

# Modèles mixtes

18 novembre 2020

Ce laboratoire doit être remis le **2 décembre à 17h sur Moodle**. Dans votre réponse pour chaque question, veuillez inclure une copie du code R utilisé (s'il y a lieu) et des résultats obtenus.

## 1. Rendement du blé en fonction de l'engrais et de l'humidité du sol

Le tableau de données `Wheat` inclus dans le package `nlme` présente le résultat d'une expérience agricole visant à mesurer l'effet de la quantité d'engrais (`fertilizer`) et de l'humidité du sol (`Moisture`) sur le rendement d'une variété de blé (`DryMatter`, la variable réponse). Les plants ont été répartis en 12 plateaux (`Tray`). L'humidité du sol est constante dans un plateau, mais chaque plateau est divisé en 4 sections recevant chacune une quantité différente d'engrais. Il y a donc 48 mesures individuelles du rendement, soit 4 par plateau.

```
library(nlme)
data(Wheat)
head(Wheat)
```

```
## Grouped Data: DryMatter ~ fertilizer | Tray
##   Tray Moisture fertilizer DryMatter
## 1     1         10         2    3.3458
## 2     1         10         4    4.3170
## 3     1         10         6    4.5572
## 4     1         10         8    5.8794
## 5     2         10         2    4.0444
## 6     2         10         4    4.1413
```

- Estimez les paramètres d'un modèle linéaire incluant les effets additifs de l'engrais et de l'humidité sur le rendement. Est-il correct d'ignorer le groupement des mesures par plateau? Justifiez votre réponse à partir d'un graphique des résidus par plateau.
- Ajustez maintenant un modèle mixte qui inclut les mêmes effets fixes, ainsi qu'un effet aléatoire représentant la structure groupée des données. Pour chacun des deux effets fixes (engrais et humidité), comment l'erreur-type des estimés change-t-elle par rapport au modèle sans effet aléatoire en (a)? Expliquez ces différences en fonction de la structure de l'expérience.
- À partir de graphiques de diagnostic appropriés, vérifiez si les résidus du modèle mixte en (b) ne montrent aucune tendance résiduelle et s'ils suivent une distribution normale.
- Calculez la corrélation intra-classe pour le modèle mixte en (b). Quelle est l'interprétation mathématique de cette valeur?

## 2. Courbes de croissance d'épinettes

Le tableau `Spruce` inclus dans le package `nlme` contient des données sur la croissance de 79 épinettes. Le logarithme du volume (`logSize`) pour chaque épinette (identifiée par un numéro d'arbre `Tree`) a été mesuré à 13 différents moments à partir du début de l'expérience (nombre de jours `days`). Les épinettes sont aussi réparties entre 4 placettes (`plot`).

```
data(Spruce)
head(Spruce)
```

```
## Grouped Data: logSize ~ days | Tree
##      Tree days logSize plot
## 1 01T01  152    4.51    1
## 2 01T01  174    4.98    1
## 3 01T01  201    5.41    1
## 4 01T01  227    5.90    1
## 5 01T01  258    6.15    1
## 6 01T01  469    6.16    1
```

- (a) Puisque la croissance n'est pas linéaire dans le temps, nous transformons le nombre de jours en facteurs, ce qui permettra d'estimer indépendamment la croissance à chaque point dans le temps.

```
Spruce$days <- as.factor(Spruce$days)
```

Estimez les paramètres du modèle linéaire de `logSize` en fonction du nombre de jours écoulé et répondez aux questions suivantes: (i) Comment interprétez-vous l'ordonnée à l'origine de ce modèle? (ii) Quelle est la variation moyenne de `logSize` entre le jour 152 et le jour 201 et quelle est son erreur-type?

- (b) Ajustez maintenant un modèle mixte qui tient compte du fait qu'il s'agit de mesures répétées sur les mêmes arbres et répondez aux questions suivantes (*Note*: Ignorez la placette `plot` pour l'instant.) (i) Pour ce modèle, que représentent l'écart-type de l'effet aléatoire et l'écart-type résiduel, respectivement? (ii) Selon ces résultats, pourquoi est-il avantageux de mesurer les mêmes arbres à chaque jour d'échantillonnage afin d'estimer une courbe de croissance?
- (c) Quelle est la corrélation intra-classe pour le modèle en (b)? Selon ce résultat, est-ce que la variation de taille parmi les arbres à un moment donné de l'expérience est due davantage (i) aux différences de taille entre les arbres au premier jour 152 ou (ii) à la variation aléatoire de croissance d'un arbre à l'autre pour chaque période de temps?
- (d) Ajoutez au modèle en (b) un effet aléatoire de la placette `plot`. Selon ce modèle, est-ce que la taille des arbres diffère de façon marquée entre les placettes?