# Régression linéaire multiple

## Exercice 1

On considère une modélisation linéaire simple $Y_i = z_i a + b + \varepsilon_i$ où $\varepsilon_i \sim \mathcal{N}(0, \sigma)$ pour $n > 2$ observations des caractères $z$ et $y$.

On choisit d'utiliser les outils de modélisation linéaire multiple et de modéliser le problème à l'aide du langage matricielle $Y = xm + \varepsilon$.

1. Donner les dimensions des matrices $Y$, $x$, $m$ et $\varepsilon$.
2. Montrer que :
   (a) ${}^t xx = \begin{pmatrix} n\overline{z^2} & n\overline{z} \\ n\overline{z} & n \end{pmatrix}$
   (b) ${}^t xy = \begin{pmatrix} n\overline{zy} \\ n\overline{y} \end{pmatrix}$
3. On peut montrer que l'inverse de la matrice $A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ est la matrice $\dfrac{1}{\det(A)} \begin{pmatrix} \delta & -\beta \\ -\gamma & \alpha \end{pmatrix}$.
   (a) Montrer que $\det({}^t xx) = n^2\sigma_z^2$.
   (b) Montrer que $({}^t xx)^{-1} = \dfrac{1}{n\sigma_z^2} \begin{pmatrix} 1 & -\overline{z} \\ -\overline{z} & \overline{z^2} \end{pmatrix}$.
   (c) En déduire la valeur de $\hat{m}$.
   (d) Comparez ces résultats avec les estimations de la modélisation linéaire simple.
4. (a) Donnez une estimation de la matrice de variance-covariance de l'estimateur de $m$.
   (b) En déduire les valeurs de $\hat{\sigma_a}^2$ et $\hat{\sigma_b}^2$. Comparez vos résultats avec les variances de la modélisation linéaire simple.
   (c) Donnez une estimation de $Cov(A_n, B_n)$ où $A_n$ est l'estimateur de $a$ et $B_n$ l'estimateur de $b$.

## Exercice 2

Le prof, ce petit malin, a demandé à python deux nombres de manière équiprobable, appelé $x$ et $y$, parmi $-1$, $0$ et $1$ puis a appliquer une petite formules $z = a + bx + cy + \varepsilon$ où $a$, $b$ et $c$ sont des constantes et $\varepsilon$ est la partie entière d'un tirage aléatoire d'une loi normale centré de variance $\sigma$. Dans ce modèle il y a donc 4 inconnues : $a$, $b$, $c$ et $\sigma$. Le prof, ce petit malin, a réalisé cette expérience aléatoire cinq fois. Voici ce qu'il a obtenue.

| $x$ | $y$ | $z$ |
|---|---|---|
| 0 | 0 | 2 |
| 0 | −1 | 1 |
| 0 | 0 | 0 |
| 0 | 0 | 4 |
| 1 | 1 | 1 |

Le but du problème est d'estimer les 4 inconnues. On utilise le langage matricielle et on modélise le problème sous la forme $Z = Am + \varepsilon$.

1. Donner la matrice $A$ ainsi que ${}^t AA$.
2. La matrice $B = \begin{pmatrix} 1 & -2 & 1 \\ -2 & 10 & -5 \\ 1 & -5 & 4 \end{pmatrix}$ n'est pas la matrice inverse de ${}^t AA$. Cependant on peut démontrer qu'il existe un réel $\lambda$ tel que $({}^t AA)^{-1} = \lambda B$.
   (a) Déterminer la valeur de $\lambda$.
   (b) En déduire $\hat{m}$ l'estimation ponctuelle de $m$.
3. Quelle opération matricielle permet d'obtenir les valeurs de $\hat{z}_i$ ? Réalisez cette opération.
4. Donner une estimation de $\sigma$.

5. On note $\overrightarrow{X}$ le vecteur aléatoire en dimension 3 de coordonnée $A_n$, $B_n$ et $C_n$ les estimateurs des paramètres $a$, $b$ et $c$.
   (a) Donnez $\mathbb{V}(\overrightarrow{X})$ la matrice de variance-covariance de $\overrightarrow{X}$.
   (b) En déduire $\mathbb{V}(B_n)$ et $\mathsf{Cov}(A_n, C_n)$.
6. Question de mémoire :
   (a) Déterminer le $R^2$ de ce modèle.
   (b) Donnez l'intervalle de confiance symétrique de niveau 95% de $B_n$ ainsi que l'estimation de cet intervalle.