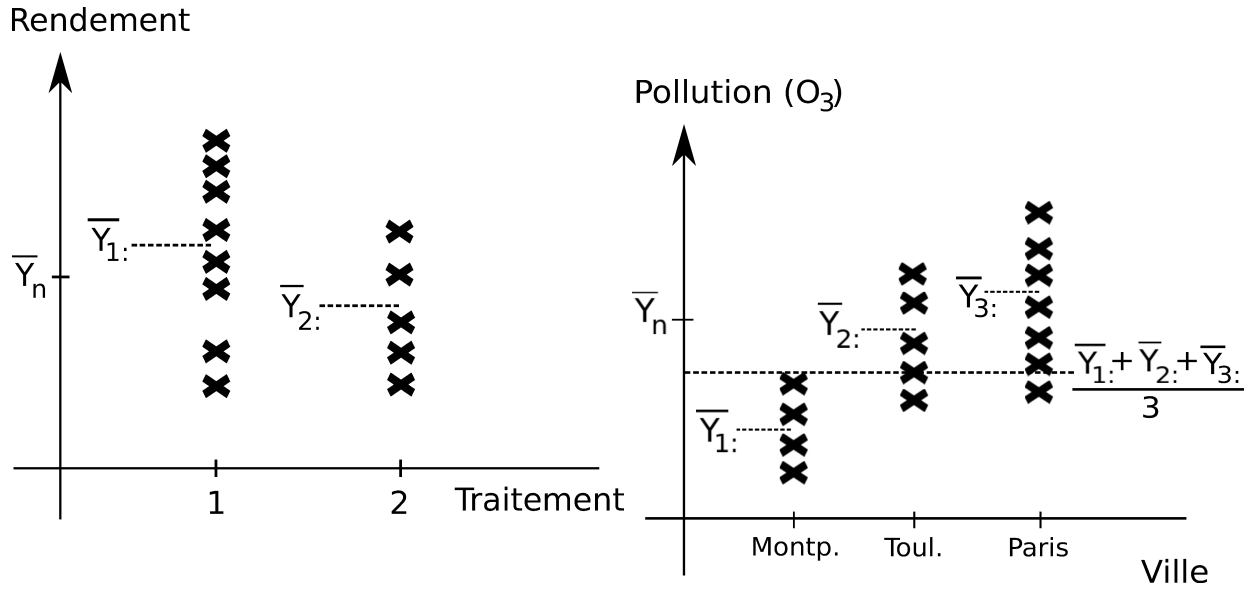
HMMA307

Modèles linéaires avancés

Cours: Joseph Salmon *Scribes: Vitus Kirchberger, Samuel Valiquette et Arielle Gantelet*

1 Anova (Analyse of variance)

Exemples



(a) $I = 2$ (b) $I = 3$

FIGURE 1 – todo improve use subfigure + sorties graphiques du notebook (en pdf)

- Nombre de modalités : I .
- Nombre total d'observations : n .
- Nombre d'observations de la i$^{\text{e}}$ modalité : n_i .

Si on a I modalités, le nombre d'observations vaut $n = n_1 + \cdots + n_I$. Le modèle s'écrit alors :

$$y_{i,j} = \mu_i^* + \varepsilon_{i,j} \ , \tag{1}$$

avec $\varepsilon_{i,j} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$ et $\text{Cov}(\varepsilon_{i,j}, \varepsilon_{i'j'}) = \delta_{ii'}\delta_{jj'}\sigma^2$.

Ici la quantité $y_{i,j}$ est la j^{e} observation de la i^{e} modalité. La notation $\delta_{ii'}$ représente le symbole de Kronecker, c'est-à-dire

$$\delta_{ii'} = \begin{cases} 1 & \text{si } i = i', \\ 0 & \text{sinon.} \end{cases} \tag{2}$$

On note la moyenne globale des observations : $\overline{y}_n = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i} y_{i,j}$, et pour tout $i \in [\![1, I]\!]$: on note la moyenne selon les modalités par $\overline{y}_{i,:} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{i,j}$.

Il est commun d'écrire le modèle comme :

$$\mu_i^* = \underbrace{\mu^*}_{\text{effet moyen}} + \underbrace{\alpha_i^*}_{\text{effet spécifique}} \tag{3}$$

Remarque 1.1. *Si l'on dispose d'estimateurs de μ^* et α_i^* notés $\hat{\mu}$ et $\hat{\alpha}_i$, alors un estimateur de μ_i^* est $\hat{\mu}_i = \hat{\alpha}_i + \hat{\mu}$. De plus, $\hat{\mu}_i$ est alors la prédiction du niveau d'expression de la modalité i. Au sens du risque quadratique, l'estimateur $\hat{\mu}_i = \overline{y}_{i,:}$ est le meilleur estimateur possible :*

$$\underset{(\mu_1,\dots,\mu_I)\in\mathbb{R}^I}{\arg\min} \sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{i,j} - \mu_i)^2 = \begin{pmatrix} \hat{\mu}_1 \\ \vdots \\ \hat{\mu}_I \end{pmatrix} = \begin{pmatrix} \overline{y}_{1,:} \\ \vdots \\ \overline{y}_{I,:} \end{pmatrix}$$

On peut donc reformuler le modèle : $y_{i,j} = \mu_i^* + \varepsilon_{i,j} = \alpha_i^* + \mu^* + \varepsilon_{i,j}$.

1.1 Somme des effets individuels est nulle

Hypothèse : $\sum_{i=1}^{I} \alpha_i^* = 0$, où l'on note $\alpha = (\alpha_1, \dots, \alpha_I)^\top$. (XXX dire pourquoi on impose cette contrainte.)

Estimateurs associés :

$$\begin{aligned} \underset{(\mu,\alpha)\in\mathbb{R}\times\mathbb{R}^I}{\arg\min} \quad & \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{i,j} - \mu - \alpha_i)^2 \\ \text{s.c.} \quad & \sum_{i=1}^{I}\alpha_i = 0 \ . \end{aligned} \tag{4}$$

Pour trouver ces estimateurs, on pose le Lagrangien suivant :

$$\mathcal{L}(\mu, \alpha, \lambda) = \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{i,j} - \mu - \alpha_i)^2 + \lambda\sum_{i=1}^{I}\alpha_i \ . \tag{5}$$

Ensuite on calcule le gradient pour chaque composante à maximiser. On note que pour n'importe quelles contraintes sur α, la dérivée par rapport à μ ne dépend pas de celles-ci. Ainsi, on a :

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mu} = 0 &\implies \sum_{i=1}^{I}\sum_{j=1}^{n_i}(\hat{\mu}+\hat{\alpha}_i - y_{i,j}) = 0 \\ &\implies \hat{\mu}\underbrace{\sum_{i=1}^{I} n_i}_{=n} + \sum_{i=1}^{I} n_i\hat{\alpha}_i - n\overline{y}_n = 0 \\ &\implies \hat{\mu} + \frac{1}{n}\sum_{i=1}^{I} n_i\hat{\alpha}_i = \overline{y}_n. \end{aligned} \tag{6}$$

Equation (6) est utile pour ce cas et pour d'autres contraintes sur α. Maintenant, pour $i_0 \in [\![1, I]\!]$ fixé, évaluons l'estimateur de α_{i_0}.

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \alpha_{i_0}} = 0 &\implies \sum_{j=1}^{n_{i_0}}(\hat{\alpha}_{i_0}+\hat{\mu} - y_{i_0,j}) + \hat{\lambda} = 0 \\ &\implies n_{i_0}\left(\hat{\alpha}_{i_0}+\hat{\mu} - \overline{y}_{i_0,:}\right) + \hat{\lambda} = 0. \end{aligned} \tag{7}$$

On somme maintenant Equation (7) pour tous les $i_0 \in [\![1, I]\!]$, on obtient :

$$\begin{aligned} 0 &= \sum_{i_0=1}^{I} n_{i_0}\left(\hat{\alpha}_{i_0}+\hat{\mu} - \overline{y}_{i_0,:}\right) + I\hat{\lambda} \\ 0 &= \sum_{i_0=1}^{I} n_{i_0}\hat{\alpha}_{i_0} + n\hat{\mu} - \sum_{i_0=1}^{I} n_{i_0}\overline{y}_{i_0,:} + I\hat{\lambda} \\ 0 &= \underbrace{\frac{\sum_{i_0=1}^{I} n_{i_0}\hat{\alpha}_{i_0}}{n}}_{(6)} + \hat{\mu} - \underbrace{\frac{\sum_{i_0=1}^{I} n_{i_0}\overline{y}_{i_0,:}}{n}}_{\overline{y}_n} + \frac{I\hat{\lambda}}{n} \\ 0 &= \frac{I\hat{\lambda}}{n} \\ 0 &= \hat{\lambda}. \end{aligned} \tag{8}$$

Ainsi, en substituant la valeur de $\hat{\lambda}$ dans Equation (7), on obtient que $\hat{\alpha}_{i_0} + \hat{\mu} = \overline{y}_{i_0}$. En sommant cette nouvelle équation et en utilisant notre contrainte, on conclut que les estimateurs vérifient :

$$\begin{aligned} &\underbrace{\sum_{i_0=1}^{I}\hat{\alpha}_{i_0}}_{=0} + \underbrace{\sum_{i_0=1}^{I}\hat{\mu}}_{I\hat{\mu}} = \sum_{i_0=1}^{I}\overline{y}_{i_0,:} \\ \implies &\hat{\mu} = \frac{\sum_{i_0=1}^{I}\overline{y}_{i_0,:}}{I} \\ \implies &\hat{\alpha}_{i_0} = \overline{y}_{i_0,:} - \frac{\sum_{i_0}^{I}\overline{y}_{i_0,:}}{I}. \end{aligned}$$

Ainsi, le résultat précédent est vrai pour tout $i_0 \in [\![1, I]\!]$. Il est aussi important de noter que $\hat{\mu}$ est la moyenne des moyennes.

 : $\hat{\mu} \neq \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i} y_{i,j}$ quand il existe $i, i' : n_i \neq n_{i'}$ (classes déséquilibrées).

1.2 La somme pondérée des effets individuels est nulle

Hypothèse : $\sum_{i_0=1}^{I} n_{i_0}\alpha_{i_0} = 0.$

Estimateurs associés :

$$\begin{aligned} \underset{(\mu,\alpha)\in\mathbb{R}\times\mathbb{R}^I}{\arg\min} \quad & \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{i,j} - \mu - \alpha_i)^2 \\ \text{s.c.} \quad & \sum_{i_0=1}^{I} n_{i_0}\alpha_{i_0} = 0 \ . \end{aligned} \tag{9}$$

Le Lagrangien est pour ce problème :

$$\mathcal{L}(\mu, \alpha, \lambda) = \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{i,j} - \mu - \alpha_i)^2 + \lambda\sum_{i_0=1}^{I} n_{i_0}\alpha_{i_0} \ . \tag{10}$$

Comme mentionné précédemment, l'équation (6) est encore valable pour ce cas. Il suffit donc de regarder pour α_{i_0} avec $i_0 \in [\![1, I]\!]$ fixé.

$$\begin{aligned} \frac{\partial\mathcal{L}}{\partial\alpha_{i_0}} &= 0 \\ \sum_{j=1}^{n_{i_0}}(\hat{\alpha}_{i_0} + \hat{\mu} - y_{i_0,j}) + \hat{\lambda}n_{i_0} &= 0 \\ n_{i_0}\hat{\alpha}_{i_0} + n_{i_0}\hat{\mu} - n_{i_0}\overline{y}_{i_0,:} + n_{i_0}\hat{\lambda} &= 0 \\ \hat{\alpha}_{i_0} + \hat{\mu} - \overline{y}_{i_0,:} + \hat{\lambda} &= 0. \end{aligned}$$

On note que la condition (6) donne $\hat{\mu} = \overline{y}_n$ avec notre contrainte. Ainsi, de la même manière que précédemment, en sommant (8) sur les k on obtient :

$$\hat{\lambda} = 0.$$

En substituant les valeurs obtenues dans (8), on conclut que nos estimateurs sont :

$$\begin{aligned} \underline{\text{Condition}} : \hat{\alpha}_{i_0} + \hat{\mu} &= \overline{y}_{i_0,:}. \\ \text{Ainsi, on obtient : } \hat{\mu} &= \overline{y}_n, \\ \hat{\alpha_{i_0}} &= \overline{y}_{i_0,:} - \overline{y}_n. \end{aligned}$$

Remarque 1.2. *L'estimateur de prédiction reste le même :* $\hat{\mu}_{i_0} = \overline{y}_{i_0,:}$.

1.3 Niveau de référence "$\alpha_{i_0} = 0$"

<u>Interprétation</u> : On choisit un modèle i_0 comme référence.

Estimateur associé :

$$\begin{aligned} \underset{(\mu,\alpha)\in\mathbb{R}\times\mathbb{R}^I}{\arg\min} \quad & \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(y_{i,j}-\mu-\alpha_i\right)^2 \\ \text{s.c.} \quad & \alpha_{i_0}=0 \ . \end{aligned} \tag{11}$$

Le Lagrangien :

$$\mathcal{L}\left(\mu,\alpha,\lambda\right)=\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(y_{i,j}-\alpha_i-\mu\right)^2+\lambda\alpha_{i_0}.$$

Calcul de $\frac{\partial\mathcal{L}}{\partial\alpha_i}$ pour $i\neq i_0$, $\frac{\partial\mathcal{L}}{\partial\lambda}$ et $\frac{\partial\mathcal{L}}{\partial\alpha_{i_0}}$:

Tout d'abord, en repartant de Equation (6) : $\hat{\mu}+\frac{1}{n}\sum_{i=1}^{I}n_i\hat{\alpha}_i=\overline{y}_n$.

On fixe $i\in[\![1,I]\!]$ et $i\neq i_0$,

$$\frac{\partial\mathcal{L}}{\partial\alpha_i}=\sum_{j=1}^{n_i}(\mu+\alpha_i-y_{i,j}) \ . \tag{12}$$

En posant $\frac{\partial\mathcal{L}}{\partial\alpha_i}=0$, on a $\forall i\neq i_0$:

$$\begin{aligned} \sum_{j=1}^{n_i}(n_i\hat{\alpha}_i+n_i\hat{\mu}-y_{i,j})&=0 \\ n_i\hat{\alpha}_i+n_i\hat{\mu}-n_i\overline{y}_{i,:}&=0 \\ \hat{\alpha}_i+\hat{\mu}-\overline{y}_{i,:}&=0 \ . \end{aligned} \tag{13}$$

Dérivons l'expression du Lagrangien par λ :

$$\frac{\partial\mathcal{L}}{\partial\lambda}=\alpha_{i_0} \ . \tag{14}$$

On en déduit directement que :

$$\hat{\alpha}_{i_0}=0. \tag{15}$$

Pour $i_0\in[\![1,I]\!]$, on a :

$$\begin{aligned} \frac{\partial\mathcal{L}}{\partial\alpha_{i_0}}(\hat{\mu},\hat{\alpha},\hat{\lambda})&=0 \\ \sum_{j=1}^{n_{i_0}}(\hat{\alpha}_{i_0}+\hat{\mu}-y_{i_0,j})+\hat{\lambda}&=0 \\ n_{i_0}\hat{\mu}-\sum_{j=1}^{n_{i_0}}y_{i_0,j}+\hat{\lambda}&=0 \\ \hat{\mu}=\overline{y}_{i_0:}+\frac{\hat{\lambda}}{n_0} \ . \end{aligned} \tag{16}$$

En sommant sur $i \neq i_0$ (XXX to finish) toutes les équations de type (13) avec Equation (6), on obtient que $\hat{\lambda} = 0$.

Cela assure donc que $\hat{\mu} = \overline{y}_{i_0:}$, d'après la dernière équation.

$$\overline{y}_{i_0,:} = \hat{\alpha}_{i_0} + \overline{y}_{i_0:} \implies \hat{\alpha}_{i_0} = \overline{y}_{i_0,:} - \overline{y}_{i_0:} \ . \tag{17}$$

En conclusion, les solutions sont :

$$\begin{cases} \hat{\alpha}_{i_0} = 0, \\ \hat{\mu} = y_{i_0}, \\ \forall i \neq i_0, \quad \hat{\alpha}_i = \overline{y}_{i,:} - \overline{y}_{i_0,:}. \end{cases} \tag{18}$$

où $\overline{y}_{i,:}$ représente le niveau moyen de la modalité i et $\overline{y}_{i_0:}$ représente le niveau moyen de la modalité de référence i_0

Remarque 1.3. *Quid de l'estimation de σ^2 ?*

$$\widehat{\sigma^2} = \frac{1}{n-I} \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{i,j} - \overline{y}_{i,:})^2 \ \textit{(variance résiduelle)} \ . \tag{19}$$

Concernant l'estimateur du niveau de bruit on peut le voir comme : $\widehat{\sigma^2} = \frac{1}{n-rg(X)} \|y - X\hat{\beta}\|^2$,

avec $X = [\mathbb{1}_{C_1}, \ldots, \mathbb{1}_{C_I}]$ *où* $\mathbb{1}_{C_i(j)} = \begin{cases} 1 \textit{ si } j \in C_i \\ 0 \textit{ sinon} \end{cases}$, $\hat{y} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \vdots \\ \hat{\mu}_I \end{bmatrix}$ *et* $\operatorname{rang}(X) = I$.