HMMA307: Advanced Linear Modeling

Chapter 6 : Random Anova

Part I: Fanchon Herman, Cassandre Lepercque, Cherif Amghar

https://github.com/fanchonherman/HMMA307_CM_Random_Anova

Part II: Yani Bouaffad, Ryma Lakehal, Loïc Sauton

https://github.com/ybouaffad/HMMA307_CM_Quantile_Regression

Université de Montpellier

Table of Contents

Motivation

Mixed models can be used in practice to deal with disordered data and allow us to use all of our data. Indeed, we can have different grouping factors but also small sample sizes. So, mixed models can process the data even when we have small sample sizes, structured data, and many covariates to fit.

Statistical model

Model equation

$$y_{ij} = \mu^* + A_j + \varepsilon_{ij}$$

- $\mu^* \in \mathbb{R}$, fixed effect,
- $A_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_A^2),\ \sigma_A^2 > 0,\ \forall j \in [\![1, J]\!]$, random effect,
- $\varepsilon_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)\ \sigma_\varepsilon^2 > 0,\ \forall i \in [\![1, I]\!]\ ,\forall j \in [\![1, J]\!]$ the noise,
- $A_j \perp\!\!\!\perp \varepsilon_{ij}\ ,\forall i, j,$
- $n = \sum_{j=1}^{J} n_j$ and n_j the number of observations of the modality J.

Esperance and covariance

- $\mathbb{E}[y_{ij}] = \mu^*$,
- $\mathbb{V}(y_{ij}) = \sigma_A^2 + \sigma_\varepsilon^2$,
- $\mathrm{cov}(y_{ij}, y_{i'j'}) = \sigma_A^2 \delta_{jj'} + \sigma_\varepsilon^2 \delta_{ii'} \delta_{jj'}$,

Matrix Model

Model equation

$$y = \mu^* \mathbb{1}_n + ZA + \varepsilon$$

- $\mu^* \in \mathbb{R}$, fixed effect
- $Z = \left[\mathbb{1}_{C_1}, \dots, \mathbb{1}_{C_J} \right] \in \mathbb{R}^{n \times J}$, design matrix,
- $C_1 \sqcup \ \dots \ \sqcup C_J = [\![1, n]\!]$, classes / modalities,
- $A = \begin{pmatrix} A_1 & \dots & A_J \end{pmatrix}^\top \in \mathbb{R}^J, \ A \sim \mathcal{N}(0, \sigma_A^2 \ I_{d_J}), \ \sigma_A^2 > 0,$ random matrix,
- $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon^2 > 0$, the noise,
- $$ZA = \sum_{j=1}^{J} A_j \mathbb{1}_{C_j} \in \mathbb{R}^n.$$

Variance calculate

- We have that, $\mathbb{V}(ZA) = \underbrace{Z}_{n\times J}\underbrace{\mathbb{V}(A)}_{J\times J}\ \underbrace{Z^\top}_{J\times n} = \sigma_A^2 ZZ^\top \ \in \mathbb{R}^{n\times n}$,
- Where,

$$ZZ^\top = \begin{bmatrix} \mathbb{1}_{C_1} & \dots & \mathbb{1}_{C_J} \end{bmatrix} \begin{bmatrix} \mathbb{1}_{C_1}^\top \\ \vdots \\ \mathbb{1}_{C_J}^\top \end{bmatrix} = \sum_{j=1}^{J} \mathbb{1}_{C_j} \mathbb{1}_{C_j}^\top,$$

- Then, $\mathbb{V}(y) = \sigma_A^2 ZZ^\top + \sigma_\varepsilon^2 I_{d_n}$.

Reminder for the statistical model

- The model : $y_{ij} = \mu^* + A_j + \varepsilon_{ij}$,
- We have that
$$\overline{y_{:j}} = \frac{1}{n_j} \sum_{i \in C_j} y_{ij},$$
- So, $\mathbb{V}(\overline{y_{:j}}) = \sigma_A^2 + \frac{\sigma_\varepsilon^2}{n_j} := \tau_j^2$ and $\mathbb{E}(\overline{y_{:j}}) = \mu^*$ without bias.

μ^*-estimator

Formula

$$\hat{\mu} = \sum_{j=1}^{J} \omega_j \overline{y_{:j}}, \quad \text{with } \omega_j \propto \frac{1}{\mathbb{V}(\overline{y_{:j}})} \text{the weighting.}$$

Remark :

In balance case, we have :

- $n_j = I$, as we already know that $n = IJ$,
- So we obtain,

$$\hat{\mu} = \frac{I}{J} \sum_{j=1}^{J} \overline{y_{:j}}.$$

Theorem

Let $X_1, \ldots, X_n$ independent random variables. We suppose that $\mathbb{E}(X_1) = \cdots = \mathbb{E}(X_n) = \mu^*$. Among the unbiased linear estimators of μ^*, the minimal variance one is given by:

$$\hat{\mu} = \sum_{i=1}^{n} \frac{X_i/\mathbb{V}(X_i)}{\sum_{i'=1}^{n} 1/\mathbb{V}(X_i)}$$

- **Proof**: We have

$$\min_{(\alpha_1,\ldots,\alpha_n)\in\mathbb{R}^n} \mathbb{V}\left(\sum_{i=1}^{n} \alpha_i X_i\right)$$

$$u.c : \sum_{i=1}^{n} \alpha_i = 1$$

$$\mathbb{E}(\sum_{i=1}^{n} \alpha_i X_i) = \mu^*$$

$$\mathbb{V}(\sum_{i=1}^{n} \alpha_i X_i) = \sum_{i=1}^{n} \alpha_i^2 \mathbb{V}(X_i)$$

Minimize the last expression amounts to minimize this expression:

$$\min_{(\alpha_1,\ldots,\alpha_n)\in\mathbb{R}^n} \sum_{i=1}^{n} \alpha_i^2 \mathbb{V}(X_i) \quad u.c \quad \sum_{i=1}^{n} \alpha_i = 1$$

- **Lagrangian** :

$$\mathcal{L}(\alpha,\lambda) = \sum_{i=1}^{n} \alpha_i^2 \mathbb{V}(X_i) + \lambda(\sum_{i=1}^{n} \alpha_i - 1)$$

Resolution of the optimization system:

$$\nabla\mathcal{L}(\hat{\alpha},\hat{\lambda}) = 0$$

$$\left\{ \begin{array}{l} \frac{\partial\mathcal{L}}{\partial\alpha} = 0 \\ \frac{\partial\mathcal{L}}{\partial\alpha_{i_0}} = 0 \ \forall i_0 \end{array} \right. \iff \left\{ \begin{array}{l} \sum_{i=1}^{n} \hat{\alpha}_i = 1 \\ 2\hat{\alpha}_{i_0}\mathbb{V}(x_{i_0} + \hat{\lambda}) = 0, \forall i_0 \end{array} \right. \quad (1)$$

$$\iff \left\{ \begin{array}{l} \hat{\alpha}_{i_0} = \frac{-\hat{\lambda}}{2\mathbb{V}(x_{i_0}} \\ \sum_{i_0=1}^{n} \hat{\alpha}_{i_0} = 1 = \frac{-\hat{\lambda}}{2}(\sum_{i_0=1}^{n} \frac{1}{\mathbb{V}(x_{i_0})}) \end{array} \right. \quad (2)$$

Finally,

$$\hat{\lambda} = -2(\frac{1}{\sum_{i=1}^{n} 1/\mathbb{V}(X_i)})$$

$$\hat{\alpha}_{i_0} = \frac{\frac{1}{\mathbb{V}(X_{i_0})}}{\sum_{i=1}^{n} 1/\mathbb{V}(X_i)} \Longrightarrow \hat{\mu} = \sum_{j=1}^{J} \alpha_j \bar{y}_{ij}$$

Variance Estimator

- σ_ε^2:

$$\mathbb{E}\left[\frac{1}{J-1}\sum_{j=1}^{n}\sum_{i=1}^{n_j}(y_{ij}-\overline{y_{:j}})^2\right]=\sigma_\varepsilon^2$$

a possible estimator unbiased for σ_ε^2 is:

$$\boxed{\hat{\sigma}_\varepsilon^2=\frac{1}{J-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij}-\overline{y_{:j}})^2}$$

- σ_A^2:
  We have:

$$\mathbb{E}\left[\frac{1}{J-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(\overline{y_{:j}}-\overline{y_n})^2\right]=\sigma_\varepsilon^2+\frac{1}{n(J-1)}\left[n^2-\sum_{j=1}^{J}n_j^2\right]\sigma_A^2$$

Therefore,

$$\boxed{\hat{\sigma}_A^2=\frac{\hat{S^2}-\hat{\sigma_\varepsilon}^2}{n^2-\sum_{j=1}^{J}n_j^2}}$$

with $\hat{S^2}=\frac{1}{J-1}\sum_{j=1}^{J}n_j(\overline{y}_{:j}-\overline{y}_n)^2$

Remarks :

- $n^2 > \sum_{j=1}^{J} n_j$,
- Nothing guarantees that $\hat{\sigma}_A^2 \geq 0$ and otherwise the model is unsuitable for data (where $\sigma_A^2 \approx 0$) . In general, it is replaced by 0 in this case.
  So $\hat{\sigma}_A^2 = max(0, \hat{\sigma}_A^2)$
- $\hat{\mu} = \sum_{j=1}^{J} w_j \overline{y_{:j}}$ is not an estimator because it is not a function of the data, it depends on σ_A^2 and σ_ε^2. We define $\tilde{\mu}$ as:

$$\tilde{\mu} = \sum_{j=1}^{J} \tilde{w}_j \overline{y_{:j}} \text{ with } \tilde{w}_j \propto \left(\hat{\sigma}_A^2 + \frac{\hat{\sigma}_\varepsilon^2}{n} \right)^{-1}$$

Another approach: Maximum Likelihood

Reminder of model :

- Equation: $y = \mu^* \mathbb{1}_n + ZA + \varepsilon$
- $Z = \left[\mathbb{1}_{C_1}, \ldots, \mathbb{1}_{C_J}\right] \in \mathbb{R}^{n \times J}$
- $A = \left[A_1, \ldots, A_J\right]^\top \in \mathbb{R}^J,\ A \sim \mathcal{N}(0, \sigma_A^2\ I_{d_J}),\ \sigma_A^2 > 0,$ random matrix,
- $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon^2 > 0$, the noise.
- And

$$\boxed{y \sim \mathcal{N}\left(\mu^* \mathbb{1}_n\ , \sigma_A^2 Z Z^\top + \sigma_\varepsilon^2 I_{d_n}\right)}$$

- **Likelihood** :

$$\mathcal{L}(y, \mu, \hat{\sigma}^2_A, \hat{\sigma}^2_\varepsilon) \ \propto |V|^{-1/2} \, exp(-\frac{1}{2} \left[y - \mu \mathbb{1}_n\right]^\top \ V^{-1} \left[y - \mu \mathbb{1}_n\right])$$

$$-\log(\mathcal{L}) = \frac{1}{2}(y - \mu\mathbb{1}_n)^\top V^{-1}(y - \mu\mathbb{1}_n) + \frac{1}{2}\log|V| + \frac{n}{2} log(2\pi)$$

We want to minimize $-\log(\mathcal{L})$

$$\implies \quad \min_{\mu, \sigma^2_A, \sigma^2_\varepsilon} \frac{1}{2}(y - \mu\mathbb{1}_n)^\top V^{-1}(y - \mu\mathbb{1}_n) + \frac{1}{2}\log|V|$$

- **first condition order on $\hat{\mu}$** :

$$\frac{\partial \log(\mathcal{L})}{\partial \hat{\mu}} = 0$$

$$\iff \quad \hat{\mu} = \frac{\mathbb{1}_n^\top V^{-1} y}{\mathbb{1}_n^\top V^{-1} \mathbb{1}_n}$$

This is the estimator $\hat{\mu}$ that we proposed before.

Remark:

Optimization in σ_A^2 and σ_ε^2 doesn't admit an explicit function.

Alternate algorithm

Beware: V is unknown $V = \sigma_\varepsilon^2 I_d + \sigma_A^2 Z Z^\top$