STT5100 – Examen Intra

(Hiver 2022)

Les calculatrices sont autorisées. Les documents sont en revanche interdits. L'examen dure 3 heures. Toute sortie avant la fin est autorisée, mais sera définitive.

La feuille propose 7 exercices et un barème approximatif est donné à titre indicatif (oui, la somme des points dépasse 100). Les réponses doivent être reportées sur le cahier joint. Si vous utilisez 2 cahiers, merci de le mentionner, en indiquant 1/2 et 2/2 respectivement.

N'hésitez pas à faire des dessins pour vous aider, mais ne considérez pas un dessin comme une preuve. Si vous utilisez un résultat du cours dans votre preuve, nommez-le aussi précisément que possible.

Si vous pensez que des hypothèses manquent pour répondre à la question, indiquez le dans le cahier. Si vous avez besoin d'introduire des objets mathématiques non définis dans l'énoncé, définissez les clairement.

Des tables de quantiles de lois usuelles sont données en annexes, après les exercices.

Exercice 1 – R^2 ***et corrélation*** [10 points]

Considérons le modèle linéaire

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_i$$

estimé par moindres carrées à partir de données $\{y_i, x_{1,i}, x_{2,i}\}$. On note $\widehat{y}_i$ la prédiction obtenue. On suppose que $\mathcal{H}_1 - \mathcal{H}_2$ sont vérifiées. Le coefficient d'ajustement, R^2, est la proportion de la variance de la variable y qui est expliquée par le modèle. Montrez que $R^2 = \text{corr}(\widehat{\boldsymbol{y}}, \boldsymbol{y})^2$.

Quelques éléments de correction. Coomme indiqué dans l'énoncé

$$R^2 = \frac{\text{Var}[\widehat{\boldsymbol{y}}]}{\text{Var}[\boldsymbol{y}]}.$$

On peut écrire

$$R^2 = \frac{\text{Var}[\widehat{\boldsymbol{y}}]}{\text{Var}[\boldsymbol{y}]} = \frac{\text{Var}[\widehat{\boldsymbol{y}}] \cdot \text{Var}[\widehat{\boldsymbol{y}}]}{\text{Var}[\boldsymbol{y}] \cdot \text{Var}[\widehat{\boldsymbol{y}}]} = \frac{\text{Var}[\widehat{\boldsymbol{y}}]^2}{\text{Var}[\boldsymbol{y}] \cdot \text{Var}[\widehat{\boldsymbol{y}}]}$$

(on reconnaît au dénominateur le carré du dénominateur de la corrélation). Pour le numérateur, comme $\widehat{\boldsymbol{y}} = \boldsymbol{y} + \widehat{\boldsymbol{\varepsilon}}$

$$\text{Var}[\widehat{\boldsymbol{y}}] = \text{Cov}[\widehat{\boldsymbol{y}}, \widehat{\boldsymbol{y}}] = \text{Cov}[\widehat{\boldsymbol{y}}, \boldsymbol{y} + \widehat{\boldsymbol{\varepsilon}}]$$

qui va s'écrire (la covariance étant bi-linéeaire)

$$\text{Var}[\widehat{\boldsymbol{y}}] = \text{Cov}[\widehat{\boldsymbol{y}}, \boldsymbol{y}] + \text{Cov}[\widehat{\boldsymbol{y}}, \widehat{\boldsymbol{\varepsilon}}]$$

et comme par construction, $\widehat{\boldsymbol{y}} \perp \widehat{\boldsymbol{\varepsilon}}$, le second terme est nul, $\text{Var}[\widehat{\boldsymbol{y}}] = \text{Cov}[\widehat{\boldsymbol{y}}, \boldsymbol{y}]$, soit

$$R^2 = \frac{\text{Var}[\widehat{\boldsymbol{y}}]^2}{\text{Var}[\boldsymbol{y}] \cdot \text{Var}[\widehat{\boldsymbol{y}}]} = \frac{\text{Cov}[\widehat{\boldsymbol{y}}, \boldsymbol{y}]^2}{\text{Var}[\boldsymbol{y}] \cdot \text{Var}[\widehat{\boldsymbol{y}}]} = \left(\frac{\text{Cov}[\widehat{\boldsymbol{y}}, \boldsymbol{y}]}{\sqrt{\text{Var}[\boldsymbol{y}] \cdot \text{Var}[\widehat{\boldsymbol{y}}]}} \right)^2 = \text{corr}(\widehat{\boldsymbol{y}}, \boldsymbol{y})^2$$

On notera que pour la régression simple, comme $\widehat{\boldsymbol{y}}$ est une transformée linéaire de $\boldsymbol{x}$, $R^2 = \text{corr}(\widehat{\boldsymbol{y}}, \boldsymbol{y})^2 = \text{corr}(\boldsymbol{x}, \boldsymbol{y})^2$.

On peut aussi détailler un peu (en utilisant des sommes et non pas des opérateurs plus abstraits), et commencer par l'écriture de la corrélation: par définition

$$\text{corr}(\boldsymbol{y}, \widehat{\boldsymbol{y}})^2 = \frac{\left(\sum_{i=1}^n (y_i - \overline{y})(\hat{y}_i - \overline{\hat{y}}) \right)^2}{\sum_{i=1}^n (y_i - \overline{y})^2 \sum_{i=1}^n (\hat{y} - \overline{\hat{y}})^2} = \frac{\left(\sum_{i=1}^n (\hat{y}_i + \hat{\varepsilon}_i - \overline{y})(\hat{y}_i - \overline{y}) \right)^2}{\sum_{i=1}^n (y_i - \overline{y})^2 \sum_{i=1}^n (\hat{y} - \overline{y})^2}$$
$$= \frac{\left(\sum_{i=1}^n \hat{\varepsilon}_i (\hat{y}_i - \overline{y}) + \sum_{i=1}^n (\hat{y}_i - \overline{y})^2 \right)^2}{\sum_{i=1}^n (y_i - \overline{y})^2 \sum_{i=1}^n (\hat{y} - \overline{y})^2}$$

or comme évoqué auparavant, par construction de l'estimateur des moindres carrés (c'est la condition du premier ordre)

$$\sum_{i=1}^n \hat{\varepsilon}_i (\hat{y}_i - \overline{y}) = 0$$

et donc

$$\text{corr}(\boldsymbol{y}, \widehat{\boldsymbol{y}})^2 = \frac{\left(\sum_{i=1}^N (\hat{y}_i - \overline{y})^2\right)^2}{\sum_{i=1}^N (y_i - \overline{y})^2 \sum_{i=1}^N (\hat{y} - \overline{y})^2} = \frac{\sum_{i=1}^N (\hat{y}_i - \overline{y})^2}{\sum_{i=1}^N (\hat{y} - \overline{y})^2} = R^2$$

Exercice 2 – ***Comparer deux estimateurs*** [15 points]

Considérons le modèle linéaire simple

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

estimé par moindres carrées à partir de données $\{y_i, x_i\}$, avec $i = 1, \cdots, 2n$. On suppose vérifiées les hypothèses $\mathcal{H}_1 - \mathcal{H}_2$ du cours. On note

$$\overline{x} = \frac{1}{2n}\sum_{i=1}^{2n} x_i,\ \overline{x}_- = \frac{1}{n}\sum_{i=1}^{n} x_i,\ \overline{x}_+ = \frac{1}{n}\sum_{i=n+1}^{2n} x_i$$

$$\overline{y} = \frac{1}{2n}\sum_{i=1}^{2n} y_i,\ \overline{y}_- = \frac{1}{n}\sum_{i=1}^{n} y_i,\ \overline{y}_+ = \frac{1}{n}\sum_{i=n+1}^{2n} y_i$$

et on pose

$$\widehat{\beta}_1 = \frac{\sum (y_i - \overline{y})(x_i - \overline{x})}{\sum (x_i - \overline{x})^2} \text{ et } \widetilde{\beta}_1 = \frac{\overline{y}_+ - \overline{y}_-}{\overline{x}_+ - \overline{x}_-}$$

1. Montrer (par le calcul) que l'estimateur par moindres carrés de β_1 est $\widehat{\beta}_1$.
2. Montrer que $\widetilde{\beta}_1$ est un estimateur sans biais de β_1.
3. Comparez $\text{Var}[\widetilde{\beta}_1]$ et $\text{Var}[\widehat{\beta}_1]$.

1. C'est du cours, je ne vais pas reprendre ici...
2. On peut écrire

$$\begin{aligned}\tilde{\beta}_1 &= \frac{\sum_{i=n+1}^{2n} y_i - \sum_{i=1}^{n} y_i}{\sum_{i=n+1}^{2n} x_i - \sum_{i=1}^{n} x_i} = \frac{\sum_{i=n+1}^{2n} (\beta_0 + \beta_1 x_i + \varepsilon_i) - \sum_{i=1}^{n} (\beta_0 + \beta_1 x_i + \varepsilon_i)}{\sum_{i=N+1}^{2n} x_i - \sum_{i=1}^{n} x_i} \\ &= \frac{n\beta_0 - n\beta_0 + \beta_1(\sum_{i=n+1}^{2n} x_i - \sum_{i=1}^{n} x_i) + (\sum_{i=n+1}^{2n} \varepsilon_i - \sum_{i=1}^{n} \varepsilon_i)}{\sum_{i=n+1}^{2n} x_i - \sum_{i=1}^{n} x_i} \\ &= \beta_1 + \frac{(\sum_{i=n+1}^{2n} \varepsilon_i - \sum_{i=1}^{n} \varepsilon_i)}{\sum_{i=n+1}^{2n} x_i - \sum_{i=1}^{n} x_i}\end{aligned}$$

On peut maintenant prendre l'espérance,

$$\begin{aligned}\mathbb{E}(\tilde{\beta}_1) &= \mathbb{E}(\beta_1) + \frac{1}{\sum_{i=N+1}^{2n} x_i - \sum_{i=1}^{n} x_i}\left(\mathbb{E}(\sum_{i=n+1}^{2n} \varepsilon_i) - \mathbb{E}(\sum_{i=1}^{n} \varepsilon_i)\right) \\ &= \beta_1 + \frac{1}{\sum_{i=n+1}^{2n} x_i - \sum_{i=1}^{n} x_i}\left(\sum_{i=n+1}^{2n} \mathbb{E}(\varepsilon_i) - \sum_{i=1}^{n} \mathbb{E}(\varepsilon_i)\right)\beta_1\end{aligned}$$

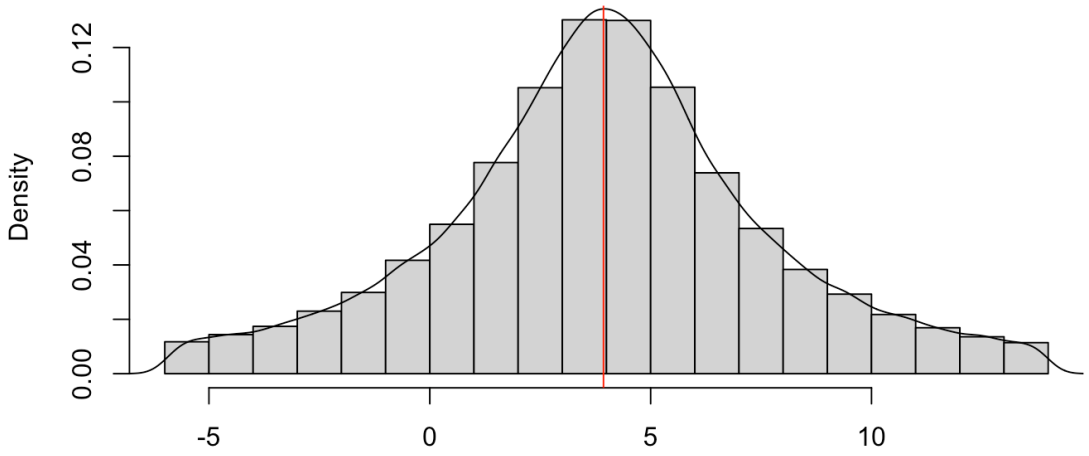
donc $\widetilde{\beta}_1$ est un estimateur sans biais de β_1.

---

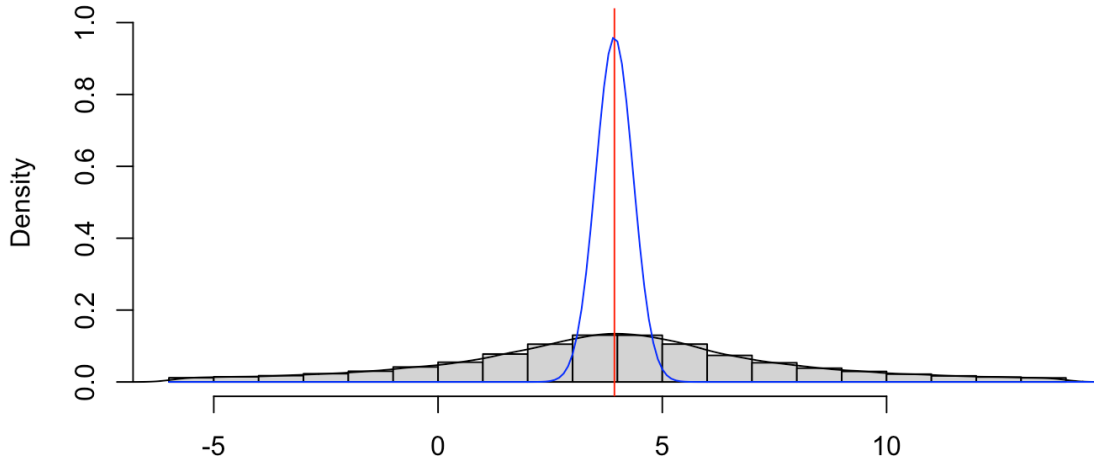
3. Il ne fallait faire aucun calcul ! c'est du cours ! c'est le théorème de Gauss-Markov ! $\widehat{\beta}_1$ est BLUE, *Best Linear Unbiased Estimator*. Or $\widetilde{\beta}_1$ est aussi un estimateur (1) linéaire (2) sans biais. Et comme *Best* signifie ici de variance minimale, forcément $\text{Var}[\widetilde{\beta}_1] \geq \text{Var}[\widehat{\beta}_1]$

On peut le vérifier en simulant. On va prendre la base de données `cars`, et la réordonner aléatoirement, pour calculer différents groupes '+' et '-'.

```
> B = rep(NA,1e5)
> n = nrow(cars)
> for(i in 1:length(B)){
+ base = cars[sample(1:n),]
+ groupe = as.factor(rep(1:2,each=n/2))
+ moyennes = aggregate(base, by = list(groupe), mean)
+ B[i] = (moyennes[1,"dist"]-moyennes[2,"dist"])/
+   (moyennes[1,"speed"]-moyennes[2,"speed"])}
> B2=B[abs(B)<5]
> hist(B2,probability=TRUE)
> lines(density(B))
> abline(v = lm(dist~speed,data=cars)$coefficients[2],col="red")
```



Je me débarrasse ici des valeurs trop grandes, obtenues pour $\widetilde{\beta}_1$, pour le dessin. On peut visualiser ci-dessous la distribution théorique de $\widehat{\beta}_1$], qui a très clairement une variance bien plus faible !



Exercice 3 – ***Variance des*** $\hat{\beta}_j$ [15 points]

On travaille ici avec le modèle linéaire homoscédastique $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ satisfaisant les hypothèses $\mathcal{H}_1 - \mathcal{H}_2$ du cours, et soit $\hat{\boldsymbol{\beta}}$ l'estimateur des moindres carrés.:

1. Si les covariables sont orthogonales, rappelez ce que vaut $\text{Var}(\hat{\beta}_j)$ pour $j = 1, \ldots, p$.
2. On ne suppose plus que les covariables sont orthogonales ici mais supposons $p = 2$. Montrez que $\text{Var}(\hat{\beta}_1) \geq \dfrac{\sigma^2}{\mathbf{x}_1^\top \mathbf{x}_1}$ (indication: calculez 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 précédents, i.e.

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

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

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

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 2×2,

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

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

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

que l'on peut simplifier

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

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

Exercice 4 – *Nouvelle observation* [15 points]

On dispose de n observations $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$, et on estime un modèle linéaire, $y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$, 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

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

2. Montrez également que dans le cas de la régression simple, autrement dit $\mathbf{x} = (1, x)$,

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

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

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

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

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

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 de la régression simple, on peut écrire (on omettra les indices sur les sommes)

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

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

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

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

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

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

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

Exercice 5 – ***Pratique de la régression*** [25 points]

On considère le modèle de régression suivant, expliquant le poids des enfants à la naissance pour 1388 ménages aux Etats-Unis, `weight` (en onces), en fonction d'un certain nombre de variables explicatives,

```
Call:
lm(formula = weight ~ male+parity+packs+white, data = naissance)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   110.22      1.676  65.763   <1e-12 ***
male            3.090     1.068   2.893   0.0038  **
parity          1.740     0.600   2.900   0.0038  **
packs         -10.460     1.791  -5.840 6.48e-09 ***
white           6.520     1.301   5.011 6.09e-07 ***
---
```

```
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19.87215 on 1383 degrees of freedom
Multiple R-squared:  0.489
```

On considère aussi un second modèle de régression, expliquant le logarithme du poids des enfants à la naissance, `log_weight`, en fonction des mêmes variables

```
Call:
lm(formula = log_weight ~ male+parity+packs+white, data = naissance)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.680      0.016   292.5   <1e-12 ***
male           0.026      0.010    2.60   0.0094  **
parity         0.016      0.006    2.67   0.0077  **
packs         -0.091      0.017   -5.35 1.01e-07 ***
white          0.062      0.012    5.17 2.73e-07 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.18654 on 1383 degrees of freedom
Multiple R-squared:  0.45
```

Pour les variables explicative, `male` est une variable indicatrice qui vaut 1 si l'enfant est un garçon, et 0 sinon; `parity` est une variable indiquant le rang de l'enfant dans la fratrie (1 pour l'ainée, 2 pour le second, etc); `packs` indique le nombre moyen de paquets de cigarettes fumés, par jour, par la mère, pendant la grossesse; `white` est une variable indicatrice qui vaut 1 si l'enfant est de peau blanche, et 0 sinon.

1. Quel serait le poids d'un garçon blanc à la naissance prédit par le premier modèle, sachant qu'il est l'ainé de la famille et que la mère fumait un paquet par jour pendant la grossesse ?
2. Quel serait le logarithme du poids d'un garçon blanc à la naissance prédit par le second modèle, sachant qu'il est l'ainé de la famille et que la mère fumait un paquet par jour pendant la grossesse ?
3. A l'aide de ce second modèle quel serait la prévision, sans biais, du poids d'un garçon blanc à la naissance prédit par le second modèle, sachant qu'il est l'ainé de la famille et que la mère fumait un paquet par jour pendant la grossesse ?

1. Étant donné:

- `male` $= 1$
- `parity` $= 1$
- `white` $= 1$
- `packs` $= 1$

la prévision sera données par

$$\widehat{\texttt{weight}} = 110.22 + 3.09 \cdot 1 + 1.74 \cdot 1 + 6.52 \cdot 1 - 10.46 \cdot 1 = 111.11$$

2. De même manière:

$$\log(\widehat{\texttt{weight}}) = 4.68 + 0.026 \cdot 1 + 0.016 \cdot 1 - 0.091 \cdot 1 + 0.062 \cdot 1 = 4.693$$

3. Pour la prévision du poids (et pas du logarithme du poids), attention à la correction, pour avoir un estimateur sans biais (cf espérance d'une loi lognorale)

$$\widehat{\texttt{weight}} = \exp\left(4.693 + \frac{0.18654^2}{2}\right) = 111.097$$

On a aussi la sortie suivante, sur un des deux modèles

```
> vcov(regression)
             (Intercept)   male  parity   packs   white
(Intercept)        2.810 -0.635 -0.633  -0.241  -1.448
male              -0.635  1.141  0.010   0.000   0.032
parity            -0.633  0.010  0.360  -0.072   0.060
packs             -0.241  0.000 -0.072   3.208   0.031
white             -1.448  0.032  0.060   0.031   1.695
```

4. A quoi correspond cette sortie informatique ?

5. On veut faire un test (de Fisher, au seuil de 5%), que le poids d'un enfant à la naissance soit le même, quel que soit le sexe et la couleur de l'enfant, toutes choses restant égales par ailleurs. Que peut-on conclure ici ?

4. C'est la matrice var-covar du model 1 (la diagonale c'est les carrés de la colonne `Std. Error` de la sortie de régression du premier modèle).

5. On cherche ici à tester $H_0 : \beta_{\text{male}} = \beta_{\text{white}} = 0$. Notons que, selon la premier sortie, $\widehat{\beta}_{\text{male}} = 3.090$ alors que $\widehat{\beta}_{\text{white}} = 6.520$. De plus, notons que la (sous) matrice de variance-covariance est ici

$$\Sigma = \begin{pmatrix} 1.141 & 0.032 \\ 0.032 & 1.695 \end{pmatrix}$$

La statistique de Fisher

$$F = \frac{1}{2}\begin{pmatrix}\widehat{\beta}_{\text{male}} - 0 & \widehat{\beta}_{\text{white}} - 0\end{pmatrix}\Sigma^{-1}\begin{pmatrix}\widehat{\beta}_{\text{male}} - 0 \\ \widehat{\beta}_{\text{white}} - 0\end{pmatrix}$$

va s'écrire ici

$$F = \frac{1}{2}\begin{pmatrix}3.09 & 6.52\end{pmatrix}\begin{pmatrix}1.141 & 0.032 \\ 0.032 & 1.695\end{pmatrix}^{-1}\begin{pmatrix}3.09 \\ 6.52\end{pmatrix} = \frac{1}{2}\begin{pmatrix}3.09 & 6.52\end{pmatrix}\begin{pmatrix}0.877 & -0.017 \\ -0.017 & 0.590\end{pmatrix}\begin{pmatrix}3.09 \\ 6.52\end{pmatrix} = 16.3689$$

Or si H_0 était vraie, on devrait avoir $F \sim \mathcal{F}(2, 1383)$ (où 2 correspond au nombre de contraintes que l'on teste, ici 2, et 1383 est le nombre de degrés de liberté, formellement $n - p$, cf slides #8 et plusieurs exercices de démo). On

va alors comparer la valeur numérique de F avec le quantile de la loi $\mathcal{F}(2, 1383)$ au niveau 95%, c'est à dire 3 (qui est la moitié du quantile de la loi du chi-deux à 2 degrés de liberté). Or $F > 3$ donc on rejette H_0.

On rajoute deux variables dans l'espoir d'améliorer la prévision, eduction correspondant au nombre d'années d'études de la mère et `income` qui est le revenu familial annuel, en milliers de dollars. Dans la sortie, les informations `t value` et `Pr(>|t|)` sont manquantes

```
Call:
lm(formula = weight ~ male+parity+packs+white+income+education, data = naissance)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   107.92      3.520
male           3.170      1.069
parity         1.800      0.602
packs         -9.730      1.837
white          5.680      1.365
income         0.059      0.033
education      0.079      0.256
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19.8520 on 1381 degrees of freedom
Multiple R-squared:  0.491
```

6. Quel serait le poids d'un garçon blanc à la naissance prédit par le premier modèle, sachant qu'il est l'ainé de la famille, que la mère fumait un paquet par jour pendant la grossesse, que le revenu familial annuel est de 105500 $ et dont la mère a fait 12 années d'études ?
7. Quelles sont ici, les variables qui sont "non-significatives" (au sens du test de Student) pour un seuil de significativité de 5% ?
8. Effectuer le test, au seuil de 5%, que le poids d'un enfant à la naissance est le même quelque soit le nombre d'années d'études de la mère et le revenu familial annuel, toutes choses restant égales par ailleurs.

6. Étant donné:

- `male` $= 1$
- `parity` $= 1$
- `white` $= 1$
- `packs` $= 1$
- `income` $= 105.5$
- `education` $= 12$

la prévision sera données par

$$\widehat{\texttt{weight}} = 107.92 + 3.17 \cdot 1 + 1.8 \cdot 1 + 5.68 \cdot 1 - 9.73 \cdot 1 + 0.059 \cdot 105.5 + 0.079 \cdot 12 = 116.0125$$

7. On calcule ici les statistiques de Student,

- `(Intercept)` $T = 107.92/3.520 \approx 30.659$
- `male` $T = 3.170/1.069 \approx 2.965$
- `parity` $T = 1.800/0.602 \approx 2.990$
- `packs` $T = -9.730/1.837 \approx -5.296$
- `white` $T = 5.680/1.365 \approx 4.161$
- `income` $T = 0.059/0.033 \approx 1.787$
- `education` $T = 0.079/0.256 \approx 0.308$

Comme on a beaucoup d'observations ($n > 1000$) on compare T aux quantiles d'une loi normale, autrement dit, à un seuil de 5%, on compare T à 1.96. Ici, les deux nouvelles variables, `income` et `education` sont (individuellement) non-significatives.

8. Contrairement à la question 5, on n'a pas ici la matrice de variance covariance des coefficients, mais on a les R^2. Il est possible d'utiliser, comme vu en démo

$$F = \frac{R_U^2 - R_R^2/2}{1 - R_U^2/1383} \approx 2.72$$

qui suivra une loi $\mathcal{F}(2, 1381)$ si H_0 est vrai. On va alors comparer au quantile à 95%, qui est encore de l'ordre de 3. Comme $F < 3$, on ne rejette pas H_0

Exercice 6 – *Régression sur deux variables corrélées* [25 points]

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 $(n, 2)$, et que $\mathbf{y}$, $\mathbf{x}_1$ et $\mathbf{x}_2$ les variables sont centrées. Le vecteur des paramètres $\boldsymbol{\beta}$ est ici de dimension 2. On notera σ^2 le paramètre de variance du bruit. Enfin, on supposera que

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

où ρ est un paramètre réel.

1. Quelle est la condition sur n et ρ pour que la matrice de design $\mathbf{X}$ soit de rang plein et que $\mathbf{X}^\top\mathbf{X}$ soit définie positive? Cette condition sera supposée vérifiée 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 $j = 1, 2$

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

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

---

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

5. Soit $\hat{\sigma}^2$ l'estimateur de la variance du bruit. Définir les statistiques des tests de significativité des paramètres β_1 et β_2 en fonction de ρ. 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 $\mathcal{H}_1$ 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 $\mathcal{H}_1 - \mathcal{H}_{2\mathbf{G}}$, en notant $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbb{I}$, la variance de $\widehat{\boldsymbol{\beta}}$ est tout simplement

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

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

4. Comme on vient de le voir

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

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 $u = 4$, $|\rho| > \sqrt{\dfrac{3}{4}}$ (~ 0.866) et si $u = 10$, $|\rho| > \sqrt{\dfrac{9}{10}}$ (~ 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 $\rho = 0.8$

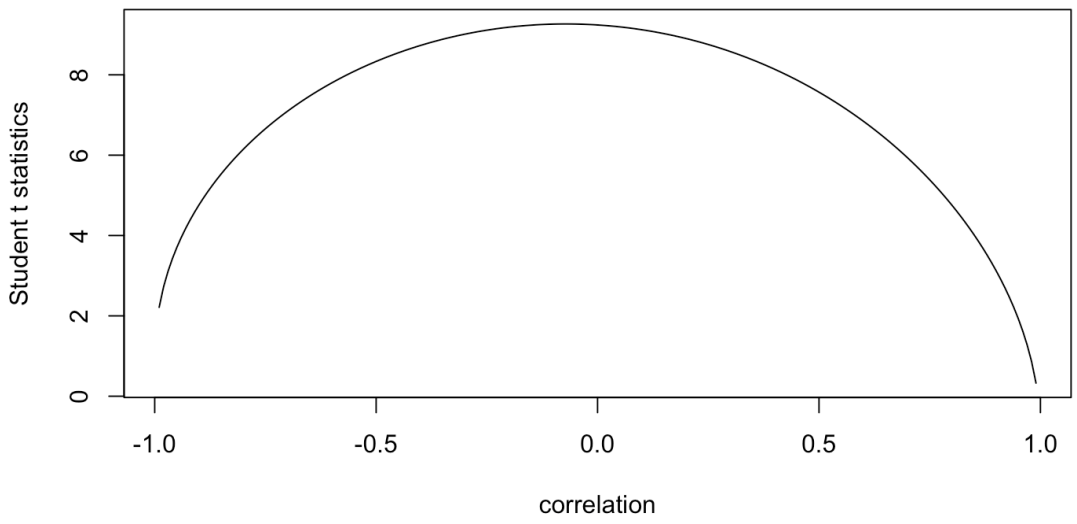
```
> rho=0.8
> set.seed(1)
> x1=rnorm(n)
> x2=rho*x1+sqrt(1-rho^2)*rnorm(n)
> y=1+x1+x2+rnorm(n)
> summary(lm(y~x1+x2))

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 ρ, 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")
```



Exercice 7 – *Centrer & centrer et réduire* [5 points]

On considère un premier modèle,

$$y_i = \alpha_0 + \alpha_1 x_{1,i} + \alpha_2 x_{2,i} + \alpha_3 x_{3,i} + \varepsilon_i$$

qui, estimée à partir de n observations $\{(y_i, x_{1,i}, x_{2,i}, x_{3,i})\}$ par moindre carrés, donne un R^2 noté R_1^2.

On centre les quatre variables: on note $\{(\widetilde{y}_i, \widetilde{x}_{1,i}, \widetilde{x}_{2,i}, \widetilde{x}_{3,i})\}$ les variables centrées,

$$\widetilde{y}_i = \beta_0 + \beta_1 \widetilde{x}_{1,i} + \beta_2 \widetilde{x}_{2,i} + \beta_3 \widetilde{x}_{3,i} + \eta_i$$

est aussi estimé par moindre carrés, et on note R_2^2 le R^2 de la régression.

On centre et on réduit les quatre variables: on note $\{(\breve{y}_i, \breve{x}_{1,i}, \breve{x}_{2,i}, \breve{x}_{3,i})\}$ les variables centrées et réduites,

$$\breve{y}_i = \gamma_0 + \gamma_1 \breve{x}_{1,i} + \gamma_2 \breve{x}_{2,i} + \gamma_3 \breve{x}_{3,i} + u_i$$

est aussi estimé par moindre carrés, et on note R_3^2 le R^2 de la régression.

1. Comparer R_1^2, R_2^2 et R_3^2.

Les trois modèles sont équivalents, $R_1^2 = R_2^2 = R_3^2$ (je laisse les plus motivés vérifier sur un jeu de données).

QUANTILES DE LA LOI NORMALE

60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
0.253	0.431	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090

QUANTILES DE LA LOI DE STUDENT A ν DEGRES DE LIBERTE

ν	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.444	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.443	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.442	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.441	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.440	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.439	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.439	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.438	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.438	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.438	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.437	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.437	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.437	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.436	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.436	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.436	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
26	0.256	0.436	0.684	0.856	1.177	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.435	0.684	0.855	1.176	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.435	0.683	0.855	1.175	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.435	0.683	0.854	1.174	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.435	0.683	0.854	1.173	1.310	1.697	2.042	2.457	2.750	3.385
35	0.255	0.434	0.682	0.852	1.170	1.306	1.690	2.030	2.438	2.724	3.340
40	0.255	0.434	0.681	0.851	1.167	1.303	1.684	2.021	2.423	2.704	3.307
45	0.255	0.434	0.680	0.850	1.165	1.301	1.679	2.014	2.412	2.690	3.281
50	0.255	0.433	0.679	0.849	1.164	1.299	1.676	2.009	2.403	2.678	3.261
55	0.255	0.433	0.679	0.848	1.163	1.297	1.673	2.004	2.396	2.668	3.245
60	0.254	0.433	0.679	0.848	1.162	1.296	1.671	2.000	2.390	2.660	3.232
∞	0.253	0.431	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090