Cours 06
Variance, covariance, corrélation

1 Définitions et propriétés élémentaires

Définition : Soient $x = (x_1, x_2, \ldots, x_n)$ et $y = (y_1, y_2, \ldots, y_n)$ deux échantillons de même taille n et de moyenne respective $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ et $\bar{y} = \frac{1}{n}\sum_{i=1}^n$.
On appelle *covariance* (estimée) de x avec y et on note $\mathrm{cov}\,(x, y)$ le nombre

$$\mathrm{cov}\,(x, y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).$$

On appelle *variance* (estimée) de x et on note $\mathrm{Var}\,(x)$ le nombre

$$\mathrm{Var}\,(x) = \mathrm{cov}\,(x, x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

On appelle *corrélation* de x avec y et on note $\mathrm{cor}\,(x, y)$ le nombre

$$\rho = \mathrm{cor}\,(x, y) = \frac{cov(x, y)}{\sqrt{\mathrm{Var}\,(x)}\sqrt{\mathrm{Var}\,(y)}}.$$

Proposition 1 *Soient x,, x', x'', y, y', et y'' des échantillons de même taille n. Notons $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ et $\bar{y} = \frac{1}{n}\sum_{i=1}^n$ la moyenne de x et de y. Soient λ, λ' et λ'' des nombres. Les relations suivantes sont satisfaites :*

1. $cov(\lambda' x' + \lambda'' x'', y) = \lambda' cov(x', y) + \lambda'' cov(x'', y)$ *(Bilinéarité : linéarité à gauche)*
2. $cov(x, \lambda' y' + \lambda'' y'') = \lambda' cov(x, y') + \lambda'' cov(x, y'')$ *(Bilinéarité : linéarité à droite)*
3. $cov(y, x) = cov(x, y)$ *(Symétrie)*
4. $Var(\lambda x) = \lambda^2 Var(x) \geq 0$ *(Homogénéité de degré 2)*

L'inégalité de Cauchy-Schwarz $(cov(x, y))^2 \leq Var(x)\,Var(y)$ montre que $\rho^2 \leq 1$.

2 Relation avec la régression linéaire

Lors du cours 05 nous avons déterminé l'équation $y = ax + b$ de la droite de régression linéaire de l'échantillon y sur l'échantillon de même taille x. Les valeurs de la pente a de cette droite et son ordonnée à l'origine b s'expriment par

$$a = \frac{\mathrm{cov}\,(x, y)}{\mathrm{Var}\,(x)} \text{ et } b = \bar{y} - a\bar{x}.$$

3 Matrice de variance-covariance

On appelle matrice de variance-covariance de x et y la matrice symétrique $\begin{pmatrix} \mathrm{Var}\,(x) & \mathrm{cov}\,(x, y) \\ \mathrm{cov}\,(y, x) & \mathrm{Var}\,(y) \end{pmatrix}$.
On appelle matrice de corrélation de x et y la matrice symétrique $\begin{pmatrix} 1 & \mathrm{cor}\,(x, y) \\ \mathrm{cor}\,(y, x) & 1 \end{pmatrix}$.

Si on a m échantillons x^1, …, x^m de même taille on définit leur matrice de variance-covariance Σ et leur matrice de corrélation R par

$$\Sigma = \begin{pmatrix} \mathrm{Var}\,(x^1) & \cdots & \mathrm{cov}\,(x^1, x^k) & \cdots & \mathrm{cov}\,(x^1, x^m) \\ \vdots & \ddots & \vdots & & \vdots \\ \mathrm{cov}\,(x^k, x^1) & \cdots & \mathrm{Var}\,(x^k) & \cdots & \mathrm{cov}\,(x^k, x^m) \\ \vdots & & \vdots & \ddots & \vdots \\ \mathrm{cov}\,(x^m, x^1) & \cdots & \mathrm{cov}\,(x^m, x^k) & \cdots & \mathrm{Var}\,(x^m) \end{pmatrix} \text{ et } R = \begin{pmatrix} 1 & \cdots & \mathrm{cor}\,(x^1, x^k) & \cdots & \mathrm{cor}\,(x^1, x^m) \\ \vdots & \ddots & \vdots & & \vdots \\ \mathrm{cor}\,(x^k, x^1) & \cdots & 1 & \cdots & \mathrm{cor}\,(x^k, x^m) \\ \vdots & & \vdots & \ddots & \vdots \\ \mathrm{cor}\,(x^m, x^1) & \cdots & \mathrm{cor}\,(x^m, x^k) & \cdots & 1 \end{pmatrix}.$$