HMMA307

Analyse de la variance à deux facteurs

Cours: Joseph Salmon *Scribes: Mégane BOYER, Mathias GOUT et Emeline TOUSTOU*

1 Cas sans interaction

Deux facteurs :

— Facteur 1 : I niveaux / I classes.

— Facteur 2 : J niveaux / J classes.

n_{ij} : nombre de répétitions / d'observations correspondant au facteur 1 dans la classe i et au facteur 2 dans la classe j.
On a alors les contraintes suivantes sur les observations totales :

$$n = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} \quad . \tag{1}$$

Exemple: Des juges en œnologie mettent des notes sur 10 à des vins.

		Facteur 2 : D		
		Vin 1	Vin 2	Vin 3
Facteur 1 : C	Juge 1	$[6,7,8]$ $n_{11}=3$	$[1,2,3,5]$ $n_{12}=4$	$[1,3]$ $n_{13}=2$
	Juge 2	$[3,8,9]$ $n_{21}=3$	$[1]$ $n_{22}=1$	$[1,2,3]$ $n_{23}=3$
	Juge 3	$[5,7,8]$ $n_{31}=3$	$[2,5]$ $n_{32}=2$	$[2]$ $n_{33}=1$

TABLE 1 – Notes attribuées aux vins par les juges.

Dans la Table 1, si le Facteur 1 est Juge 1 et le Facteur 2 est Vin 1, on a : $y_{111} = 6$, $y_{112} = 7$, $y_{113} = 8$.

Modèle : On a :

$$y_{i,j,k} \overset{iid}{\sim} \mathcal{N}(\mu_{ij}, \sigma^2), \qquad \forall i \in [\![1, I]\!], \forall j \in [\![1, J]\!], \forall k \in [\![1, n_{ij}]\!] \tag{2}$$

$$y_{i,j,k} = \mu + \alpha_i + \beta_j + \varepsilon_{i,j,k}. \tag{3}$$

Avec :

$$\mathrm{Cov}(\varepsilon_{i,j,k}, \varepsilon_{i',j',k'}) = \sigma^2 \delta_{i,i'} \delta_{j,j'} \delta_{k,k'} \tag{4}$$

où $\mu \in \mathbb{R}$ représente l'effet moyen, α_i l'effet spécifique du niveau i pour le premier facteur, et β_j l'effet spécifique du niveau j pour le deuxième facteur.

⚠ : Si le plan d'expérience n'est pas équilibré (*i.e.*, les n_{ij} sont différents), l'analyse mathématique est ardue. Ici, on supposera donc pour faciliter l'analyse :

$$\boxed{\forall i \in [\![1, I]\!], \quad \forall j \in [\![1, J]\!], \quad n_{ij} = K} \quad . \tag{5}$$

Ainsi on a donc au final $n = IJK$ observations.

En suivant une approche habituelle par moindre carrés on peut écrire le modèle sous forme matricielle :

$$X = [\mathbb{1}_n \quad \mathbb{1}_{C_1} \quad \dots \quad \mathbb{1}_{C_I} \quad \mathbb{1}_{D_1} \quad \dots \quad \mathbb{1}_{D_J}] \in \mathbb{R}^{n\times(1+I+J)} \quad . \tag{6}$$

où $\mathrm{rang}(X) = I + J + 1 - 2 = I + J - 1$ et $\mathbb{1}_n = (1, \dots, 1)^\top$.

Définition 1.1.

$$\begin{aligned} \underset{(\mu,\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^I\times\mathbb{R}^J}{\arg\min} \ & \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{i,j,k} - \mu - \alpha_i - \beta_j)^2 \\ s.c. \quad & \sum_{i=1}^{I}\alpha_i = 0 \ , \\ & \sum_{j=1}^{J}\beta_j = 0 \ . \end{aligned} \tag{7}$$

On obtient pour ce problème le Lagrangien suivant :

$$\mathcal{L}(\mu, \alpha, \beta, \lambda_\alpha, \lambda_\beta) = \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{i,j,k} - \mu - \alpha_i - \beta_j)^2 + \lambda_\alpha\left(\sum_{i=1}^{I}\alpha_i\right) + \lambda_\beta\left(\sum_{j=1}^{J}\beta_j\right) . \tag{8}$$

On cherche à résoudre le système suivant :

$$\begin{cases} \dfrac{\partial\mathcal{L}}{\partial\mu} &= 0 \\ \dfrac{\partial\mathcal{L}}{\partial\alpha} &= 0 \\ \dfrac{\partial\mathcal{L}}{\partial\beta} &= 0 \\ \dfrac{\partial\mathcal{L}}{\partial\lambda_\alpha} &= 0 \\ \dfrac{\partial\mathcal{L}}{\partial\lambda_\beta} &= 0 \end{cases}$$

On obtient ainsi les résultats suivants :

$$\frac{\partial\mathcal{L}}{\partial\mu} = 0 \implies n\widehat{\mu} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{i,j,k} \implies \widehat{\mu} = \overline{y}_n. \tag{9}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = 0 \implies \forall i \in [\![1, I]\!], \quad \widehat{\alpha}_i = \underbrace{\overline{y}_{i,:,:}}_{=\frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{i,j,k}} - \widehat{\mu} \tag{10}$$

$$\frac{\partial \mathcal{L}}{\partial \beta} = 0 \implies \forall j \in [\![1, J]\!], \widehat{\beta}_j = \underbrace{\overline{y}_{:,j,:}}_{=\frac{1}{IK}\sum_{i=1}^{I}\sum_{k=1}^{K} y_{i,j,k}} - \widehat{\mu} \tag{11}$$

<u>Prédicteur associé :</u>

$$\begin{aligned}\widehat{y_{ij}} &= \widehat{\mu} + \widehat{\alpha}_i + \widehat{\beta}_j \\ &= \overline{y}_{i,:,:} + \overline{y}_{:,j,:} - \widehat{\mu} \\ &= \overline{y}_{i,:,:} + \overline{y}_{:,j,:} - \overline{y}_n\end{aligned}$$

Un estimateur (sans biais) de σ^2 est lui donné :

$$\widehat{\sigma^2} = \frac{1}{n-(I+J-1)} \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} (\widehat{y_{ij}} - y_{i,j,k})^2 \quad . \tag{12}$$

Remarque 1.1. *En effet, on peut vérifier facilement que*

$$\mathbb{E}(\widehat{\sigma^2}) = \sigma^2 \quad . \tag{13}$$

Test global (d'influence d'un facteur) :
<u>Facteur 1 :</u> H_0 : "$\alpha_1 = \alpha_2 = \cdots = \alpha_I$".
<u>Facteur 2 :</u> H_0 : "$\beta_1 = \beta_2 = \cdots = \beta_J$".

Pour le facteur 1 :

$$F_{obs} = \frac{1}{\widehat{\sigma^2}} \frac{KJ}{(I-1)} \sum_{i=1}^{I} (\overline{y}_{i,:,:} - \overline{y}_n)^2 \sim \mathcal{F}^{I-1}_{n-(I+J-1)}. \tag{14}$$

Test : Si $F_{obs} > \mathcal{F}^{I-1}_{n-(I+J-1)}(1-\alpha)$, on rejette H_0 (au niveau α). (XXX quantile de la loi de Fisher au niveau $1-\alpha$)