Corrigé du TD 7

1. On a ajouté la contrainte $\alpha_1 = 0$. On veut se placer dans le cadre de la régression linéaire multiple. Pour cela on écrit la i^e observation comme ceci : $X_i = (1, 0, \dots, 0, 1, 0, \dots, 0)$

toujours un 1 là

1 à la q^e place si on est dans le q^e groupe (et rien si on est dans le 1er groupe)

On a alors : $y_i = X_i \times \begin{pmatrix} \mu \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix}$.

La matrice des observations est :

$$\begin{array}{c} n_1 \\ \\ n_2 \\ \\ \\ n_p \\ \\ \end{array} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & 0 & \cdots & 0 \\ \vdots & 1 & 0 \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & 1 & 0 \cdots & 0 \\ \vdots & & \vdots & \\ \vdots & 0 & \cdots 0 & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots 0 & 1 \end{bmatrix} = X$$

On a n_i individus dans chaque dans le i^e groupe et $n_1 + \dots + n_p = n$.

Calculons $X'X$:

$$\begin{bmatrix} 1 & \cdots & & & & & \cdots & 1 \\ 0 & \cdots 0 & 1 & \cdots & 1 & & 0 \cdots & 0 \\ \vdots & & 0 & \cdots & 0 & & & \vdots \\ \vdots & & \vdots & & \vdots & & 0 \cdots & 0 \\ 0 & \cdots 0 & 0 & \cdots & 0 & & 1 \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & 0 & \cdots & 0 \\ \vdots & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & 1 & 0 \cdots & 0 \\ \vdots & & & \vdots \\ \vdots & 0 & \cdots 0 & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots 0 & 1 \end{bmatrix} = \begin{bmatrix} n & n_2 & \cdots & n_p \\ n_2 & n_2 & & 0 \\ \vdots & & \ddots & \\ n_p & 0 & & n_p \end{bmatrix}$$

$$
\begin{bmatrix}
\frac{1}{n_1} & -\frac{1}{n_1} & \cdots & -\frac{1}{n_1} \\
-\frac{1}{n_1} & \frac{1}{n_1}+\frac{1}{n_2} & \frac{1}{n_1} \cdots & \frac{1}{n_1} \\
\vdots & \frac{1}{n_1} & \ddots & \frac{1}{n_1} \\
-\frac{1}{n_1} & \frac{1}{n_1} \cdots & \frac{1}{n_1} & \frac{1}{n_1}+\frac{1}{n_p}
\end{bmatrix}
$$

donc on a ici $(XX')^{-1}$

$$
\begin{bmatrix}
n & n_2 & \cdots & n_p \\
n_2 & n_2 & & 0 \\
\vdots & & \ddots & \\
n_p & & 0 & n_p
\end{bmatrix}
\begin{bmatrix}
1 & & 0 \\
& \ddots & \\
0 & & 1
\end{bmatrix}
$$

On sait que les estimateurs sont les coordonnées du vecteur $(XX')^{-1}X'Y$.

$$
\begin{bmatrix}
1 & \cdots & & & & \cdots & 1 \\
0 \cdots 0 & 1 \cdots 0 & & & 0 \cdots 0 \\
\vdots & 0 \cdots 0 & & & 0 \cdots 0 \\
0 \cdots 0 & 0 \cdots 0 & & & 1 \cdots 1
\end{bmatrix}
\begin{bmatrix}
y_1 \\ \vdots \\ \vdots \\ y_n
\end{bmatrix}
$$

$$
\begin{bmatrix}
\frac{1}{n_1} & -\frac{1}{n_1} & \cdots & -\frac{1}{n_1} \\
-\frac{1}{n_1} & \frac{1}{n_1}+\frac{1}{n_2} & \frac{1}{n_1} \cdots & \frac{1}{n_1} \\
\vdots & \frac{1}{n_1} & \ddots & \frac{1}{n_1} \\
-\frac{1}{n_1} & \frac{1}{n_1} \cdots & \frac{1}{n_1} & \frac{1}{n_1}+\frac{1}{n_p}
\end{bmatrix}
\begin{bmatrix}
\frac{1}{n_1} \cdots \frac{1}{n_1} & 0 & \cdots & & 0 \\
-\frac{1}{n_1} \cdots -\frac{1}{n_1} & \frac{1}{n_2} \cdots \frac{1}{n_2} & 0 & & 0 \\
\vdots & & & & \\
-\frac{1}{n_1} \cdots -\frac{1}{n_1} & 0 \cdots & 0 & & \frac{1}{n_p} \cdots \frac{1}{n_p}
\end{bmatrix}
\begin{bmatrix}
\overline{y}_1 \\ \overline{y}_2 - \overline{y}_1 \\ \vdots \\ \overline{y}_p - \overline{y}_1
\end{bmatrix}
$$

($\overline{y}_i$ = moyenne empirique dans le i^e groupe)

Estimateur de μ : $\overline{y}_1$

Pour $i \geq 2$, estimateur de α_i : $\overline{y}_i - \overline{y}_1$.

2. Notons $y_{i1}, \ldots, y_{in_i}$ les observations du i^e groupe.

On sait que
$$
F = \frac{\sum_{i=1}^{p} n_i (\overline{y}_i - \overline{y})^2}{\sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i)^2} \times \frac{n-p}{p-1} \sim F_{p-1,n-p}
$$

Le test est donc : on rejette (H_0) si $F \geq f_{p-1, n-p}$

(le fractile tel que $\mathbb{P}(X \leq f_{p-1,n-p}) = 0,95$

pour $X \sim \mathcal{F}_{p-1, n-p}$)

Dans le cas des données fournies :

$\bar{y}_1 = 83$, $\bar{y}_2 = 91$, $\bar{y}_3 = 94$, $\bar{y}_4 = 65$

$\bar{y} = 84,3$

$\bar{y}_1 - \bar{y} = -1,3$, $\bar{y}_2 - \bar{y} = 6,7$, $\bar{y}_3 - \bar{y} = 9,7$, $\bar{y}_4 - \bar{y} = -1$[illegible]

$F = 0,45$ $\quad p-1 = 3$, $n-p = 6$

Donc le résultat du test est : on garde (H_0).

3. On sait que sous (H_2)

$$T_2 = \frac{\hat{\alpha}_2 - 0}{\hat{\sigma}\sqrt{[(X'X)^{-1}]_{22}}} \sim \mathcal{T}(n-p)$$

avec $\hat{\sigma}^2$ estimateur de la variance $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2}{n-p}$

Calculons : $\hat{\sigma}^2 = 891$, $\hat{\sigma} = 29,8$

$$T_2 = \frac{8}{29,8 \times \sqrt{\frac{1}{3} + \frac{1}{4}}} = 0,35$$

On rejette (H_2) si $|T_2| \geq 2,447$ (ce qui est le cas)

Rappel : pourquoi T_2 est de loi de Student.

c'est le rapport entre $\dfrac{\hat{\alpha}_2}{\sigma\sqrt{(X'X)^{-1}_{22}}}$ (loi normale centrée réduite)

et $\sqrt{\left(\dfrac{\hat{\sigma}}{\sigma}\right)^2 (n-p) \times \dfrac{1}{(n-p)}}$

avec ceci $\perp\!\!\!\perp$ $\hat{\alpha}_2$ et de loi $\chi^2(n-p)$