

TD n° 6 : Test de Fisher

Exercice 1. On observe n valeurs d'un *vecteur* (Y, X_1, X_2) notées $(y_1, x_{1,1}, x_{2,1}), \dots, (y_n, x_{1,n}, x_{2,n})$ telles que, pour tout $i \in \{1, \dots, n\}$, y_i est une réalisation de

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \epsilon_i,$$

où $\epsilon_1, \dots, \epsilon_n$ sont n *var iid* suivant la loi normale $\mathcal{N}(0, \sigma^2)$. Les paramètres $\beta_0, \beta_1, \beta_2$ et σ sont des réels inconnus.

1. Écrire le modèle linéaire associé sous la forme matricielle usuelle : $Y = X\beta + \epsilon$, en indiquant ce que sont ici Y, X, β et ϵ .

En posant $y = (y_1, \dots, y_n)^t$ et en adoptant les notations de la question 1, les données recueillies donnent :

$$X^t X = \begin{pmatrix} 8 & 4 & 0 \\ 4 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad X^t y = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$$

et $\|y\|^2 = 10$, où $\|\cdot\|$ désigne la norme euclidienne dans $\mathbb{R}^n$. Soit $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)^t$ l'*emco* de $\beta = (\beta_0, \beta_1, \beta_2)^t$.

2. Quel est le nombre d'observations n ?
3. Exprimer $\hat{\beta}$ en fonction de X et Y . Donner l'*emco* ponctuel b de β .
4. Quelle est la loi de $\hat{\beta}$, ainsi que celle de chacune de ses composantes : $\hat{\beta}_0, \hat{\beta}_1$ et $\hat{\beta}_2$?
5. Donner un estimateur sans biais de σ^2 en fonction de X et Y .
6. Montrer, à l'aide de formules, que $X^t(y - Xb) = 0$, puis que $\|y - Xb\|^2 = \|y\|^2 - (X^t y)^t b$. En déduire l'estimation ponctuelle de σ^2 .
7. On considère les hypothèses :

$$H_0 : \beta_0 - \beta_1 + \beta_2 = 2 \quad \text{contre} \quad H_1 : \beta_0 - \beta_1 + \beta_2 \neq 2.$$

Peut-on rejeter H_0 au risque 5% ? Utiliser le test de Fisher (f_{obs} , p-valeur...).

Quel autre test aurions-nous pu utiliser ?

Exercice 2. Dans une étude en sciences sociales, on souhaite expliquer une variable Y à partir de 3 variables quantitatives X_1, X_2 et X_3 . On observe $n = 100$ valeurs de (Y, X_1, X_2, X_3) notées $(y_1, x_{1,1}, x_{2,1}, x_{3,1}), \dots, (y_n, x_{1,n}, x_{2,n}, x_{3,n})$. On adopte le modèle de *rlm* : pour tout $i \in \{1, \dots, n\}$, y_i est une réalisation de

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \epsilon_i,$$

où $\epsilon_1, \dots, \epsilon_n$ sont n *var iid* suivant la loi $\mathcal{N}(0, \sigma^2)$. Les paramètres $\beta_0, \beta_1, \beta_2, \beta_3$ et σ sont des réels inconnus. On considère le modèle mis sous la forme matricielle usuelle : $Y = X\beta + \epsilon$.

En posant $y = (y_1, \dots, y_n)^t$, les données recueillies donnent :

$$(X^t X)^{-1} = \frac{1}{6} \begin{pmatrix} 0.06 & 0 & 0 & 0 \\ 0 & -0.1 & 0.8 & -2 \\ 0 & 0.8 & -0.4 & 4 \\ 0 & -2 & 4 & -10 \end{pmatrix}, \quad X^t y = \begin{pmatrix} 25 \\ -20 \\ 40 \\ 20 \end{pmatrix}$$

et $\|y\|^2 = 640$, où $\|\cdot\|$ désigne la norme euclidienne dans $\mathbb{R}^n$.

On considère les hypothèses :

$$H_0 : \begin{cases} 4\beta_0 + \beta_1 = 2 \\ 2\beta_2 - 5\beta_3 = 14 \end{cases} \quad \text{contre} \quad H_1 : 4\beta_0 + \beta_1 \neq 2 \text{ ou } 2\beta_2 - 5\beta_3 \neq 14.$$

Peut-on rejeter H_0 au risque 5% ? Utiliser le test de Fisher avec les techniques de calculs introduites dans l'exercice précédent.

Exercice 3. Soient $A_1, \dots, A_k, B_1, \dots, B_l, C_1, \dots, C_m$, $n = k + l + m$ var indépendantes telles que

- $A_1, \dots, A_k$ suivent chacune la loi $\mathcal{N}(\mu_1, \sigma^2)$,
- $B_1, \dots, B_l$ suivent chacune la loi $\mathcal{N}(\mu_2, \sigma^2)$,
- $C_1, \dots, C_m$ suivent chacune la loi $\mathcal{N}(\mu_3, \sigma^2)$.

Les paramètres μ_1, μ_2, μ_3 et σ sont inconnus. On pose $Y = (A_1, \dots, A_k, B_1, \dots, B_l, C_1, \dots, C_m)^t$,

$$\bar{A} = \frac{1}{k} \sum_{i=1}^k A_i, \quad \bar{B} = \frac{1}{l} \sum_{i=1}^l B_i, \quad \bar{C} = \frac{1}{m} \sum_{i=1}^m C_i.$$

1. Écrire le modèle linéaire associé sous la forme matricielle usuelle : $Y = X\beta + \epsilon$, en indiquant ce que sont ici Y , X , β et ϵ .
2. Calculer l'emco $\hat{\beta}$ de β , ainsi qu'un estimateur $\hat{\sigma}^2$ sans biais de σ^2 . On les exprimera en fonction des composantes de Y , et de $\bar{A}$, $\bar{B}$ et $\bar{C}$.

Une application du modèle avec $k = l = m = 5$ donne les estimations ponctuelles de β et σ^2 suivantes : $b = (49, 56, 51)^t$ et $s^2 = 7.83$. On considère les hypothèses :

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad \text{contre} \quad H_1 : \text{"il existe au moins deux moyennes qui diffèrent"}.$$

3. Montrer que l'on a

$$H_0 : Q\beta = r \quad \text{contre} \quad H_1 : Q\beta \neq r,$$

$$\text{avec } Q = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \text{ et } r = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

4. Calculer la matrice $(Q(X^t X)^{-1} Q^t)^{-1}$ et le réel $(Qb - r)^t (Q(X^t X)^{-1} Q^t)^{-1} (Qb - r)$.
5. Est-ce que l'on peut dire, au risque 5%, qu'il existe au moins deux moyennes qui diffèrent ?