parfaitement negative
$r = 1$: " " " " " " positive
$r < 0$: " " " est (forte / moyenne / faible) negative
$r \to 0$: " " " est (" / " / ") positive
$r = 0$: absence de liaison lineaire

Les coefficients de la droite

$$\begin{cases} a = \frac{Cov(x,y)}{V(x)} \\ b = \bar{y} - a\bar{x} \end{cases}$$

$\bar{x} = 0 \Rightarrow X$ variable centrée
$\sigma x = 1 \Rightarrow X$,, réduite
$X^* = \frac{x_i - \bar{x}}{\sigma x}$

Centrage et réduction

On pose $z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma x_j}$

ind	X_1^*	X_2^*	....	X_p^*
ind_1	z_{11}	z_{12}	....	z_{1p}
⋮				
ind_n	z_{n1}	..	....	z_{np}
$\bar{x}_j$	0	0	--	0
σx	1	1	...	1

Le tableau ~~sous~~ sous forme matricielle

$$Z = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & & \\ \cdots & \cdots & \cdots & \end{pmatrix} \in \mathbb{R}^{n \times p}$$

Z est une matrice centrée réduite.

variable centrée réduite :
- Centre de gravité $g = (0, 0, 0, \dots, 0)^T$
- Inertie total : $I_{tot}(g = 0) = \sum_{j=1}^{p} V(X_j^*)$

$$= \underbrace{1 + + \cdots + 1}_{p \text{ fois}} = p$$

Matrice de corrélation R.

+ $R = \frac{1}{n} Z^T . Z$
+ ou bien en utilisant $\bar{x}_j^* = 0$ et $\sigma x_j^* = 1$

$r(X_l^*, X_k^*) = \frac{1}{n} \sum_{i=1}^{n} z_{il} z_{ik}$ avec $r_{p,l}$ les valeurs de Z

$$r_{ij} = r(X_i, X_j) = \frac{Cov(X_i, X_j)}{\sigma X_i \times \sigma X_j}$$

$$r_{ij} = r(X_i, X_j) = r(X_j, X_i) = r_{ji}$$

$$r_{ii} = \frac{Cov(X_i, X_i)}{\sigma^2 X_i} = \frac{V(X_i)}{V(X_i)} = 1$$

Produit Matriciel

$A \in R^{n \times m}$ $\quad B \in R^{m \times p}$

$A . B \in \mathbb{R}^{n \times p}$ $\quad (n \times \cancel{m})(\cancel{m} \times p)$

Interprétation de corrélation

−1 fort −0,5 faible 0 0,5 fort 1

Les composantes principales (facteurs)

$Y_j = Z \cdot v_j$

avec Z matrice centrée reduite

v_j : Les vecteurs propres de R.

Choix des composantes principales

valeur propre	Inertie expliqué (%)	Inertie cumulé
λ_1	$\frac{\lambda_1}{p} \times 100$	$\frac{\lambda_1}{p} \times 100$
λ_2	$\frac{\lambda_2}{p} \times 100$	$\left(\frac{\lambda_1}{p} + \frac{\lambda_2}{p}\right) \times 100$
$\vdots$	$\vdots$	
λ_p	$\frac{\lambda_p}{p} \times 100$	$\left(\frac{\lambda_1}{p} + \frac{\lambda_2}{p} \cdots + \frac{\lambda_p}{p}\right) \times 100$

+ Comment choisir les facteurs à Retenir?
  Extraire les facteurs de façons à expliquer au moins 80% de la variance totale.

$$R v_j = \lambda_j v_j$$

ACP :

Matrice : $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{ij} & \cdots & a_{1m} \\ a_{21} & & & & & \\ \vdots & & & & & \\ a_{n1} & \cdots & \cdots & \cdots & \cdots & a_{nm} \end{pmatrix} \in \mathbb{R}^{n \times m}$

avec n : nombre de ligne
m : nbre de colonne

Matrice transposé : A^T

exp $A = \begin{pmatrix} 2 & 6 & 4 \\ 5 & 19 & 10 \end{pmatrix}$ $A^T = \begin{pmatrix} 2 & 5 \\ 6 & 19 \\ 4 & 10 \end{pmatrix}$

transposé d'un vecteur-colonne est un vecteur ligne.
$x \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $x^T (x_1 \cdots x_n)$

Tableau individus/variables

ind	X_1	X_2	...	X_p
ind 1				
ind n				
moyenne				
ecart				

Le tableau sous forme matriciel
$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & & & \\ x_{n1} & \cdots & \cdots & x_{np} \end{pmatrix} \in \mathbb{R}^{n \times p}$

Chaque individu est decrit par p variables, formant un vecteur de dimension p :
$\text{ind } i = (x_{i1}, x_{i2} \cdots x_{ip})^T = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix}$

La distance entre deux individus :

$d(\text{ind}_\ell, \text{ind}_k) = \sqrt{\sum_{j=1}^{p} (x_{\ell j} - x_{kj})^2}$

$\text{ind}_\ell = (x_{\ell 1}, x_{\ell 2}, \cdots, x_{\ell p})^T$

$\text{ind}_k = (x_{k1}, \cdots\cdots, x_{kp})^T$

exp : $\text{ind}_1 (5, 17, 6)$ $\text{ind}_2 (10, 2, 6)$

$d(\text{ind}_1, \text{ind}_2) = \sqrt{(10-5)^2 + (2-17)^2 + (6-6)^2} = 15,81$

Le centre de gravité : $g = (\bar{x}_1, \bar{x}_2, \bar{x}_3)^T$

Inertie totale : $I_{tot}(g) = \sum_{j=1}^{p} V(X_j)$