



UNIVERSITÉ
DE MONTPELLIER

Université de Montpellier
Faculté des Sciences

PROJET HMMA307

Spécialité : *MIND/SIAD*

Etudiant :
Cherif AMGHAR

Sous le thème

Modèles mixtes linéaires

Enseignant :

Pr. Joseph Salmon

08 Novembre 2020

Table des matières

Introduction	3
1 Modèles mixtes linéaires avec un effet aléatoire	4
1.1 Propriétés des modèles mixtes	4
1.2 Ajustement du modèle : un effet fixe et aléatoire	5
2 Modèles mixtes linéaires avec des facteurs croisés ou imbriqués	6
2.1 Ajuster un modèle avec un effet fixe et plusieurs effets aléatoires	7
2.2 Vérification des hypothèses	7

Introduction

Un modèle linéaire mixte est un modèle pour lequel le modèle comprend à la fois des effets fixes et des effets aléatoires. Les MLM incluent des variables à effets fixes et aléatoires. Le mélange entre les deux est à l'origine du nom. Les effets fixes décrivent les relations entre les covariables et la variable dépendante pour une population entière, les effets aléatoires sont spécifiques à l'échantillon.

En d'autres termes, un effet aléatoire est un effet dont nous ne voulons pas généraliser les propriétés (les modalités ont été choisies de manière aléatoire dans quelque chose de plus grand) et un effet fixe est un effet dont on veut généraliser les propriétés. Il s'agit de la variable manipulée dont nous avons choisi les niveaux spécifiques.

Les données qu'on analyse ici sont des dénombrements d'invertébrés sur 3-4 sites dans chacun des 7 estuaires (choisis au hasard). Ici, les estuaires sont l'effet aléatoire, car il existe un grand nombre d'estuaires possibles, et on n'échantillonne que quelques-uns d'entre eux au hasard, mais on aime faire des inférences sur les estuaires en général.

Tout au long de ce projet, on s'intéresse à deux principaux types de modèles linéaires mixtes, modèle mixte à un effet aléatoire et le deuxième pour deux effets aléatoires, on explique les sorties sous python.

Chapitre 1

Modèles mixtes linéaires avec un effet aléatoire

On analyse un ensemble de données visant à tester l'effet de la pollution de l'eau sur l'abondance de certains invertébrés marins subtidiaux en comparant des échantillons d'estuaires modifiés et vierges. Comme le nombre total est important, on suppose que les données sont continues.

Modèles mixtes à un effet aléatoire est une introduction aux modèles mixtes pour une réponse continue avec un effet aléatoire . On apprend à vérifier les hypothèses et à faire des inférences, y compris le bootstrap paramétrique.

l'équation du modèle est :

$$y_{ij} = \beta_0 + u_{0i} + \beta_1 \times t_{ij} + \varepsilon_{ij}$$

i indexe les individus et j indexe les variables(modification, estuaries, site, hydriod,...) . Ainsi, $i \in 1, \dots, 54$ et $j \in 1, \dots, 6$.

1.1 Propriétés des modèles mixtes

Les modèles mixtes font des hypothèses importantes :

1. $(y|x)$ sont i.i.d et suit loi normale
2. $V(y|x)$ est constante
3. y s'écrit sous forme linéaire en fonction de x et z (l'effet aléatoire) .
4. y est indépendant de z
5. z suit loi normale.

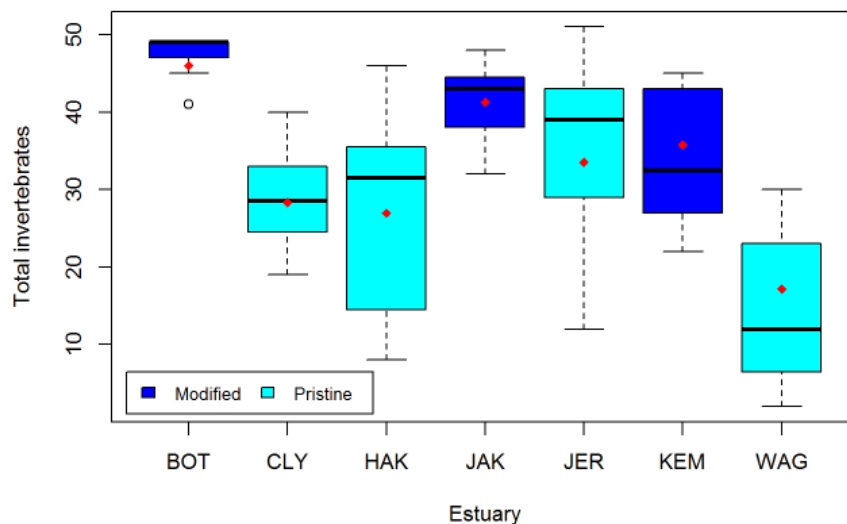
Il est important de vérifier que nos données répondent aux hypothèses du modèle qu'on a utilisé. on regarde toutes les hypothèses dans l'ordre.

1.2 Ajustement du modèle : un effet fixe et aléatoire

Dans cet ensemble de données, hypothèses vérifiées, on a un effet fixe (Modification ; modifié vs vierge) et un effet aléatoire (Estuaire). Pour ajuster un modèle d'abondance totale. On ajuste le modèle complet par maximum de vraisemblance, et un deuxième modèle qui n'a pas l'effet fixe de Modification. On obtient que l'effet fixe c'est à dire l'effet de modification permet d'affecter sur le modèle et que le modèle qui contient la plus petit valeur de AIC est le modèle complet. On compare ces deux modèles avec un test de rapport de vraisemblance.

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
ft.estu.0	3	415.02	420.99	-204.51	409.02			
ft.estu	4	411.92	419.87	-201.96	403.92	5.1055	1	0.02385 *

On constate qu'il existe des preuves d'un effet de modification ($p = 0,02385$). Et pour tester les variables aléatoires soit on utilise le test d'anova, mais les p-value sont très approximatives, on utilise un bootstrap paramétrique. Il s'agit d'une méthode basée sur la simulation.



Pour un modèle mixte simple avec un effet aléatoire(Estuary), un graphique des données brutes avec les moyens du modèle superposés est une possibilité de voir les choses.

Chapitre 2

Modèles mixtes linéaires avec des facteurs croisés ou imbriqués

Deux facteurs sont croisés lorsque chaque catégorie (niveau) d'un facteur coexiste dans la conception avec chaque catégorie de l'autre facteur. En d'autres termes, il y a au moins une observation dans chaque combinaison de catégories pour les deux facteurs.

Un facteur est imbriqué dans un autre facteur lorsque chaque catégorie du premier facteur coexiste avec une seule catégorie de l'autre. En d'autres termes, une observation doit être dans une catégorie du facteur 2 pour avoir une catégorie spécifique de facteur 1. Toutes les combinaisons de catégories ne sont pas représentées. Il existe également des plans intermédiaires qui sont partiellement croisés, où certains niveaux d'un facteur se produisent dans plusieurs niveaux (mais pas tous) du deuxième facteur. Ces conceptions ont souvent été enseignées comme des problèmes séparés avec différentes façons d'effectuer des analyses de variance (ANOVA) selon que vous avez des facteurs croisés ou imbriqués.

on analyse un ensemble de données visant à tester l'effet de la pollution de l'eau sur l'abondance de certains invertébrés marins subtidiaux en comparant des échantillons d'estuaires modifiés et vierges. Comme le nombre total est important, nous supposons que les données sont continues.

2.1 Ajuster un modèle avec un effet fixe et plusieurs effets aléatoires

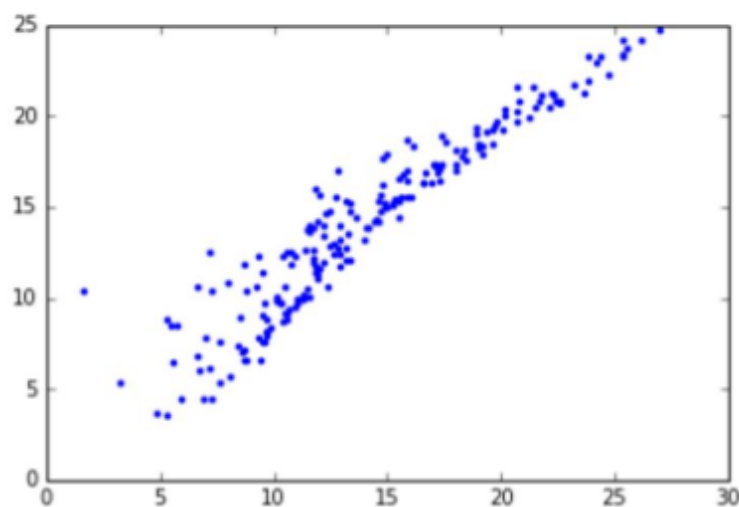
Dans cet ensemble de données, on a un effet fixe (Modification; modifié vs vierge) et deux effets aléatoires (Estuaire et Site). Le site est imbriqué dans l'estuaire car chaque site ne peut appartenir qu'à un seul estuaire. On remarque que l'estuaire JAK et l'estuaire JER ont chacun des sites numérotés 1, même si ces sites ne sont en aucun cas connectés. c'est pour ça on imbrique la variable site dans le Etuary on peut faire cette solution dans le premier temps c'est à dire dans l'étape de saisie de données mais si les choses sont les mêmes elle devraient être étiquetées de la même façon et dans le cas inverse elles devraient être différemment.

2.2 Vérification des hypothèses

Les hypothèses sont les mêmes que pour un facteur aléatoire.

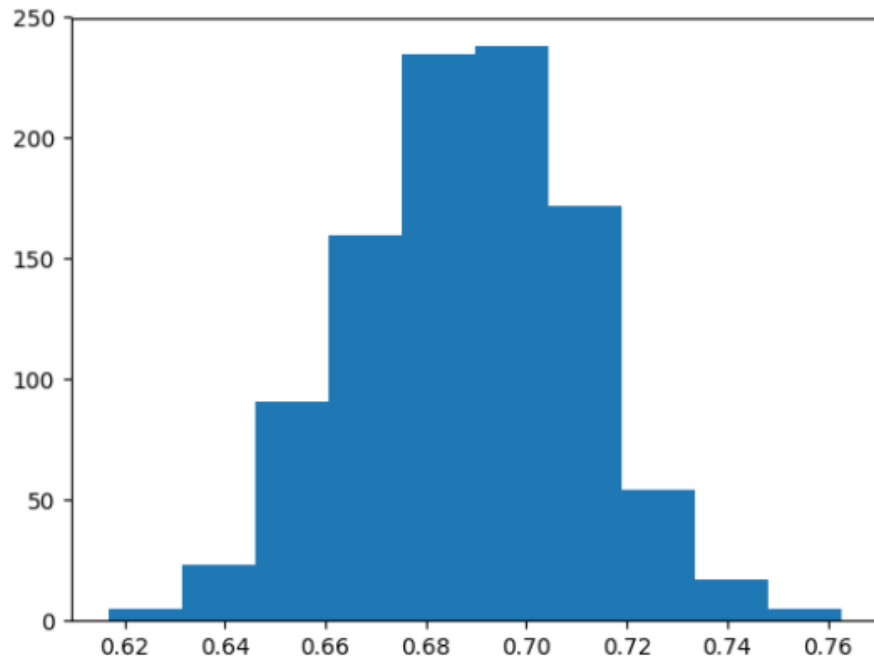
En bref, les hypothèses 1 et 5 ne peuvent pas être vérifiées, mais peuvent être assurées en prélevant des échantillons aléatoires, et l'hypothèse 6 n'est pas cruciale et difficile à vérifier. Pour vérifier l'hypothèse 2, on recherche une relation en ligne droite sur le graphique des résidus. Pour vérifier les hypothèses 3 et 4, on recherche une forme en éventail et une forme en U sur le tracé résiduel par rapport à l'ajustement. Ajustement avec anova.

```
Out[21]: [<matplotlib.lines.Line2D at 0xbb520b8>]
```



Test d'hypothèse pour les effets aléatoires :

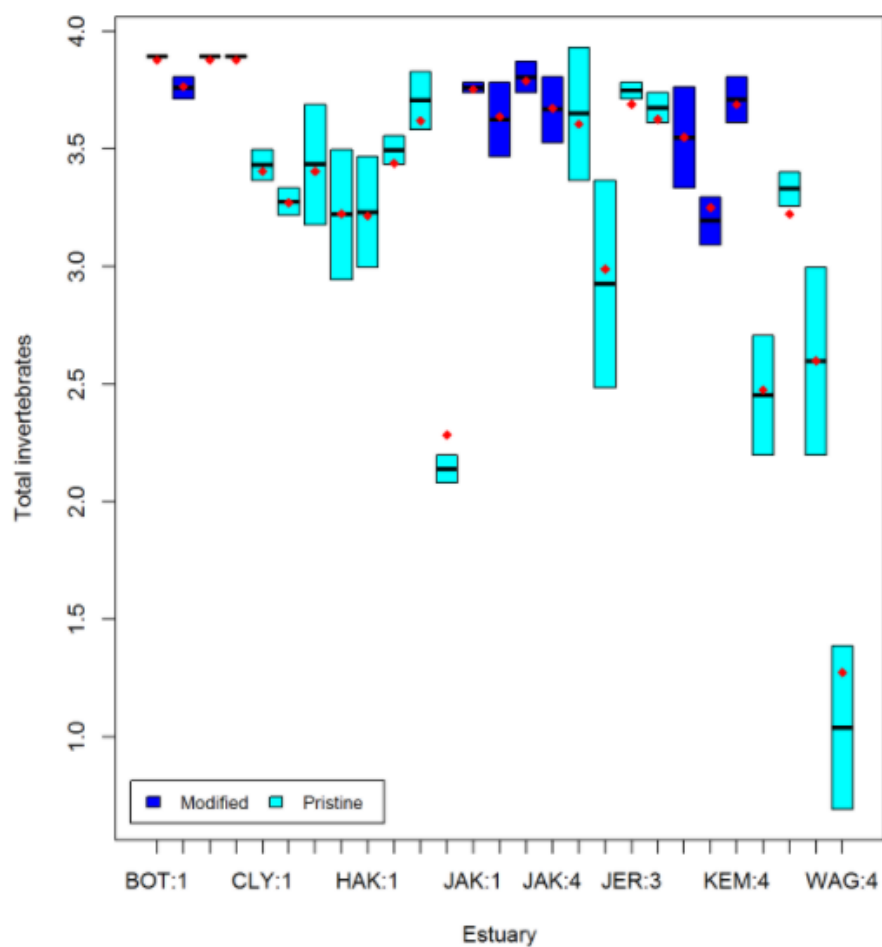
Comme dans les modèles mixtes 1, on utilise un bootstrap paramétrique. On va tester si on doit avoir un effet aléatoire pour Site étant donné que on a un effet aléatoire pour Estuaire dans le modèle.



Distribution de l'exactitude de la classification à l'aide du bootstrap

les intervalles de confiance sont rapportés, montrant qu'il existe une probabilité de 95(percent) que les intervalles de confiance 64,4(percent) et 73,0(percent) couvrent la véritable compétence du modèle.

On peut interpréter les résultats :en utilisant :



On peut tracer dans chaque site, mais c'est un peu étrange pour un boxplot car il n'y a que deux observations par site, on peut utiliser si on a plus d'observations dans chaque site.