Analyse de données de panel

Modèles linéaires mixtes

1 - Introduction

- **Données de panel** : 2 dimensions = 2 sources d'aléa

  $\hookrightarrow$ coupe transversale sur les individus (plusieurs individus)
  $\hookrightarrow$ coupe longitudinale dans le temps (données temporelles)

  = données **longitudinales**
  = observations **répétées** sur les individus au cours du temps

⚠ on va supposer dans ce cours une **réponse Y continue** (distribution gaussienne)

Notation : y_{it} : réponse de l'individu i au temps t

Exemple : Suivi du PIB sur un ensemble de T périodes pour N pays.

Pour la modélisation :

- **Variables explicatives** (X) : différentes natures possibles (qualitative = "facteur" / quantitative)

  $\hookrightarrow$ caractéristiques des individus : **constantes dans le temps**

  Ex : Sexe

  $\hookrightarrow$ caractéristiques du temps de mesure : **identiques pour tous les individus**

  Ex : Montant du SMIC à la date t

  $\hookrightarrow$ caractéristiques **variables dans le temps et au travers les individus**

  Ex : Salaire

- **Particularités** :

  - toutes les mesures sur un même individu partagent un même effet, lié à l'individu (ex : capacités physiques)

  $\hookrightarrow$ **effet de l'individu** i, noté u_i

  - toutes les mesures au même temps partagent un même environnement, lié au temps (ex : contexte économique)

  $\hookrightarrow$ **effet du temps** t, noté v_t

- **Type de modélisation envisagée :**

$$Y_{it} = a + bx_{it} + u_i + v_t + \varepsilon_{it}$$

$\hookrightarrow$ régression linéaire avec

$$\varepsilon_{it} \sim \mathcal{N}(0, \sigma_0^2)$$

$\Rightarrow$ **Quel statut pour u_i et v_t ? Quelle nature de leur effet ?**

2 - La notion d'effet aléatoire

Exemple :

On étudie la qualité du pain selon sa température de cuisson. Pour cela on extrait 4 pains de chacune de 6 fournées, cuites à 3 températures différentes.

Description :

- individu :
- variable réponse :
- variable explicative :

$\hookrightarrow$ des répétitions ?

Notation :

y_{ijk} : réponse observée de l'individu k dans la fournée j à la température i

Modèle proposé :

Définition :
Un *"effet aléatoire"* est un facteur dont les niveaux sont des réalisations d'une variable aléatoire.

Effet fixe / effet aléatoire

- un facteur à effet fixe est défini avec un **nombre fini** de niveaux possibles

  $\hookrightarrow$ on est alors intéressé par l'impact de chacun de ces niveaux sur la variable réponse

- un facteur à effet aléatoire est défini avec un **nombre infini** de niveaux possibles (avec évidemment seulement un nombre fini présent dans l'expérience observée)

  $\hookrightarrow$ on est alors intéressé par la variabilité induite par ces niveaux

Hypothèse :
La variable aléatoire dont sont issus les différents niveaux du facteur, est supposée distribuée selon la loi normale :

$$U \sim \mathcal{N}(0, \sigma_1^2)$$

$\hookrightarrow$ Le paramètre σ_1^2 est un paramètre inconnu à estimer.

Il traduit la variabilité entre les niveaux de la variable à effet aléatoire :

- variabilité inter-individuelle
- variabilité inter-temporelle

3 - Impact sur la modélisation

3.1 - Les objets :

En oubliant l'effet temporel, et en considérant :

$$Y_{it} = a + bx_{it} + u_i + \varepsilon_{it}$$

- les objets aléatoires :

$$\begin{array}{lcl} \varepsilon_{it} & \sim & \mathcal{N}(0, \sigma_0^2) \\ U_i & \sim & \mathcal{N}(0, \sigma_1^2) \\ Y_{it} & \sim & \mathcal{N}(\qquad , \qquad) \end{array}$$

- les paramètres à estimer :

  - les paramètres mesurant l'"impact" des variables à effet fixe : a, b
  - le paramètre de variance de l'effet aléatoire : σ_1^2
  - le paramètre de variance de l'erreur : σ_0^2

3.2 - Loi marginale et conditionnelle :

Toujours en considérant : $Y_{it} = a + bx_{it} + u_i + \varepsilon_{it}$

- effets moyens :

$$\begin{aligned} E(Y_{it}) &= \\ E(Y_{it}|U_i = u_i) &= \end{aligned}$$

- composantes de la variance :

$$\begin{aligned} V(Y_{it}) &= \\ V(Y_{it}|U_i = u_i) &= \end{aligned}$$

$\hookrightarrow$ **loi conditionnelle**

$$Y_{it}|U_i = u_i \sim \mathcal{N}(\quad , \quad)$$

$\hookrightarrow$ **loi marginale**

$$Y_{it} \sim \mathcal{N}(\quad , \quad)$$

3.3 - Dépendance :

Calculons la covariance entre 2 observations du même individu i, l'une au temps t_1 et l'autre au temps t_2 : $\text{cov}(Y_{it_1}, Y_{it_2})$.

Rappel : les erreurs sont indépendantes entre elles.

$$\begin{aligned} \text{cov}(Y_{it_1}, Y_{it_2}) &= \text{cov}(u_i + \varepsilon_{it_1} \ , \ u_i + \varepsilon_{it_2}) \\ &= \text{cov}(u_i \ , \ u_i) \\ &= \text{var}(u_i) \\ &= \sigma_1^2 \end{aligned}$$

$\hookrightarrow$ la présence d'un effet aléatoire dans la modélisation introduit de la **dépendance** entre toutes les observations reliées à une même réalisation de l'effet aléatoire (ici à un même individu).

3.4 - Écriture matricielle :

Pour l'individu i, le vecteur de ses répétitions a pour matrice de variance-covariance :

$$\text{Var}\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iT} \end{pmatrix} = \begin{pmatrix} \sigma_0^2+\sigma_1^2 & \sigma_1^2 & \dots & \sigma_1^2 \\ \sigma_1^2 & \sigma_0^2+\sigma_1^2 & \dots & \sigma_1^2 \\ \vdots & & \ddots & \\ \sigma_1^2 & \sigma_1^2 & \dots & \sigma_0^2+\sigma_1^2 \end{pmatrix}$$

$$= V_s$$

$\hookrightarrow$ notée V_s.

Les individus restent indépendants entre eux.

$$\mathrm{Var}\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1T} \\ \cdots \\ Y_{21} \\ \vdots \\ Y_{2T} \\ \cdots \\ \vdots \\ \cdots \\ Y_{n1} \\ \vdots \\ Y_{nT} \end{pmatrix} = \left(\begin{array}{c:c:c:c} V_s & 0_{TT} & \cdots & 0_{TT} \\ \hdashline 0_{TT} & V_s & \ddots & \vdots \\ \hdashline \vdots & \ddots & \ddots & 0_{TT} \\ \hdashline 0_{TT} & \cdots & 0_{TT} & V_s \end{array}\right)$$

4 - Quelques modèles à effets aléatoires

Ou encore appelés ...

... modèles mixtes

... modèles à composantes de la variance

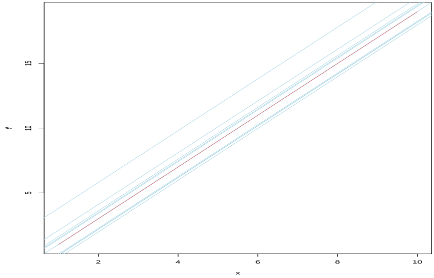
$$E(Y_{it}) = a + bx_{it} \qquad (\mathcal{M}_0)$$

$i = 1, ..., I$ et $t = 1, ..., T$

4.1 - Modèle à intercept aléatoire

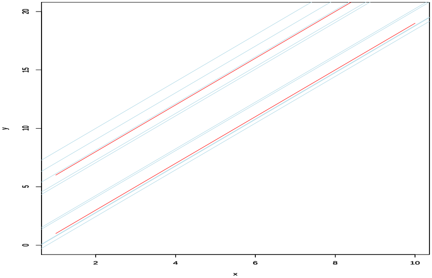
$I = 10$ et $T = 5$

$$E(Y_{it}|U_i = u_i) = a + bx_{it} + u_i$$



Avec un facteur à effet fixe :
à 2 niveaux.

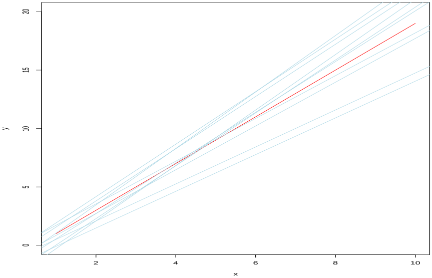
$$E(Y_{ijt}|U_i = u_i) = a + \beta_j + bx_{ijt} + u_i$$



4.2 - Modèle à pente aléatoire

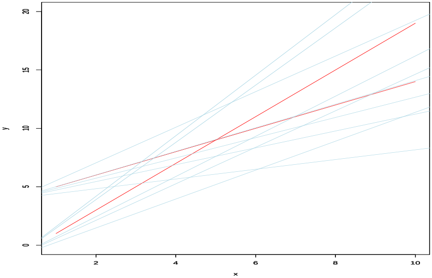
$I = 10$ et $T = 5$

$$E(Y_{it}|U_i = u_i) = a + bx_{it} + u_i x_{it}$$



Avec un facteur à effet fixe sur l'intercept et sur la pente :
à 2 niveaux.

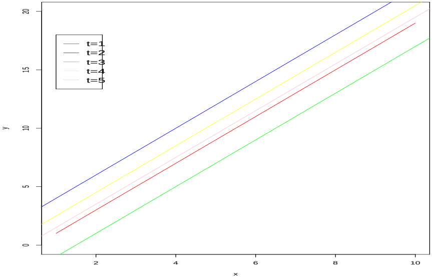
$$E(Y_{ijt}|U_i = u_i) = a + \beta_j + bx_{ijt} + \gamma_j x_{ijt} + u_i x_{ijt}$$



4.3 - Modèle à effet fixe temporel

$I = 10$ et $T = 5$

$$E(Y_{it}) = a + bx_{it} + v_t$$



$\hookrightarrow$ représentation graphique plus compliquée

Avec intercept individuel aléatoire

$$E(Y_{it}|U_i = u_i) = a + bx_{it} + u_i + v_t$$

4.4 - Modèle à effet aléatoire temporel

- intercept aléatoire temporel

$$E(Y_{it}|U_i = u_i, V_t = v_t) = a + bx_{it} + v_t$$

- intercept aléatoire temporel et individuel

$$E(Y_{it}|U_i = u_i, V_t = v_t) = a + bx_{it} + u_i + v_t$$

- intercept aléatoire temporel et pente aléatoire individuelle

$$E(Y_{it}|U_i = u_i, V_t = v_t) = a + bx_{it} + u_i x_{it} + v_t$$

- intercept aléatoire individuel et pente aléatoire temporelle

$$E(Y_{it}|U_i = u_i, V_t = v_t) = a + bx_{it} + u_i + v_t x_{it}$$

... toutes les combinaisons sont possibles et imaginables aussi avec plusieurs variables explicatives.