1 Théorie

Les données proviennent de l'observation d'un échantillon statistique de taille n de vecteurs de $\mathbb{R}^{p+1}$:

$$(Z_i^1, \ldots, Z_i^p, Y_i) \quad i = 1, \ldots, n.$$

On cherche à expliquer une variable quantitative Y_i (réponse) par p variables $Z_i^1, \ldots, Z_i^p$ dites *variables explicatives* (ou encore *régresseurs*).

On pose $\theta := (\beta_1, \ldots, \beta_p, \sigma^2) \in \Theta := \mathbb{R}^p \times \mathbb{R}_+^*$. Le *modèle linéaire* consiste à supposer que, pour tout $\theta \in \Theta$, les variables

$$\varepsilon_i(\theta) = \sigma^{-1}\left\{Y_i - (\beta_1 Z_i^1 + \cdots + \beta_p Z_i^p)\right\}, \quad i = 1, 2, \ldots, n,$$

sont des variables indépendantes et identiquement distribuées. Il est plus pratique dans ce cas de représenter le modèle sous la forme matricielle

$$\mathbf{Y} = \mathbf{Z}\beta + \sigma\varepsilon(\theta), \tag{1}$$

où

$$\mathbf{Y} := \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \beta := \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \varepsilon(\theta) := \begin{bmatrix} \varepsilon_1(\theta) \\ \vdots \\ \varepsilon_n(\theta) \end{bmatrix}, \quad \mathbf{Z} := \begin{bmatrix} Z_1^1 & \cdots & Z_1^p \\ Z_2^1 & \cdots & Z_2^p \\ \vdots & & \vdots \\ Z_n^1 & \cdots & Z_n^p \end{bmatrix};$$

$\mathbf{Y}$ est le vecteur des observations, β est le vecteur des paramètres de régression et $\mathbf{Z}$ est la matrice de régression de taille $n \times p$.

Pour estimer le paramètre $\beta \in \mathbb{R}^p$ dans le modèle de régression linéaire, la *méthode des moindres carrés* consiste à chercher un estimateur $\hat{\beta} \in \mathbb{R}^p$ qui minimise le risque quadratique empirique i.e. minimise la fonction

$$\mathbf{u} \mapsto J_n(\mathbf{u}) := \sum_{i=1}^n (Y_i - u_1 Z_i^1 - \cdots - u_p Z_i^p)^2 = \|\mathbf{Y} - \mathbf{Z}\mathbf{u}\|^2, \qquad \text{où } \mathbf{u} := \begin{bmatrix} u_1 \\ \vdots \\ u_p \end{bmatrix}.$$

1. Montrer que toute solution $\hat{\mathbf{u}} \in \arg\min_{\mathbf{u} \in \mathbb{R}^p} J_n(\mathbf{u})$ est solution des *équations d'estimation* :

$$\mathbf{Z}^T\mathbf{Y} = \mathbf{Z}^T\mathbf{Z}\mathbf{u},$$

en $\mathbf{u} \in \mathbb{R}^p$.

Correction :

La fonction J_n est dérivable (car quadratique) et on a

$$\nabla_{\mathbf{u}} J_n(\mathbf{u}) = -2\mathbf{Z}^T(\mathbf{Y} - \mathbf{Z}\mathbf{u}).$$

De plus J_n est convexe (car, entre autre, 2 fois dérivable de Hessienne positive) donc tout point qui annule le gradient est bien un minimum. Or un point qui annule le gradient est unique car c'est un polynôme de degré 1 et donc toute solution de $\mathbf{Z}^T\mathbf{Y} = \mathbf{Z}^T\mathbf{Z}\mathbf{u}$ est le minimum global de J_n.

Nous supposons dans la suite que les hypothèses de *Gauss-Markov* sont vérifiées :

GM1 $n > p$ et la matrice $\mathbf{Z}$ est de rang p.

GM2 les erreurs de régression sont *homoscédastiques* , pour tout $\theta \in \Theta$, $\mathbb{E}_\theta[\varepsilon(\theta)] = 0$ et $\mathrm{Var}_\theta(\varepsilon(\theta)) = \mathrm{I}_n$.

Notons que l'hypothèse (GM1) implique que la matrice de Gram $\mathbf{Z}^T\mathbf{Z}$ est inversible. Nous posons

$$\mathbf{Z}^\# = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T,$$

qui est appelée la *pseudo-inverse* de $\mathbf{Z}$.

2. Montrer que $\mathbf{Z}^\#\mathbf{Z} = \mathrm{I}_p$ et $\mathbf{Z}\mathbf{Z}^\# = \mathrm{H}$ où H est le projecteur orthogonal sur l'espace vectoriel engendré par les colonnes de la matrice $\mathbf{Z}$.

   Correction :
   $\mathbf{Z}^\#\mathbf{Z} = \mathrm{I}_p$ est trivial. Montrons que H (de taille $n \times n$) est le projecteur orthogonal sur l'espace vectoriel engendré par les colonnes de la matrice $\mathbf{Z}$. On note $\mathbf{Z}^i$ la i-ème colonne de $\mathbf{Z}$.
   On remaque d'abord que pour tout $v \in \mathbb{R}^n$, $Hv = \mathbf{Z}(\mathbf{Z}^\# v) \in vect(\mathbf{Z}^1, \dots, \mathbf{Z}^n)$. Il reste à vérifier que pour tout $v \in vect(\mathbf{Z}^1, \dots, \mathbf{Z}^n)$, alors $Hv = v$. En effet, si $v \in vect(\mathbf{Z}^1, \dots, \mathbf{Z}^n)$, alors $v = \mathbf{Z}w$ avec $w \in \mathbb{R}^p$. On a alors

   $$\mathbf{Z}^\# v = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Z}w = w,$$

   et donc $Hv = \mathbf{Z}w = v$.

3. Montrer que l'estimateur des moindres carrés est unique et a pour expression :

   $$\hat{\beta} := (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\,\mathbf{Y} = \mathbf{Z}^\#\mathbf{Y}\,. \tag{2}$$

   Correction :
$\mathbf{Z}^T\mathbf{Z}$ étant inversible on a

$$\mathbf{Z}^T\mathbf{Y} = \mathbf{Z}^T\mathbf{Z}\mathbf{u} \Leftrightarrow (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Y} = \mathbf{u}.$$

Il existe donc bien une unique solution au problème des moindres carrés données par $\hat{\beta}$.

L'estimateur $\hat{\beta}$ est dit *linéaire*, car il est obtenu en calculant une combinaison linéaire des observations $Y_1, \dots, Y_n$.

4. Montrer que l'estimateur des moindres carrés est un estimateur sans biais de β.

   Correction :
   On a (par linéarité de l'espérance) $\mathbb{E}_\theta[\hat{\beta}] = \mathbf{Z}^\#\mathbb{E}_\theta[\mathbf{Y}]$. De plus, $\mathbb{E}_\theta[\varepsilon(\theta)] = 0$ donc $\mathbb{E}_\theta[\mathbf{Y}] = \mathbf{Z}\beta$. On conclut, en remarquant que $\mathbf{Z}^\#\mathbf{Z} = I_n$, que $\mathbb{E}_\theta[\hat{\beta}] = \beta$, *i.e.*, l'estimateur est sans biais.

5. Montrer que la matrice de covariance de cet estimateur est donnée par :

   $$\mathrm{Var}_\theta(\hat{\beta}) = \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1}\,. \tag{3}$$

   Correction :
   On a $\hat{\beta} = \mathbf{Z}^\#\mathbf{Y}$ avec $\mathbf{Y}$ un vecteur gaussien de matrice de covariance $\sigma^2 I_n$. On a donc $\mathrm{Var}_\theta(\hat{\beta}) = \mathbf{Z}^\#(\mathbf{Z}^\#)^T\sigma^2 I_n = \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1}$.

Soit $\mathbf{B}$ une matrice déterministe de taille $p \times n$; on pose $\tilde{\beta} := \mathbf{B}\mathbf{Y}$.

6. Montrer que l'estimateur $\tilde{\beta}$ est sans biais si et seulement si $\mathbf{B}\mathbf{Z} = \mathrm{I}_p$.

Correction :
On a (par linéarité de l'espérance) $\mathbb{E}[\tilde{\beta}] = \mathbf{B}\mathbb{E}[\mathbf{Y}]$. De plus (comme précédemment) $\mathbb{E}_\theta[\varepsilon(\theta)] = 0$ donc $\mathbb{E}[\mathbf{Y}] = \mathbf{Z}\beta$ et $\mathbb{E}[\tilde{\beta}] = \mathbf{BZ}\beta$. L'estimateur est donc sans biais si et seulement si $\mathbf{BZ} = I_p$.

Dans la suite du problème, nous supposons que $\mathbf{BZ} = \mathrm{I}_p$.

7. Montrer que pour tout $\theta \in \Theta$, la matrice de covariance de l'estimateur $\tilde{\beta}$ est

$$\mathrm{Var}_\theta(\tilde{\beta}) = \sigma^2 \mathbf{BB}^T \ .$$

Correction :
On a $\tilde{\beta} = \mathbf{BY}$ avec $\mathbf{Y}$ un vecteur gaussien de matrice de covariance $\sigma^2 I_n$. On a donc $\mathrm{Var}_\theta(\hat{\beta}) = \mathbf{BB}^T\sigma^2 I_n = \sigma^2\mathbf{BB}^T$.

8. Montrer que pour tout $\theta \in \Theta$,

$$\mathbb{E}_\theta\left[(\tilde{\beta}-\beta)(\hat{\beta}-\beta)^T\right] = \sigma^2\mathbf{B}(\mathbf{Z}^\#)^T = \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1}.$$

Correction :
Comme les estimateur sont sans biais on a, par linéarité de l'espérance,

$$\begin{aligned}\mathbb{E}_\theta\left[(\tilde{\beta}-\beta)(\hat{\beta}-\beta)^T\right] =& \mathbb{E}_\theta\left[\tilde{\beta}\hat{\beta}^T\right] - \beta\beta^T \\ =& \mathbb{E}_\theta\left[\mathbf{BY}(\mathbf{Z}^\#\mathbf{Y})^T\right] - \beta\beta^T. \\ =& \mathbf{B}\mathbb{E}_\theta\left[\mathbf{YY}^T\right](\mathbf{Z}^\#)^T - \beta\beta^T\end{aligned}$$

Enfin, on a $\mathbb{E}_\theta\left[\mathbf{YY}^T\right] = \mathbf{Z}\beta\beta^T\mathbf{Z}^T + \sigma^2 I_n$ ce qui permet de conclure.

Si A et B sont deux matrices symétriques $p\times p$, nous notons $A \succeq B$ si et seulement si , pour tout $x \in \mathbb{R}^p$, $x^TAx \geq x^TBx$.

9. Montrer que pour tout $\theta \in \Theta$, $\mathrm{Var}_\theta(\tilde{\beta}) \succeq \mathrm{Var}_\theta(\hat{\beta})$.
Correction :
Soit $\theta \in \Theta$. D'après les questions 5 et 8 (où on a calculé $\mathrm{Cov}_\theta(\hat{\beta},\beta)$), comme la covarivance éest une forme bilinéaire symmétrique,

$$\mathrm{Cov}_\theta(\hat{\beta},\hat{\beta}-\tilde{\beta}) = \mathrm{Var}_\theta(\hat{\beta}) - \mathrm{Cov}_\theta(\hat{\beta},\tilde{\beta}) = 0.$$

De même,

$$\begin{aligned}\mathrm{Var}_\theta(\hat{\beta}) &= \mathrm{Var}_\theta(\tilde{\beta}) + \mathrm{Var}_\theta(\hat{\beta}-\tilde{\beta}) + 2\mathrm{Cov}_\theta(\hat{\beta},\hat{\beta}-\tilde{\beta}) \\ &= \mathrm{Var}_\theta(\tilde{\beta}) + \mathrm{Var}_\theta(\hat{\beta}-\tilde{\beta})\end{aligned}$$

$\mathrm{Var}_\theta(\hat{\beta}-\tilde{\beta})$ étant une matrice de covariance, elle est symmétrique positive et donc $\mathrm{Var}_\theta(\tilde{\beta}) \succeq \mathrm{Var}_\theta(\hat{\beta})$.

Nous appelons $\hat{\mathbf{Y}} = \mathbf{Z}\hat{\beta}$ la *prédiction* des observations $\mathbf{Y}$. En observant que $\mathbf{Z}\hat{\beta} = H\mathbf{Y}$, la prédiction $\hat{\mathbf{Y}}$ est la projection orthogonale de $\mathbf{Y}$ sur l'espace engendré par les colonnes de la matrice de régression $\mathbf{Z}$. Nous appelons les *résidus de régression* les composantes du vecteur

$$\mathbf{Y} - \mathbf{Z}\hat{\beta} = \mathbf{Y} - \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Y} = (\mathrm{I}_n - \mathrm{H})\mathbf{Y}. \tag{4}$$

Considérons la statistique définie comme la somme des carrés des résidus (appelée *Sum of Squared Errors of prediction* ou *SSE* dans la littérature anglo-saxonne) :

$$\mathrm{SSE} := \|\mathbf{Y} - \mathrm{H}\mathbf{Y}\|^2. \tag{5}$$

10. Montrer que pour tout $\theta \in \Theta$:

$$\|(\mathrm{I}_n - \mathrm{H})\varepsilon(\theta)\|^2 = \mathrm{Tr}\left((\mathrm{I}_n - \mathrm{H})\varepsilon(\theta)\varepsilon^T(\theta)(\mathrm{I}_n - \mathrm{H})\right).$$

Correction :
Il suffit de remarquer que pour $A \in \mathbb{R}^{n\times 1}$, $\|A\|^2 = A^T A = \mathrm{Tr}(AA^T)$ et que H est symmetrique ($H^T = H$).

11. En déduire que

$$\hat{\sigma}^2 := (n-p)^{-1}\mathrm{SSE}$$

est un estimateur sans biais de la variance σ^2.
Correction :
On remarque que $\mathbb{E}_\theta[\varepsilon(\theta)\varepsilon^T(\theta)] = \sigma^2 I_n$. Par linéarité de l'espérance

$$\begin{aligned}
\mathbb{E}_\theta[\hat{\sigma}^2] =& (n-p)^{-1}\mathrm{Tr}\left((\mathrm{I}_n - \mathrm{H})\mathbb{E}[\varepsilon(\theta)\varepsilon^T(\theta)](\mathrm{I}_n - \mathrm{H})\right)\\
=& (n-p)^{-1}\mathrm{Tr}\left((\mathrm{I}_n - \mathrm{H})\mathbb{E}[\varepsilon(\theta)\varepsilon^T(\theta)](\mathrm{I}_n - \mathrm{H})\right)\\
=& (n-p)^{-1}\sigma^2\mathrm{Tr}\left((\mathrm{I}_n - \mathrm{H})(\mathrm{I}_n - \mathrm{H})\right)\\
=& (n-p)^{-1}\sigma^2\|\mathrm{I}_n - \mathrm{H}\|^2.
\end{aligned}$$

H étant le porjecteur orthogonal sur $vect(\mathbf{Z}^1, \ldots, \mathbf{Z}^p)$ et $\mathbf{Z}$ étant de rang p on a $\|\mathrm{I}_n - \mathrm{H}\|^2 = n-p$ et on obteint le résultat.

12. Montrer que, pour tout $\theta \in \Theta$,

$$\mathbb{E}_\theta[\hat{\mathbf{Y}}(\mathbf{Y} - \hat{\mathbf{Y}})^T] = 0.$$

Correction :
On rappelle que $\hat{\mathbf{Y}} = \mathrm{H}\mathbf{Y}$, $\mathbb{E}_\theta\left[\mathbf{Y}\mathbf{Y}^T\right] = \mathbf{Z}\beta\beta^T\mathbf{Z}^T + \sigma^2 I_n$ et $H(\mathrm{I}_n - \mathrm{H})^T = 0$. On a donc

$$\begin{aligned}
\mathbb{E}_\theta[\hat{\mathbf{Y}}(\mathbf{Y} - \hat{\mathbf{Y}})^T] =& \mathrm{H}\,\mathbb{E}_\theta[\mathbf{Y}\mathbf{Y}^T](\mathrm{I}_n - \mathrm{H})^T\\
=& \mathrm{H}(\mathbf{Z}\beta\beta^T\mathbf{Z}^T + \sigma^2 I_n)(\mathrm{I}_n - \mathrm{H})^T\\
=& \mathrm{H}\,\mathbb{E}[\mathbf{Z}\hat{\beta}\hat{\beta}^T\mathbf{Z}^T](\mathrm{I}_n - \mathrm{H})^T\\
=& \mathrm{H}\mathrm{H}\,\mathbb{E}_\theta[\mathbf{Y}\mathbf{Y}^T]\mathrm{H}^T(\mathrm{I}_n - \mathrm{H})^T = 0
\end{aligned}$$

13. Déterminer la matrice de Fisher et la borne Cramér-Rao.
Correction :
La vraisemblance est donnée par

$$p_\theta(y) = \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{\|y - H\beta\|^2}{2\sigma^2}\right).$$

On a alors

$$\partial_\beta(\ln(p_\theta(\mathbf{Y}))) = \frac{-1}{2\sigma^2}\nabla_u J_n(\beta) = \frac{\mathbf{Z}^T(\mathbf{Y} - \mathbf{Z}\beta)}{\sigma^2}$$

et

$$\partial_{\sigma^2}(\ln(p_\theta(\mathbf{Y}))) = -\frac{n}{4\pi\sigma^2} + \frac{\|\mathbf{Y} - H\beta\|^2}{2\sigma^4}$$

On déduit ainsi que la matrice d'infromation de Fisher est donnée par

$$\begin{aligned}\mathbf{I}_n =& \mathbb{E}[\nabla_\beta(\ln(p_\theta(\mathbf{Y})))\nabla_\beta(\ln(p_\theta(\mathbf{Y})))^T] \\ =& diag((\mathbf{Z}^T\mathbf{Z})/\sigma^2, n/(2\sigma^4))\end{aligned}$$

Maintenant on prend $g : \Theta \to \mathbb{R}, (\beta, \sigma^2) \mapsto \beta$. La borne de Cramer Rao est donnée par

$$\nabla g(\theta)^T \mathbf{I}_n^{-1} \nabla g(\theta) = \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1}.$$

Remarquons que l'estimateur des moindres carrés atteint donc la variance minimale et $(\mathbf{Z}^T\mathbf{Z})^{-1}$ est une borne inférieure pour tout estimateur de la forme **BY**.

14. Calculer l'estimateur du maximum de vraisemblance du paramètre θ. Qu'en déduire ?

Correction :
La densité de Y est donnée par

$$p_\theta(y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|y - H\beta\|^2}{2\sigma^2}\right).$$

On en déduit que (en maximisant la vraissemblance, *i.e.* en annulant le gradient par rapport à θ) l'emv $\hat{\theta}$ est donnée par $(\hat{\beta}, \frac{n-p}{n}\hat{\sigma}^2)$.

Considérons finalement la somme des carrés de régression (ou *régression sum of squares*, RSS) :

$$\mathrm{RSS} := \|\mathrm{H}\mathbf{Y}\|^2. \tag{6}$$

15. Montrer que

$$\mathrm{TSS} := \|\mathbf{Y}\|^2 = \mathrm{RSS} + \mathrm{SSE}. \tag{7}$$

Correction :
Il suffit de remarquer que H est la projection orthogonale sur $vect(\mathbf{Z}^1, \ldots, \mathbf{Z}^p)$ tandis que $I_n - H$ est la projection orthogonale sur $vect(\mathbf{Z}^1, \ldots, \mathbf{Z}^p)^\perp$.

On suppose maintenant que $\mathbf{Y}$ est l'observation canonique d'un modèle gaussien

$$(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \{\mathrm{N}(\mathbf{Z}\beta, \sigma^2\mathrm{I}_n) : \theta = (\beta, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}_+^*\}).$$

16. Déterminer la distribution de l'estimateur des moindres carrés $\hat{\beta}$. **Correction** :
Y étant Gaussien on a $\hat{\beta} \sim \mathscr{N}(\beta, \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1})$.

17. Déterminer la distribution de $\hat{\sigma}^2$. **Correction** :
En utilisant le théorème de Cochran, on a $\hat{\sigma}^2/(n-p) \sim \chi^2_{n-p}$.

18. Montrer que $\hat{\beta}$ et $\hat{\sigma}^2$ sont indépendants. **Correction** :
C'est le théorème de Cochran car H est la projection orthogonale sur $vect(\mathbf{Z}^1, \ldots, \mathbf{Z}^p)$ tandis que $I_n - H$ est la projection orthogonale sur $vect(\mathbf{Z}^1, \ldots, \mathbf{Z}^p)^\perp$.

19. Soit $\mathbf{x} \in \mathbb{R}^p$. Montrer que

$$\frac{\mathbf{x}^T\hat{\beta} - \mathbf{x}^T\beta}{\hat{\sigma}\sqrt{\mathbf{x}^T(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{x}}}$$

suit une loi de Student à $(n-p)$ degrés de liberté.
Correction :
Un quotient de Gaussienne et d'une Chi2 indépendantes est Student.

20. Déterminer un intervalle de confiance bilatéral de niveau de couverture $1-\alpha$ pour $\beta^T \mathbf{x}$ pour $\alpha \in]0,1[$.

    Correction :
    On note t_α le quantile d'ordre α de la la loi student et on a alors

    $$\mathbb{P}(\beta^T \mathbf{x} \in [\mathbf{x}^T \hat{\beta} + t_{\alpha/2}\hat{\sigma}\sqrt{\mathbf{x}^T(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{x}}, \mathbf{x}^T \hat{\beta} + t_{1-\alpha/2}\hat{\sigma}\sqrt{\mathbf{x}^T(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{x}}]) = 1-\alpha$$

21. Construire un test de l'hypothèse

    $$H_0 : \beta^T \mathbf{x} = 0, \qquad \text{contre} \qquad H_1 : \beta^T \mathbf{x} \neq 0$$

    de niveau α. **Correction** :
    On pose $T = \frac{\mathbf{x}^T \hat{\beta}}{\hat{\sigma}\sqrt{\mathbf{x}^T(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{x}}}$. Si $T \in I_\alpha := [t_{\alpha/2}, t_{1-\alpha/2}]$, on ne rejette pas H_0 sinon on rejette.

22. Déterminer la p-valeur de ce test.

    Correction :
    On note $\phi_{S,k}$ la fonction de repartition de la loi de Student à k degré de liberté. On a alors (comme la loi de Student est symmétrique) $p(T) := \inf\{\alpha \in [0,1], T \in I_\alpha\} = 2|\phi_{S,n-p}(T) - \phi_{S,n-p}(-T)| = 2(1-\phi_{S,n-p}(|T|))$.

23. Soit R une matrice de taille $q \times p$ de rang $q \leq p$. Montrer que

    $$\frac{1}{q\hat{\sigma}^2}\{R(\hat{\beta}-\beta)\}^T [R(\mathbf{Z}^T\mathbf{Z})^{-1}R^T]^{-1}\{R(\hat{\beta}-\beta)\}$$

    suit une loi de Fisher q,$n-q$ degrés de liberté.

    Correction :
    Il s'agit d'un quotient de deux chi2 indépendants.

24. Déterminer une région de confiance pour le vecteur (β_1, β_2).

    Correction :
    On a l'ellipsoïde de confiance suivant :

    $$\mathbb{P}\Big(\hat{\beta} \in \big\{\beta \in \mathbb{R}^2, (\beta - \widehat{\beta})^T R^T [R(\mathbf{Z}^T\mathbf{Z})^{-1}R^T]^{-1} R(\beta - \widehat{\beta}) \leq q f_\alpha \widehat{\sigma}^2\big\}\Big) = \alpha.$$

    où f_α est le $(1-\alpha)$-quantile de la loi de Fisher $\mathscr{F}_{2,n-2}$.

2 Pratique

Nous allons maintenant illustrer l'utilisation de ces résultats en analysant des données journalières de concentration en ozone. Nous conseillons d'utiliser `Jupyter` et pour la régression linéaire, la fonction `R` `lm` (il y a beaucoup de tutoriels intéressants en ligne sur la régression linéaire que nous vous conseillons de consulter).

La variable à expliquer est la concentration en ozone notée `O3` et les variables explicatives sont la température notée `T12`, le vent noté `Vx` et la nébulosité notée `Ne12`. Nous rajouterons dans la matrice de régression le vecteur constant $(1,\cdots,1)^T$ que nous appelons `intercept`.

1. Estimer l'estimateur des moindres carrés du paramètre.

2. Déterminer les intervalles de confiance bilatères à 95% pour chaque valeur des paramètres.
3. Visualiser les régions de confiance à 95 % pour (β_1, β_2) et (β_1, β_3).

Nous cherchons à répondre aux questions suivantes :

(i) est-ce que la valeur de `O3` est influencée par `Vx` ?

(ii) y a-t-il un effet nébulosité ?

(iii) est-ce que la valeur de `O3` est influencée par `Vx` ou `T12` ?

4. Formuler les différentes questions comme des tests d'hypothèses.
5. Construire des procédures de tests pour ces trois hypothèses.
6. Conclure.