# Examen Intra `MAT7381`

(2 heures)

Mars 2020

**Exercice 1.** *On a simulé les données suivantes : $x_i = i$ pour $i = 1, \cdots, 50$, les $\varepsilon_i$ sont tirés suivant des lois $\mathcal{N}(0,1)$, et $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ avec $\beta_0 = 2$ et $\beta_1 = 1$. On obtient la sortie donnée ci-dessous. Que vaut le biais de $\widehat{\beta}_1$, estimé par moindres carrés (associé à la variable $x$) ?*

```
> x = 1:50
> epsilon = rnorm(50)
> y = 2+x+epsilon
> summary(lm(y~x))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.250      0.281    7.99 2.26e-10 ***
x              0.980      0.014    69.9  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Le *biais* n'est pas quelque chose que l'on observe sur les données, c'est une propriété théorique d'un estimateur. On a ici les hypothèses $\mathcal{H}_\mathbf{1}$ (plus d'observations que de variables et identifiabilité du modèle - ce qui arrive toujours avec une seule variable explicative qui n'est pas catégorielle) et $\mathcal{H}_\mathbf{2}$ (bruit centré, de variance constante et indépendent) donc – cf cours – *l'estimateur $\widehat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$ obtenu par moindres carrés est sans biais* – au sens où $\mathbb{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$. Donc a fortiori, le biais de $\widehat{\beta}_1$ est nul !

**Exercice 2.** *On travaille ici avec le modèle linéaire homoscédastique $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ satisfaisant les hypothèses $\mathcal{H}_\mathbf{1} - \mathcal{H}_\mathbf{2}$ du cours, et soit $\hat{\boldsymbol{\beta}}$ l'estimateur des MCO:*

1. *Si les covariables sont orthogonales, rappelez ce que vaut $\mathrm{Var}(\hat{\beta}_j)$ pour $j = 1, \ldots, p$.*

2. *Supposons $p = 2$, montrer que* $\mathrm{Var}(\hat{\beta}_1) \geq \dfrac{\sigma^2}{\mathbf{x}_1^\top \mathbf{x}_1}$ *(indication: calculer le 1er terme diagonal de* $(\mathbf{X}^\top \mathbf{X})^{-1}$*).*

1. Si les covariables $\mathbf{x}_j$ et $\mathbf{x}_{j'}$ sont orthogonales, $\mathbf{x}_j^\top \mathbf{x}_{j'} = 0$, et donc tous les termes de $\mathbf{X}^\top \mathbf{X}$ hors diagonale sont nuls. Pour les termes diagonaux, on retrouve $\mathbf{x}_j^\top \mathbf{x}_j$. Comme on a une matrice diagonale, son inverse est la matrice diagonale dont les termes sont les inverses des précédants, i.e.

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \left(\mathsf{diag}\big(\mathbf{x}_j^\top \mathbf{x}_j\big)\right)^{-1} = \mathsf{diag}\left(\frac{1}{\mathbf{x}_j^\top \mathbf{x}_j}\right)$$

et donc, comme $\mathrm{Var}(\hat{\beta}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$,

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\mathbf{x}_1^\top \mathbf{x}_1}$$

2. On ne suppose plus que les covariables sont orthogonales ici, et on se limite au cas où $p = 2$. Rappelons que pour une matrice $2 \times 2$,

$$\text{si } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ alors } A^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

avec ici un cas particulier puisque la matrice est symmétrique, donc $b = c$. Ou dit autrement, $bc \geq 0$. $\mathrm{Var}(\hat{\beta}_1)$ est simplement un terme diagonal de la matrice inverse, donc

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\mathbf{x}_1^\top \mathbf{x}_1 \cdot \mathbf{x}_2^\top \mathbf{x}_2 - (\mathbf{x}_1^\top \mathbf{x}_2)^2} \cdot \mathbf{x}_2^\top \mathbf{x}_2 \geq \frac{\sigma^2}{\mathbf{x}_1^\top \mathbf{x}_1 \cdot \mathbf{x}_2^\top \mathbf{x}_2} \cdot \mathbf{x}_2^\top \mathbf{x}_2$$

que l'on peut simplifier

$$\mathrm{Var}(\hat{\beta}_1) \geq \frac{\sigma^2 \cdot \mathbf{x}_2^\top \mathbf{x}_2}{\mathbf{x}_1^\top \mathbf{x}_1 \cdot \mathbf{x}_2^\top \mathbf{x}_2} = \frac{\sigma^2}{\mathbf{x}_1^\top \mathbf{x}_1}$$

avec l'égalité dans le cas particulier où $\mathbf{x}_1^\top \mathbf{x}_2 = 0$, ce qui correspond à la question précédante...

**Exercice 3.** *On dispose de $n$ observations $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$, et on estime un modèle linéaire par moindres carrés. Soit $\mathbf{X}$ la matrix $n \times p$ associée. On suppose que $\mathcal{H}_1 - \mathcal{H}_2$ sont vérifiées. On obtient une nouvelle observation $(y_{n+1}, bx_{n+1})$.*

1. *Montrez que l'erreur de prédiction $e_{n+1} = Y_{n+1} - \hat{Y}_{n+1}$ vérifie*

$$e_{n+1} = \varepsilon_{n+1} - \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\varepsilon}, \quad \mathrm{Var}(e_{n+1}) = \sigma^2 \left(1 + \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1}\right).$$

2. *Montrez également que dans le cas $p = 2$, i.e. $\mathbf{x} = (1, x)$*

$$\mathrm{Var}(e_{n+1}) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum (x_i - \overline{x})^2}\right).$$

1) La prévision est ici $\hat{Y}_{n+1} = \mathbf{x}_{n+1}^\top \widehat{\boldsymbol{\beta}}$ soit

$$\hat{Y}_{n+1} = \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{x}_{n+1}^\top \boldsymbol{\beta} + \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\varepsilon}$$

alors que $Y_{n+1} = \mathbf{x}_{n+1}^\top \boldsymbol{\beta} + \varepsilon_{n+1}$. En faisant la différence entre les deux, on obtient la réponse demandée.

Notons $e_{n+1} = \varepsilon_{n+1} - \mathbf{a}^\top \boldsymbol{\varepsilon}$, de telle sorte que

$$\mathsf{Var}(e_{n+1}) = \mathsf{Var}(\varepsilon_{n+1}) - 2\mathsf{Cov}(\varepsilon_{n+1}, \mathbf{a}^\top \boldsymbol{\varepsilon}) + \mathbf{a}^\top \mathsf{Var}(\boldsymbol{\varepsilon}) \mathbf{a}$$

où $\mathsf{Cov}(\varepsilon_{n+1}, \mathbf{a}^\top \boldsymbol{\varepsilon}) = \mathbf{a}^\top \mathsf{Cov}(\varepsilon_{n+1}, \boldsymbol{\varepsilon})$ où $\mathsf{Cov}(\varepsilon_{n+1}, \boldsymbol{\varepsilon})$ est simplement le vecteur des $\mathsf{Cov}(\varepsilon_{n+1}, \varepsilon_i)$, or les résidus sont supposés indépendants (hypothèse $\mathcal{H}_2$) donc $\mathsf{Cov}(\varepsilon_{n+1}, \mathbf{a}^\top \boldsymbol{\varepsilon} = 0$. Aussi,

$$\mathsf{Var}(e_{n+1}) = \mathsf{Var}(\varepsilon_{n+1}) + \mathbf{a}^\top \mathsf{Var}(\boldsymbol{\varepsilon}) \mathbf{a} = \sigma^2 \left(1 + \mathbf{a}^\top \mathbf{a}\right)$$

or

$$\mathbf{a}^\top \mathbf{a} = \underbrace{(\mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top)}_{=\mathbf{a}^\top} \cdot \underbrace{(\mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top)^\top}_{=\mathbf{a}}$$

soit

$$\mathbf{a}^\top \mathbf{a} = \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \cdot \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1} = \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1}$$

ce qui correspond exactement à la réponse attendue.

2) dans le cas $p = 2$,

$$\mathbf{X}^\top\mathbf{X} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$$

dont l'inverse est

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1} = \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}$$

donc ici

$$1+\mathbf{x}_{n+1}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_{n+1} = 1+\frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} 1 & x_{n+1} \end{pmatrix} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \begin{pmatrix} 1 \\ x_{n+1} \end{pmatrix}$$

Notons tout d'abord que le terme au dénominateur est

$$n\sum x_i^2 - (\sum x_i)^2 = n^2 \left( \frac{1}{n}\sum x_i^2 - \frac{1}{n}\sum x_i \right)^2 = n^2 \cdot \frac{1}{n}\sum [x_i - \overline{x}]^2$$

Le terme de droite peut s'écrire

$$\begin{pmatrix} 1 & x_{n+1} \end{pmatrix} \begin{pmatrix} \sum x_i^2 - x_{n+1}\sum x_i \\ -\sum x_i + n x_{n+1} \end{pmatrix} = \sum x_i^2 - x_{n+1}\sum x_i - x_{n+1}\sum x_i + n x_{n+1}^2$$

soit $\sum (x_i - x_{n+1})^2$, qui peut s'écrire aussi

$$\sum (x_i - x_{n+1})^2 = \sum \left( [x_i - \overline{x}] + [\overline{x} - x_{n+1}] \right)^2$$

On a alors

$$\sum [x_i - \overline{x}]^2 + 2\sum [x_i - \overline{x}][\overline{x} - x_{n+1}] + \sum [\overline{x} - x_{n+1}]^2$$

or le terme du centre est nul, puisque

$$\sum [x_i - \overline{x}][\overline{x} - x_{n+1}] = [\overline{x} - x_{n+1}] \cdot \sum [x_i - \overline{x}] = 0 \text{ car } \sum x_i = n\overline{x} = \sum \overline{x}$$

alors que le terme de droite est $n[\overline{x} - x_{n+1}]^2$ Si on remet tout ensemble, on a

$$1 + \frac{1}{n}\frac{\sum [x_i - \overline{x}]^2 + n[\overline{x} - x_{n+1}]^2}{\sum [x_i - \overline{x}]^2} = 1 + \frac{1}{n} + \frac{n[\overline{x} - x_{n+1}]^2}{n\sum [x_i - \overline{x}]^2}$$

ce qui correspond exactement à l'expression demandée

$$\text{Var}(e_{n+1}) = \sigma^2 \left( 1 + \frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum (x_i - \overline{x})^2} \right).$$

**Exercice 4.** *Considérons* $\mathbf{X}$ *une matrice* $n \times p$ *de covariances, et notons* $\mathbf{x}_i^\top$ *la ième ligne de la matrice* $\mathbf{X}$*, et* $\mathbf{X}_{(i)}$ *la matrice* $(n-1) \times p$ *la matrice* $\mathbf{X}$ *privée de la ligne* $i$*. De manière similaire,* $\mathbf{y}$ *est un vecteur de taille* $n$*,* $y_i$ *désigne la ième observation, et* $\mathbf{y}_{(i)}$ *le vecteur de taille* $n-1$ *obtenu à partir de* $\mathbf{y}$ *en enlevant la ième observation. On notera* $H_{ii}$ *le ième terme sur la diagonale de la matrice* $H$ *de projection (tel que les prévision obtenues par moindres carrés s'écrivent* $\widehat{\mathbf{y}} = H\mathbf{y}$*). On appelle* $M$ *le modèle construit à partir des* $n$ *observations* $\{\mathbf{y}, \mathbf{X}\}$ *et* $M_i$ *le modèle construit à partir des* $n-1$ *observations* $\{\mathbf{y}_{(i)}, \mathbf{X}_{(i)}\}$

1. *Montrer que* $\mathbf{X}^\top\mathbf{X} = \mathbf{X}_{(i)}^\top\mathbf{X}_{(i)} + \mathbf{x}_i\mathbf{x}_i^\top$ *pour tout* $i = 1, \ldots, n$
2. *Montrer que* $\mathbf{X}_{(i)}^\top\mathbf{y}_{(i)} = \mathbf{X}^\top\mathbf{y} + \mathbf{x}_i y_i$ *pour tout* $i = 1, \ldots, n$
3. *Montrez que*
$$\left(\mathbf{X}_{(i)}^\top\mathbf{X}_{(i)}\right)^{-1} = \left(\mathbf{X}^\top\mathbf{X}\right)^{-1} + \frac{1}{1-H_{ii}}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{x}_i\mathbf{x}_i^\top\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$$
4. *Montrez que la prévision de l'observation* $\mathbf{x}_i$ *à l'aide du modèle* $M_i$ *est*
$$\widehat{y}_i^{(i)} = \frac{1}{1-H_{ii}}\widehat{y}_i - \frac{H_{ii}}{1-H_{ii}}y_i$$
*où* $\widehat{y}_i$ *est la prévision obtenue par* $M$*.*
5. *En notant* $\widehat{\varepsilon}_i$ *les résidus estimés du modèle* $M$*, les résidus studentisés sont*
$$\widehat{t}_i = \frac{\widehat{\varepsilon}_i}{\widehat{\sigma}\sqrt{1-H_{ii}}}$$
*Montrez qu'on peut les écrire*
$$\widehat{t}_i = \frac{\widehat{y}_i - \widehat{y}_i^{(i)}}{\widehat{\sigma}\sqrt{1 + \mathbf{x}_i^\top(\mathbf{X}_{(i)}^\top\mathbf{X}_{(i)})^{-1}\mathbf{x}_i}}$$
6. *Si* $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbb{I})$*, quelle est la loi des* $\widehat{t}_i$

**Indice***: on admettra que si* $\mathbf{M}$ *est une matrice symétrique inversible*$p \times p$*, et si* $\mathbf{u}$ *et* $\mathbf{v}$ *sont deux vecteurs de dimension* $p$*,*
$$\left(\mathbf{M} + \mathbf{u}\mathbf{v}^\top\right)^{-1} = \mathbf{M}^{-1} - \frac{\mathbf{M}^{-1}\mathbf{u}\mathbf{v}^\top\mathbf{M}^{-1}}{1 + \mathbf{u}^\top\mathbf{M}\mathbf{v}}$$

1) C'est simplement l'écriture du produit matriciel. Dans notre cas, $\mathbf{X}$ est une matrice $n \times p$ et $\mathbf{x}_i$ est un vecteur de taille $p$. On peut montrer que $\mathbf{X}^\top \mathbf{X} = \mathbf{x}_1 \mathbf{x}_1^\top + \cdots + \mathbf{x}_i \mathbf{x}_i^\top + \cdots + \mathbf{x}_n \mathbf{x}_n^\top$, ce qui suffit à obtenir l'égalité

2) C'est la aussi juste du calcul matriciel...

3) Pour cette question, on avait un indice... on va l'utiliser avec $\mathbf{u} = -\mathbf{x}_i$ et $\mathbf{v} = \mathbf{x}_i^\top$, de telle sorte que

$$\left(\mathbf{X}_{(i)}^\top \mathbf{X}_{(i)}\right)^{-1} = \left(\mathbf{X}^\top \mathbf{X} - \mathbf{x}_i \mathbf{x}_i^\top\right)^{-1} = (\mathbf{X}^\top \mathbf{X})^{-1} + \frac{(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1}}{1 - \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i}$$

On utilise alors le fait que $H_{ii} = \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i$ (le terme générique de $H = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ étant $H_{ij} = \mathbf{x}_i (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_j^\top$).

4) la prévision de l'observation $\mathbf{x}_i$ à l'aide du modèle $M_i$ est

$$\widehat{y}_i^{(i)} = \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_{(i)} = \mathbf{x}_i^\top (\mathbf{X}_{(i)}^\top \mathbf{X}_{(i)})^{-1} \mathbf{X}_{(i)}^\top \mathbf{y}_{(i)}$$

que l'on peut réécrire avec la relation précédante

$$\widehat{y}_i^{(i)} = \mathbf{x}_i^\top \left[ (\mathbf{X}^\top \mathbf{X})^{-1} + \frac{(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1}}{1 - H_{ii}} \right] \underbrace{\mathbf{X}_{(i)}^\top \mathbf{y}_{(i)}}_{\mathbf{X}^\top \mathbf{y} - \mathbf{x}_i^\top y_i}$$

i.e.

$$\widehat{y}_i^{(i)} = \mathbf{x}_i^\top \underbrace{\widehat{\boldsymbol{\beta}}}_{(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}} + \frac{H_{ii}}{1 - H_{ii}} \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}} - H_{ii} \mathbf{y}_i - \frac{H_{ii}^2}{1 - H_{ii}} y_i$$

qui se réécrit simplement sous la forme

$$\widehat{y}_i^{(i)} = \frac{1}{1 - H_{ii}} \widehat{y}_i - \frac{H_{ii}}{1 - H_{ii}} y_i$$

5) On note $\widehat{\varepsilon}_i$ les résidus estimés du modèle $M$, autrement dit, $\widehat{\varepsilon}_i = y_i - \widehat{y}_i$ et donc en utilisant l'écriture précédante

$$\widehat{\varepsilon}_i = (1 - H_{i,i}) \big(y_i - \widehat{y}_i^{(i)}\big)$$

et donc les résidus studentisés $\widehat{t}_i$ s'écrivent

$$\widehat{t}_i = \frac{\widehat{\varepsilon}_i}{\widehat{\sigma} \sqrt{1 - H_{ii}}} = \frac{\sqrt{1 - H_{ii}} \big(y_i - \widehat{y}_i^{(i)}\big)}{\widehat{\sigma}}$$

On n'est pas trop loin... il faut juste réécrire le $\sqrt{1-H_{ii}}$. On avait vu que

$$\left(\mathbf{X}_{(i)}^\top \mathbf{X}_{(i)}\right)^{-1} = (\mathbf{X}^\top\mathbf{X})^{-1} + \frac{(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1}}{1-H_{i,i}}$$

car $H_{ii} = \mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_i$, et en multipliant (à gauche) par $\mathbf{x}_i^\top$ et (à droite) par $\mathbf{x}_i$, on obtient

$$\mathbf{x}_i^\top\left(\mathbf{X}_{(i)}^\top \mathbf{X}_{(i)}\right)^{-1}\mathbf{x}_i = \mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_i + \frac{\mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{X}\mathbf{x}_i^\top)^{-1}}{1-H_{i,i}}$$

soit

$$\mathbf{x}_i^\top\left(\mathbf{X}_{(i)}^\top \mathbf{X}_{(i)}\right)^{-1}\mathbf{x}_i = H_{ii} + \frac{H_{ii}^2}{1-H_{ii}} = \frac{H_{ii}}{1-H_{ii}}$$

de telle sorte que

$$1 + \mathbf{x}_i^\top\left(\mathbf{X}_{(i)}^\top \mathbf{X}_{(i)}\right)^{-1}\mathbf{x}_i = 1 + \frac{H_{ii}}{1-H_{ii}} = \frac{1}{1-H_{ii}}$$

et on obtient alors l'expression demandée.

**Exercice 5.** *Soit $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$ un modèle de régression linéaire homoscédastique gaussien. On suppose que la matrice de design $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2)$ est de taille $(n, 2)$. Le vecteur des paramètres $\boldsymbol{\beta}$ est donc de dimension 2. On notera $\sigma^2$ le paramètre de variance du bruit. Enfin, on supposera en outre que $\mathbf{x}_1$, $\mathbf{x}_2$ sont centrés et que*

$$\mathbf{X}^\top\mathbf{X} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

*où $\rho$ est un paramètre réel.*

1. *Quelle est la condition sur $n$ et $\rho$ pour que la matrice de design $\mathbf{X}$ soit de rang plein et que $\mathbf{X}^\top\mathbf{X}$ soit définie positive? Cette condtion sera supposée par la suite.*

2. *Soit $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)$ l'estimateur obtenu par moindres carrés ordinaires de ce modèle. Montrez que pour $j = 1, 2$*

$$\hat{\beta}_j = \frac{1}{1-\rho^2}\mathbf{z}_j^\top\mathbf{Y}, \qquad \textit{avec } \mathbf{z}_1 = \mathbf{x}_1 - \rho\mathbf{x}_2 \textit{ et } \mathbf{z}_2 = \mathbf{x}_2 - \rho\mathbf{x}_1.$$

3. *Calculez $\mathrm{Var}(\hat{\beta}_j)$ en fonction de $\sigma^2$ et $\rho$, pour $j = 1, 2$.*

4. *On définit le critère du facteur d'augmentation de la variance (VIF) du $j$ème régresseur* $\mathrm{VIF}_j$ *par* $\mathrm{VIF}_j = \left((\mathbf{X}^\top\mathbf{X})^{-1}\right)_{jj}$. *Pour quelle condition sur* $\rho$, $\mathrm{VIF}_j$ *est-il supérieur à 4? supérieur à 10?*

5. *Soit* $\hat{\sigma}^2$ *l'estimateur de la variance du bruit* ($= RSS/(n-2)$). *Définir les statistiques des tests de significativité des paramètres* $\beta_1$ *et* $\beta_2$ *en fonction de* $\rho$. *Que se passe-t-il lorsque* $|\rho| \to 1$? *Commentez.*

1) la matrice $\mathbf{X}$ est de rang plein si $\mathbf{X}^\top\mathbf{X}$ est inversible. Or le déterminant de $\mathbf{X}^\top\mathbf{X}$ vaut $1-\rho^2$ donc la condition nécessaire et suffisante est que $\rho^2 \neq 1$, ou encore $\rho \in (-1,+1)$ (car $\rho \in [-1,+1]$ et les deux bords sont exclus).

2) L'estimateur par moindres carrés est donné, comme $\mathcal{H}_{\mathbf{1}}$ est vérifiée si $\rho \in (-1,_1)$, par

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{x}_1^\top\mathbf{y} \\ \mathbf{x}_2^\top\mathbf{y} \end{pmatrix} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x}_1^\top\mathbf{y} \\ \mathbf{x}_2^\top\mathbf{y} \end{pmatrix}$$

ce qui s'écrit

$$\widehat{\boldsymbol{\beta}} = \frac{1}{1-\rho^2} \begin{pmatrix} \mathbf{x}_1^\top\mathbf{y} - \rho\mathbf{x}_2^\top\mathbf{y} \\ -\rho\mathbf{x}_1^\top\mathbf{y} + \mathbf{x}_2^\top\mathbf{y} \end{pmatrix} = \frac{1}{1-\rho^2} \begin{pmatrix} (\mathbf{x}_1^\top - \rho\mathbf{x}_2^\top)\mathbf{y} \\ (-\rho\mathbf{x}_1^\top + \mathbf{x}_2^\top)\mathbf{y} \end{pmatrix}$$

donc finallement

$$\begin{cases} \widehat{\beta}_1 = (\mathbf{x}_1^\top - \rho\mathbf{x}_2^\top)\mathbf{y} = \mathbf{z}_1^\top\mathbf{y} \text{ si } \mathbf{z}_1 = \mathbf{x}_1 - \rho\mathbf{x}_2 \\ \widehat{\beta}_2 = (-\rho\mathbf{x}_1^\top + \mathbf{x}_2^\top)\mathbf{y} = \mathbf{z}_2^\top\mathbf{y} \text{ si } \mathbf{z}_2 = \mathbf{x}_2 - \rho\mathbf{x}_1 \end{cases}$$

ce qui correspond bien à l'écriture de l'exercice.

3) Comme on $\mathcal{H}_{\mathbf{1}} - \mathcal{H}_{\mathbf{2G}}$, en notant $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbb{I}$, la variance de $\widehat{\boldsymbol{\beta}}$ est tout simplement

$$\mathrm{Var}[\widehat{\boldsymbol{\beta}}] = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1} = \frac{\sigma^2}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$$

Comme on veut juste les termes diagonaux, on a $\mathrm{Var}(\hat{\beta}_j) = \dfrac{\sigma^2}{1-\rho^2}$

4) Comme on vient de le voir

$$\mathrm{VIF}_j = \left((\mathbf{X}^\top\mathbf{X})^{-1}\right)_{jj} = \frac{1}{1-\rho^2}.$$

Vouloir $\text{VIF}_j \geq u$ signifie $1-\rho^2 \leq 1/u$ ou encore $\rho^2 \geq 1-1/u = (u-1)/u$, soit

$$|\rho| > \sqrt{\frac{u-1}{u}}$$

Aussi, si $u=4$, $|\rho| > \sqrt{\frac{3}{4}}$ $(\sim 0.866)$ et si $u=10$, $|\rho| > \sqrt{\frac{9}{10}}$ $(\sim 0.9486)$.

5) La *statistiques des tests de significativité des paramètres* est ici la statistique de Student,

$$t_j = \frac{\widehat{\beta}_j}{\sqrt{\widehat{\text{Var}}(\widehat{\beta}_j)}} = \frac{\widehat{\beta}_j \cdot \sqrt{1-\rho^2}}{\widehat{\sigma}}$$

Si $|\rho| \to 1$, la statistique de test tend vers 0, et on aura alors davantage tendance à accepter $H_0$, et à voir la variable $j$ comme *non-significative*: cela s'explique simplement car si $|\rho| \to 1$, les deux variables sont alors très très corrélés, et apportent la même information. Le test de Student est un test simple, et on se demande si une variable est (ou pas) significative sachant que l'autre reste présente. La valeur de la variable $j$ diminue alors avec la corrélation.

La simulation ci-dessous permet de le visualiser, avec $\rho = 0.1$

```
> n=78
> rho=0.1
> set.seed(1)
> x1=rnorm(n)
> x2=rho*x1+sqrt(1-rho^2)*rnorm(n)
> y=1+x1+x2+rnorm(n)
> summary(lm(y~x1+x2))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.0514     0.1240   8.480 1.46e-12 ***
x1            1.2356     0.1355   9.122 8.74e-14 ***
x2            1.1332     0.1391   8.149 6.22e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.077 on 75 degrees of freedom
Multiple R-squared:  0.6826,Adjusted R-squared:  0.6741
F-statistic: 80.64 on 2 and 75 DF,  p-value: < 2.2e-16
```

puis $\rho = 0.8$

```
> rho=0.8
> set.seed(1)
> x1=rnorm(n)
> x2=rho*x1+sqrt(1-rho^2)*rnorm(n)
> y=1+x1+x2+rnorm(n)
```

```
> summary(lm(y~x1+x2))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.0514     0.1240   8.480 1.46e-12 ***
x1            1.0722     0.2254   4.757 9.31e-06 ***
x2            1.2208     0.2306   5.294 1.15e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.077 on 75 degrees of freedom
Multiple R-squared:  0.7715,Adjusted R-squared:  0.7654
F-statistic: 126.6 on 2 and 75 DF,  p-value: < 2.2e-16
```

Si on regarde juste la statistique de test pour la première variable, en fonction de $\rho$, on obtient

```
> ttest = function(rho){
  set.seed(1)
  x1=rnorm(n)
  x2=rho*x1+sqrt(1-rho^2)*rnorm(n)
  y=1+x1+x2+rnorm(n)
  summary(lm(y~x1+x2))$coefficients[2,3]
}
> r = seq(-.99,.99,by=.01)
> t = Vectorize(ttest)(r)
> plot(r,t,type="l",xlab="correlation",ylab="Student t statistics")
```