Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

# Données longitudinales et modèles mixtes

Etienne Dantan

*(Etienne.Dantan@univ-nantes.fr)*

**Master 2 Recherche**
**Modélisation en Pharmacologie Clinique et Épidémiologie**

*(UE Données longitudinales et données de survie)*

12, 13, 14 et 25 Novembre 2013

# Plan



1. Introduction
2. Objectifs
3. Méthodes historiques
4. Modèle linéaire mixte
5. Logiciels
6. Stratégie de modélisation
7. Application
8. Discussion
9. Références

# Plan

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

1. Introduction
2. Objectifs
3. Méthodes historiques
4. Modèle linéaire mixte
5. Logiciels
6. Stratégie de modélisation
7. Application
8. Discussion
9. Références

# Données groupées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Fréquentes en épidémiologie

  $\rightarrow$ Étude de facteurs familiaux ou environnementaux

- Résultat d'un plan de sondage où chaque observation n'est pas tirée au sort indépendamment des autres

- + facile de sélectionner des sujets par zone géographique concentrée

  $\rightarrow$ Exemples : hôpitaux, écoles, communes,...

$\Rightarrow$ Corrélation intra-groupe

$\Rightarrow$ Méthodes statistiques adaptées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Définition

- ou Données répétées
- Variable réponse mesurée de façon répétée au cours du temps sur les mêmes sujets
- Traitées comme des données groupées

  $\rightarrow$ toutes les observations d'un même sujet constituent un groupe

$\Rightarrow$ Intérêt majeur en épidémiologie :

$\rightarrow$ Étude de l'évolution de sujets face à une pathologie

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Intérêt

- Étude d'un ou plusieurs marqueurs au cours du temps
- Reflet de variations liées

  → à la dégradation de l'état de santé du patient

  → à l'amélioration en réaction à la prise d'un médicament

# Exemple 1 (Données de Captopril)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- *Hand, Daly, Lunn, Mc Conway and Otrowski (1994)*
- Patients souffrant d'hypertension (N=15)
- Critère de jugement :
  - Pression Artérielle Systolique,
  - Pression Artérielle Dyastolique

avant et après traitement (Captopril)

$\Rightarrow$ Effet du traitement sur la PAS, PAD ?

# Exemple 1 (Données de Captopril)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

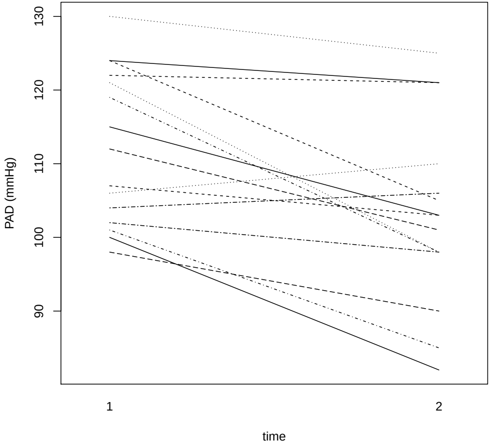
| Patient | Avant | | Après | |
|---|---|---|---|---|
| | PAS | PAD | PAS | PAD |
| 1 | 210 | 130 | 201 | 125 |
| 2 | 169 | 122 | 165 | 121 |
| 3 | 187 | 124 | 166 | 121 |
| 4 | 160 | 104 | 157 | 106 |
| 5 | 167 | 112 | 147 | 101 |
| 6 | 176 | 101 | 145 | 85 |
| 7 | 185 | 121 | 168 | 98 |
| 8 | 206 | 124 | 180 | 105 |
| 9 | 173 | 115 | 147 | 103 |
| 10 | 146 | 102 | 136 | 98 |
| 11 | 174 | 98 | 151 | 90 |
| 12 | 201 | 119 | 168 | 98 |
| 13 | 198 | 106 | 179 | 110 |
| 14 | 148 | 107 | 129 | 103 |
| 15 | 154 | 100 | 131 | 82 |

# Exemple 1 (Données de Captopril)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

$\neq$ Profils individuels

## Remarques :



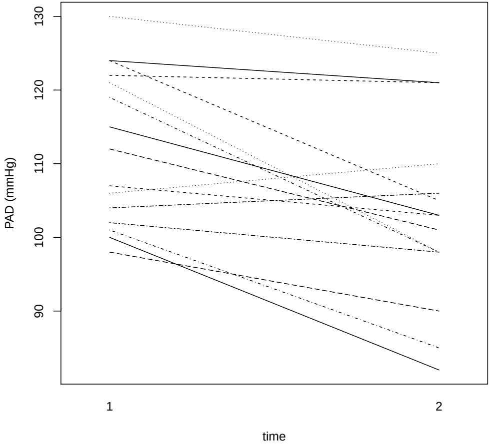
# Exemple 1 (Données de Captopril)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

$\neq$ Profils individuels

## Remarques :

- Observations appariées
- Variabilité importante entre patients



# Exemple 2 (Courbes de croissance)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- *Goldstein et al. (1979)*
- La croissance est-elle liée à la taille de la mère ?
- Filles de 6 ans (N=20) issues de familles où la taille de la mère est considérée comme étant
  - petite,
  - moyenne,
  - ou grande
- Critère de jugement : Croissance (taille) des filles au cours de 4 années (de 6 à 10 ans)

# Exemple 2 (Courbes de croissance)

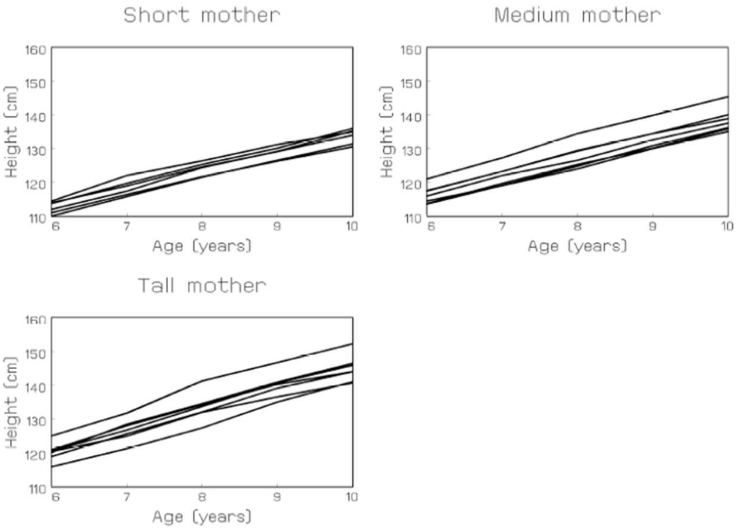


- Répartition des 20 jeunes filles selon la taille de leur mère

| | Taille de la mère (cm) | Nombre de jeunes filles |
|---|---|---|
| Petites | $< 155$ | 6 |
| Moyennes | $[155; 164]$ | 7 |
| Grandes | $> 164$ | 7 |

# Exemple 2 (Courbes de croissance)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



# Exemple 2 (Courbes de croissance)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

Remarques :

# Exemple 2 (Courbes de croissance)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarques :

- Relation linéaire entre l'âge et la taille
- Variabilité inter-sujet moyenne
- Variabilité intra-sujet faible
- Nombre de mesures/sujet fixé
- Temps de mesure fixés

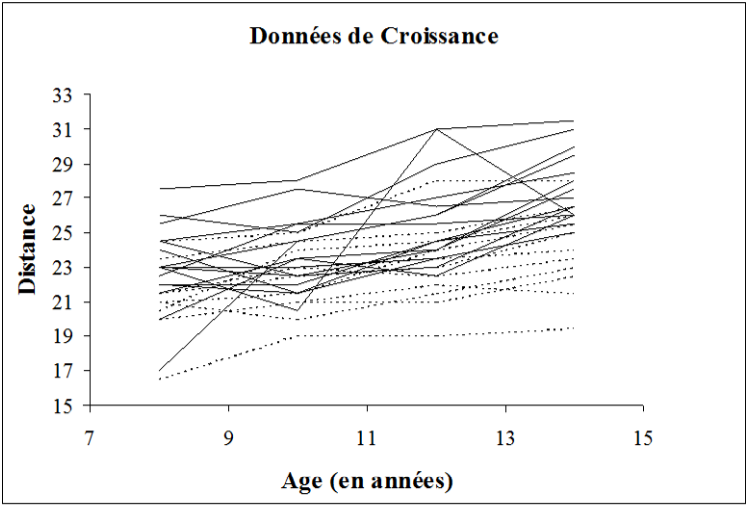
# Exemple 3 (Croissance faciale)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- *Pothoff and Roy (1964)*
- La croissance maxillaire est-elle liée au sexe ?
- Filles (N=11), Garçons (N=16)
- Critère de jugement : distance du centre de la fente pituitaire à la fente ptérymaxillaire (unités : $10^{-4}$m) mesurée à 8, 10, 12 et 14 ans chez les filles et les garçons

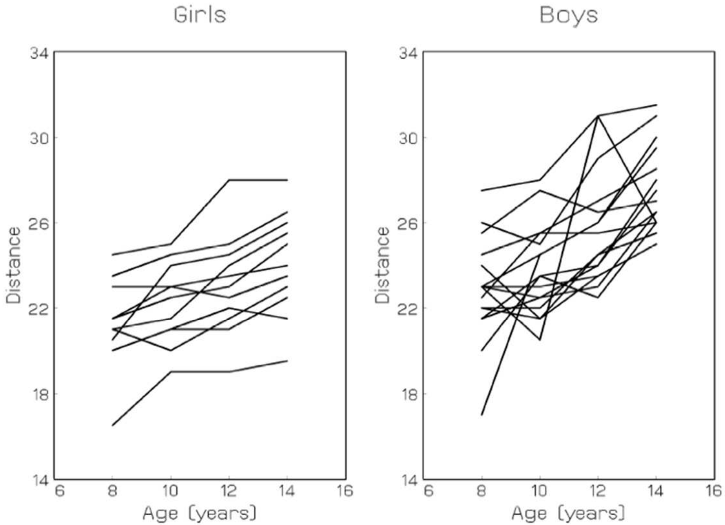
# Exemple 3 (Croissance faciale)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



# Exemple 3 (Croissance faciale)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



# Exemple 3 (Croissance faciale)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

Remarque :

# Exemple 3 (Croissance faciale)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarque :

- Relation linéaire
- Variabilité entre filles et garçons
- Variabilité intra-sujet (pour les 2 genres) importante
- Nombre de mesures/sujet fixé
- Temps de mesure fixés

# Exemple 4 (Étude pré-clinique “Rats”)

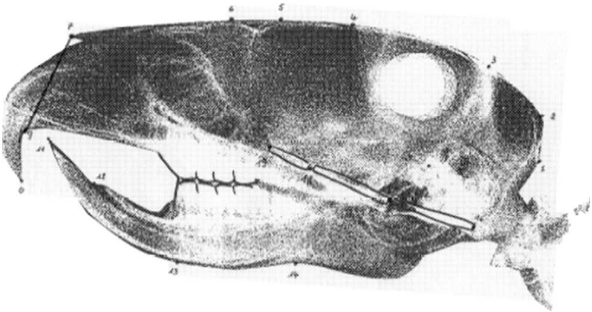
Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- *Verbeke and Molenberghs (2003)*
- Étude pré-clinique thérapeutique randomisée (50 rats)
- Étude de la relation entre la production de testostérone et la croissance crânienne
- Effet de l'augmentation des doses d'un médicament réduisant la production de testostérone (Decapeptyl) $\rightarrow$ croissance crânienne chez les rats Wistar mâles
- 3 groupes : témoins, dose faible, dose forte (administré à 45 jours)
- Critère de jugement : taille du crâne selon un axe prédéterminé, mesures réalisées à partir de 50 jours puis tous les 10 jours

# Exemple 4 (Étude pré-clinique “Rats”)

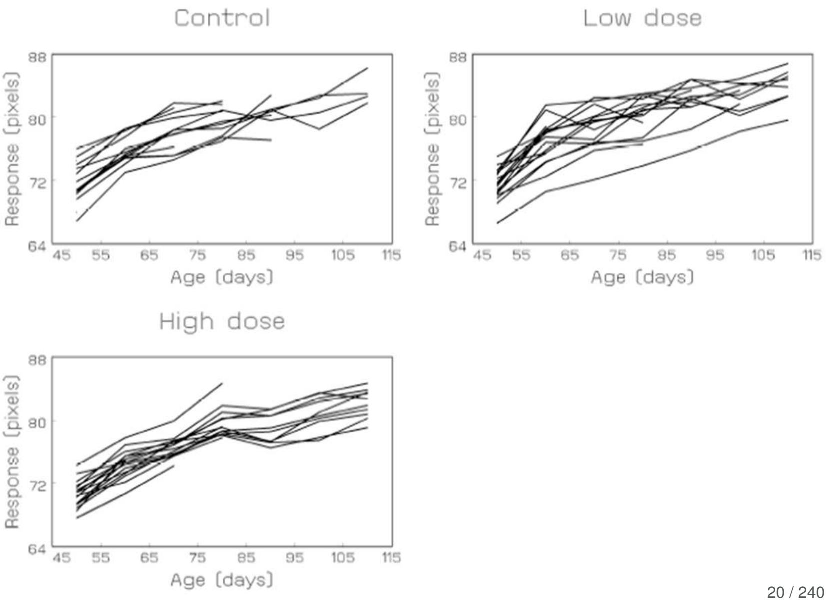
Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d’inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- The responses are distances (pixels) between well defined points on x-ray pictures of the skull of each rat:



# Exemple 4 (Étude pré-clinique “Rats”)





# Exemple 4 (Étude pré-clinique “Rats”)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarques

# Exemple 4 (Étude pré-clinique “Rats”)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarques

- Relation linéaire
- Variabilité entre les rats importante
- Variabilité intra-sujet faible
- Temps de mesure fixés
- Données manquantes pour raison connue (anesthésie)

# Exemple 5 (Cohorte PAQUID)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Cohorte française prospective sur le vieillissement cognitif
- *Dartigues et al. (1992)*
- 3777 personnes âgées de plus de 65 ans
- Critère de jugement : Mesures répétées de tests psychométriques (MMSE, Isaacs, Benton, etc.) pour étudier le vieillissement cognitif des personnes âgées

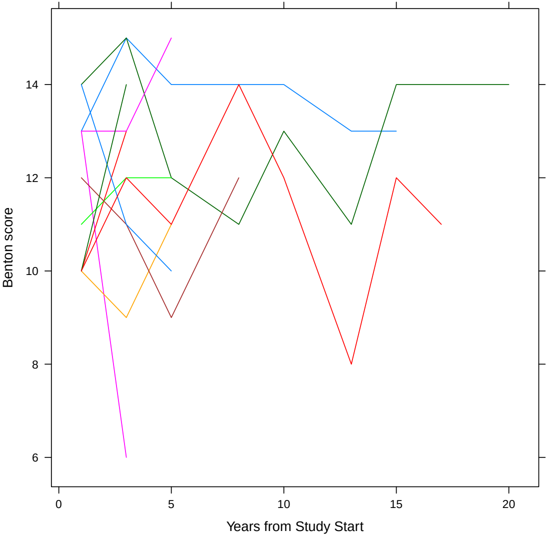

$\Rightarrow$ 8 visites

- Démence évaluée à chaque visite
- Information sur le décès obtenue en temps continu

# Exemple 5 (Cohorte PAQUID)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 12 sujets âgés de plus de 65 ans



# Exemple 5 (Cohorte PAQUID)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarques

# Exemple 5 (Cohorte PAQUID)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarques

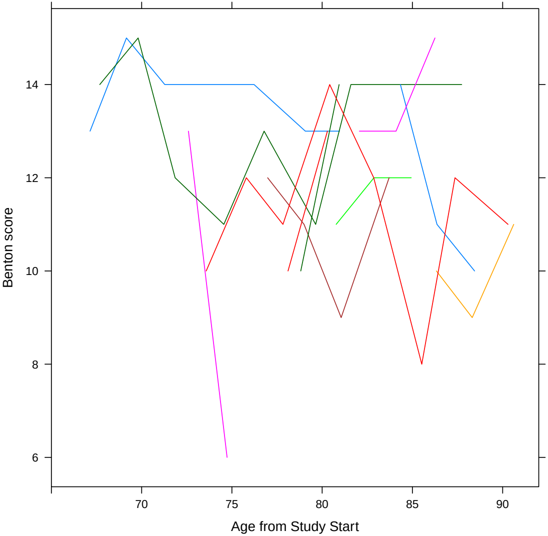
- Variabilité inter-sujet élevé
- Variabilité intra-sujet élevé
- Données manquantes (plan déséquilibré)
  $\rightarrow$ Cause parfois inconnue (sortie d'étude)
- Temps de base : années depuis début d'étude

$\Rightarrow$ Temps de mesures fixés

# Exemple 5 (Cohorte PAQUID)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 12 sujets âgés de plus de 65 ans



# Exemple 5 (Cohorte PAQUID)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarques

# Exemple 5 (Cohorte PAQUID)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarques

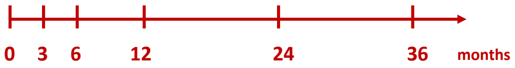
- Variabilité inter-sujet élevé
- Variabilité intra-sujet élevé
- Données manquantes (plan déséquilibré)
  → Cause parfois inconnue (données manquantes intermittentes, sorties d'étude)
- **Temps de base : âge depuis début d'étude**

⇒ Temps de mesures fixés

# Exemple 6 (Cohorte DIVAT)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

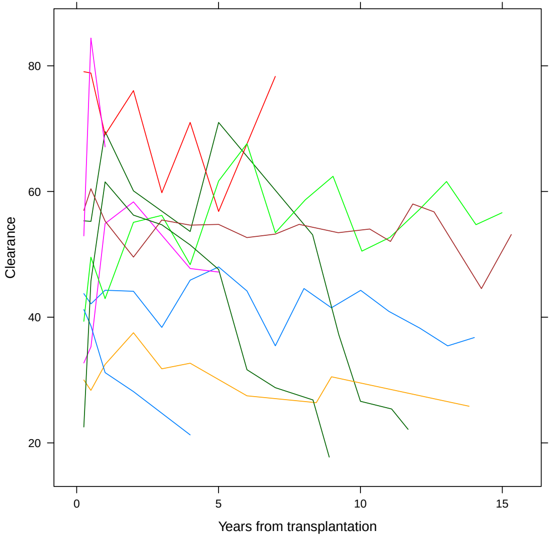
- Cohorte française prospective sur la transplantation rénale
- *Ladrière et al. (2010)*
- plus de 15000 greffes rénales depuis 1990
- Critère de jugement : Mesures répétées de clairance à la créatinine pour le suivi de patients transplantés rénaux



# Exemple 6 (Cohorte DIVAT)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 10 patients transplantés



# Exemple 6 (Cohorte DIVAT)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarques

# Exemple 6 (Cohorte DIVAT)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarques

- Variabilité inter-sujet élevé
- Variabilité intra-sujet élevé
- Données manquantes (plan déséquilibré)

  $\rightarrow$ Cause parfois inconnue (données manquantes intermittentes, sortie d'étude)
- Temps de base : années depuis début d'étude

$\Rightarrow$ Temps de mesures fixés

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Différents types d'étude

- Données expérimentales/observationnelles
- Étude appariée
- Étude pré-clinique ou essais cliniques
- Étude observationnelle

  $\rightarrow$ cohorte prospective

$\Rightarrow$ Étude d'un processus d'évolution

# Particularités des données longitudinales

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Variabilité inter-sujet
- Variabilité intra-sujet (souvent importante)
- Mesures ordonnées
- Temps de mesure (fixes ou non)
- Nombre de mesures (fixes ou non)
- Espacement entre mesures (fixes ou non)
- Présence éventuelle de données manquantes

# Particularités des données longitudinales

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Mesures répétées au cours du temps d'un sujet sont souvent fortement corrélées

  $\rightarrow$ 2 observations chez un même sujet plus proche que 2 mesures prises chez deux individus distincts

$\Rightarrow$ Modélisation de données répétées : Prise en compte de cette corrélation

  $\rightarrow$ sinon Inférence incorrecte des paramètres du modèle d'évolution

- Corrélation entre 2 mesures peut dépendre de l'intervalle de temps qui les sépare

$\Rightarrow$ Structure de Variance-Covariance

# Plan

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

1. Introduction
2. Objectifs
3. Méthodes historiques
4. Modèle linéaire mixte
5. Logiciels
6. Stratégie de modélisation
7. Application
8. Discussion
9. Références

# Objectifs

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Étudier l'évolution d'une variable réponse en fonction du temps
- Identifier les facteurs influençant cette évolution
- à l'aide d'approches permettant de tenir compte de la particularité des données longitudinales
- en particulier le modèle linéaire mixte
  - $\rightarrow$ Introduire les notations mathématiques nécessaires
  - $\rightarrow$ Illustrer son usage à partir d'une application

# Plan



1. Introduction
2. Objectifs
3. Méthodes historiques
4. Modèle linéaire mixte
5. Logiciels
6. Stratégie de modélisation
7. Application
8. Discussion
9. Références

# Retour à l'exemple 2 (Courbes de croissance)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- La croissance est-elle liée à la taille de la mère ?
- Filles de 6 ans (N=20) issues de familles où la taille de la mère est considérée comme étant petite, moyenne, ou grande
- Critère de jugement : Croissance (taille) des filles au cours de 4 années (de 6 à 10 ans)

- Matrice des corrélations des 5 mesures réalisées

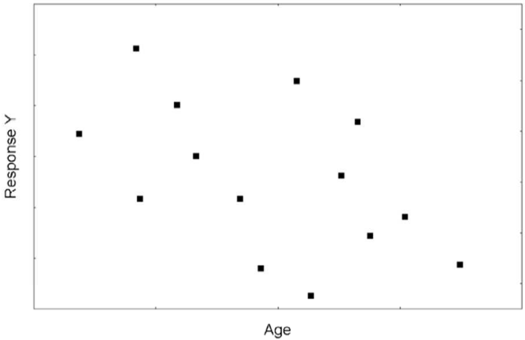
$$\begin{pmatrix} 1.00 & 0.95 & 0.96 & 0.93 & 0.87 \\ 0.95 & 1.00 & 0.97 & 0.96 & 0.89 \\ 0.96 & 0.97 & 1.00 & 0.98 & 0.94 \\ 0.93 & 0.96 & 0.98 & 1.00 & 0.98 \\ 0.87 & 0.89 & 0.94 & 0.98 & 1.00 \end{pmatrix}$$

$\Rightarrow$ Cette structure de corrélation ne doit pas être ignorée dans les analyses

# Données transversales vs. données longitudinales (1)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

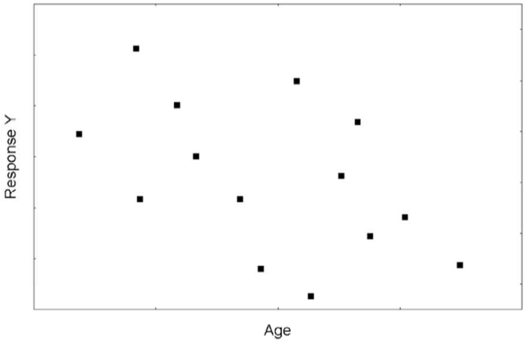
## Étude de la relation entre une variable réponse $Y$ et l'âge à partir de données transversales



# Données transversales vs. données longitudinales (1)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

Étude de la relation entre une variable réponse $Y$ et l'âge à partir de données transversales



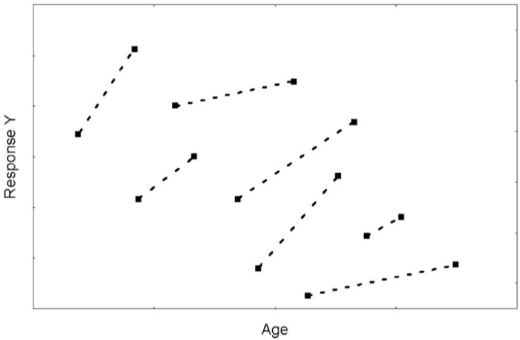

$\Rightarrow$ Relation négative entre l'âge et $Y$ ?

# Données transversales vs. données longitudinales (2)



## Étude de la relation entre une variable réponse $Y$ et l'âge à partir de données longitudinales

On aurait pu avoir les même données avec 2 mesures par sujet ! !

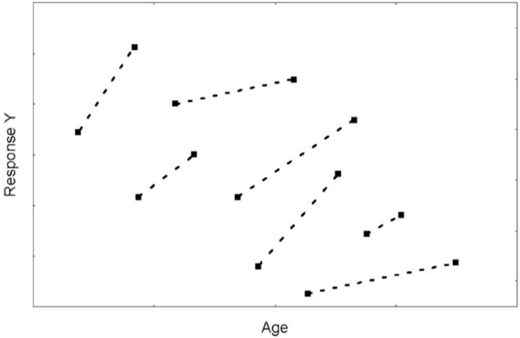


# Données transversales vs. données longitudinales (2)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Étude de la relation entre une variable réponse $Y$ et l'âge à partir de données longitudinales

On aurait pu avoir les même données avec 2 mesures par sujet ! !

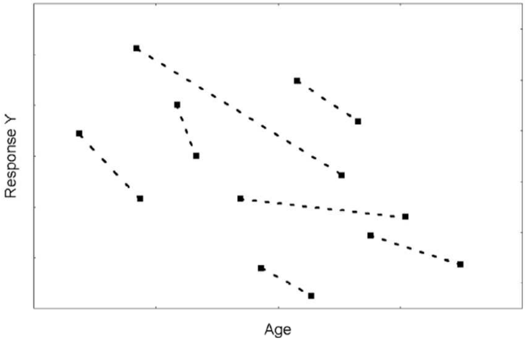



$\Rightarrow$ Relation transversale négative mais relation longitudinale positive ?

# Données transversales vs. données longitudinales (3)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

On aurait pu se trouver dans un autre cas de figure ...




$\Rightarrow$ Relation transversale et longitudinale négatives ?

$\Rightarrow$ Les données longitudinales permettent de distinguer les différences **entre** sujets des changements **intra** sujets

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Prise en compte de la corrélation inter-mesure (intra-sujet)

- t-test apparié : $\Delta_i = y_{i1} - y_{i2}$
  $\Rightarrow$ réduit le nb de mesure à 1 par sujet
  mais valable que pour 2 mesures
- Analyse séparée réalisée à chaque temps
- Analyse de l'aire sous la courbe
- Analyse de la dernière mesure effectuée
- Analyse du changement depuis la 1ère mesure
- ANCOVA

# Méthode simple 1

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Analyse séparée réalisée à chaque temps

- Avantages :
  - simple à interpréter
  - utilise l'ensemble des données disponibles
- Inconvénients :
  - ne considère pas l'évolution globale
  - ne permet pas d'étudier des différences d'évolution
  - problème de la multiplicité des tests

# Méthode simple 2

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Analyse de l'aire sous la courbe d'évolution

Pour chaque sujet *i* :

$$AUC_i = (t_{i2} - t_{i1}) * \frac{y_{i1} + y_{i2}}{2} + (t_{i3} - t_{i2}) * \frac{y_{i2} + y_{i3}}{2} + \dots$$

$\Rightarrow AUC_i$ sont analysées

- Avantages :
  - pas de problème de tests multiples
  - OK pour des plans déséquilibrés
  - considère les $\neq$ globales
- Inconvénients :
  - utilise une information partielle : $AUC_i$

# Méthode simple 3

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Analyse de la dernière mesure effectuée

Utilisés dans les essais randomisés où il n'y a de différences systématiques à l'inclusion

$\Rightarrow$ Évaluation de l'effet du traitement en comparant uniquement la dernière mesure effectuée

- Avantages :
  - pas de problème de tests multiples
  - OK pour des plans déséquilibrés
- Inconvénients :
  - utilise une information partielle : $y_{in_i}$

# Méthode simple 4

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Analyse du changement depuis la 1ère mesure

Méthode simple pour comparer des évolutions entre sujets, en ajustant sur les $\neq$ à l'inclusion

$\Rightarrow$ Analyser les changements spécifiques à chaque individu $(y_{in_i} - y_{i1})$

- Avantages :
  - pas de problème de tests multiples
  - OK pour des plans déséquilibrés
- Inconvénients :
  - utilise une information partielle : $y_{in_i} - y_{i1}$

# Méthode simple 5

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## ANCOVA

Analyse de Variance avec Covariable

$\Rightarrow$ 1ère mesure souvent utilisée comme variable dans le modèle de régression

- Avantages :
  - pas de problème de tests multiples
  - OK pour des plans déséquilibrés
- Inconvénients :
  - utilise une information partielle : $y_{in_i}$ et $y_{i1}$
  - non prise en compte variabilité de $y_{i1}$

# Un peu de discussion

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

$\neq$ Méthodes présentées :

- basées sur des “summary statistics”
- résumant le vecteur de mesures répétées pour chaque sujet séparément
- puis analyse de cette “summary statistic” par méthodes classiques

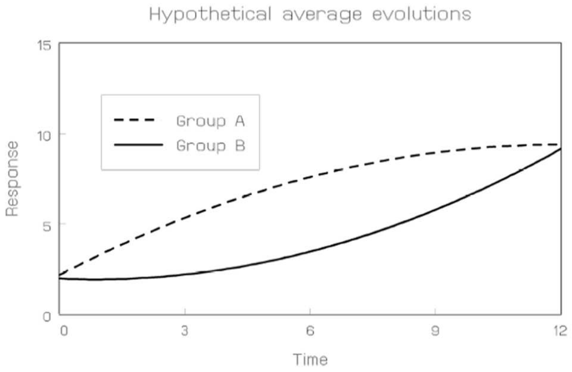
$\Rightarrow$ Analyse de données longitudinales devient analyse de données indépendantes

MAIS

# Un peu de discussion

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Perte d'information importante
- Ne s'attachent pas à la dynamique/trajectoire du processus d'évolution



# Méthode 6 : ANOVA à mesures répétées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- ANOVA à mesures répétées : univariée
- MANOVA à mesures répétées : multivariée

- Applicable dans le cas de données parfaitement équilibrées
  - $\rightarrow$ mesures pour chaque sujet aux mêmes *n* moments
  - $\rightarrow$ aucune données manquantes

- Modéliser chaque observation individuelle
- Hypothèse forte sur la covariance entre mesures

# Méthode 6 : ANOVA à mesures répétées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Principe :

- Similaire à l'ANOVA classique
- Tenir compte et modéliser les 2 sources de variation possibles :
  - Variation inhérente à la population d'où est issu l'échantillon (variation biologique,...)
    $\Rightarrow$ Variations inter-individus
  - Variation propre à l'individu (fluctuations intra-sujets et erreurs de mesure)
    $\Rightarrow$ Variations intra-individus

# Méthode 6 : ANOVA à mesures répétées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Notations :

- $Y_{ijk}$ : observation faite pour le sujet $i$ au temps $j$ dans le groupe $k$
  - $i = 1, ..., r_k$ où $r_k$ : effectif du groupe $k$
  - $j = 1, ..., n_i$ mesures
  - $k = 1, ..., q$ groupes
- Effectif total : $m = \sum_{k=1}^{q} r_k$
- $\forall\ i,\ \ n_i = n$

$\Rightarrow$ Chaque individu est observé aux $n$ temps

# Méthode 6 : ANOVA à mesures répétées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Modèle :

$$
\begin{aligned}
Y_{ijk} &= \mu + \tau_k + \gamma_j + (\tau\gamma)_{jk} + b_{ik} + \epsilon_{ijk} \\
&= \mu_{jk} + \varepsilon_{ijk}
\end{aligned}
$$

- $\mu$ : moyenne générale
- $\tau_k$ : déviation par rapport à $\mu$ due au groupe $k$
- $\gamma_j$ : déviation associé au temps $j$
- $(\tau\gamma)_{jk}$ : interaction entre groupe $k$ et temps $j$
- $b_{ik}$ : effet aléatoire avec $E(b_{ik}) = 0$
  $\rightarrow$ source de variations aléatoires due aux variations inter-individus
- $\epsilon_{ijk}$ : erreur aléatoire avec $E(\epsilon_{ijk}) = 0$
  $\rightarrow$ source de variations aléatoires due aux variations intra-individus

# Méthode 6 : ANOVA à mesures répétées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarques :

- $E(b_{ik}) = 0$ et $E(\epsilon_{ijk}) = 0$ , donc

$$E(Y_{ijk}) = \mu + \tau_k + \gamma_j + (\tau\gamma)_{jk} = \mu_{jk}$$

  $\Rightarrow$ réponse moyenne de l'individu $i$ au temps $j$ du groupe $k$

- $\varepsilon_{ijk} = b_{ik} + \epsilon_{ijk}$ : somme des effets aléatoires

  $Y_{ijk}$ fluctue autour de $\mu$ au temps $j$ pour le sujet $i$ du groupe $k$

  $\Rightarrow \varepsilon_{ijk}$ représente l'ensemble des sources de variations aléatoires

# Méthode 6 : ANOVA à mesures répétées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarques :

- temps de mesures n'apparaissent pas explicitement dans le modèle

$\rightarrow$ $\gamma_j$ et interaction $(\tau\gamma)_{jk}$ associés à chaque temps

- Non prise en compte de la chronologie et de l'espacement entre temps de mesures

# Méthode 6 : ANOVA à mesures répétées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Hypothèses :

- $b_{ik}$ : écart $/\mu$ du sujet $i$ du groupe $k$ identique pour $\forall\ j$

  $\rightarrow$ hypothèse du modèle

- $b_{ik} \sim \mathcal{N}(0, \sigma_b^2)$ + mutuellement indépendants
  - normalité supposée $\forall\ k$ : variation similaire au sein de chaque groupe
  - indépendance entre les individus

# Méthode 6 : ANOVA à mesures répétées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Hypothèses

- $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma_e^2)$ + mutuellement indépendants
  - normalité supposée $\forall\ j, k$ : effet des sources de variations intra-individuelles identiques au sein de chaque groupe et pour chaque temps
  - indépendance des erreurs

  $\epsilon_{ijk}$ : ensemble des fluctuations intra-individuelles, or corrélation inter-mesure

  $\Rightarrow$ Hypothèse non raisonnable ?

## $\neq$ hypothèses $\Rightarrow$ normalité des $Y_{ijk}$

$$Y_{ik} = \mu_k + \varepsilon_{ik} \sim \mathcal{N}(\mu_k, \sigma_b^2 J_n + \sigma_e^2 I_n)$$

où $I_n$ : matrice identité et $J_n$ : matrice carré de 1 de dimension $n * n$

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Table d'ANOVA :

- $\bar{Y}_{i.k} = \frac{1}{n}\sum_{j=1}^{n} Y_{ijk}$ : moyenne estimée des mesures du sujet $i$ du groupe $k$
- $\bar{Y}_{.jk} = \frac{1}{r_k}\sum_{i=1}^{r_k} Y_{ijk}$ : moyenne estimée des $r_k$ sujets du groupe $k$ au temps $j$
- $\bar{Y}_{..k} = \frac{1}{r_k n}\sum_{i=1}^{r_k}\sum_{j=1}^{n} Y_{ijk}$ : moyenne estimée de l'ensemble des observations du groupe $k$
- $\bar{Y}_{.j.} = \frac{1}{m}\sum_{k=1}^{q}\sum_{i=1}^{r_k} Y_{ijk}$ : moyenne estimée de l'ensemble des observations au temps $j$
- $\bar{Y}_{...}$ : moyenne estimée de l'ensemble des $m * n$ observations

# Méthode 6 : ANOVA à mesures répétées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Table d'ANOVA :

$$SC_G = \sum_{k=1}^{q} nr_k(\bar{Y}_{..k} - \bar{Y}_{...})^2$$

$$SC_{Tot,ind} = n\sum_{k=1}^{q}\sum_{i=1}^{r_k}(\bar{Y}_{i.k} - \bar{Y}_{...})^2$$

$$SC_T = m\sum_{j=1}^{n}(\bar{Y}_{.j.} - \bar{Y}_{...})^2$$

$$SC_{GT} = \sum_{j=1}^{n}\sum_{k=1}^{q} r_k(\bar{Y}_{.jk} - \bar{Y}_{...})^2 - SC_T - SC_G$$

$$SC_{Tot,all} = \sum_{k=1}^{q}\sum_{i=1}^{r_k}\sum_{j=1}^{n} r_k(\bar{Y}_{ijk} - \bar{Y}_{...})^2$$

# Méthode 6 : ANOVA à mesures répétées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Table d'ANOVA :

| Source | SC | ddl | CM | F |
|---|---|---|---|---|
| Groupe | $SC_G$ | $q-1$ | $CM_G$ | $F_G = CM_G / CM_{err\ inter}$ |
| Erreur inter | $SC_{Tot,ind} - SC_G$ | $m-q$ | $CM_{err\ inter}$ | |
| Temps | $SC_T$ | $n-1$ | $CM_T$ | $F_T = CM_T / CM_{err\ intra}$ |
| Groupe*Temps | $SC_{GT}$ | $(q-1)(n-1)$ | $CM_{GT}$ | $F_{GT} = CM_{GT} / CM_{err\ intra}$ |
| Erreur intra | $SC_{err\ intra}$ | $(m-q)(n-1)$ | $CM_{err\ intra}$ | |
| Total | $SC_{Tot,all}$ | $n*m-1$ | | |

avec $SC_{err\ intra} = SC_{Tot,all} - SC_{GT} - SC_T - SC_{Tot,ind}$

- les $F$ sont distribués selon loi de Fisher avec les ddl correspondant

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Tests d'hypothèses :

- Parallélisme (interaction groupe*temps)

$$H_0 : (\tau\gamma)_{jk} = 0 \;\; \forall\, j, k$$

$$H_1 : \exists \text{ au moins 1 } (\tau\gamma)_{jk} \neq 0$$

$\Rightarrow$ Rejet de $H_0$ si $F_{GT} > \mathcal{F}_{(q-1)(n-1),(m-q)(n-1),\alpha}$

- Effet temps

$$H_0 : \gamma_j = 0 \;\; \forall\, j$$

$$H_1 : \exists \text{ au moins 1 } \gamma_j \neq 0$$

$\Rightarrow$ Rejet de $H_0$ si $F_T > \mathcal{F}_{(n-1),(m-q)(n-1),\alpha}$

# Méthode 6 : ANOVA à mesures répétées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Effet groupe

$$H_0 : \tau_k = 0 \ \ \forall \ k$$

$$H_1 : \exists \text{ au moins } 1 \ \tau_k \neq 0$$

$\Rightarrow$ Rejet de $H_0$ si $F_G > \mathcal{F}_{(q-1),(m-q),\alpha}$

# Méthode 6 : MANOVA à mesures répétées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

$$Y_{ik} = \mu_k + \varepsilon_{ik} \sim \mathcal{N}(\mu_k, \Sigma = \sigma_b^2 J_n + \sigma_e^2 I_n)$$

où $I_n$ : matrice identité et $J_n$ : matrice carré de 1 de dimension $n * n$

- hypothèse de l'ANOVA sur la matrice $\Sigma$ trop restrictive

$\Rightarrow$ Matrice non structurée

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \sigma_{1n} & \sigma_{n2} & \dots & \sigma_1^2 \end{pmatrix}$$

- MANOVA : Extension de l'ANOVA

# Intérêts et Limites (ANOVA, MANOVA pour mesures répétées)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- nécessite un plan parfaitement équilibré

  $\rightarrow$ sujet observé aux mêmes temps, pas de DM

- dans certains cas : pas de soucis
  - phases précoce I, II, domaine de l'agriculture, ...
  - car degré de contrôle élevé de l'expérience (planification, réalisation,...)

- mais

# Intérêts et Limites (ANOVA, MANOVA pour mesures répétées)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- nécessite un plan parfaitement équilibré

  $\rightarrow$ sujet observé aux mêmes temps, pas de DM

- dans certains cas : pas de soucis
  - phases précoce I, II, domaine de l'agriculture, ...
  - car degré de contrôle élevé de l'expérience (planification, réalisation,...)

- mais en Recherche clinique/ épidémiologique :
  - présence de DM
  - temps et nombre de mesures pas fixés

$\Rightarrow$ ANOVA pas adaptée

# Intérêts et Limites (ANOVA, MANOVA pour mesures répétées)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Matrice de covariance $\Sigma$ de chaque vecteur $Y_i$
  - identique $\forall$ $i$ (indépendamment du groupe)
  - hypothèse classique d'homoscédasticité, mais peut être trop forte
- structure de $\Sigma$ pour ANOVA
  - hypothèse d'indépendance des erreurs intra
  - $\Rightarrow$ non réaliste pour des données longitudinales car corrélation inter-mesure
- structure de $\Sigma$ pour MANOVA : non structurée
  - pas d'hypothèse sur la structure de corrélation
  - mais méthodes peu puissantes
  - $n(n+1)/2$ paramètres supplémentaires à estimer

# Intérêts et Limites (ANOVA, MANOVA pour mesures répétées)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Matrice de covariance $\Sigma$ identique $\forall$ groupe
  - probablement faux
  - $\rightarrow$ ex : peu probable que $Y$ varie de manière identique dans le groupe placebo et traitement

- Inclusion de covariables
  - Fixes : ANCOVA
  - Dépendantes du temps : possible mais compliqué
  - Covariable continues : Non ! ! !

# Intérêts et Limites (ANOVA, MANOVA pour mesures répétées)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Tests et interprétation
  - Test d'hypothèses $\Rightarrow$ rejet ou non de $H_0$
  - $\rightarrow$ évolution moyenne de la variable réponse est-elle $\neq$ selon le groupe ?
- Autres questions d'intérêt
  - Vitesse d'évolution de la variable réponse est-elle $\neq$ selon le groupe ?
  - Prédiction : quel est la trajectoire de la variable réponse pour un homme de 45 ans ?

$\Rightarrow$ Inclusion explicite des temps de mesure dans le modèle

# Plan



1. Introduction
2. Objectifs
3. Méthodes historiques
4. Modèle linéaire mixte
5. Logiciels
6. Stratégie de modélisation
7. Application
8. Discussion
9. Références

# Plan

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

4. Modèle linéaire mixte
Rappels sur le modèle linéaire
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle

# Rappels

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Objectif souvent recherché

- Étudier et établir les relations (potentielles) entre un ensemble de variables

## Situation fréquente

- On considère 1 variable dépendante à expliquer/prédire et des variables indépendantes ou explicatives

## Modèles de régression

- Relation entre une variable dépendante et des variables explicatives

  $\Rightarrow$ Étudier les associations, faire des prévisions

# Exemple (1)

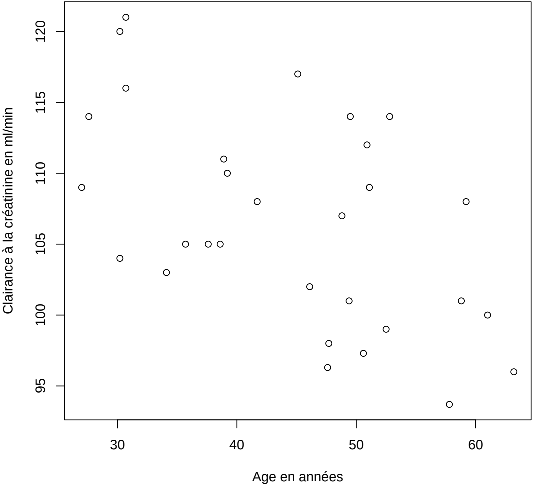
Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Étude de la fonction rénale ($Y$) selon l'âge ($X$)

| Sujet | $Y$ | $X$ | Sujet | $Y$ | $X$ |
|---|---|---|---|---|---|
| 1 | 114.0 | 27.6 | 16 | 96.0 | 63.2 |
| 2 | 110.0 | 39.2 | 17 | 99.0 | 52.5 |
| 3 | 114.0 | 52.8 | 18 | 103.0 | 34.1 |
| 4 | 102.0 | 46.1 | 19 | 98.0 | 47.7 |
| 5 | 101.0 | 49.4 | 20 | 109.0 | 27.0 |
| 6 | 107.0 | 48.8 | 21 | 120.0 | 30.2 |
| 7 | 109.0 | 51.1 | 22 | 111.0 | 38.9 |
| 8 | 117.0 | 45.1 | 23 | 105.0 | 38.6 |
| 9 | 108.0 | 41.7 | 24 | 116.0 | 30.7 |
| 10 | 97.3 | 50.6 | 25 | 96.3 | 47.6 |
| 11 | 101.0 | 58.8 | 26 | 112.0 | 50.9 |
| 12 | 104.0 | 30.2 | 27 | 105.0 | 35.7 |
| 13 | 105.0 | 37.6 | 28 | 93.7 | 57.8 |
| 14 | 121.0 | 30.7 | 29 | 108.0 | 59.2 |
| 15 | 114.0 | 49.5 | 30 | 100.0 | 61.0 |

# Exemple (2)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



# Définition du modèle de régression linéaire

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

On cherche à expliquer les variations de $Y$ par la variable $X$
Si la relation est linéaire, le modèle s'écrit :

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

- $\beta_0$ est l'ordonnée à l'origine (moyenne de $Y_i$ quand $x_i = 0$)
- $\beta_1$ est la pente (changement moyen de $Y_i$ quand $x_i$ augmente d'une unité)
- $\epsilon_i$ est le résidu (différence entre la valeur prédite et celle observée)
- Les résidus sont distribués selon une loi normale de moyenne nulle et de variance $\sigma_\epsilon^2$ (variance résiduelle)

# Estimation de la droite de régression (1)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Objectif : Trouver la meilleure droite pour un nuage de points
- Minimisation des valeurs des résidus
- Critère des Moindres Carrés :

$$CMC = \sum_i (y_i - \widehat{y}_i)^2 = \sum_i (y_i - \beta_0 - \beta_1 x_i)^2$$

- Calcul des dérivées partielles de CMC :

$$\begin{cases} \partial CMC / \partial \beta_0 = -1 \times 2 \times \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) \\ \partial CMC / \partial \beta_1 = -x_i \times 2 \times \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) \end{cases}$$

$$\begin{cases} \partial CMC / \partial \beta_0 = -2 \sum_i y_i + 2n\beta_0 + 2\beta_1 \sum_i x_i \\ \partial CMC / \partial \beta_1 = -2 \sum_i y_i x_i + 2\beta_0 \sum_i x_i + 2\beta_1 \sum_i x_i^2 \end{cases}$$

# Estimation de la droite de régression (2)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Les valeurs optimales, $\widehat{\beta}_0$ et $\widehat{\beta}_1$, minimisent le CMC :

$$\begin{cases} \sum_i y_i - n\widehat{\beta}_0 - \widehat{\beta}_1 \sum_i x_i = 0 \\ \sum_i y_i x_i - \widehat{\beta}_0 \sum_i x_i - \widehat{\beta}_1 \sum_i x_i^2 = 0 \end{cases}$$

- Le CMC est minimum pour :

$$\begin{cases} \widehat{\beta}_1 = \dfrac{\sum_i x_i y_i - (\sum_i x_i)(\sum_i y_i)/n}{\sum_i x_i^2 - (\sum_i x_i)^2/n} \\ \widehat{\beta}_0 = (\sum_i y_i)/n - \widehat{\beta}_1 (\sum_i y_i)/n \end{cases}$$

# Estimation des autres paramètres importants (1)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Variance résiduelle

$$\widehat{\sigma}^2 = \frac{\sum_i (y_i - \widehat{y}_i)^2}{n-2} = \frac{\sum_i (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)^2}{n-2}$$

# Estimation des autres paramètres importants

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- $R^2$

On cherche à présent à expliquer les variations de $X$ par la variable $Y$

Si la relation est linéaire, le modèle s'écrit :

$$X_i = \beta'_0 + \beta'_1 Y_i + \epsilon_i$$

$\Rightarrow$ Raisonnement similaire pour estimer $\beta'_0$ et $\beta'_1$

$$\begin{cases} \widehat{\beta'}_1 = \dfrac{\sum_i x_i y_i - (\sum_i x_i)(\sum_i y_i)/n}{\sum_i y_i^2 - (\sum_i y_i)^2/n} \\ \widehat{\beta'}_0 = (\sum_i x_i)/n - \widehat{\beta}_1(\sum_i x_i)/n \end{cases}$$

# Estimation des autres paramètres importants

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Si $\widehat{\beta}_1$ représente la pente de $Y$ en fonction de $X$ et que $\widehat{\beta}'_1$ représente la pente de $X$ en fonction de $Y$, alors on montre que :

$$\widehat{\beta}_1 \widehat{\beta}'_1 = \frac{\sum_i x_i.y_i - \frac{\sum_i x_i \sum_i y_i}{n}}{\sum_i x_i^2 - \frac{(\sum_i x_i)^2}{n}} \times \frac{\sum_i x_i.y_i - \frac{\sum_i x_i \sum_i y_i}{n}}{\sum_i y_i^2 - \frac{(\sum_i y_i)^2}{n}}$$

$$= \frac{\left(\sum_i x_i.y_i - \frac{\sum_i x_i \sum_i y_i}{n}\right)^2}{\left(\sum_i x_i^2 - \frac{(\sum_i x_i)^2}{n}\right) \times \left(\sum_i y_i^2 - \frac{(\sum_i y_i)^2}{n}\right)} = r^2$$

$\Rightarrow$ $r^2$ représente la proportion de variation de $Y$ expliquée par $X$
($r$ est le coefficient de corrélation linéaire)

# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Étude de la fonction rénale ($Y$) en fonction de l'âge ($X$)

- Données

$$n = 30$$

$$\sum_i y_i = 3196.3$$

$$\sum_i y_i^2 = 342125.7$$

$$\sum_i x_i = 1334.3$$

$$\sum_i x_i^2 = 62626.5$$

$$\sum_i x_i y_i = 140943.0$$

# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- $\widehat{\beta}_0 = 123.0$

$\Rightarrow$ La fonction rénale d'un nouveau né ($x = 0$) est estimée à 123.0 $ml/min$ en moyenne. Attention : cette valeur n'est pas fiable (aucun enfant dans l'étude)

- $\widehat{\beta}_1 = -0.37$

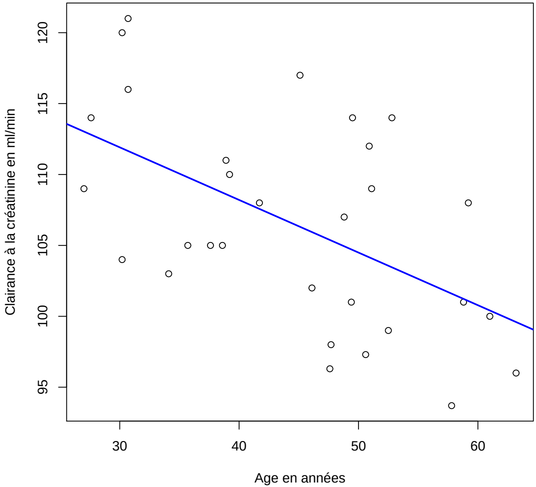
$\Rightarrow$ La fonction rénale chute en moyenne de 3.7 $ml/min$ tous les 10 ans

- $\widehat{\sigma}^2 = 6.35$

- $\widehat{r}^2 = 0.29$

$\Rightarrow$ 29% de la variation de $Y$ est expliquée par $X$

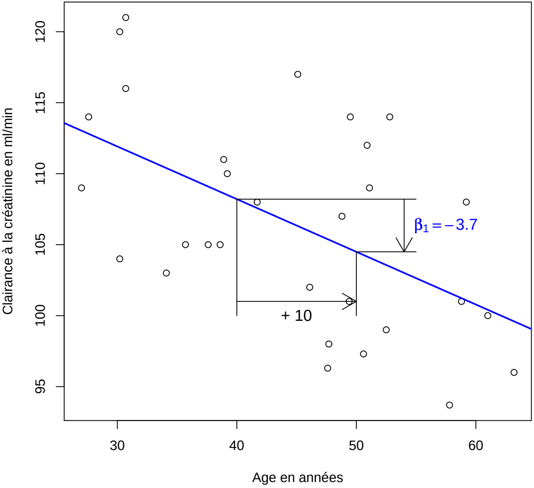
# Exemple




Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

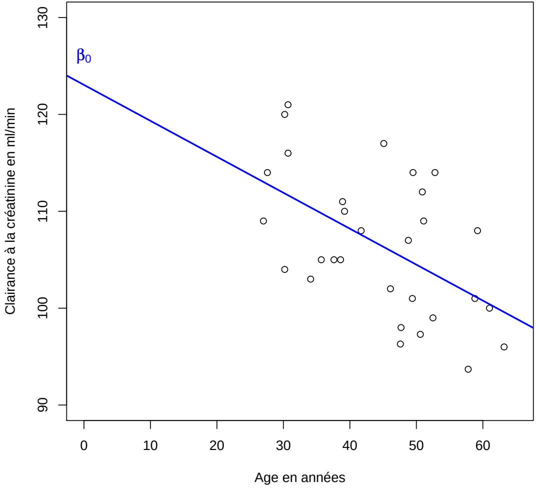
# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



# Intervalle de confiance de $\beta_1$

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

$$IC_{(1-\alpha)} = \left[\widehat{\beta}_1 \pm t_{\alpha,n-2} s(\widehat{\beta}_1)\right]$$

- $t_{\alpha,n-2}$ : fractile de la loi de Student à $n-2$ ddl
- $s(\widehat{\beta}_1)$ : écart-type estimé de la pente
  - $s(\widehat{\beta}_1) = \widehat{\sigma}/(\widehat{\sigma}_x\sqrt{n-1})$, où $\widehat{\sigma}_x$ est l'écart-type de $X$
- Si $0 \in IC_{(1-\alpha)}$ : nous ne pouvons pas conclure que la pente est significativement $\neq$ de 0
- Si $0 \notin IC_{(1-\alpha)}$ : on conclura que la pente est significativement différente de 0

# Test de $\beta_1$ (méthode 1 : Test de Student)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Définition des hypothèses

- $H_0 : \beta_1 = 0$
- $H_1 : \beta_1 \neq 0$

$\rightarrow$ Statistique de test : $T = \beta_1 / s(\beta_1) \sim T_{n-2\ ddl}$

$\rightarrow$ Définition de la région critique

$\Rightarrow$ Si $t$ appartient à la région critique, on rejette $H_0$, sinon on ne peut pas rejeter $H_0$ au seuil $\alpha$

# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Étude de la fonction rénale ($Y$) en fonction de l'âge ($X$)

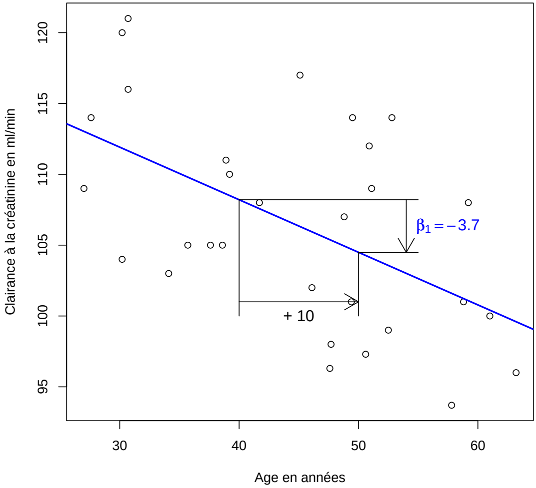
- Intervalle de confiance à 95% de $\beta_1$ ($t_{5\%;28} = 2.048$) :

$$IC_{95\%} = [-0.37 \pm 2.048 \times 0.11] = [-0.48; -0.26]$$

- L'intervalle de confiance ne comprend pas la valeur 0, il semble donc que la pente soit significativement différente de zéro
- Test $\rightarrow H_0 : \beta_1 = 0$ contre $H_1 : \beta_1 \neq 0$
- $\alpha = 0.05$, $t_{5\%;28} = 2.048$
- $t = -0.37/0.11 = -3.35$
- $|t| > 2.048$, on rejette $H_0$

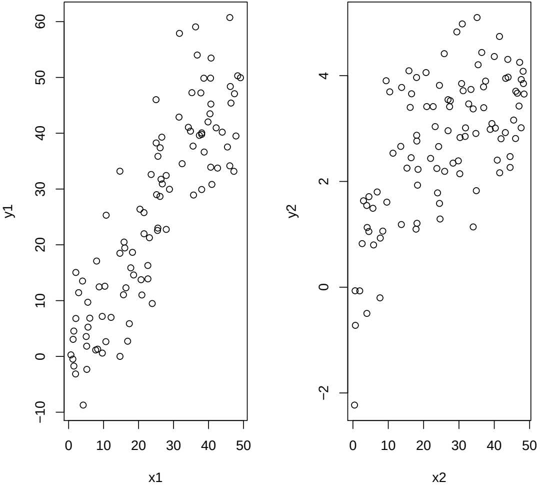
# Linéarité

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



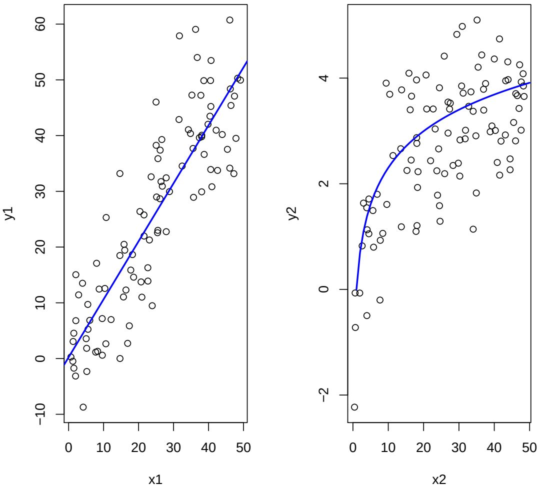
# Linéarité

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



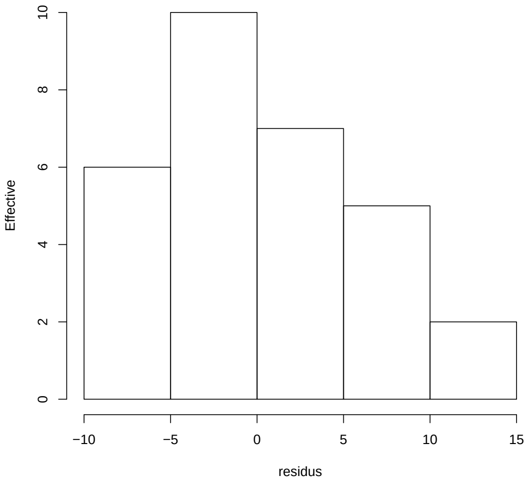
# Linéarité

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



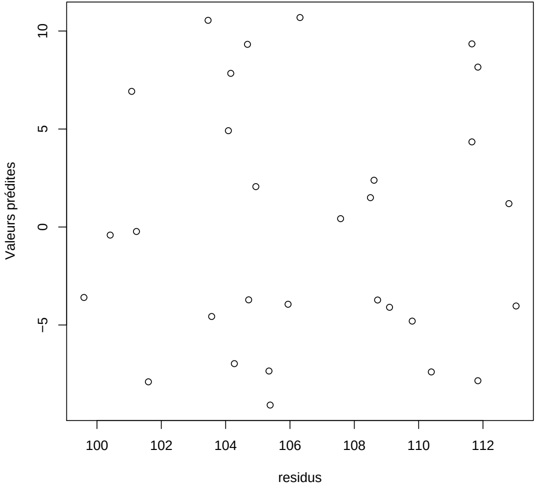
# Normalité des résidus

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



# Homoscédasticité des résidus

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



# Plan



4. Modèle linéaire mixte
- Rappels sur le modèle linéaire
- Modèle linéaire mixte
- Inclusion de covariables
- Exercices
- Généralisation
- Estimation
- Tests d'inférence
- Adéquation du modèle

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Objectif :

Étudier l'évolution de la variable réponse $Y$ en fonction du temps $t$

## Notations :

- $Y_{ij}$ : variable réponse/dépendante quantitative gaussienne

  avec $i = 1, ..., N$ individus

  et $j = 1, ..., n_i$ mesures
- $t_{ij}$ : temps
- $X_i$ : variable explicative invariante avec le temps

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 1 Modèle linéaire simple :

$$\begin{cases} Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 X_i + \epsilon_{ij} \\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2) \\ \epsilon_{ij} \perp \epsilon_{ij'} \end{cases}$$

Remarque : $\perp$ : indépendant de

## Problèmes :

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 1 Modèle linéaire simple :

$$\begin{cases} Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 X_i + \epsilon_{ij} \\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2) \\ \epsilon_{ij} \perp \epsilon_{ij'} \end{cases}$$

Remarque : $\perp$ : indépendant de

Problèmes : Mesures répétées = mesures corrélées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 1 Modèle linéaire simple :

$$\begin{cases} Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 X_i + \epsilon_{ij} \\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2) \\ \epsilon_{ij} \perp \epsilon_{ij'} \end{cases}$$

Remarque : $\perp$ : indépendant de

Problèmes : Mesures répétées = mesures corrélées

- $\widehat{\beta}$ moins efficace (grande variance)
- $\widehat{Var}(\widehat{\beta})$ biaisées
- Tests des $\beta$ biaisées

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 1 solution naïve :

Considérer l'effet sujet au travers d'un intercept spécifique aux sujets :

$$\begin{cases} Y_{ij} = \beta_{0i} + \beta_1 t_{ij} + \beta_2 X_i + \epsilon_{ij} \\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2) \\ \epsilon_{ij} \perp \epsilon_{ij'} \end{cases}$$

## Problèmes :

- $N$ paramètres supplémentaires
- $\beta_2$ non identifiable
- Conclusions difficilement généralisables à d'autres sujets

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 1 solution naïve :

Considérer l'effet sujet au travers d'un intercept et d'une pente spécifiques aux sujets :

$$\begin{cases} Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \beta_2 X_i + \epsilon_{ij} \\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2) \\ \epsilon_{ij} \perp \epsilon_{ij'} \end{cases}$$

## Problèmes :

- $2N$ paramètres supplémentaires
- $\beta_2$ toujours non identifiable
- Conclusions difficilement généralisables à d'autres sujets

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Solution 1 : Suppression $X_i$

$$\begin{cases} Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \epsilon_{ij} \\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2) \\ \epsilon_{ij} \perp \epsilon_{ij'} \end{cases}$$

- interprétation épidémiologique limitée

  $\rightarrow$ Étude de l'association entre variables explicatives et variable dépendante

$\Rightarrow$ Solution non souhaitée

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Solution 2 : Analyse en 2 temps

1. Estimation du modèle pour chaque sujet $i$

   c'est-à-dire pour les $n_i$ observations du sujet $i$

$$Y_{ij} = \widehat{\beta}_{0i} + \widehat{\beta}_{1i} t_{ij} + \epsilon_{ij}$$

2. Analyse les $\widehat{\beta}_{0i}$ et $\widehat{\beta}_{1i}$ en fonction $X_i$

$$\widehat{\beta}_{0i} = \gamma_0 + \gamma_1 X_i + e_i$$
$$\widehat{\beta}_{1i} = \alpha_0 + \alpha_1 X_i + e_i$$

## Problèmes :

- Imprécision sur $\widehat{\beta}_{0i}$ et $\widehat{\beta}_{1i}$
- $Var(\widehat{\beta}_i)$ négligée dans la seconde étape
- Pas de variables explicatives dépendantes du temps
- perte d'information

# Modèle linéaire mixte

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- introduit par Harville (1977)
- popularisé par Laird et Ware (1982)
- aujourd'hui : méthode de référence pour l'étude d'un marqueur gaussien au cours du temps
  - considère la corrélation des observations dans un contexte de données répétées
  - permet de décrire l'évolution moyenne au cours du temps
  - permet de décrire les évolutions individuelles à l'aide d'effets aléatoires mesurant écart de chacun des individus à l'évolution moyenne
  - permet d'étudier associations entre facteurs de risque et évolution

# Modèle linéaire mixte

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Idée

- variable réponse suit un modèle de régression linéaire
- avec des coefficients de régression spécifique à chaque sujet

$\Rightarrow$ Modèle linéaire mixte = Modèle linéaire à effets aléatoires

# Modèle linéaire à intercept aléatoire

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

Soit $Y_{ij}$ la variable réponse du sujet $i$ ($i = 1, ..., k$) pour le temps $j$ ($j = 1, ..., n_i$) et $N = \sum_{i=1}^{k}$

$$
\begin{cases}
Y_{ij} = (\beta_0 + \gamma_{0i}) + \beta_1 t_{ij} + \beta_2 X_i + \epsilon_{ij} \\
\gamma_{0i} \sim \mathcal{N}(0, \sigma_0^2) \\
\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2) \\
\gamma_{0i} \perp \epsilon_{ij} \quad \forall\, i, j \\
\epsilon_{ij'} \perp \epsilon_{i'j'} \quad \forall\, i, j, i', j'
\end{cases}
$$

## Intérêt

- un seul paramètre supplémentaire à estimer $\sigma_0^2$
- $\beta_2$ identifiable

# Modèle linéaire à intercept et pente aléatoires

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

$$\begin{cases} Y_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 X_i + \epsilon_{ij} \\ \gamma_{0i} \sim \mathcal{N}(0, \sigma_0^2) \\ \gamma_{1i} \sim \mathcal{N}(0, \sigma_1^2) \\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2) \\ \gamma_{0i} \perp \epsilon_{ij} \quad \forall\ i, j \\ \epsilon_{ij'} \perp \epsilon_{i'j'} \quad \forall\ i, j, i', j' \end{cases}$$

- $\beta_0$, $\beta_1$ : effets fixes
- $\gamma_{0i}$, $\gamma_{1i}$ : effets aléatoires

# Modèle linéaire à intercept et pente aléatoires

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Modèle linéaire à intercept et pente aléatoires s'écrit également :

$$\begin{cases} Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \beta_2 X_i + \epsilon_{ij} \\ \beta_{0i} \sim \mathcal{N}(\beta_0, \sigma_0^2) \\ \beta_{1i} \sim \mathcal{N}(\beta_1, \sigma_1^2) \\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2) \end{cases}$$

avec $\beta_{0i} = \beta_0 + \gamma_{0i}$
et $\beta_{1i} = \beta_1 + \gamma_{1i}$

# Plan

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 4. Modèle linéaire mixte

Rappels sur le modèle linéaire
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle

# Approche stratifiée

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 1. Variables explicatives qualitatives/catégorielles

- Estimation d'un modèle linéaire mixte pour chaque sous-groupe de sujets défini par les modalités de la variable explicative

$$Y_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \epsilon_{ij}$$

# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Étude de l'évolution de la clairance à la créatinine chez des patients transplantés rénaux en fonction du genre du donneur

- Échantillon de 925 patients
- Chez 302 femmes

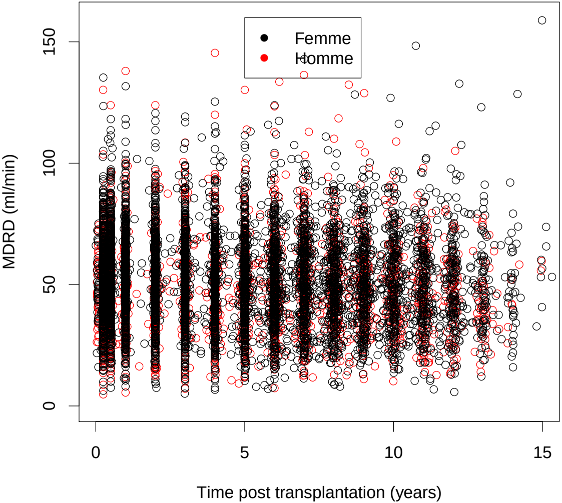
$$MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \epsilon_{ij}$$

- Chez 623 hommes

$$MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \epsilon_{ij}$$

# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Chez les femmes

```
Linear mixed-effects model fit by maximum likelihood
 Data: dataFemme
       AIC      BIC    logLik
  22390.78 22426.84 -11189.39

Random effects:
 Formula: ~1 + tpsbio | repere
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 15.740353 (Intr)
tpsbio       2.115033 -0.205
Residual     7.565436

Fixed effects: MDRD ~ tpsbio
               Value Std.Error   DF  t-value p-value
(Intercept) 49.11550 0.9379574 2707 52.36432       0
tpsbio      -0.59706 0.1424307 2707 -4.19197       0
 Correlation:
       (Intr)
tpsbio -0.252

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-4.17427558 -0.50641145 -0.01711828  0.50348803  6.03585779

Number of Observations: 3010
Number of Groups: 302
```

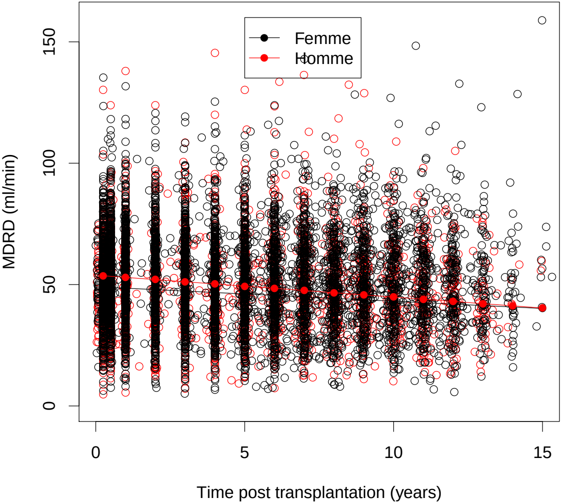
Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Chez les hommes

```
Linear mixed-effects model fit by maximum likelihood
 Data: dataHomme
       AIC      BIC    logLik
  48738.49 48779.08 -24363.25

Random effects:
 Formula: ~1 + tpsbio | repere
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 15.949301 (Intr)
tpsbio       1.846495 -0.179
Residual     8.492231

Fixed effects: MDRD ~ tpsbio
               Value Std.Error   DF   t-value p-value
(Intercept) 53.85869 0.6649845 5779  80.99239       0
tpsbio      -0.90101 0.0885983 5779 -10.16963       0
 Correlation:
       (Intr)
tpsbio -0.243

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-12.08955217  -0.48624769  -0.00947898   0.48177898   6.59969828

Number of Observations: 6403
Number of Groups: 623
```

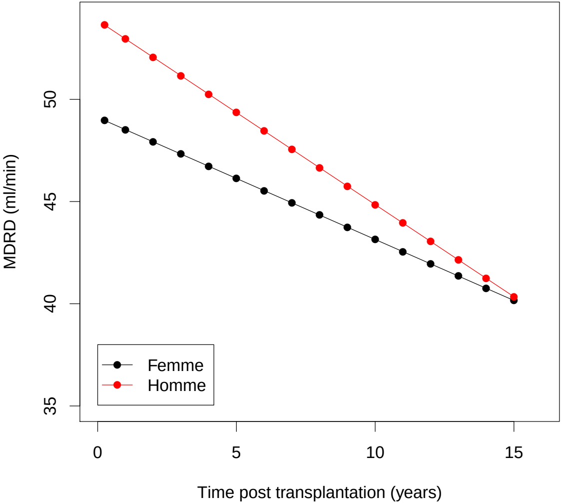
# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



# Approche stratifiée

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Problèmes

# Approche stratifiée

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Problèmes

- une seule covariable
- pas de test formel de la signification de la covariable
- pas de covariable quantitative continue
- perte de puissance

# Approche ajustée

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 2 covariables

$$Y_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_{ij}$$

## Interprétation des paramètres :

# Approche ajustée

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 2 covariables

$$Y_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_{ij}$$

### Interprétation des paramètres :

- $\beta_0$ : intercept (niveau moyen du marqueur pour $t_{ij} = 0$ et $X_{2i} = X_{3i} = 0$)
- $\beta_1$ : augmentation/diminution moyenne pour 1 unité de temps
- $\beta_0 + \beta_2$ : niveau initial moyen du marqueur chez les sujets $X_{2i} = 1$ et $X_{3i} = 0$
- $\beta_0 + \beta_3$ : niveau initial moyen du marqueur chez les sujets $X_{2i} = 0$ et $X_{3i} = 1$
- $\beta_0 + \beta_2 + \beta_3$ : niveau initial moyen du marqueur chez les sujets $X_{2i} = 1$ et $X_{3i} = 1$

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

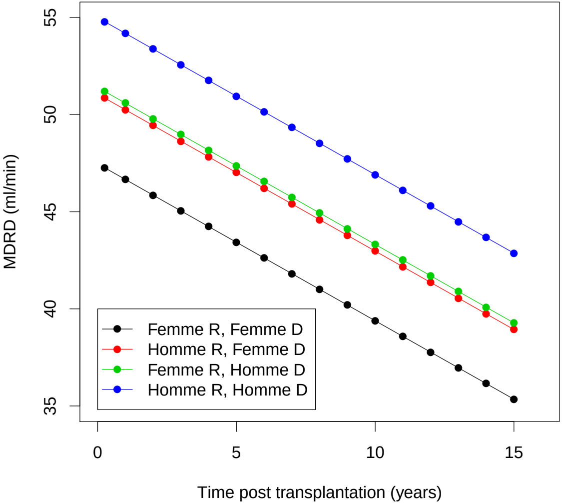
## Évolution de MDRD en fonction du genre du receveur et du donneur

$$MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 SexeR_i + \beta_3 SexeD_i + \epsilon_{ij}$$

```
Linear mixed-effects model fit by maximum likelihood
Data: dataPDR

Fixed effects: MDRD ~ tpsbio + sexe + sexeD
                Value Std.Error   DF   t-value p-value
(Intercept) 47.47009 1.1130009 8487  42.65054   0e+00
tpsbio      -0.80833 0.0759506 8487 -10.64279   0e+00
sexe         3.58957 1.0765828  922   3.33423   9e-04
sexeD        3.93341 1.1175563  922   3.51965   5e-04
 Correlation:
       (Intr) tpsbio sexe
tpsbio -0.123
sexe   -0.555  0.001
sexeD  -0.641 -0.003 -0.060

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-12.51673113  -0.48986881  -0.01560432   0.48556782   6.86849025

Number of Observations: 9413
Number of Groups: 925
```

# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



# Approche ajustée

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 2 covariables + 1 interaction

$$Y_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + \epsilon_{ij}$$

Interprétation des paramètres :

- $\beta_0$ : intercept (niveau moyen du marqueur pour $t_{ij} = 0$ et $X_{2i} = X_{3i} = 0$)
- $\beta_1$ : augmentation/diminution moyenne pour 1 unité de temps
- $\beta_0 + \beta_2$ : niveau initial moyen du marqueur chez les sujets $X_{2i} = 1$ et $X_{3i} = 0$
- $\beta_0 + \beta_3$ : niveau initial moyen du marqueur chez les sujets $X_{2i} = 0$ et $X_{3i} = 1$
- $\beta_0 + \beta_2 + \beta_3 + \beta_4$ : niveau initial moyen du marqueur chez les sujets $X_{2i} = 1$ et $X_{3i} = 1$

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Évolution de MDRD en fonction du genre du receveur et du donneur avec 1 interaction

$$MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 SexeR_i + \beta_3 SexeD_i + \beta_4 SexeR_i SexeD_i + \epsilon_{ij}$$

```
Linear mixed-effects model fit by maximum likelihood
Data: dataPDR

Fixed effects: MDRD ~ tpsbio + sexe + sexeD + sexe * sexeD
               Value Std.Error   DF   t-value p-value
(Intercept) 47.65894 1.4088872 8487  33.82736  0.0000
tpsbio      -0.80821 0.0759533 8487 -10.64089  0.0000
sexe         3.26099 1.8520421  921   1.76075  0.0786
sexeD        3.63695 1.7548845  921   2.07247  0.0385
sexe:sexeD   0.49716 2.2761202  921   0.21842  0.8271
 Correlation:
            (Intr) tpsbio sexe   sexeD
tpsbio      -0.096
sexe        -0.754 -0.001
sexeD       -0.795 -0.003  0.605
sexe:sexeD   0.613  0.002 -0.814 -0.771

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-12.51649105  -0.48971387  -0.01580078   0.48543137   6.86881756

Number of Observations: 9413
Number of Groups: 925
```

# Approche ajustée

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 1 covariable + 1 interaction avec le temps

$$Y_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 X_{2i} + \beta_3 X_{2i} t_{ij} + \epsilon_{ij}$$

### Interprétation des paramètres :

- $\beta_0$ : intercept (niveau moyen du marqueur pour $t_{ij} = 0$ et $X_{2i} = 0$)
- $\beta_1$ : augmentation/diminution moyenne pour 1 unité de temps chez les sujets $X_{2i} = 0$
- $\beta_0 + \beta_2$ : niveau initial moyen du marqueur chez les sujets $X_{2i} = 1$
- $\beta_1 + \beta_3$ : : augmentation/diminution moyenne pour 1 unité de temps chez les sujets $X_{2i} = 1$

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Évolution de MDRD en fonction du genre du donneur en interaction avec le temps

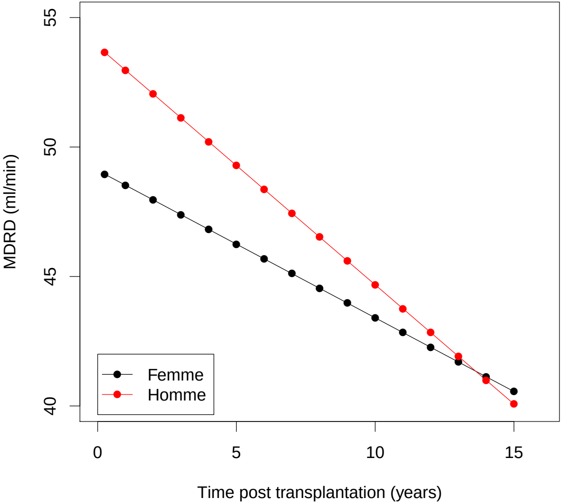
$$MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 SexeD_i + \beta_3 SexeD_i t_{ij} + \epsilon_{ij}$$

```
Linear mixed-effects model fit by maximum likelihood
Data: dataPDR

Fixed effects: MDRD ~ tpsbio + sexeD + sexeD * tpsbio
                  Value Std.Error   DF  t-value p-value
(Intercept)   49.09557 0.9514397 8486 51.60134  0.0000
tpsbio        -0.56913 0.1344378 8486 -4.23341  0.0000
sexeD          4.79563 1.1587521  923  4.13861  0.0000
tpsbio:sexeD  -0.35207 0.1627760 8486 -2.16293  0.0306
 Correlation:
             (Intr) tpsbio sexeD
tpsbio       -0.250
sexeD        -0.821  0.205
tpsbio:sexeD  0.207 -0.826 -0.250

Standardized Within-Group Residuals:
        Min          Q1          Med           Q3          Max
-12.52157590  -0.48981767  -0.01406612   0.48610603   6.86450258

Number of Observations: 9413
Number of Groups: 925
```

# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
- Rappels
- Modèle linéaire mixte
- Inclusion de covariables
- Exercices
- Généralisation
- Estimation
- Tests d'inférence
- Adéquation du modèle

Logiciels
Stratégie de modélisation
Application
Discussion
Références



Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 2. Variables explicatives quantitatives

$$Y_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 X_{2i} + \beta_3 X_{2i} t_{ij} + \epsilon_{ij}$$

### Interprétation des paramètres :

- $\beta_0$ : intercept (niveau moyen du marqueur pour $t_{ij} = 0$ et $X_{2i} = 0$)
- $\beta_1$ : augmentation/diminution moyenne pour 1 unité de temps chez les sujets $X_{2i} = 0$
- $\beta_2$ : variation du niveau initial moyen du marqueur pour 1 augmentation de 1 unité de $X_{2i}$
- $\beta_3$ : variation de l'augmentation/diminution moyenne pour 1 unité de temps et pour une augmentation de 1 unité de $X_{2i}$

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Évolution de MDRD en fonction de l'âge du donneur en interaction avec le temps

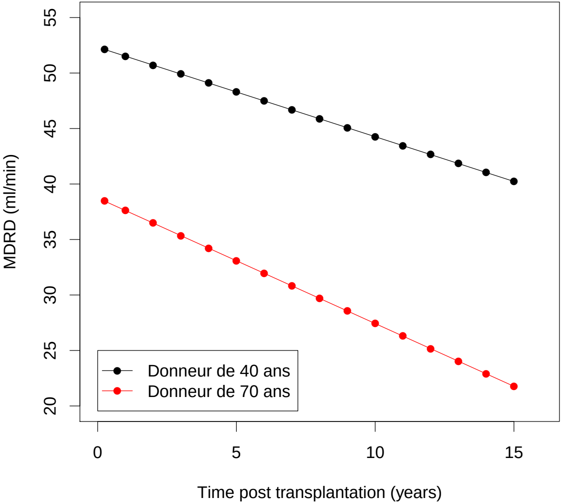
$$MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 AgeD_i + \beta_3 AgeD_i t_{ij} + \epsilon_{ij}$$

```
Linear mixed-effects model fit by maximum likelihood
Data: dataPDR

Fixed effects: MDRD ~ tpsbio + ageD + ageD * tpsbio
                Value Std.Error   DF   t-value p-value
(Intercept) 70.43808 1.4002425 8464  50.30420  0.0000
tpsbio      -0.37258 0.2106563 8464  -1.76869  0.0770
ageD        -0.45279 0.0326921  922 -13.85004  0.0000
tpsbio:ageD -0.01084 0.0050057 8464  -2.16622  0.0303
 Correlation:
            (Intr) tpsbio ageD
tpsbio      -0.324
ageD        -0.934  0.306
tpsbio:ageD  0.300 -0.933 -0.325

Standardized Within-Group Residuals:
          Min             Q1            Med             Q3            Max
-12.511476729   -0.488246327   -0.009175349    0.488296117    6.863599839

Number of Observations: 9390
Number of Groups: 924
```

Remarque : moyenne AgeD=39 ans

# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



# Notations matricielles

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Cas d'un modèle avec 1 covariable + 1 interaction avec le temps

$$
\begin{aligned}
Y_{ij} &= (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 X_{2i} + \beta_3 X_{2i} t_{ij} + \epsilon_{ij} \\
&= \beta_0 + \beta_1 t_{ij} + \beta_2 X_{2i} + \beta_3 X_{2i} t_{ij} + \gamma_{0i} + \gamma_{1i} t_{ij} + \epsilon_{ij} \\
&= (1, t_{ij}, X_{2i}, X_{2i} t_{ij}) \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + (1, t_{ij}) \begin{pmatrix} \gamma_{0i} \\ \gamma_{1i} \end{pmatrix} + \epsilon_{ij} \\
&= X_{ij}^T \beta + Z_{ij}^T \gamma_i + \epsilon_{ij}
\end{aligned}
$$

# Notations matricielles

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

Pour l'ensemble des mesures $j$, on a le vecteur $Y_i = (Y_{i1}, Y_{i2}, ..., Y_{in_i})^T$ :

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \dots \\ Y_{in_i} \end{pmatrix} = \begin{pmatrix} 1, t_{i1}, X_{2i}, X_{2i}t_{i1} \\ 1, t_{i2}, X_{2i}, X_{2i}t_{i2} \\ \dots \\ 1, t_{in_i}, X_{2i}, X_{2i}t_{in_i} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} 1, t_{i1} \\ 1, t_{i2} \\ \dots \\ 1, t_{in_i} \end{pmatrix} \begin{pmatrix} \gamma_{0i} \\ \gamma_{1i} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \dots \\ \epsilon_{in_i} \end{pmatrix}$$

$$Y_i = X_i\beta + Z_i\gamma_i + \epsilon_i$$

# Notations matricielles

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarques :

- Formulation matricielle utilisée pour l'estimation
- $\epsilon_i = \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \dots \\ \epsilon_{in_i} \end{pmatrix} \sim MVN(0, \sigma_\epsilon^2 I_{n_i})$

  $I_{n_i}$ matrice identité
- les dimensions des $\neq$ matrices ou vecteurs

  $X_i$ : dimension $(n_i * 4)$
  $\beta$ : dimension $(4 * 1)$
  $Z_i$ : dimension $(n_i * 2)$
  $\epsilon_i$ : dimension $(2 * 1)$
- $Z_i \subset X_i$

$\Rightarrow$ Modèle multivariée complexe envisageable

# En résumé

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Modèle

$$\begin{cases} Y_{ij} = X_{ij}^T \beta + Z_{ij}^T \gamma_i + \epsilon_{ij} \\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2) \\ \gamma_i \sim \mathcal{N}(0, B) \end{cases}$$

## Forme vectorielle

$$\begin{cases} Y_i = X_i \beta + Z_i \gamma_i + \epsilon_i \\ \epsilon_i \sim \mathcal{MVN}(0, \sigma_\epsilon^2 I_{n_i}) \\ \gamma_i \sim \mathcal{N}(0, B) \end{cases}$$

# En résumé

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Forme marginale

$$\begin{cases} Y_i = X_i\beta + \varepsilon_i \\ \varepsilon_i \sim \mathcal{MVN}(0, Z_iBZ_i^T + \sigma_\epsilon^2 I_{n_i}) \end{cases}$$

avec $\varepsilon_i = Z_i\gamma_i + \epsilon_i$

- $E(\varepsilon_i) =$
- $V(\varepsilon_i) =$

# En résumé

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Forme marginale

$$\begin{cases} Y_i = X_i\beta + \varepsilon_i \\ \varepsilon_i \sim \mathcal{MVN}(0, Z_iBZ_i^T + \sigma_\epsilon^2 I_{n_i}) \end{cases}$$

avec $\varepsilon_i = Z_i\gamma_i + \epsilon_i$

- $E(\varepsilon_i) = E(Z_i\gamma_i + \epsilon_i) = Z_iE(\gamma_i) + E(\epsilon_i) = 0$
- $V(\varepsilon_i) =$

# En résumé

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Forme marginale

$$\begin{cases} Y_i = X_i\beta + \varepsilon_i \\ \varepsilon_i \sim \mathcal{MVN}(0, Z_iBZ_i^T + \sigma_\epsilon^2 I_{n_i}) \end{cases}$$

avec $\varepsilon_i = Z_i\gamma_i + \epsilon_i$

- $E(\varepsilon_i) = E(Z_i\gamma_i + \epsilon_i) = Z_iE(\gamma_i) + E(\epsilon_i) = 0$
- $V(\varepsilon_i) = V(Z_i\gamma_i + \epsilon_i) = V(Z_i\gamma_i) + V(\epsilon_i) = Z_iV(\gamma_i)Z_i^T + \sigma_\epsilon^2 I_{n_i} = Z_iBZ_i^T + \sigma_\epsilon^2 I_{n_i}$

  Remarque : $\gamma_i \perp \epsilon_{ij} \quad \forall\, i, j \Rightarrow Cov(\gamma_i, \epsilon_i) = 0$

# En résumé

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Forme marginale

$$\begin{cases} Y_i = X_i\beta + \varepsilon_i \\ \varepsilon_i \sim \mathcal{MVN}(0, Z_iBZ_i^T + \sigma_\epsilon^2 I_{n_i}) \end{cases}$$

avec $\varepsilon_i = Z_i\gamma_i + \epsilon_i$

- $E(\varepsilon_i) = E(Z_i\gamma_i + \epsilon_i) = Z_iE(\gamma_i) + E(\epsilon_i) = 0$
- $V(\varepsilon_i) = V(Z_i\gamma_i + \epsilon_i) = V(Z_i\gamma_i) + V(\epsilon_i) = Z_iV(\gamma_i)Z_i^T + \sigma_\epsilon^2 I_{n_i} = Z_iBZ_i^T + \sigma_\epsilon^2 I_{n_i}$

  Remarque : $\gamma_i \perp \epsilon_{ij} \;\; \forall\, i,j \Rightarrow Cov(\gamma_i, \epsilon_i) = 0$

$$\Rightarrow Y_i \sim \mathcal{MVN}(X_i\beta, Z_iBZ_i^T + \sigma_\epsilon^2 I_{n_i})$$

Hypothèse d'indépendance des réponses entre sujets (groupes) : $Y_i \perp Y_{i'}$

# Plan

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
- Rappels
- Modèle linéaire mixte
- Inclusion de covariables
- Exercices
- Généralisation
- Estimation
- Tests d'inférence
- Adéquation du modèle

Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 4. Modèle linéaire mixte

Rappels sur le modèle linéaire
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle

# Exercice 1

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Écrire le modèle linéaire à intercept aléatoire
- Calculer $E(Y_{ij})$, $E(Y_{ij}|\gamma_{0i})$, $Var(Y_{ij})$ et $Var(Y_{ij}|\gamma_{0i})$
- Calculer le coefficient de corrélation de Pearson $Corr(Y_{ij}, Y_{ij'})$
- Calculer le coefficient de corrélation intra-classe pour ce modèle :

$$\rho = \frac{Var[E(Y_{ij}|\gamma_{0i})]}{Var(Y_{ij})}$$

Remarque : $Var(Y_{ij}) = E[Var(Y_{ij}|\gamma_{0i})] + Var[E(Y_{ij}|\gamma_{0i})]$

# Exercice 1

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

$$Y_{ij} = (\beta_0 + \gamma_{0i}) + \beta_1 t_{ij} + \beta_2 X_i + \epsilon_{ij}$$

- Espérance populationnelle

$$E(Y_{ij}) = E(X_{ij}^T\beta + \gamma_{0i} + \epsilon_{ij}) = E(X_{ij}^T\beta) + E(\gamma_{0i}) + E(\epsilon_{ij}) = X_{ij}^T\beta$$

- Espérance spécifique au sujet

$$E(Y_{ij}|\gamma_{0i}) = X_{ij}^T\beta + \gamma_{0i}$$

- Variance populationnelle

$$Var(Y_{ij}) = V(X_{ij}^T\beta + \gamma_{0i} + \epsilon_{ij}) = V(X_{ij}^T\beta) + \sigma_0^2 + \sigma_\epsilon^2 = \sigma_0^2 + \sigma_\epsilon^2$$

- Variance spécifique au sujet

$$Var(Y_{ij}|\gamma_{0i}) = \sigma_\epsilon^2$$

# Exercice 1

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Coefficient de corrélation de Pearson

$$Corr(Y_{ij}, Y_{ij'}) = \rho_{ijj'} = \frac{Cov(Y_{ij}, Y_{ij'})}{\sqrt{V(Y_{ij})V(Y_{ij'})}}$$

or

$$\begin{aligned}
Cov(Y_{ij}, Y_{ij'}) &= E[(Y_{ij} - E(Y_{ij}))(Y_{ij'} - E(Y_{ij'}))] \\
&= E[(\gamma_{0i} + \epsilon_{ij})(\gamma_{0i} + \epsilon_{ij'})] \\
&= E(\gamma_{0i}^2 + \gamma_{0i}\epsilon_{ij'} + \epsilon_{ij}\gamma_{0i} + \epsilon_{ij}\epsilon_{ij'}) \\
&= E(\gamma_{0i}^2) + E(\gamma_{0i}\epsilon_{ij'}) + E(\epsilon_{ij}\gamma_{0i}) + E(\epsilon_{ij}\epsilon_{ij'}) \\
&= E((\gamma_{0i} - E(\gamma_{0i}))^2) \\
&\quad + E((\gamma_{0i} - E(\gamma_{0i}))(\epsilon_{ij'} - E(\epsilon_{ij'}))) \\
&\quad + E((\gamma_{0i} - E(\gamma_{0i}))(\epsilon_{ij} - E(\epsilon_{ij}))) \\
&\quad + E((\epsilon_{ij} - E(\epsilon_{ij}))(\epsilon_{ij'} - E(\epsilon_{ij'}))) \\
&= V(\gamma_{0i}) + Cov(\gamma_{0i}, \epsilon_{ij'}) + Cov(\gamma_{0i}, \epsilon_{ij}) + Cov(\epsilon_{ij}, \epsilon_{ij'}) \\
&= V(\gamma_{0i}) = \sigma_0^2
\end{aligned}$$

# Exercice 1

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

donc

$$Corr(Y_{ij}, Y_{ij'}) = \rho_{ijj'} = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\epsilon^2}$$

- Coefficient de corrélation intra-classe

$$\rho = \frac{V[E(Y_{ij}|\gamma_{0i})]}{V(Y_{ij})} = \frac{\text{Var. inter-sujet}}{\text{Var . totale}}$$
$$= \frac{V(X_{ij}^T\beta + \gamma_{0i})}{\sigma_0^2 + \sigma_\epsilon^2} = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\epsilon^2}$$
$$= Corr(Y_{ij}, Y_{ij'})$$

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarques

- $\forall i, j, j', Corr(Y_{ij}, Y_{ij'})$ constant
- Modèle linéaire à intercept aléatoire = Modèle de corrélation uniforme
- $Cov(Y_{ij}, Y_{ij'}|\gamma_{0i}) = E[(Y_{ij} - E(Y_{ij}|\gamma_{0i}))(Y_{ij'} - E(Y_{ij'}|\gamma_{0i}))] = E(\epsilon_{ij}\epsilon_{ij'}) = 0$
- Interprétation de $\gamma_{0i}$ :
  - $\rightarrow$ "effet "sur le niveau initial du marqueur de l'ensemble des variables explicatives spécifiques au sujet $i$ non incluses dans le modèle
  - $\sigma_0^2 = 0$ ? $\Rightarrow \gamma_{0i} = 0 \Rightarrow$ pas d'effet sujet (pas de corrélation intra-sujet)

# Remarque

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Coefficient de corrélation de Pearson dans un modèle sans effet aléatoire

$$Corr(Y_{ij}, Y_{ij'}) = \rho_{ijj'} = \frac{Cov(Y_{ij}, Y_{ij'})}{\sqrt{V(Y_{ij})V(Y_{ij'})}}$$

or

$$\begin{aligned} Cov(Y_{ij}, Y_{ij'}) &= E[(Y_{ij} - E(Y_{ij}))(Y_{ij'} - E(Y_{ij'}))] \\ &= E[(\epsilon_{ij})(\epsilon_{ij'})] \\ &= E(\epsilon_{ij}\epsilon_{ij'}) \\ &= E((\epsilon_{ij} - E(\epsilon_{ij}))(\epsilon_{ij'} - E(\epsilon_{ij'}))) \\ &= Cov(\epsilon_{ij}, \epsilon_{ij'}) = 0 \end{aligned}$$

donc

$$Corr(Y_{ij}, Y_{ij'}) = \rho_{ijj'} = \frac{0}{\sigma_\epsilon^2} = 0$$

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Évolution de MDRD en fonction du genre du donneur

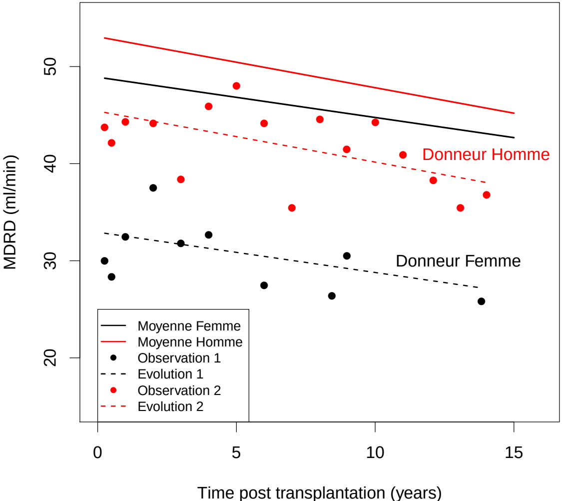
$$MDRD_{ij} = (\beta_0 + \gamma_{0i}) + \beta_1 t_{ij} + \beta_2 SexeD_i + \beta_3 SexeD_i t_{ij} + \epsilon_{ij}$$

```
Linear mixed-effects model fit by maximum likelihood
 Data: dataPDR
       AIC      BIC    logLik
  72994.24 73037.14 -36491.12

Random effects:
 Formula: ~1 | repere
        (Intercept) Residual
StdDev:    15.76693 9.994176

Fixed effects: MDRD ~ tpsbio + sexeD + sexeD * tpsbio
                Value Std.Error   DF  t-value p-value
(Intercept)  48.90503 0.9554553 8486 51.18506  0.0000
tpsbio       -0.41493 0.0543973 8486 -7.62786  0.0000
sexeD         4.16794 1.1635489  923  3.58209  0.0004
tpsbio:sexeD -0.11001 0.0655429 8486 -1.67850  0.0933
 Correlation:
             (Intr) tpsbio sexeD
tpsbio       -0.218
sexeD        -0.821  0.179
tpsbio:sexeD  0.181 -0.830 -0.219

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-9.903012383 -0.512553801  0.000268795  0.514145819  5.845333103

Number of Observations: 9413
Number of Groups: 925
```

# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Observations et évolutions estimées de 2 sujets : 1 donneur Femme et 1 donneur Homme)



# Exercice 2

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Écrire le modèle linéaire à intercept et pente aléatoires
- Calculer $E(Y_{ij})$, $E(Y_{ij}|\gamma_i)$, $Var(Y_{ij})$ et $Var(Y_{ij}|\gamma_i)$
- Calculer la covariance $Cov(Y_{ij}, Y_{ij'})$

# Exercice 2

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

$$Y_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 X_i + \epsilon_{ij}$$

- Espérance populationnelle

$$E(Y_{ij}) = E(X_{ij}^T\beta + \gamma_{0i} + \epsilon_{ij}) = E(X_{ij}^T\beta) + E(\gamma_{0i}) + E(\epsilon_{ij}) = X_{ij}^T\beta$$

- Espérance spécifique au sujet

$$E(Y_{ij}|\gamma_i) = X_{ij}^T\beta + \gamma_{0i} + \gamma_{1i}t_{ij}$$

# Exercice 2

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Variance populationnelle

$$Var(Y_{ij}) = V(X_{ij}^T\beta + \gamma_{0i} + \gamma_{1i}t_{ij} + \epsilon_{ij}) = \sigma_0^2 + t_{ij}^2\sigma_1^2 + \sigma_\epsilon^2$$

$\rightarrow$ Var. $\nearrow$ avec le temps (en pratique, c'est souvent le cas)

- Variance spécifique au sujet

$$Var(Y_{ij}|\gamma_i) = \sigma_\epsilon^2$$

$\rightarrow$ conditionnellement à $\gamma_i$, hypothèse d'homoscédasticité est vérifiée (Var. constante)

# Exercice 2

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Covariance entre 2 mesures du même sujet

$$
\begin{aligned}
Cov(Y_{ij}, Y_{ij'}) &= E[(Y_{ij} - E(Y_{ij}))(Y_{ij'} - E(Y_{ij'}))] \\
&= E[(\gamma_{0i} + \gamma_{1i} t_{ij} + \epsilon_{ij})(\gamma_{0i} + \gamma_{1i} t_{ij'} + \epsilon_{ij'})] \\
&= E(\gamma_{0i}^2) + t_{ij} t_{ij'} E(\gamma_{1i}^2) + E(\epsilon_{ij}\epsilon_{ij'}) + \ldots \\
&= \sigma_0^2 + t_{ij} t_{ij'} \sigma_1^2 + 0 + \ldots \\
&= \sigma_0^2 + t_{ij} t_{ij'} \sigma_1^2
\end{aligned}
$$

$\rightarrow$ Covariance dépend du temps

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

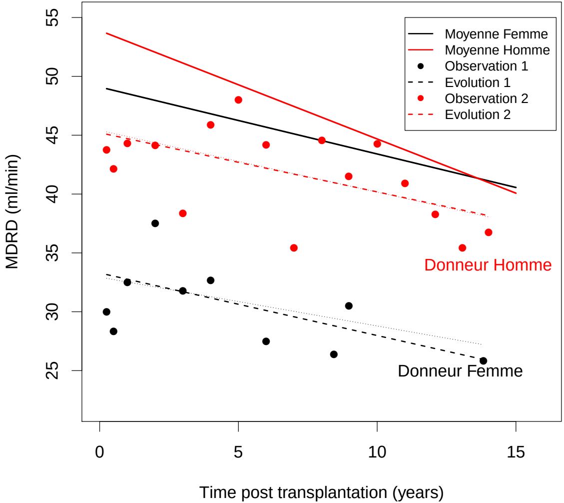
## Évolution de MDRD en fonction du genre du donneur

$$MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 SexeD_i + \beta_3 SexeD_i t_{ij} + \epsilon_{ij}$$

```
Linear mixed-effects model fit by maximum likelihood
 Data: dataPDR
       AIC      BIC    logLik
  71164.26 71221.45 -35574.13

Random effects:
 Formula: ~1 + tpsbio | repere
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 15.895851 (Intr)
tpsbio       1.942811 -0.194
Residual     8.203701

Fixed effects: MDRD ~ tpsbio + sexeD + sexeD * tpsbio
                Value Std.Error   DF  t-value p-value
(Intercept)  49.09557 0.9514397 8486 51.60134  0.0000
tpsbio       -0.56913 0.1344378 8486 -4.23341  0.0000
sexeD         4.79563 1.1587521  923  4.13861  0.0000
tpsbio:sexeD -0.35207 0.1627760 8486 -2.16293  0.0306
 Correlation:
             (Intr) tpsbio sexeD
tpsbio       -0.250
sexeD        -0.821  0.205
tpsbio:sexeD  0.207 -0.826 -0.250

Standardized Within-Group Residuals:
        Min          Q1          Med           Q3          Max
-12.52157590 -0.48981767 -0.01406612  0.48610603  6.86450258
```

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Observations et évolutions estimées de 2 sujets : 1 donneur Femme et 1 donneur Homme)



# Plan

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 4. Modèle linéaire mixte

Rappels sur le modèle linéaire
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle

# Généralisation

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- $Cov(Y_{ij}, Y_{ij'}|\gamma_i) = E[(Y_{ij} - E(Y_{ij}|\gamma_i))(Y_{ij'} - E(Y_{ij'}|\gamma_i))] = E(\epsilon_{ij}\epsilon_{ij'}) = 0$
  - Hypothèse d'indépendance des réponses $Y_{ij}$ conditionnellement aux effets aléatoires

$\Rightarrow$ Hypothèse trop forte ?

$\Rightarrow$ Structure de la corrélation entre les $Y_{ij}$ réaliste ?

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Extension du modèle

$$Y_i = X_i\beta + Z_i\gamma_i + \epsilon_i \qquad \epsilon \sim \mathcal{N}(0, \Sigma_i = \sigma^2_\epsilon I_{n_i})$$

- corrélation des erreurs : $\Sigma_i$ non diagonale
- par ex : erreur autoregressive

$$\begin{cases} \epsilon_{ij} = \omega(t_{ij}) + \nu_{ij} \\ \nu_{ij} \sim \mathcal{N}(0, \sigma^2_\nu) \\ \nu_{ij} \perp \nu_{ij'} \\ \omega(t_{ij}) \sim \mathcal{N}(0, \sigma^2_\omega) \\ Cov(\omega(t), \omega(s)) = \exp(-\kappa|t-s|) \end{cases}$$

$X_i\beta$ : profil moyen pour la population ; $\gamma_i$ : tendance individuelle à long terme ; $\omega(t_{ij})$ : erreur autoregressive, i.e. variations individuelles à court terme ; $\nu_{ij}$ : erreur de mesure

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Modèle linéaire mixte (le plus général)

$$\begin{cases} Y_i = X_i\beta + \epsilon_i \\ \epsilon_i \sim \mathcal{MVN}(0, V_i) \end{cases}$$

## Choix de la structure de $V_i$

- Fréquence des mesures
- Durée de suivi

  si $n_i$ grand $\rightarrow$ structure de covariance + complexe
- Objectif
  - Estimation $\widehat{\beta}$, $\widehat{V}(\widehat{\beta})$ $\rightarrow$ structure de covariance simple
  - Estimation $\widehat{\gamma}_i$, $E(Y_i|\widehat{\gamma}_i)$ $\rightarrow$ structure de covariance + complexe

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Quelques exemples

- $V_i$ échangeable

$$Var(Y_{ij}) = \sigma^2 \;\; \forall \; i, j$$

$$Cov(Y_{ij}, Y_{ij'}) = \delta \;\; \forall \; i, j, j'$$

- $V_i$ diagonale par bande

$$Var(Y_{ij}) = \sigma^2 \;\; \forall \; i, j$$

$$Cov(Y_{ij}, Y_{i(j+l)}) = \sigma_l \;\; \forall \; i, j, l$$

Pertinent lorsqu'il y a un ordre entre les mesures et délais entre les mesures constants

# Généralisation

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- $V_i$ non structurée

$$Var(Y_{ij}) = \sigma_j^2 \ \ \forall \ i$$

$$Cov(Y_{ij}, Y_{ij'}) = \sigma_{jj'} \ \ \forall \ i$$

Pertinent lorsqu'il y a un ordre entre les mesures

# Plan

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 4. Modèle linéaire mixte

Rappels sur le modèle linéaire
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle

# Vraisemblance

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Utilisation de la formulation marginale du modèle linéaire mixte

$$Y_i \sim \mathcal{N}(X_i\beta, V_i = Z_iBZ_i^T + \Sigma_i)$$

- Estimation des effets fixes $\beta$ et des paramètres $\phi$ de variance-covariance de $V_i$ par méthode du maximum de vraisemblance

$$L(\theta) = \prod_{i=1}^{k} f(Y_i) = \prod_{i=1}^{k} \int f(Y_i, \gamma_i) d\gamma_i$$

$$= \prod_{i=1}^{k} \int f(Y_i|\gamma_i) f(\gamma_i) d\gamma_i$$

avec $f(Y_i)$ : densité marginale du marqueur

$\theta = (\beta, \phi)$

# Vraisemblance

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## i.e. Vraisemblance marginale

$$L(\theta) = \prod_{i=1}^{k} \left(\frac{1}{2\pi}\right)^{n_i/2} |V_i|^{-1/2} \exp\left(-\frac{1}{2}(Y_i - X_i\beta)^T V_i^{-1}(Y_i - X_i\beta)\right)$$

## ⇒ log-Vraisemblance marginale

$$\begin{aligned} l(\theta) &= \ln(L(\theta)) \\ &= \sum_{i=1}^{k} -\frac{n_i}{2}\ln(2\pi) - \frac{1}{2}\ln(|V_i|) - \frac{1}{2}(Y_i - X_i\beta)^T V_i^{-1}(Y_i - X_i\beta) \end{aligned}$$

# Maximum de vraisemblance

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Généralités sur la recherche du maximum de la fonction $f$

Soit $f$ une fonction continue et dérivable définie sur un domaine $D$ allant dans $\mathbb{R}$ et dont la dérivée est également continue et dérivable.

$f(x)$ est un maximum de $f$, $x \in D$, si

- $f'(x) = 0$ (pour avoir un optimum)
- $f''(x) < 0$ (pour avoir une courbe concave et donc un maximum)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Annulation de la dérivée 1ère

$$\frac{\partial l}{\partial \beta} = -\sum_{i=1}^{k} X_i^T V_i^{-1}(Y_i - X_i\beta) = 0$$

- si paramètres $\phi$ de variance-covariance connus, alors

$$\Rightarrow \widehat{\beta} = \left(\sum_{i=1}^{k} X_i^T V_i^{-1} X_i\right)^{-1} \sum_{i=1}^{k} X_i^T V_i^{-1} Y_i \qquad (1)$$

$$V(\widehat{\beta}) = \left[-E\left(\frac{\partial^2 l}{\partial\beta\partial\beta^T}\right)\right]^{-1} = \left(\sum_{i=1}^{k} X_i^T V_i^T X_i\right)^{-1} \qquad (2)$$

# Maximum de vraisemblance

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- si paramètres $\phi$ de variance-covariance inconnus, alors
  - on remplace $\beta$ par $\widehat{\beta}$ (formule (1)) dans log-vraisemblance marginale, on obtient une fonction de vraisemblance notée (3)
  - puis on maximise la vraisemblance (3) à l'aide d'un algorithme itératif

# Maximum de vraisemblance restreinte

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- paramètres de covariance par Max de Vrais. biaisés

  $\rightarrow$ estimation ne tient pas compte $\searrow$ ddl liée à estimation $\beta$

- Solution : Maximum de vraisemblance restreinte/résiduelle

$$l_{REML}(\phi) = l(\widehat{\beta}, \phi) - \frac{1}{2} \ln \left| \sum_{i=1}^{k} X_i^T V_i^{-1} X_i \right|$$

- ne permet pas de comparer des modèles avec des effets fixes $\neq$

- si échantillon grand, $\widehat{\phi} \approx \widehat{\phi_{REML}}$

# Algorithme d'optimisation

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Estimation des paramètres

- par approche vraisemblance marginale : utilisation d'algorithme itératif
  - algorithme EM

    méthode itérative d'estimation en présence de données incomplètes, basée sur la maximisation de l'espérance conditionnelle de la log-vraisemblance des données complètes $(y, u)$ sachant les données observées $y$
  - algorithme de type Newton-Raphson
- par approche bayésienne
  - algorithme MCMC

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Algorithme "Expectation-Maximisation"

- Algorithme itératif où chaque itération $k$ se décompose en 2 étapes :

  **E :** Calcul de l'espérance de la log-vraisemblance complète sachant les données incomplètes $y$ et l'estimation courante du vecteur de paramètres $\theta^{(k)}$

$$E(\log(L(\theta; y, u)|\theta^{(k)})) = \int \log(L(\theta; y, u))p(u|y; \theta^{(k)})du$$

  **M :** Maximisation de cette espérance $E(\log(L(\theta; y, u)|\theta^{(k)}))$ avec

$$\theta^{(k+1)} = \underset{\theta}{\text{Argmax}}\ E(\log(L(\theta; y, u)|\theta^{(k)}))$$

# Algorithme EM

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Avantages/Inconvénients

- convergence, bonne propriété asymptotique des estimateurs
- existence de maximas locaux
- critères d'arrêt peu restrictifs
- assez lent
- pas d'estimation directe de la variance des paramètres

# Algorithme de type Newton-Raphson

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Principe

- Algorithme d'optimisation classique pour trouver les racines d'une fonction d'une ou plusieurs dimensions
- Algorithme itératif basé sur gradient de la fonction $\nabla(\theta^{(k)})$ et matrice hessienne $H^{(k)}$ (dérivées secondes)
- A l'itération $k$, vecteur de paramètres $\theta$ mis à jour :

$$\theta^{(k+1)} = \theta^{(k)} - \left(H^{(k)}\right)^{-1} \nabla\left(L(\theta^{(k)})\right)$$

- Critères d'arrêt stricts
- Estimation de la variance des paramètres

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Principe

- Paramètres $\theta$ considérés comme v.a. et non vecteur de paramètres inconnus
- Distribution *a priori* de $\theta$ est notée $\pi(\theta)$
- Inférence basée sur loi de distribution de $\theta$ conditionnelle aux observations $Y$ (distribution *a posteriori*)

$$\pi(\theta|Y) = \frac{f(Y|\theta)\pi(\theta)}{\int f(Y|\theta)\pi(\theta)d\theta}$$

avec $f(Y|\theta) = L(\theta)$

# Principe bayésien et algorithme MCMC

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

$\Rightarrow$ Estimation en considérant loi *a posteriori*

- vraisemblance sans expression analytique
- constante de normalisation $\int f(Y|\theta)\pi(\theta)d\theta$ pas toujours calculable

$\Rightarrow$ algorithme spécifique MCMC

- Très long
- Critère de convergence peu restrictif
- Structure de covariance complexe possible
- Effets aléatoires non gaussiens possible

# Estimation des effets aléatoires

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Objectif 1 :
- Objectif 2 :

# Estimation des effets aléatoires

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Objectif 1 : Estimations des paramètres du modèle linéaire marginal $\beta$
- Objectif 2 :

# Estimation des effets aléatoires

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Objectif 1 : Estimations des paramètres du modèle linéaire marginal $\beta$
- Objectif 2 : Estimation des effets aléatoires utiles pour prédiction individuelles
- Distribution *a priori* : $\gamma_i \sim f(\gamma_i) = \mathcal{N}(0, B)$
- Effets aléatoires $\gamma_i$ estimés par espérance de distribution *a posteriori* $E(\gamma_i | Y_i = y_i)$

$$E(\gamma_i | Y_i = y_i) = \int \gamma_i f(\gamma_i | Y_i) d\gamma_i$$

$$\begin{pmatrix} Y_i \\ \gamma_i \end{pmatrix} \sim \mathcal{MVN}\left( \begin{pmatrix} X_i\beta \\ 0 \end{pmatrix}, \begin{pmatrix} V_i & Z_i\beta \\ \beta Z_i^T & B \end{pmatrix} \right)$$

# Estimation des effets aléatoires

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- avec propriétés de loi $\mathcal{MVN}$ :

$$\begin{aligned} E(\gamma_i | Y_i) &= E(\gamma_i) + Cov(\gamma_i, Y_i) V(Y_i)^{-1} (Y_i - E(Y_i)) \\ &= B Z_i^T V_i^{-1} (Y_i - X_i \beta) \end{aligned}$$

or

$$\begin{aligned} Cov(\gamma_i, Y_i) &= E\left[(Y_i - X_i\beta)\gamma_i^T\right] = E\left[(Z_i\gamma_i + \epsilon_i)\gamma_i^T\right] \\ &= Z_i E(\gamma_i \gamma_i^T) + E(\epsilon_i \gamma_i^T) \\ &= Z_i V(\gamma_i) + 0 \\ &= Z_i B \end{aligned}$$

# Estimation des effets aléatoires

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

$\Rightarrow$ pour avoir estimateur bayésien empirique $\widehat{\gamma_i}$ [1], on utilise $\widehat{\beta}$ et $\widehat{\phi}$ :

$$\widehat{\gamma_i} = \widehat{B} Z_i^T \widehat{V_i}^{-1} (Y_i - X_i \widehat{\beta})$$

On peut aussi avoir $V(\widehat{\gamma_i})$

1. Best Linear Unbiased Prediction, BLUP

# Prédictions individuelles

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

Soit un modèle linéaire mixte général

$$\begin{cases} Y_i = X_i\beta + Z_i\gamma_i + \epsilon_i \\ \epsilon_i \sim \mathcal{MVN}(0, \Sigma_i) \\ \gamma_i \sim \mathcal{MVN}(0, B) \end{cases}$$

On pose $V_i = V(Y_i) = Z_i B Z_i^T + \Sigma_i$

## Prédictions

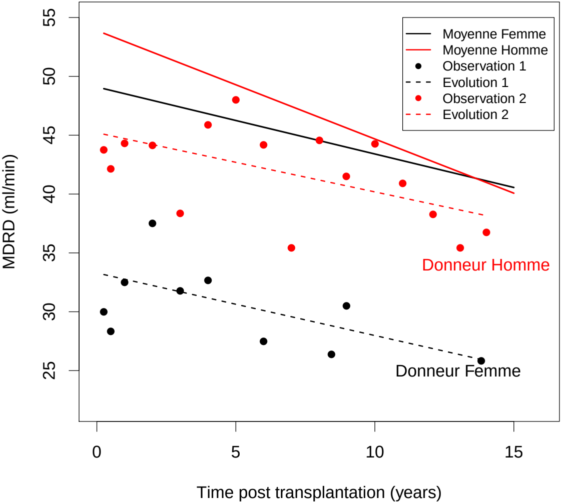
- Profil moyen/marginal de la population : $\widehat{E(Y_i)} = X_i\widehat{\beta}$
- Profil conditionnel/spécifique pour le sujet $i$ :

$$\widehat{Y_i} = E(Y_i|\widehat{\beta}, \widehat{\gamma_i}) = X_i\hat{\beta} + Z_i\hat{\gamma}_i = \ldots = \Sigma_i V_i^{-1} X_i\widehat{\beta} + (I_{n_i} - \Sigma_i V_i^{-1}) Y_i$$

# Exemple
## Modèle linéaire à intercept et pente aléatoires

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

### Observations et évolutions estimées de 2 sujets : 1 donneur Femme et 1 donneur Homme)



# Prédictions individuelles

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarques

- Prédiction individuelle= moyenne pondérée de moyenne marginale $X_i\widehat{\beta}$ et valeur observée $Y_i$
- Pondération dépend du rapport $\frac{\text{Variance intra-sujet } \Sigma_i}{\text{Variance totale } V_i}$
- Si $\Sigma_i \approx V_i$, alors $\rho \approx 0$

$$\widehat{Y}_i \approx \widehat{E(Y_i)}$$

$\rightarrow$ prédiction individuelle proche du profil moyen

- Si $\Sigma_i \approx Y_i$, alors $\rho \approx 1$
  $\rightarrow$ prédiction individuelle proche observation

# Plan

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 4. Modèle linéaire mixte

- Rappels sur le modèle linéaire
- Modèle linéaire mixte
- Inclusion de covariables
- Exercices
- Généralisation
- Estimation
- **Tests d'inférence**
- Adéquation du modèle

# Tests d'inférence (1)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Ajout d'une variable explicative

- Modèle avec $(p-1)$ variables explicatives $(X_1, ..., X_{p-1})$
- Test pour l'addition d'une variable $X_p$ dans le modèle

  $\rightarrow$ le modèle contenant la variable $X_p$ fournit-il plus d'information sur l'évolution du marqueur que le modèle sans cette variable ?

  $\Rightarrow$ $H_0 : \beta_p = 0$

$\Rightarrow$ 3 Tests basés sur la vraisemblance $L(\theta)$

# Tests d'inférence (2)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Test du rapport des vraisemblance

$$\chi^2_L = -2[l(\beta^0) - l(\beta)] \sim \chi^2_{1ddl}$$

$$\beta^0 = (\beta_1, ..., \beta_{p-1}, 0) \quad \text{et} \quad \beta = (\beta_1, ..., \beta_{p-1}, \beta_p)$$

- Test de Wald

$$\chi^2_W = \frac{\widehat{\beta}^2_p}{\widehat{var}(\widehat{\beta}_p)} \sim \chi^2_{1ddl}$$

- Test du score

$$\chi^2_S = \left[\frac{\delta \log L(\widehat{\beta}_0|X)}{\delta \widehat{\beta}_0}\right]' \left[-E\left(\frac{\delta^2 \log L(\widehat{\beta}_0|X)}{\delta \widehat{\beta}^2_0}\right)\right]^{-1} \left[\frac{\delta \log L(\widehat{\beta}_0|X)}{\delta \widehat{\beta}_0}\right] \sim \chi^2_{1ddl}$$

où $\widehat{\beta}_0$ estimateur de $\beta$ pour le modèle sans la $p$-ième covariable

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

# Évolution de MDRD en fonction du genre du donneur

$$MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 SexeD_i + \beta_3 SexeD_i t_{ij} + \epsilon_{ij}$$

```
Linear mixed-effects model fit by maximum likelihood
 Data: dataPDR
       AIC      BIC    logLik
  71164.26 71221.45 -35574.13

Random effects:
 Formula: ~1 + tpsbio | repere
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 15.895851 (Intr)
tpsbio       1.942811 -0.194
Residual     8.203701

Fixed effects: MDRD ~ tpsbio + sexeD + sexeD * tpsbio
                Value Std.Error   DF  t-value p-value
(Intercept)  49.09557 0.9514397 8486 51.60134  0.0000
tpsbio       -0.56913 0.1344378 8486 -4.23341  0.0000
sexeD         4.79563 1.1587521  923  4.13861  0.0000
tpsbio:sexeD -0.35207 0.1627760 8486 -2.16293  0.0306
 Correlation:
             (Intr) tpsbio sexeD
tpsbio       -0.250
sexeD        -0.821  0.205
tpsbio:sexeD  0.207 -0.826 -0.250

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-12.52157590 -0.48981767 -0.01406612  0.48610603  6.86450258
```

# Tests d'inférence (3)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Cas d'une variable explicative à plusieurs modalités

Soit

$$\begin{cases} Y_i = \beta_{0i} + \beta_{1i} t_i + \beta_2 X_i + \epsilon_i \\ \epsilon_i \sim \mathcal{MVN}(0, \Sigma_i) \\ (\beta_{0i}, \beta_{1i})^T \sim \mathcal{MVN}(\beta, B) \end{cases}$$

exemple : avec $X_i \in \{1, 2, 3\}$

$X_i \in$ {- de 20 ans, 30-40 ans, 40 ans et + }

Que représente $\beta_2$ ?

# Tests d'inférence (3)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Cas d'une variable explicative à plusieurs modalités

Soit

$$\begin{cases} Y_i = \beta_{0i} + \beta_{1i} t_i + \beta_2 X_i + \epsilon_i \\ \epsilon_i \sim \mathcal{MVN}(0, \Sigma_i) \\ (\beta_{0i}, \beta_{1i})^T \sim \mathcal{MVN}(\beta, B) \end{cases}$$

exemple : avec $X_i \in \{1, 2, 3\}$

$X_i \in$ {- de 20 ans, 30-40 ans, 40 ans et + }

Que représente $\beta_2$ ?

Variation du niveau initial du marqueur pour une augmentation de 1 unité de $X$

# Tests d'inférence (4)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

Problème :

# Tests d'inférence (4)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Problème :

- Variation identique d'un groupe à l'autre

$\Rightarrow$ Hypothèse restrictive

$$\begin{cases} Y_i = \beta_{0i} + \beta_{1i} t_i + \beta_2 X_i + \epsilon_i \\ \epsilon_i \sim \mathcal{MVN}(0, \Sigma_i) \\ (\beta_{0i}, \beta_{1i})^T \sim \mathcal{MVN}(\beta, B) \end{cases}$$

Modèle incorrect !

# Tests d'inférence (5)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

Solution :

# Tests d'inférence (5)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Solution :

- Recoder la variable explicative $X$ à l'aide de variables binaires
- Exemple : avec 2 variables $i_{X_1}$ et $i_{X_2}$

| Catégories | Codage $X$ | $i_{X_1}$ | $i_{X_2}$ |
|---|---|---|---|
| - de 20 ans | 1 | 0 | 0 |
| 30-40 ans | 2 | 1 | 0 |
| 40 ans et + | 3 | 0 | 1 |

- On considère le modèle :

$$\begin{cases} Y_i = \beta_{0i} + \beta_{1i} t_i + \beta_2 i_{X_1} + \beta_3 i_{X_2} + \epsilon_i \\ \epsilon_i \sim \mathcal{MVN}(0, \Sigma_i) \\ (\beta_{0i}, \beta_{1i})^T \sim \mathcal{MVN}(\beta, B) \end{cases}$$

# Tests d'inférence (6)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Question :

- Quel est le niveau initial du marqueur pour les sujets âgés entre 20 et 30 ans ?
- Quel est le niveau initial du marqueur pour les sujets âgés entre 30 et 40 ans ?
- Quel est le niveau initial du marqueur pour les sujets âgés de plus 40 ans ?

# Tests d'inférence (6)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Question :

- Quel est le niveau initial du marqueur pour les sujets âgés entre 20 et 30 ans ?

$$E(Y_i|t_{ij} = 0, i_{x_1} = 0, i_{x_2} = 0) = \beta_{0i} + \beta_{1i} * 0 + \beta_2 * 0 + \beta_3 * 0 = \beta_{0i}$$

- Quel est le niveau initial du marqueur pour les sujets âgés entre 30 et 40 ans ?
- Quel est le niveau initial du marqueur pour les sujets âgés de plus 40 ans ?

# Tests d'inférence (6)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Question :

- Quel est le niveau initial du marqueur pour les sujets âgés entre 20 et 30 ans ?

$$E(Y_i|t_{ij} = 0, i_{x_1} = 0, i_{x_2} = 0) = \beta_{0i} + \beta_{1i} * 0 + \beta_2 * 0 + \beta_3 * 0 = \beta_{0i}$$

- Quel est le niveau initial du marqueur pour les sujets âgés entre 30 et 40 ans ?

$$E(Y_i|t_{ij} = 0, i_{x_1} = 1, i_{x_2} = 0) = \beta_{0i} + \beta_{1i} * 0 + \beta_2 * 1 + \beta_3 * 0 = \beta_{0i} + \beta_2$$

- Quel est le niveau initial du marqueur pour les sujets âgés de plus 40 ans ?

# Tests d'inférence (6)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Question :

- Quel est le niveau initial du marqueur pour les sujets âgés entre 20 et 30 ans ?

$$E(Y_i|t_{ij} = 0, i_{x_1} = 0, i_{x_2} = 0) = \beta_{0i} + \beta_{1i} * 0 + \beta_2 * 0 + \beta_3 * 0 = \beta_{0i}$$

- Quel est le niveau initial du marqueur pour les sujets âgés entre 30 et 40 ans ?

$$E(Y_i|t_{ij} = 0, i_{x_1} = 1, i_{x_2} = 0) = \beta_{0i} + \beta_{1i} * 0 + \beta_2 * 1 + \beta_3 * 0 = \beta_{0i} + \beta_2$$

- Quel est le niveau initial du marqueur pour les sujets âgés de plus 40 ans ?

$$E(Y_i|t_{ij} = 0, i_{x_1} = 0, i_{x_2} = 1) = \beta_{0i} + \beta_{1i} * 0 + \beta_2 * 0 + \beta_3 * 1 = \beta_{0i} + \beta_3$$

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Évolution de MDRD en fonction de l'âge du donneur en 3 classes avec interaction avec le temps

$$MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 \mathbb{1}_{ageD \in [20-30)} + \beta_3 \mathbb{1}_{ageD \in [40et+)} + \beta_4 \mathbb{1}_{ageD \in [20-30)} t_{ij} + \beta_5 \mathbb{1}_{ageD \in [40et+)} t_{ij} + \epsilon_{ij}$$

```
Linear mixed-effects model fit by maximum likelihood
 Data: dataPDR
       AIC      BIC    logLik
  70855.89 70927.36 -35417.94

Random effects:
 Formula: ~1 + tpsbio | repere
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 14.867270 (Intr)
tpsbio       1.926501 -0.26
Residual     8.213020

Fixed effects: MDRD ~tpsbio+i.ageD3040+i.ageD40+i.ageD3040*tpsbio+i.ageD40*tpsbio
                      Value Std.Error   DF   t-value p-value
(Intercept)        60.43755 0.9344210 8463  64.67914  0.0000
tpsbio             -0.46551 0.1337015 8463  -3.48170  0.0005
i.ageD3040         -4.51007 1.6466743  921  -2.73890  0.0063
i.ageD40          -13.40581 1.1577296  921 -11.57939  0.0000
tpsbio:i.ageD3040  -0.60430 0.2384404 8463  -2.53437  0.0113
tpsbio:i.ageD40    -0.45832 0.1685987 8463  -2.71842  0.0066
```

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Test de l'interaction âge du donneur avec le temps

$$H_0 : \beta_4 = \beta_5 = 0$$

$\rightarrow$ Évolution de MDRD en fonction de l'âge du donneur en 3 classes sans interaction avec le temps

```
Linear mixed-effects model fit by maximum likelihood
 Data: dataPDR
       AIC      BIC    logLik
  70861.37 70918.55 -35422.69

Random effects:
 Formula: ~1 + tpsbio | repere
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 14.890471 (Intr)
tpsbio       1.945881 -0.269
Residual     8.212004

Fixed effects: MDRD ~ tpsbio + i.ageD3040 + i.ageD40
                Value Std.Error   DF   t-value p-value
(Intercept)  61.17493 0.9029383 8465  67.75095   0e+00
tpsbio       -0.80022 0.0758944 8465 -10.54384   0e+00
i.ageD3040   -5.81814 1.5654798  921  -3.71652   2e-04
i.ageD40    -14.40267 1.0993027  921 -13.10164   0e+00
```

# Exemple
## Modèle linéaire à intercept et pente aléatoires

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

### Test de l'interaction âge du donneur avec le temps

$$H_0 : \beta_4 = \beta_5 = 0$$

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5) \quad \beta^0 = (\beta_0, \beta_1, \beta_2, \beta_3)$$

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Test de l'interaction âge du donneur avec le temps

$$H_0 : \beta_4 = \beta_5 = 0$$

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5) \quad \beta^0 = (\beta_0, \beta_1, \beta_2, \beta_3)$$

$\Rightarrow$ Test du rapport de vraisemblance

- log-vraisemblance “modèle complet” = -35417.94
- log-vraisemblance “modèle incomplet”= -35422.69

$$\chi^2_{calc} = -2[l(\beta^0) - l(\beta)] = -2(-35422.69 - (-35417.94)) \approx 9.48$$

avec $\chi^2_{2ddl;5\%} = 5.99 \rightarrow$ région critique RC : $[5.99; +\infty[$

$\Rightarrow \chi^2_{calc} \in \text{RC} \leftrightarrow p < \alpha = 5\%$ Remarque $p = 0.0087$

$\Rightarrow$ Rejet de $H_0$ au seuil 5% : il semble y avoir une interaction des classes d'âge du donneur avec la pente d'évolution

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Test de la covariance

- Test des variances des e.a. = 0 $\rightarrow$ Impossible

$$H_0 : \sigma^2 = 0$$

$\Rightarrow$ Limite de l'espace des paramètres ($\sigma^2 \geq 0$)

- Test de $q$ e.a. corrélés vs. $(q+1)$ e.a.

$$H_0 : B = \begin{pmatrix} B_{qq} & 0 \\ 0 & 0 \end{pmatrix} \; vs. \; H_1 : B = \begin{pmatrix} B_{qq} & C \\ C^T & \sigma^2_{q+1} \end{pmatrix}$$

où $B_{qq}$ : matrice var-cov des $q$ 1er e.a. ; $\sigma^2_{q+1}$ : variance du $(q+1)$-ième e.a. : $C$ : vecteur des $q$ covariances entre $(q+1)$-ième e.a. et les $q$ 1er e.a.

# Tests d'inférence (8)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Distribution de la stat du Rapport de Vraisemblance

$$LRS = -2\ln\left(\frac{L(\theta_{H_0})}{L(\theta_{H_1})}\right) = -2l(\theta_{H_0}) + 2l(\theta_{H_1})$$

- Sous $H_0$, distribution asymptotique de $LRS$ : mélange de 2 distributions de $\chi^2_{q+1}$ et $\chi^2_q$ à proba égale

$$\begin{aligned} \text{p-value} &= Pr(\chi^2_{q+1:q} < LRS) \\ &= 0.5 * Pr(\chi^2_{q+1} < LRS) + 0.5 * Pr(\chi^2_q < LRS) \end{aligned}$$

# Exemple



## Modèle linéaire à intercept et pente aléatoires vs. Modèle linéaire à intercept aléatoire

$$H_0 : \text{Variance pente aléatoire} = \sigma_1^2 = 0$$

### Modèle linéaire à intercept et pente aléatoires

```
Linear mixed-effects model fit by maximum likelihood
 Data: dataPDR
       AIC      BIC    logLik
  71164.26 71221.45 -35574.13

Random effects:
 Formula: ~1 + tpsbio | repere
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 15.895851 (Intr)
tpsbio       1.942811 -0.194
Residual     8.203701

Fixed effects: MDRD ~ tpsbio + sexeD + sexeD * tpsbio
                Value Std.Error   DF  t-value p-value
(Intercept)  49.09557 0.9514397 8486 51.60134  0.0000
tpsbio       -0.56913 0.1344378 8486 -4.23341  0.0000
sexeD         4.79563 1.1587521  923  4.13861  0.0000
tpsbio:sexeD -0.35207 0.1627760 8486 -2.16293  0.0306
```

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Modèle linéaire à intercept aléatoire

```
Linear mixed-effects model fit by maximum likelihood
 Data: dataPDR
       AIC      BIC    logLik
  72994.24 73037.14 -36491.12

Random effects:
 Formula: ~1 | repere
        (Intercept) Residual
StdDev:    15.76693 9.994176

Fixed effects: MDRD ~ tpsbio + sexeD + sexeD * tpsbio
                Value Std.Error   DF  t-value p-value
(Intercept)  48.90503 0.9554553 8486 51.18506  0.0000
tpsbio       -0.41493 0.0543973 8486 -7.62786  0.0000
sexeD         4.16794 1.1635489  923  3.58209  0.0004
tpsbio:sexeD -0.11001 0.0655429 8486 -1.67850  0.0933
```

# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Modèle linéaire à intercept et pente aléatoires vs. Modèle linéaire à intercept aléatoire

$$H_0 : \text{Variance pente aléatoire} = \sigma_1^2 = 0$$

# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Modèle linéaire à intercept et pente aléatoires vs. Modèle linéaire à intercept aléatoire

$$H_0 \text{ : Variance pente aléatoire} = \sigma_1^2 = 0$$

$\Rightarrow$ Test du rapport de vraisemblance

- log-vraisemblance sous $H_0$ = -36491.12
- log-vraisemblance sous $H_1$ = -35574.13

$$\chi^2_{calc} = -2(-36491.12 - (-35574.13)) \approx 1833.981$$

$$\begin{aligned} \rightarrow p &= 0.5 * Pr(\chi^2_2 < LRS) + 0.5 * Pr(\chi^2_1 < LRS) \\ &= 1 - 0.5 * Pr(\chi^2_2 < LRS) - 0.5 * Pr(\chi^2_1 < LRS) \\ &\approx 0 \end{aligned}$$

$p < \alpha = 5\% \Rightarrow$ Rejet de $H_0$ au seuil 5% : il semble que la pente d'évolution de MDRD soit spécifique au sujet

# Tests d'inférence (9)



## Comparaison de modèles non emboités

- Modèles $M_1$ et $M_2$ sont emboités si

$$\theta_{M_1} \subset \theta_{M_2} \quad \text{ou l'inverse}$$

$\theta_{M_1}$ : vecteur de paramètres à estimer de $M_1$ ; $\theta_{M_2}$ : vecteur de paramètres à estimer de $M_2$

- Critère d'Akaike

$$AIC = -2 * l(\theta) + 2 * \kappa$$

$\Rightarrow$ Choix du modèle avec AIC minimum

- Bayesian Information Criterion

$$BIC = -2 * l(\theta) + \kappa * \ln(n)$$

$\Rightarrow$ Choix du modèle avec BIC minimum

$\kappa$ : nombre de paramètres estimés ; $n$ : nombre d'observations

# Exemple

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Évolution de MDRD en fonction de l'âge du donneur en 3 classes avec interaction avec le temps

$$MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 \mathbb{1}_{ageD \in [20-30)} + \beta_3 \mathbb{1}_{ageD \in [40et+)} + \beta_4 \mathbb{1}_{ageD \in [20-30)} t_{ij} + \beta_5 \mathbb{1}_{ageD \in [40et+)} t_{ij} + \epsilon_{ij}$$

```
Linear mixed-effects model fit by maximum likelihood
 Data: dataPDR
       AIC      BIC    logLik
  70855.89 70927.36 -35417.94
```

## Évolution de MDRD en fonction du genre du donneur

$$MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 SexeD_i + \epsilon_{ij}$$

```
Linear mixed-effects model fit by maximum likelihood
 Data: dataPDR
       AIC      BIC    logLik
  71164.26 71221.45 -35574.13
```

$\Rightarrow$ Choix du modèle : Donneur en 3 classes

# Plan



4. Modèle linéaire mixte
Rappels sur le modèle linéaire
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle

# Adéquation du modèle

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Principales hypothèses du modèle linéaire mixte

- Linéarité des relations avec le marqueur longitudinal
- Continuité du marqueur longitudinal
- Données manquantes
- Normalité des effets aléatoires
- Normalité des erreurs
- Homogénéité de la population d'étude (Homoscédasticité de l'erreur de mesure)

# Adéquation du modèle

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Linéarité

- Hypothèse sensible aux données
- Évaluation similaire à celle des modèles linéaires classiques : appréciation graphique

$\Rightarrow$ Transformation de la variable réponse $Y_i$

- $\ln(Y_i)$,...

$\Rightarrow$ Modèle linéaire généralisé à effets aléatoires

# Adéquation du modèle

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Continuité du marqueur

- $Y_i$ : variable aléatoire quantitative continue

  Que faire avec $Y_i$ v.a. quantitative discrète ?

  Que faire avec $Y_i$ v.a. binaire ?

- Solution possible :

  $\Rightarrow$ Modèle linéaire généralisé

# Adéquation du modèle

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Données manquantes

- Plan déséquilibré OK :

  $\rightarrow$ présence de données manquantes intermittentes ou de sorties d'étude

- Hypothèse majeure :

  Données manquantes distribuées aléatoirement

  $\rightarrow$ Écriture de la vraisemblance

$\Rightarrow$ Modèle linéaire mixte robuste à la présence de DM aléatoires

# Adéquation du modèle

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Normalité des effets aléatoires

- Estimation modèle linéaire mixte requiert spécification de la distribution conditionnelle de $Y_i$

Quelle loi pour les effets aléatoires $\gamma_i$ ?

- Étude de la distribution des $\widehat{\gamma}_i$ avec histogramme

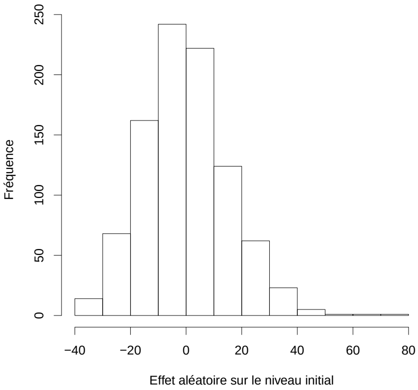
  $\Rightarrow$ Insuffisant

$\Rightarrow$ Modèle linéaire mixte robuste à mauvaise spécification

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

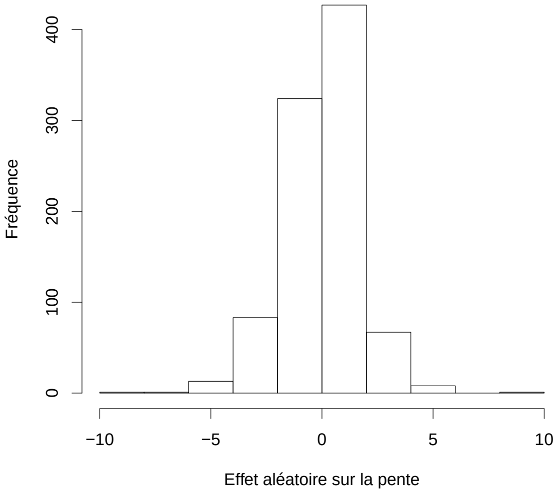
## Évolution de MDRD en fonction du genre du donneur

$$MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 SexeD_i + \beta_3 SexeD_i t_{ij} + \epsilon_{ij}$$



Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Évolution de MDRD en fonction du genre du donneur



# $\neq$ type de résidus

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

Avec Modèle linéaire mixte, $\exists \neq$ types de prédiction

$\Rightarrow \neq$ types de résidus

## Résidus marginaux

$$Y_i - \widehat{E(Y_i)} = Y_i - X_i\widehat{\beta}$$

de variance $\tilde{V}_i = V_i - X_i \left( \sum_{i=1}^{k} X_i^T V_i^{-1} X_i \right)^{-1} X_i^T$

$\Rightarrow$ Résidus marginaux standardisés $R_{ij} \sim \mathcal{N}(0, 1)$

# $\neq$ type de résidus

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Résidus spécifiques aux sujets

$$
\begin{aligned}
\widehat{\epsilon_i} &= Y_i - \widehat{E(Y_i|\gamma_i)} \\
&= Y_i - X_i\widehat{\beta} - Z_i\widehat{\gamma_i} \\
&= Y_i - X_i\widehat{\beta} - Z_iBZ_i^T V_i^{-1}(Y_i - X_i\widehat{\beta}) \\
&= (I - Z_iBZ_i^T V_i^{-1})(Y_i - X_i\widehat{\beta}) \\
&= \Sigma_i V_i^{-1}(Y_i - X_i\widehat{\beta})
\end{aligned}
$$

de variance $V(\widehat{\epsilon_i}) = \Sigma_i = \sigma_\epsilon^2 I$

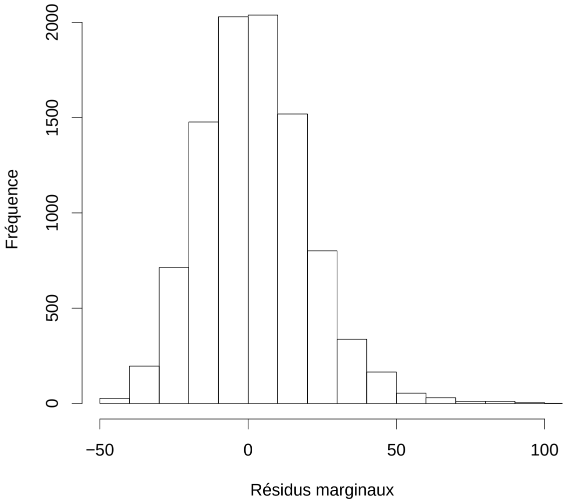
$\Rightarrow$ Résidus conditionnels standardisés $R_{ij}^* \sim \mathcal{N}(0,1)$

Si normalité non respectée, ou bien hétéroscédasticité $\Rightarrow$ Transformation de $Y_i$

$\Rightarrow$ Modèle linéaire mixte robuste à l'hypothèse de normalité des résidus et à celle d'une variance constante

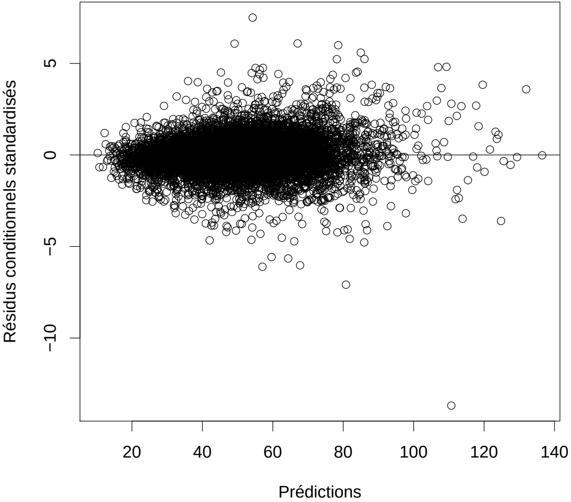
Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Évolution de MDRD en fonction du genre du donneur



Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

# Évolution de MDRD en fonction du genre du donneur



# Plan



1. Introduction
2. Objectifs
3. Méthodes historiques
4. Modèle linéaire mixte
5. Logiciels
6. Stratégie de modélisation
7. Application
8. Discussion
9. Références

# Présentations des données disponibles

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 1 ligne par observation

| Sujets | Temps | MDRD | sexeD | AgeD | ... |
|---|---|---|---|---|---|
| 1 | 0.25 | 43.75 | 0 | 19 | ... |
| 1 | 0.50 | 42.13 | 0 | 19 | ... |
| 1 | 1.00 | 44.29 | 0 | 19 | ... |
| 1 | 2.00 | 38.38 | 0 | 19 | ... |
| 1 | 2.98 | 45.89 | 0 | 19 | ... |
| ... | ... | ... | ... | ... | ... |
| 2 | 0.25 | 36.76 | 1 | 54 | ... |
| 2 | 2.00 | 32.71 | 1 | 54 | ... |
| 2 | 3.00 | 35.30 | 1 | 54 | ... |
| 2 | 6.00 | 32.12 | 1 | 54 | ... |
| 2 | 6.98 | 29.14 | 1 | 54 | ... |
| ... | ... | ... | ... | ... | ... |
| N | . | . | . | . | . |

# Programmation sous SAS

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

$$\begin{cases} MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 SexeD_i + \epsilon_{ij} \\ \gamma_i = \begin{pmatrix} \gamma_{0i} \\ \gamma_{1i} \end{pmatrix} \sim \mathcal{MVN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}\right) \\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2) \\ (\gamma_{0i}, \gamma_{1i}) \perp (\epsilon_1, \ldots, \epsilon_n) \end{cases}$$

## Procédure MIXED de SAS

```
Proc Mixed data=dataPDR method= ml  COVTEST;
class repere;
model MDRD= tpsbio sexeD/s outpred=p outpm=marg;
random intercept tpsbio/sub=repere type=UN G SOLUTION:
run;
```

## Diagnostic des résidus conditionnels

```
Proc univariate data=p normal;
Var resid; Histogram resid;
Proc gplot data=p; plot resid*pred;
```

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Multiples options :

- method : REML (par défaut), ML
- covtest : test des covariances
- outpred : résidus issus estimateurs bayésiens
- outpm : résidus marginaux
- type : matrice de var-cov UN (non structuré), VC (diagonale), AR(1) (auto-régressive),...
- G : affichage de la matrice de var-cov ;
- NOCLPRINT : si variable explicative catégorielle (à préciser dans class), pas d'impression $\neq$ modalités de la variable
- /s : affichage des estimations effets fixes
- SOLUTION : estimation a posteriori des effets aléatoires

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

$$\begin{cases} MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 SexeD_i + \epsilon_{ij} \\ \gamma_i = \begin{pmatrix} \gamma_{0i} \\ \gamma_{1i} \end{pmatrix} \sim \mathcal{MVN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}\right) \\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2) \\ (\gamma_{0i}, \gamma_{1i}) \perp (\epsilon_1, \ldots, \epsilon_n) \end{cases}$$

## library(nlme)

```
library(nlme)
modaj<-lme(MDRD ~ tpsbio+sexeD,data=dataPDR,
random=~1+tpsbio|repere,
na.action = na.omit,
control=list(maxIter=1000,opt="optim"),
method="ML")
summary(modaj)
```

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Diagnostic

```
hist(modaj$residuals[,1],xlim=c(-50,100),main="",
xlab="Résidus marginaux",ylab="Fréquence")
```

```
plot(modaj$fitted[,2],modaj$residuals[,2]/sd(modaj$residuals[,2]),
main="",xlab="Prédictions",ylab="Résidus conditionnels standardisés")
abline(h=0)
```

```
hist(modaj$coeff$random$repere[,1],main="",
xlab="Effet aléatoire sur le niveau initial",ylab="Fréquence")
```

```
hist(modaj$coeff$random$repere[,2],main="",
xlab="Effet aléatoire sur la pente",ylab="Fréquence")
```

# Plan

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

1. Introduction
2. Objectifs
3. Méthodes historiques
4. Modèle linéaire mixte
5. Logiciels
6. Stratégie de modélisation
7. Application
8. Discussion
9. Références

# Procédure de sélection des covariables

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Si nombre de variables explicatives grand :

  $\rightarrow$ pas raisonnable d'étudier l'ensemble des modèles possibles

- Définition d'un critère de sélection

  - $p$ (ou $R^2$, $AIC$, ...)

- Il existe $\neq$ algorithmes de sélection de variables explicatives

- mais ne pas perdre de vue

  Choix du meilleur modèle dépend de l'objectif ! !

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Procédure de sélection pas à pas ascendante

- estimation des $\neq$ modèles univariés
- modèle initial (celui sans covariable) : modèle M0
- ajout de la 1ère variable explicative (celle avec le + petit *p* ($< \alpha$) en univarié) : modèle M1
- ajout de la 2de variable explicative (celle avec le + petit *p* ($< \alpha$) en univarié parmi les modèles univariés $\neq$ de M1) : modèle M2
  $\rightarrow$ 2de variable conservée si $p < \alpha$ dans M2
- etc.
- jusqu'à ce que plus aucune variable ne soit ajoutée

# Procédure de sélection des covariables

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Procédure de sélection pas à pas descendante

- estimation du modèle initial (modèle avec toutes les variables explicatives) : modèle M0
- élimination de la 1ère covariable (celle avec le *p* $(> \alpha)$ le plus grand dans M0) : modèle M1
- élimination de la 2de covariable (celle avec le *p* $(> \alpha)$ le plus grand dans M1) : modèle M2
- etc.
- jusqu'à ce que plus aucune variable ne soit éliminée

# Procédure de sélection des covariables

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Procédure de sélection pas à pas mixtes

- choix d'une stratégie de sélection ascendante ou descendante
- estimation du modèle initial : modèle M0
- ajout/suppression de la 1ère covariable : modèle M1
- élimination d'une covariable de M1 si $p$ est + petit/grand que celui de l'étape précédente : modèle M2
- ajout/suppression de la 2de covariable : modèle M3
- élimination d'une covariable de M3 si $p$ est + petit/grand que celui de l'étape précédente : modèle M4
- etc.
- jusqu'à ce que plus aucune variable ne soit ajoutée/éliminée

# 1 suggestion de stratégie de modélisation

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 1. Définition de la question d'intérêt

- Définition du critère de jugement

  $\rightarrow$ Évolution d'un marqueur quantitatif gaussien

- Choix du modèle

  $\rightarrow$ Modèle de régression linéaire mixte

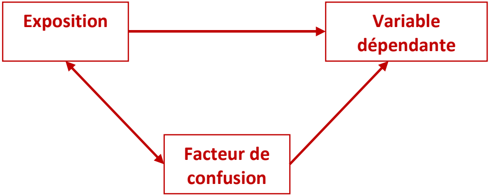
$\Rightarrow$ 1 variable explicative d'intérêt ?

  $\rightarrow$ Variable forcée dans le modèle quelque soit $p$

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Remarque : Biais de confusion

- Distorsion de l'estimation de l'effet la variable explicative principale sur la variable réponse, liée à un tiers facteur qui
  - est associé à l'exposition
  - est un facteur de risque de la maladie indépendamment de l'exposition
  - n'est pas dans le chemin causal entre le déterminant principal et la variable dépendante



# 1 suggestion de stratégie de modélisation

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Distinction entre confusion et interaction (modification de l'effet)
  - Situation dans laquelle l'association entre le facteur d'exposition et la variable dépendante diffère selon le niveau de la tiers variable

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 2. Facteurs de confusion potentiels

- Choix de variables forcées dans le modèle car *a priori* clinique comme facteurs de confusion
- Associations entre variable explicative principale et autres covariables : analyse descriptive
- Ensemble des modèles de régression univariés : analyse univariée

  $\Rightarrow$ Facteurs de confusion définis sur *a priori* épidémiologique :

  - Covariables significatives en analyse descriptive ($p < 0.05$) et en analyse univariée ($p < 0.20$)

# 1 suggestion de stratégie de modélisation

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 3. Interactions

- Pas raisonnable de tester la totalité des interactions entre covariables
- Définition des interactions cliniquement pertinentes

## 4. Variables incluses dans le modèle

- Variable explicative principale
- Variables de confusion forcées
- Variables significatives ($p < 0.2$) en univarié
- Interactions

# 1 suggestion de stratégie de modélisation

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 5. Sélection du modèle final

- Sélection descendante des interactions significatives à partir du modèle complet
- Sélection descendantes des effets simples à partir du modèle multivarié retenu à l'étape précédente

# Plan

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

1. Introduction
2. Objectifs
3. Méthodes historiques
4. Modèle linéaire mixte
5. Logiciels
6. Stratégie de modélisation
7. Application
8. Discussion
9. Références

# Contexte générale

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Insuffisance Rénale Chronique (IRC)

- Diminution progressive et irréversible du débit de filtration du rein

$\Rightarrow$ Solutions thérapeutiques :

- épuration rénale (hémodialyse, dialyse péritonéale)
  $\rightarrow$ très contraignant
- greffe rénale

# Contexte générale

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- $\neq$ facteurs associés à l'échec de greffe :
  - créatinine du donneur,
  - âge et genre du receveur et du donneur,...
- et masse néphronique fonctionnelle $\Leftrightarrow$ État de santé du rein
- Intérêt : meilleur allocation des greffons aux receveurs

# Contexte générale

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

Impact du ratio $\frac{\text{Poids du greffon}}{\text{Poids du receveur}}$ (PgPr) mis en évidence (Giral et al. 2005 et 2010) sur :

- la survie patient/greffon

  Augmentation du risque d'échec de greffe au delà de 2 ans après la greffe chez les faibles PgPr ($< 2.3g/kg$) par rapport aux ratios élevés ($\geq 2.3g/kg$)

$\Rightarrow$ Qu'en est-il sur l'évolution du débit de filtration glomérulaire ?

# Objectif

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Étudier la relation entre le ratio “Poids du greffon/Poids du receveur sur l'évolution du débit de filtration glomérulaire
- à l'aide d'un modèle linéaire à effets aléatoires

# Population d'étude

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

- Cohorte de 954 patients transplantés rénaux issus d'un réseau composé de 5 centres hospitaliers (Caen, Grenoble, Nancy, Nantes et Tours)

  Remarque : une partie de la cohorte DIVAT

- Critère d'exclusion :
  - Donneur vivant,
  - Age du receveur < 18 ans,
  - greffe du rein et du pancréas

# Population d'étude

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Variables pré-greffe disponibles :

- PgPr (ratio $< 2.3$ vs. $\geq 2.3$)
- âge du receveur et âge du donneur
- créatinine du donneur
- genre du receveur et du donneur (femme vs. homme)
- nombre de transplantations rénales antérieures (0 vs. 1 ou +)
- nombre d'incompatibilités HLA ($< 4$ vs. $\geq 4$)
- présence de PRA (0 vs. $> 0$)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 1. Définition de la question d'intérêt

- Étude de l'évolution du débit de filtration glomérulaire en fonction du ratio "Poids du greffon/Poids du receveur

- Critère de jugement :

  $\rightarrow$ Débit de filtration glomérulaire (ou clairance) estimé par la formule MDRD à partir de valeurs de la créatinine mesurées au cours du suivi


- 1 variable explicative d'intérêt

  $\rightarrow$ ratio "Poids du greffon/Poids ($< 2.3$ vs. $\geq 2.3$)

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 2. Facteurs de confusion potentiels

- Variables forcées sur *a priori* clinique

  $\rightarrow$ aucune

- Associations entre variable explicative principale et autres covariables :

  $\Rightarrow$ analyse descriptive

- Ensemble des modèles de régression univariés :

  $\Rightarrow$ analyse univariée

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Variables quantitatives

| | | Global N=954 | | PgPr$<$ 2.3 N=207 | | PgPr$\geq$ 2.3 N=747 | | p |
|---|---|---|---|---|---|---|---|---|
| | NA | *M* | sd | *M* | sd | *M* | sd | |
| Age receveur | 0 | 45.58 | 12.99 | 46.24 | 12.87 | 45.4 | 13.03 | 0.4108 |
| Age donneur | 4 | 40.04 | 15.33 | 36.08 | 17.67 | 41.14 | 14.43 | $<$0.0001 |
| Créat. donneur | 18 | 98.3 | 54.63 | 91.78 | 54.29 | 100.08 | 54.62 | 0.0567 |

NA : non attributed
M : Moyenne
sd : standard deviation

## Variables qualitatives

| | | Global N=954 | | PgPr$<$ 2.3 N=207 | | PgPr$\geq$ 2.3 N=747 | | p |
|---|---|---|---|---|---|---|---|---|
| | NA | Eff. | % | Eff. | % | Eff. | % | |
| Receveur Homme | 0 | 588 | 61.64 | 153 | 73.91 | 435 | 58.23 | $<$0.0001 |
| Donneur Homme | 2 | 642 | 67.44 | 104 | 50.24 | 538 | 72.21 | $<$0.0001 |
| au - 1 transplantation antérieure | 0 | 122 | 12.79 | 21 | 10.14 | 101 | 13.52 | 0.2393 |
| + de 4 incompatibilités | 19 | 356 | 38.07 | 87 | 42.65 | 269 | 36.8 | 0.1422 |
| Présence de PRA | 8 | 234 | 24.74 | 45 | 21.84 | 189 | 25.54 | 0.3153 |

# Analyse univariée

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Modèle à intercept et pente aléatoires : 1 covariable $X_i$ seulement sur le niveau initial de MDRD

$$MDRD_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})t_{ij} + \beta_2 X_i + \epsilon_{ij}$$

| | $\beta_2$ | p |
|---|---|---|
| Ratio PgPr ($\geq$ 2.3 vs. $<$ 2.3) | 2.44 | 0.0583 |
| Age receveur | -0.82 | $<$ 0.0001 |
| Age donneur | -0.48 | $<$ 0.0001 |
| Creat. donneur | -0.02 | 0.1243 |
| Genre receveur (homme vs. femme) | 3.97 | 0.0003 |
| Genre donneur (homme vs. femme) | 4.17 | 0.0002 |
| Transplantation antérieure (au - 1 vs. 0) | 6.85 | $<$ 0.0001 |
| Incompatibilité HLA ($\geq$ 4 vs. $<$ 4) | 0.15 | 0.8912 |
| Présence de PRA | 0.86 | 0.4887 |

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 2. Facteurs de confusion potentiels

- Variables forcées sur *a priori* clinique

  $\rightarrow$ aucune

$\Rightarrow$ Facteurs de confusion définis sur *a priori* épidémiologique :

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 2. Facteurs de confusion potentiels

- Variables forcées sur *a priori* clinique

  $\rightarrow$ aucune

$\Rightarrow$ Facteurs de confusion définis sur *a priori* épidémiologique :

$\rightarrow$ Age du donneur
$\rightarrow$ Créatinine du donneur
$\rightarrow$ Genre du receveur
$\rightarrow$ Genre du donneur

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 3. Interactions

- Définition des interactions cliniquement pertinentes

  → Ratio PgPr * temps
  → Age du donneur * temps
  → Créatinine du donneur donneur * temps
  → genre du receveur * temps
  → genre du donneur * temps

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## 4. Variables incluses dans le modèle

- Variable explicative principale :
  - $\rightarrow$ Ratio PgPr ($\geq$ 2.3 vs $<$ 2.3) et Ratio PgPr * temps
- Variables de confusion forcées :
  - $\rightarrow$ Age du donneur
  - $\rightarrow$ Créatinine du donneur donneur
  - $\rightarrow$ Genre du receveur
  - $\rightarrow$ Genre du donneur
- Variables significatives ($p < 0.2$) en univarié :
  - $\rightarrow$ Age du receveur
  - $\rightarrow$ Transplantation antérieure
- Interactions
  - $\rightarrow$ Age du donneur * temps
  - $\rightarrow$ Créatinine du donneur donneur * temps
  - $\rightarrow$ genre du receveur * temps
  - $\rightarrow$ genre du donneur * temps

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

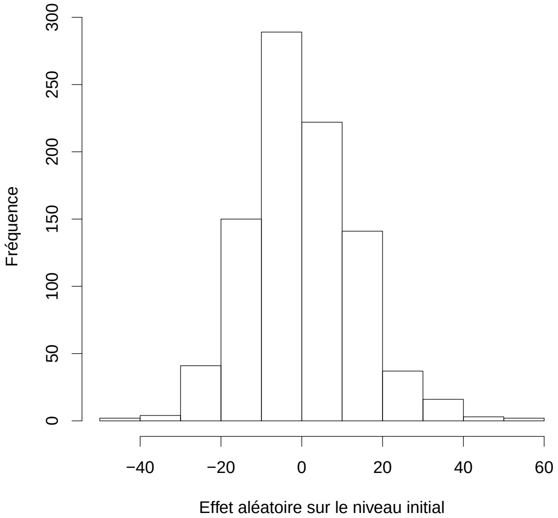
## Modèle à intercept et pente aléatoires : 1 covariable $X_i$ seulement sur le niveau initial de MDRD

$$MDRD_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \beta_2 iratio_i + \beta_3 AgeD_i + \beta_4 CreatD_i + \beta_5 SexeD_i + \beta_6 sexeR_i$$
$$+ \beta_7 itransplant + \beta_8 iratio_i * t_{ij} + \beta_9 ageD * t_{ij} + \beta_{10} sexeD_i * t_{ij} + \epsilon_{ij}$$

```
Random effects:
 Formula: ~1 + tpsbio | repere
 Structure: General positive-definite, Log-Cholesky parametrization
             StdDev    Corr
(Intercept) 13.901213 (Intr)
tpsbio       1.938013 -0.26
Residual     8.227326

Fixed effects: MDRD ~ tpsbio + iratio + iratio * tpsbio + ageD + creatD + sexeR +
                 Value Std.Error   DF   t-value p-value
(Intercept)   63.78606 1.9643390 8335  32.47202  0.0000
tpsbio        -0.01555 0.2724958 8335  -0.05705  0.9545
iratio         5.27288 1.2355953  900   4.26748  0.0000
ageD          -0.46447 0.0323770  900 -14.34564  0.0000
creatD        -0.01889 0.0085457  900  -2.21014  0.0273
sexeR          4.40669 0.9607068  900   4.58693  0.0000
sexeD          2.31850 1.0830391  900   2.14074  0.0326
itranspl       5.09771 1.4061677  900   3.62525  0.0003
tpsbio:iratio -0.02492 0.1936157 8335  -0.12870  0.8976
tpsbio:ageD   -0.01231 0.0051439 8335  -2.39342  0.0167
tpsbio:sexeD  -0.41900 0.1689676 8335  -2.47980  0.0132

Number of Observations: 9246
Number of Groups: 907
```

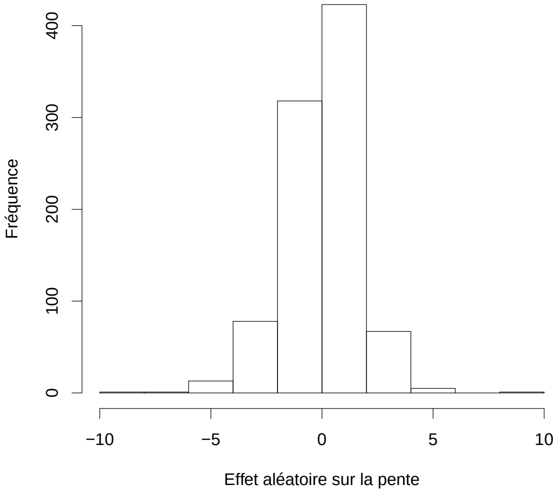
# Adéquation

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



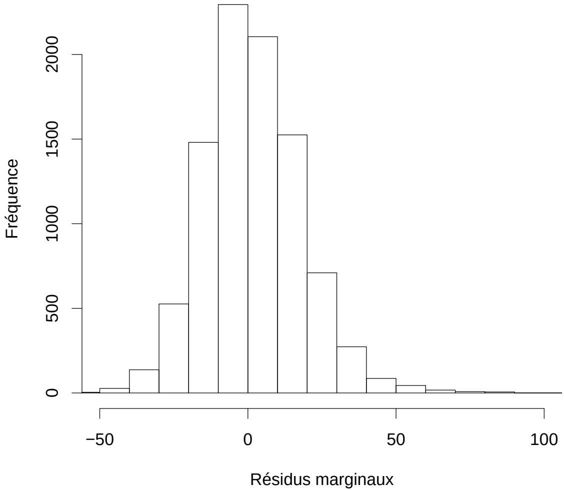
# Adéquation

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



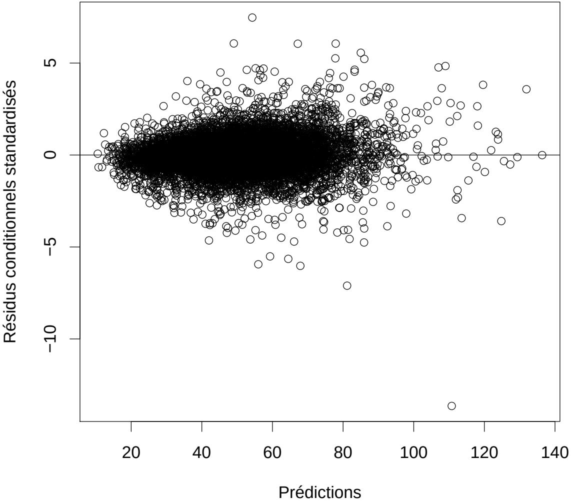
# Adéquation

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



# Adéquation

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références



# Conclusion

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Question d'intérêt :

- Effet du ratio "Poids du greffon/Poids du receveur" sur l'évolution du débit de filtration glomérulaire

$\Rightarrow$ Niveau de MDRD initial plus bas chez les sujets avec faible ratio / aux sujets avec un ratio élevé, indépendamment des autres covariables

$\Rightarrow$ Pas d'effet du ratio sur l'évolution du débit de filtration glomérulaire au cours du temps

# Plan

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

1. Introduction
2. Objectifs
3. Méthodes historiques
4. Modèle linéaire mixte
5. Logiciels
6. Stratégie de modélisation
7. Application
8. Discussion
9. Références

# Discussion

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Introduction/Objectif

- Présentation de $\neq$ exemples
- Mise en évidence des particularités de données répétées

## Méthodes historiques

- Présentation de $\neq$ approches méthodologiques simples mais limitées dans l'interprétation
- Présentation de l'ANOVA pour mesures répétées

  $\rightarrow$ possible en cas de plan équilibré

# Discussion

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Modèle linéaire mixte

- Prise en compte de plusieurs facteurs de variation

  $\rightarrow$ variables continues

  $\rightarrow$ variables catégorielles

  $\rightarrow$ interactions

- Analyse appropriée en cas de plan déséquilibré

  $\rightarrow$ présence de données manquantes

- Possibilité de modéliser des variances hétérogènes

# Discussion

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Limites du Modèle linéaire mixte

- Hypothèses du modèle
  - linéarité,
  - normalité,
  - variable réponse quantitative gaussienne

$\Rightarrow$ Modèle non linéaire mixte

- Présence de données manquantes informatives ?

# Discussion

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

## Discussion de la Stratégie de modélisation

- Proposition et non vérité absolue
- Raisonnement généralisable à d'autres modèles de regression
  - Régression linéaire simple
  - Modèles de survie
  - Modèles GLM
  - etc.

# Plan

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

1. Introduction
2. Objectifs
3. Méthodes historiques
4. Modèle linéaire mixte
5. Logiciels
6. Stratégie de modélisation
7. Application
8. Discussion
9. Références

# Reférences

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

Dartigues JF, Gagnon M, Letenneur L, Barbeger-Gateau P, Commenges D, Evaldre M and Salamon R (1992). Principal lifetime occupation and cognitive impairment in a french elderly cohort (Paquid). *American Journal of Epidemiology*, **135,** 981–988.

Dempster A, Laird N and Rubin D (1977). Maximum likelihood from incomplete data via EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B Methodological*, **39,** 1–38.

Fletcher R (2000). *Practical methods of optimization second ed*. Wiley and Sons, New York.

Giral M, Nguyen JM, Karam G, Kessler M, Hurault de Ligny B, Buchler M, Bayle F, Meyer C, Foucher Y, Martin ML, Daguin P, Soulillou JP (2005). Impact of graft mass on the clinical outcome of kidney transplants. *Journal of the American Society of Nephrology*, **16(1)**, 261–268.

Giral M, Foucher Y, Karam G, Labrune Y, Kessler M, de Ligny BH, Büchler M, Bayle F, Meyer C, Trehet N, Daguin P, Renaudin K, Moreau A and Soulillou JP (2010). Kidney and recipient weight incompatibility reduces long-term graft survival. *Journal of the American Society of Nephrologyn* **21(6)**, 1022–1029.

Goldstein H (1979). *The Design and Analysis of Longitudinal Studies*. London, Academic Press

Hand DJ, Daly F, Lunn AD, McConway KJ and Ostrowski E (1994). *A handbook of small data sets*. Chapman and Hall.

Harville D (1977). Maximum Likelihood Approaches to Variance Component Estimation and Related Problems. *Journal of the American Statistical Association*, **72,** 320–339.

Ladrière M, Foucher Y, Legendre C, Kamar N, Garrigue V, Morélon E, Kessler M, Soulillou JP and Giral M. (2010). The western europe cohort of kidney transplanted recipients - the DIVAT network. *Clinical transplants*, 460–461.

Laird NM and Ware JH (1982). Random-effects models for longitudinal data. *Biometrics*, **38,** 963–974.

Pothoff S and Roy FN (1964). A generalized multivariate analysis of variance model useful especially for growth curve problem. *Biometrika*, **51,** 313–326

Verbeke G et Molenberghs G (2000). *Linear mixed models for longitudinal data*. Springer Series in Statistics : New York.

Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, **59,** 254–262.

Introduction
Objectifs
Méthodes historiques
Modèle linéaire mixte
Rappels
Modèle linéaire mixte
Inclusion de covariables
Exercices
Généralisation
Estimation
Tests d'inférence
Adéquation du modèle
Logiciels
Stratégie de modélisation
Application
Discussion
Références

# Merci de votre attention