Sorbonne Université Année 2023/2024

M2 Statistique Premier semestre

Feuille de TD 4
Correction

Exercice 1 : Soit $X = (X_1, X_2)^T$ un vecteur gaussien de $\mathbb{R}^2$ qui suit la loi $\mathcal{N}(m, V)$. Déterminer la loi de $X_1 + X_2$.
On pose $u = (1,1)^T$ et on a donc

$$X_1 + X_2 = u^T X \sim \mathcal{N}\left(u^T m, u^T V u\right).$$

Exercice 2 : Soient X et Y deux variables aléatoires indépendantes telles que $X \sim \mathcal{N}(0,1)$ et Y suit une loi de Rademacher de paramètre $p \in (0,1)$, i.e $\mathbb{P}[Y=1] = p$ et $\mathbb{P}[Y=-1] = 1-p$.

1. Montrer que $XY \sim \mathcal{N}(0,1)$.
   On a, par symétrie de la loi normale, pour tout x,
   $$\begin{aligned}\mathbb{P}[XY \leq x] &= \mathbb{P}[XY \leq x | Y = 1]\,\mathbb{P}[Y=1] + \mathbb{P}[XY \leq x | Y = -1]\,\mathbb{P}[Y=-1] \\ &= p\mathbb{P}[X \leq x] + (1-p)\mathbb{P}[-X \leq x] \\ &= p\mathbb{P}[X \leq x] + (1-p)\mathbb{P}[X \leq x] \\ &= \mathbb{P}[X \leq x]\end{aligned}$$

2. Montrer que $\mathrm{Cov}(X, XY) = 2p - 1$ On a $\mathbb{E}[Y] = 2p - 1$ et $\mathbb{E}[XY] = 0$, et donc
   $$\mathrm{Cov}(X, XY) = \mathbb{E}[XXY] = \mathbb{E}\left[X^2\right]\mathbb{E}[Y] = 2p - 1$$

3. Les variables X et XY sont-elles indépendantes ?
   Pour $p \neq 1/2$, elles ne peuvent pas être indépendantes. Pour $p = 1/2$, on a pour tout x, y
   $$\mathbb{P}[X \leq x, XY \leq y] = p\mathbb{P}[X \leq x, X \leq y] + (1-p)\mathbb{P}[X \leq x, X \geq -y]$$
   En prenant par exemple $x = 1$ et $y = -2$, on obtient donc
   $\mathbb{P}[X \leq 1, XY \leq -2] = 1/2\mathbb{P}[X \leq -2] \neq \mathbb{P}[X \leq 1]\,\mathbb{P}[XY \leq -2] = \mathbb{P}[X \leq 1]\,\mathbb{P}[X \leq -2]$.
   On a donc un exemple de variables aléatoires avec une covariance nulle mais qui ne sont pas indépendantes.

4. Montrer que le vecteur (X, XY) n'est pas gaussien.

   On considère la combinaison linéaire $Z = X + XY = X(1+Y)$. A noter que $\mathbb{E}[Z] = 0$. On suppose par l'absurde que Z est gaussien, et par symétrie sa médiane doit donc valoir $\mathbb{E}[Z] = 0$. Or

$$\begin{aligned}\mathbb{P}[Z \leq 0] &= \mathbb{P}\left[X(1+Y) \leq 0 | Y = 1\right]\mathbb{P}\left[Y = 1\right] + \mathbb{P}\left[X(1+Y) \leq 0 | Y = -1\right]\mathbb{P}\left[Y = -1\right] \\ &= \mathbb{P}\left[2X \leq 0\right]\mathbb{P}\left[Y = 1\right] + 1 - p \\ &= p\mathbb{P}\left[X \leq 0\right] + 1 - p = 1 - \frac{1}{2}p \neq \frac{1}{2}.\end{aligned}$$

   A noter que pour $p = 1/2$, on a covariance nulle et pas indépendance, donc le vecteur ne pouvait pas être gaussien.

Exercice 3 : On considère le modèle

$$Y_i = m + \sigma\epsilon_i, \qquad 1 \leq i \leq n$$

où les v.a. ϵ_i sont i.i.d. de loi commune $\mathcal{N}(0,1)$, pour des paramètres $m \in \mathbb{R}$ et $\sigma > 0$. On note $\overline{Y}_n = n^{-1}\sum_{i=1}^n Y_i$.

1. On suppose que σ est connu.

   (a) Déterminer un intervalle de confiance symétrique pour m de niveau $1-\alpha$.

   Comme $\overline{Y}_n \sim \mathcal{N}(m, \sigma^2/n)$, on a $\sqrt{n}(\bar{Y}_n - m)/\sigma \sim \mathcal{N}(0,1)$. On a donc pour tout $u \geq 0$,

$$\mathbb{P}\left(\sqrt{n}\left|\frac{\bar{Y}_n - m}{\sigma}\right| \leq u\right) = 2\Phi(u) - 1.$$

   avec Φ la fonction de répartition de la loi normale centrée réduite. Ainsi, on obtient l'intervalle de confiance de niveau $1-\alpha$:

$$I(m) = \left[\bar{Y}_n - \sigma\frac{\Phi^{-1}(1-\alpha/2)}{\sqrt{n}}\ ;\ \bar{Y}_n + \sigma\frac{\Phi^{-1}(1-\alpha/2)}{\sqrt{n}}\right].$$

   (b) Pour $\sigma = 3$, combien d'observations doit-on avoir pour que la longueur de l'intervalle de confiance de niveau 95% soit inférieure à 2 ? Donner la forme de cet intervalle au niveau 95% pour $\sigma = 3$, $n = 25$ et $\bar{y}_{25} = \bar{Y}_{25}(\omega) = 20$. (Indication : $\Phi^{-1}(0.975) \approx 2$.)

   La longueur de l'intervalle de confiance de niveau $1-\alpha$ est donnée par la fonction

$$L(n) = 2\frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha/2).$$

   Pour $\sigma = 3$ et $1-\alpha = 0.95$, on a pour tout n,

$$L(n) \leq 2 \Leftrightarrow \frac{3}{\sqrt{n}}\Phi^{-1}(0.975) \leq 1 \Leftrightarrow (3\Phi^{-1}(0.975))^2 \leq n.$$

De $\Phi^{-1}(0.975) \simeq 2$ on déduit que $(3\Phi^{-1}(0.975))^2 \simeq 36$. Ainsi, il faut au minimum $n_0 = 36$ observations pour que l'intervalle de confiance $I(m)$ soit de longueur plus petite que 2.

Pour $1-\alpha = 0.95$, $\sigma = 3$, $n = 25$ et $\bar{y}_{25} = 20$, la réalisation de l'intervalle de confiance $I(m)$ est :

$$\begin{aligned}
& \left[\bar{y}_n - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha/2)\ ;\ \bar{y}_n + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha/2)\right] \\
= \quad & \left[20 - \frac{3}{\sqrt{25}}\Phi^{-1}(0.975)\ ;\ 20 + \frac{3}{\sqrt{25}}\Phi^{-1}(0.975)\right] \\
\simeq \quad & [18.8\ ;\ 21.2].
\end{aligned}$$

(c) Soit $\alpha \in [0,1]$. Proposer un test de niveau α pour l'hypothèse $H_0 : m = m_0$ contre $H_1 : m \neq m_0$. Pour $\sigma = 3$, $n = 25$, $\bar{y}_{25} = \bar{Y}_{25}(\omega) = 20$ et $m_0 = 18.9$, quelle est la p-valeur de ce test ? Peut-on accepter l'hypothèse H_0 aux niveaux 1%, 5% et 10% ? (Indication : $\Phi\left(\frac{5.5}{3}\right) \simeq \Phi(1.83) \simeq 0.97$.)

On rejette H_0 si

$$m_0 \notin \left[\bar{Y}_n - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha/2),\ \bar{Y}_n + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha/2)\right],$$

c'est-à-dire

$$\sqrt{n}\left|\frac{\bar{Y}_n - m_0}{\sigma}\right| > \Phi^{-1}(1-\alpha/2),$$

ou encore

$$\alpha > 2\left(1 - \Phi\left(\sqrt{n}\left|\frac{\bar{Y}_n - m_0}{\sigma}\right|\right)\right).$$

La p-valeur du test est donc donnée par

$$\alpha_0 = 2\left(1 - \Phi\left(\sqrt{n}\left|\frac{\bar{Y}_n - m_0}{\sigma}\right|\right)\right).$$

Pour $\sigma = 3$, $n = 25$, $\bar{y}_{25} = 20$ et $m_0 = 18.9$, on obtient donc la p-valeur

$$\alpha_0 = 2\left(1 - \Phi\left(\sqrt{25}\left|\frac{20 - 18.9}{3}\right|\right)\right) \simeq 2(1 - \Phi(1.83)) \simeq 2(1 - 0.97) \simeq 0.06.$$

On rejette donc H_0 au niveau 10%, et on accepte H_0 aux niveaux 5% et 1%.

2. On ne suppose plus que σ est connu. On pose $\hat{\sigma}_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}$.

(a) Écrire le modèle de régression associé aux hypothèses énoncées plus haut et donner l'estimateur des moindres carrés.

Le modèle de régression associé est $Y = X\beta + \sigma\varepsilon$, avec $X = [1, \cdots, 1]^t \in \mathbb{R}^n$, $\beta = m \in \mathbb{R}$ et $\varepsilon = [\varepsilon_1, \cdots, \varepsilon_n]^t \in \mathbb{R}^n$. L'estimateur des moindres carrés est ainsi

$\hat{\beta} = (X^t X)^{-1} X^t Y = \bar{Y}_n$.

(b) Énoncer le théorème de Cochran dans ce cas.

Étant donné le modèle de régression, $Y \sim \mathcal{N}(m(1 \cdots 1)^t, \sigma^2 I_n)$ et la matrice de projection associée est :

$$P = X(X^t X)^{-1} X^t = \frac{1}{n} \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix}.$$

De plus, $PY = (\bar{Y}_n \cdots \bar{Y}_n)^t$, $P_\perp Y = (I_n - P)Y = Y - (\bar{Y}_n \cdots \bar{Y}_n)^t$, $P(m(1 \cdots 1)^t) = m(1 \cdots 1)^t$ et $P_\perp(m(1 \cdots 1)^t) = 0$. D'après le théorème de Cochran,

— $PY \sim \mathcal{N}(m(1 \cdots 1)^t), \sigma^2 P)$, $P_\perp Y \sim \mathcal{N}(0, \sigma^2 P_\perp)$;

— PY et $P_\perp Y$ sont indépendants ;

— $\frac{\|P(Y - m(1 \cdots 1)^t)\|^2}{\sigma^2} = \frac{n(\bar{Y}_n - m)^2}{\sigma^2} \sim \chi^2(1)$, $\frac{\|Y - (\bar{Y}_n \cdots \bar{Y}_n)^t\|^2}{\sigma^2} = \frac{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}{\sigma^2} \sim \chi^2(n-1)$.

(c) Montrer que $\hat{\sigma}_n^2$ est un estimateur sans biais et consistant de σ^2.

On vient de voir que

$$(n-1)\hat{\sigma}_n^2/\sigma^2 = \sum_{i=1}^n (Y_i - \bar{Y}_n)^2/\sigma^2 \sim \chi^2(n-1).$$

Ainsi, $\mathbb{E}[(n-1)\hat{\sigma}_n^2/\sigma^2] = n - 1$, c'est-à-dire $\mathbb{E}[\hat{\sigma}_n^2] = \sigma^2$ donc $\hat{\sigma}_n^2$ est sans biais. De plus, la loi des grands nombres et le théorème de continuité impliquent que

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2 = \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}_n^2 \xrightarrow[n \to \infty]{p.s.} \mathbb{E}[Y_1^2] - m^2 = \sigma^2,$$

i.e. s_n^2 est un estimateur convergent de σ^2. Ainsi, $\hat{\sigma}_n^2 = \frac{n}{n-1} s_n^2 \xrightarrow{P} \sigma^2$.

(d) Donner la loi exacte de $\sqrt{n} \frac{\bar{Y}_n - m}{\hat{\sigma}_n}$.

On a :

$$\frac{\hat{\beta} - \beta}{\hat{\sigma}_n \sqrt{1/n}} = \sqrt{n} \frac{\bar{Y}_n - m}{\hat{\sigma}_n} \sim \mathcal{T}(n-1).$$

(e) Tester l'hypothèse $H_0 : \sigma^2 = 3$ contre $H_1 : \sigma^2 \neq 3$ au niveau α.

un intervalle de confiance de niveau $(1 - \alpha)$ pour σ^2 est donné par

$$\left[\frac{(n-1)\hat{\sigma}_n^2}{c_{n-1}(1 - \alpha/2)}, \frac{(n-1)\hat{\sigma}_n^2}{c_{n-1}(\alpha/2)} \right]$$

où $c_{n-1}(\alpha/2)$ et $c_{n-1}(1 - \alpha/2)$ sont les quantiles d'ordres $\alpha/2$ et $1 - \alpha/2$ d'une loi χ_{n-1}^2 (ici $p = 1$ si p est le nombre de variables explicatives). Ainsi, le test consistant à rejeter H_0 si

$$3 \notin \left[\frac{(n-1)\hat{\sigma}_n^2}{c_{n-1}(1 - \alpha/2)}, \frac{(n-1)\hat{\sigma}_n^2}{c_{n-1}(\alpha/2)} \right]$$

i.e., si

$$\hat{\sigma}_n^2 > 3\frac{c_{n-1}(1-\alpha/2)}{n-1} \quad \text{ou} \quad \hat{\sigma}_n^2 < 3\frac{c_{n-1}(\alpha/2)}{n-1}$$

est de niveau α.

(f) Déterminer un intervalle de confiance de niveau $1-\alpha$ pour m. En déduire un test pour l'hypothèse $H_0 : m = m_0$ contre $H_1 : m \neq m_0$ au niveau α.

un intervalle de confiance de niveau $1-\alpha$ pour m est donné par

$$J(m) = \left[\bar{Y}_n - \hat{\sigma}_n\frac{t_{n-1}(1-\alpha/2)}{\sqrt{n}}, \bar{Y}_n + \hat{\sigma}_n\frac{t_{n-1}(1-\alpha/2)}{\sqrt{n}}\right].$$

où $t_{n-1}(1-\alpha/2)$ est le quantile d'ordre $1-\alpha/2$ d'une loi de Student $\mathcal{T}(n-1)$.

On peut construire un test de niveau α en rejetant H_0 si $m_0 \notin J(m)$, c'est-à-dire si

$$\sqrt{n}\left|\frac{\bar{Y}_n - m_0}{\hat{\sigma}_n}\right| > t_{n-1}(1-\alpha/2).$$

(g) Tester maintenant $H_0 : m \geq m_0$ contre $H_1 : m < m_0$ au niveau α. Calculer la p-valeur lorsque $m_0 = 12.5$, $n = 25$, $\bar{y}_{25} = \bar{Y}_{25}(\omega) = 12$ et $\hat{\sigma}_n^2(\omega) = 1.69$? Peut-on accepter l'hypothèse H_0 au niveau 5% ? (Indication : $F_{\mathcal{T}(24)}(-1.92) \simeq 0.03$.)

Pour ce test unilatéral, on cherche une région de rejet de la forme $\mathcal{R} =]-\infty, c_\alpha[$ pour $\bar{Y}_n$ telle que :

$$\begin{aligned}
\alpha &= \sup_{m \geq m_0} \P(\bar{Y}_n < c_\alpha) \\
&= \sup_{m \geq m_0} \P\left(\sqrt{n}\frac{\bar{Y}_n - m}{\hat{\sigma}_n} < \sqrt{n}\frac{c_\alpha - m}{\hat{\sigma}_n}\right) \\
&= \sup_{m \geq m_0} F_{\mathcal{T}(n-1)}\left(\sqrt{n}\frac{c_\alpha - m}{\hat{\sigma}_n}\right) \\
&= F_{\mathcal{T}(n-1)}\left(\sqrt{n}\frac{c_\alpha - m_0}{\hat{\sigma}_n}\right),
\end{aligned}$$

i.e.

$$c_\alpha = m_0 + \frac{\hat{\sigma}_n}{\sqrt{n}} t_{n-1}(\alpha).$$

Pour ce test,

$$\begin{aligned}
T(Y) = 1 &\Leftrightarrow \bar{Y}_n < c_\alpha \\
&\Leftrightarrow \sqrt{n}\frac{\bar{Y}_n - m_0}{\hat{\sigma}_n} < t_{n-1}(\alpha) \\
&\Leftrightarrow F_{\mathcal{T}(n-1)}\left(\sqrt{n}\frac{\bar{Y}_n - m_0}{\hat{\sigma}_n}\right) < \alpha.
\end{aligned}$$

Ainsi, la p-valeur du test $T(Y) = 1 \Leftrightarrow \bar{Y}_n < c_\alpha$ est :

$$\alpha_0 = F_{\mathcal{T}(n-1)}\left(\sqrt{n}\frac{\bar{Y}_n - m_0}{\hat{\sigma}_n}\right).$$

Pour $m_0 = 12,5$, $n = 25$, $\bar{y}_{25} = 12$ et $\hat{\sigma}_n^2 = 1,69$, la p-valeur est :

$$\alpha_0 = F_{\mathcal{T}(24)}\left(\sqrt{25}\left(\frac{12 - 12.5}{\sqrt{1.69}}\right)\right) \simeq F_{\mathcal{T}(24)}(-1,92) \simeq 0.03.$$

On rejette donc H_0 au niveau 5%.

Exercice 4 : On considère le modèle suivant

$$Y_i = a + bt_i + \varepsilon_i, \quad 1 \leq i \leq n,$$

où les variables aléatoires ε_i sont i.i.d. $\mathcal{N}(0, \sigma^2)$, les réels $(t_i)_{1\leq i\leq n}$ sont connus et a, b, σ^2 sont trois paramètres réels inconnus. On suppose que $\sum_{i=1}^n t_i = 0$ et on note

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i, \quad v_t = \frac{1}{n}\sum_{i=1}^n t_i^2 > 0, \quad v_Y = \frac{1}{n}\sum_{i=1}^n Y_i^2 - \bar{Y}^2 \quad \text{et} \quad \rho = \frac{1}{n}\sum_{i=1}^n Y_i t_i.$$

1. Préciser les conditions d'identifiabilité du modèle.

   Le modèle est identifiable si et seulement si X est de rang $p = 2$, c'est-à-dire si et seulement si les vecteurs $(t_i)_{1\leq i\leq n}$ et $(1)_{1\leq i\leq n}$ ne sont pas colinéaires. Ceci est équivalent à l'existence de $i \neq j$ tel que $t_i \neq t_j$, ou, considérant l'hypothèse $\sum_{i=1}^n t_i = 0$, à l'existence de $i \in \{1, \ldots, n\}$ tel que $t_i \neq 0$.

2. Calculer les estimateurs des moindres carrés $\hat{a}$, $\hat{b}$ et $\hat{\sigma}^2$ de a, b et σ^2 en fonction de $\bar{Y}$, v_t, v_Y et ρ. Quelle est leur loi jointe ?

   L'estimateur des moindres carrés de β est donné par :

$$\begin{aligned}\hat{\beta} &= (X^tX)^{-1}X^tY = \begin{pmatrix} n & \sum_i t_i \\ \sum_i t_i & \sum_i t_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_i Y_i \\ \sum_i t_i Y_i \end{pmatrix} \\ &= \begin{pmatrix} n & 0 \\ 0 & nv_t \end{pmatrix}^{-1} \begin{pmatrix} n\bar{Y} \\ n\rho \end{pmatrix} = \begin{pmatrix} \bar{Y} \\ \frac{\rho}{v_t} \end{pmatrix},\end{aligned}$$

   donc

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} \bar{Y} \\ \frac{\rho}{v_t} \end{pmatrix}.$$

Par ailleurs, l'estimateur de σ^2 est défini par :

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\left\|Y - X\hat{\beta}\right\|^2 = \frac{1}{n-2}\sum_i \left(Y_i - \bar{Y} - \frac{\rho}{v_t}t_i\right)^2 \\
&= \frac{1}{n-2}\Big(\sum_i (Y_i - \bar{Y})^2 + \frac{\rho^2}{v_t^2}\sum_i t_i^2 - 2\frac{\rho}{v_t}\sum_i t_i Y_i + 0\Big) \\
&= \frac{1}{n-2}\Big(nv_Y + \frac{\rho^2}{v_t^2}nv_t - 2\frac{\rho}{v_t}\,n\rho\Big) = \frac{n}{n-2}\Big(v_Y - \frac{\rho^2}{v_t}\Big).
\end{aligned}$$

Pour rappel, $\hat{\beta}$ consiste en les coordonnées dans la base des colonnes de X, de la projection de Y sur l'espace engendré par les colonnes de X, noté $\mathcal{M}(X)$, et

$$\hat{\sigma}^2 = \frac{1}{n-2}\|Y - X\hat{\beta}\|^2 = \frac{1}{n-2}\|P_{\mathcal{M}(X)^\perp}Y\|^2.$$

D'après le cours, les estimateurs $\hat{\beta}$ et $\hat{\sigma}^2$ sont indépendants et ont pour loi :

$$\begin{aligned}
\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix}a\\ b\end{pmatrix}, \sigma^2(X^tX)^{-1}\right) = \mathcal{N}\left(\begin{pmatrix}a\\ b\end{pmatrix}, \sigma^2\begin{pmatrix}\frac{1}{n} & 0\\ 0 & \frac{1}{nv_t}\end{pmatrix}\right); \\
\frac{(n-2)\hat{\sigma}^2}{\sigma^2} &\sim \chi^2_{(n-2)}.
\end{aligned}$$

3. Soit $\alpha \in [0,1]$. Donner un intervalle de confiance de niveau $1-\alpha$ pour chacun des paramètres a et b. En déduire un rectangle de confiance de niveau 95% pour le paramètre (a,b).

   D'après le cours, on a les intervalles de confiance de niveau $1-\alpha$ suivants :
   - pour a
   $$\left[\hat{a} - \frac{t_{n-2}(1-\alpha/2)\hat{\sigma}}{\sqrt{n}}, \hat{a} + \frac{t_{n-2}(1-\alpha/2)\hat{\sigma}}{\sqrt{n}}\right]$$
   - pour b
   $$\left[\hat{b} - \frac{t_{n-2}(1-\alpha/2)\hat{\sigma}}{\sqrt{nv_t}}, \hat{b} + \frac{t_{n-2}(1-\alpha/2)\hat{\sigma}}{\sqrt{nv_t}}\right]$$

   Comme $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$, on en déduit que

$$\begin{aligned}
&\mathbb{P}\left(|\hat{a} - a| > \frac{\hat{\sigma}}{\sqrt{n}}t_{n-2}(1-\alpha/4) \text{ ou } |\hat{b} - b| > \frac{\hat{\sigma}}{\sqrt{nv_t}}t_{n-2}(1-\alpha/4)\right) \\
&\leq \mathbb{P}\left(|\hat{a} - a| > \frac{\hat{\sigma}}{\sqrt{n}}t_{n-2}(1-\alpha/4)\right) + \mathbb{P}\left(|\hat{b} - b| > \frac{\hat{\sigma}}{\sqrt{nv_t}}t_{n-2}(1-\alpha/4)\right) \\
&\leq \alpha/2 + \alpha/2 = \alpha.
\end{aligned}$$

   Donc

$$\mathbb{P}\left(|\hat{a} - a| \leq \frac{\hat{\sigma}}{\sqrt{n}}t_{n-2}(1-\alpha/4) \text{ et } |\hat{b} - b| \leq \frac{\hat{\sigma}}{\sqrt{nv_t}}t_{n-2}(1-\alpha/4)\right) \geq 1-\alpha.$$

Ceci implique que la région rectangulaire donnée par

$$\left\{(a,b) \in \mathbb{R}^2 \mid |\hat{a}-a| \leq \frac{\hat{\sigma}}{\sqrt{n}} t_{n-2}(0.9875) \text{ et } |\hat{b}-b| \leq \frac{\hat{\sigma}}{\sqrt{nv_t}} t_{n-2}(0.9875)\right\}$$

est une région de confiance de niveau 0.95 pour (a,b).

4. Montrer que

$$\frac{1}{2\hat{\sigma}^2}\left(\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}-\begin{pmatrix}a\\ b\end{pmatrix}\right)^T \mathbf{X}^T\mathbf{X}\left(\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}-\begin{pmatrix}a\\ b\end{pmatrix}\right) \sim \mathcal{F}_{2,n-2}.$$

On a d'après la question précédente

$$\frac{1}{\sigma^2}\left(\mathbf{X}^T\mathbf{X}\right)^{1/2}\left(\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}-\begin{pmatrix}a\\ b\end{pmatrix}\right) \sim \mathcal{N}(0, I_2)$$

et donc

$$\left\|\frac{1}{\sigma^2}\left(\mathbf{X}^T\mathbf{X}\right)^{1/2}\left(\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}-\begin{pmatrix}a\\ b\end{pmatrix}\right) \sim \mathcal{N}(0, I_2)\right\|^2 = \left(\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}-\begin{pmatrix}a\\ b\end{pmatrix}\right)^T \mathbf{X}^T\mathbf{X}\left(\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}-\begin{pmatrix}a\\ b\end{pmatrix}\right) \sim \chi^2_2$$

On peut donc écrire

$$\frac{1}{2\hat{\sigma}^2}\left(\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}-\begin{pmatrix}a\\ b\end{pmatrix}\right)^T \mathbf{X}^T\mathbf{X}\left(\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}-\begin{pmatrix}a\\ b\end{pmatrix}\right)$$

comme la fraction de deux chi deux indépendantes divisées par leurs degrés de liberté et on obtient donc le résultat.

5. En déduire une ellipse de confiance de niveau 95% pour le paramètre (a,b).
On otient donc l'ellipse

$$\mathcal{E} = \left\{(a',b'), \frac{1}{2\hat{\sigma}^2}\left(\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}-\begin{pmatrix}a'\\ b'\end{pmatrix}\right)^T \mathbf{X}^T\mathbf{X}\left(\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}-\begin{pmatrix}a'\\ b'\end{pmatrix}\right) < k_{2,n-2,1-\alpha}\right\}$$

où $f_{2,n-2,1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi de fisher de paramètres $2, n-2$.

6. Donner un intervalle de confiance pour $5a-8b$, de niveau 95%, lorsque $n=18$. On pose $B=(5,-8)$, de sorte que $B\beta = 5a-8b$. D'après le cours, on en déduit l'intervalle de confiance pour $5a-8b$ de niveau 95% :

$$\left[(5\hat{a}-8\hat{b}) - t_{16}(0.975)\sqrt{\left(\frac{25}{n}+\frac{64}{nv_t}\right)\hat{\sigma}^2},\ (5\hat{a}-8\hat{b}) + t_{16}(0.975)\sqrt{\left(\frac{25}{n}+\frac{64}{nv_t}\right)\hat{\sigma}^2}\right].$$

7. Tester l'hypothèse H_0 : "$a=b$" contre H_1 : "$a \neq b$" au niveau 1% lorsque $n=22$.

H_0 s'écrit encore $a-b=0$, et faire le test revient à vérifier que 0 appartient à l'intervalle de confiance de $a-b$. On pose donc $B=(1,-1)$, et on obtient (cours) l'IC

$$\left[(\hat{a}-\hat{b})-t_{16}(0.995)\sqrt{\left(\frac{1}{n}+\frac{1}{nv_t}\right)\hat{\sigma}^2},\ (\hat{a}-\hat{b})+t_{16}(0.995)\sqrt{\left(\frac{1}{n}+\frac{1}{nv_t}\right)\hat{\sigma}^2}\right].$$

Exercice 6 : On se place dans le cadre de l'ANOVA à deux facteurs :

1. Proposer un test pour tester l'absence d'effet globbale du facteur A, i.e pour tester

$$H0 : ''\forall i, \alpha_i = 0,\ \text{et}\ \forall i,j, \gamma_{i,j} = 0'' \quad \text{contre} \quad H1 : ''\exists i, \alpha_i \neq 0,\ \text{ou}\ \exists (i,j), \gamma_{i,j} \neq 0''$$

Tester l'absence globale d'effet du facteur A revient à tester

$$H0 : ''\mathbb{E}[\mathbf{X}] \in \mathcal{M}_0 \oplus \mathcal{M}_B'' \qquad \text{contre} \qquad H1 : \mathbb{E}[\mathbf{X}] \in D$$

et on obtient donc la statistique de test

$$\mathcal{F}_{A,C} = \frac{\dim\left(D^{\perp}\right)\left\|P_D\mathbf{X} - P_{\mathcal{M}_0 \oplus \mathcal{M}_B}\mathbf{X}\right\|^2}{\left(\dim(D) - \dim\left(\mathcal{M}_0 \oplus \mathcal{M}_B\right)\right)\left\|P_{D^{\perp}}\mathbf{X}\right\|^2}$$

A noter que l'on a

$$\left\|P_D\mathbf{X} - P_{\mathcal{M}_0 \oplus \mathcal{M}_B}\mathbf{X}\right\|^2 = SCE_A + SCE_{inter}$$

et on a donc

$$\mathcal{F}_{A,C} = \frac{IJ(K-1)\left(SCE_A + SCE_{inter}\right)}{(IJ-J)SCE_{residu}} \sim \mathcal{F}_{IJ-J,IJ(K-1)}$$

2. Proposer un test pour tester l'absence d'effet globbale du facteur B, i.e pour tester

$$H0 : ''\forall j, \beta_j = 0,\ \text{et}\ \forall i,j, \gamma_{i,j} = 0'' \quad \text{contre} \quad H1 : ''\exists j, \beta_j \neq 0,\ \text{ou}\ \exists (i,j), \gamma_{i,j} \neq 0''$$

Tester l'absence globale d'effet du facteur B revient à tester

$$H0 : ''\mathbb{E}[\mathbf{X}] \in \mathcal{M}_0 \oplus \mathcal{M}_A'' \qquad \text{contre} \qquad H1 : \mathbb{E}[\mathbf{X}] \in D$$

et on obtient donc la statistique de test

$$\mathcal{F}_{B,C} = \frac{\dim\left(D^{\perp}\right)\left\|P_D\mathbf{X} - P_{\mathcal{M}_0 \oplus \mathcal{M}_A}\mathbf{X}\right\|^2}{\left(\dim(D) - \dim\left(\mathcal{M}_0 \oplus \mathcal{M}_A\right)\right)\left\|P_{D^{\perp}}\mathbf{X}\right\|^2}$$

A noter que l'on a

$$\left\|P_D\mathbf{X} - P_{\mathcal{M}_0 \oplus \mathcal{M}_A}\mathbf{X}\right\|^2 = SCE_B + SCE_{inter}$$

et on a donc

$$\mathcal{F}_{B,C} = \frac{IJ(K-1)\,(SCE_B + SCE_{inter})}{(IJ-I)SCE_{residu}} \sim \mathcal{F}_{IJ-I,IJ(K-1)}$$