

Les Modèles Linéaires Mixtes

HMMA 307 : Modèles Linéaires Avancés

KANDOUCI Walid

[https://github.com/WalidKandouci/
HMMA307--How-Linear-Mixed-Model-Works.git](https://github.com/WalidKandouci/HMMA307--How-Linear-Mixed-Model-Works.git)

Université de Montpellier



Sommaire

Introduction

Exemple: sleepstudy

Application

Conclusion

Introduction

- ▶ **Modèles mixtes:** des modèles comportant à la fois des facteurs à effets fixes (ces effets entrant dans la définition de la moyenne du modèle) et des facteurs à effets aléatoires (ces effets entrant, quant à eux, dans la définition de la variance du modèle)

$$y = X\beta + \epsilon$$

- X matrice ($N \times p$)
- β ($p \times 1$): les coefficients de régression pour chaque variable indépendante du modèle
- ϵ ($N \times p$) contient les erreurs (résidus) du modèle

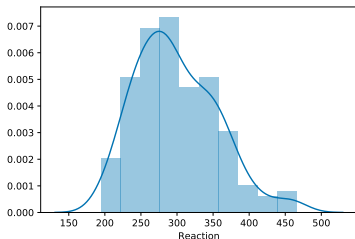
Exemple: sleepstudy

	Reaction	Days	Subject
1	249.5600	0	308
2	258.7047	1	308
3	250.8006	2	308
4	321.4398	3	308
5	356.8519	4	308

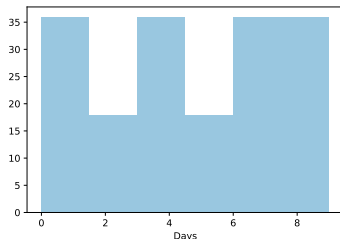
► **180 individus** × **3 variables**

- **Reaction:** Temps de réaction en moyenne (en milisecondes)
- **Days:** Nombre de jours de privation de sommeil
- **Subject:** Numéro du sujet sur lequel l'observation a été faite.

Nos variables "Reactions" et "Days"



(a) Distribution des réactions



(b) Distribution des jours

Figure: Analyse de nos variables "Reaction" et "Days"

- ▶ **Reaction:** Temps de réaction en moyenne (en millisecondes)
- ▶ **Days:** Nombre de jours de privation de sommeil

Évolution du temps de réaction en fonction des jours

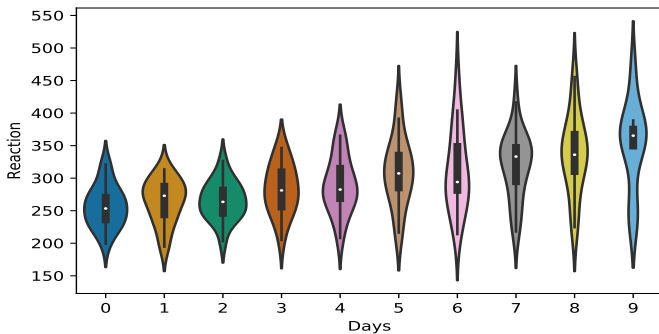


Figure: violins-plots de nos données, jours /réactions.

- Le temps réaction par rapport aux jours a une tendance à la hausse avec des variations entre les jours et les individus.

Évolution du temps de réaction en fonction des jours

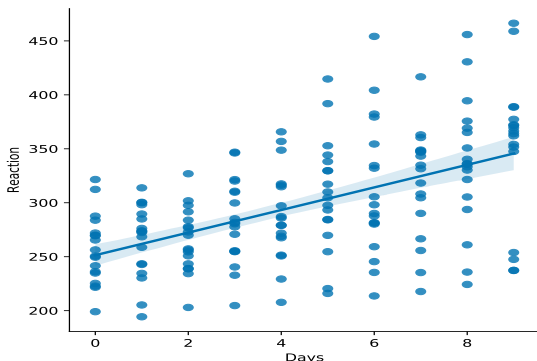


Figure: Représentation du temps de réaction en fonctions des jours

- **Reaction:** Temps de réaction en moyenne (en milisecondes)
- **Days:** Nombre de jours de privation de sommeil

Application

- ▶ Nous allons implémenter maintenant les méthodes **MLL**, **OLS** et **GLM**
- ▶ Nous allons comparer leurs erreur quadratique moyenne (**RMSE**)
- ▶ Analyser le résultat obtenue grace a la méthode **MLL**

Application

```
# OLS
modelOLS = smf.ols("Reaction ~ Days",
                   data, groups=data["Subject"])
resultOLS = modelOLS.fit()
print(resultOLS.summary())

# GLM
modelGLM = smf.glm("Reaction ~ Days",
                   data, groups=data["Subject"])
resultGLM = modelGLM.fit()
print(resultGLM.summary())

#LMM
md = smf.mixedlm("Reaction ~ Days",
                 data, groups=data["Subject"])
mdf = md.fit()
print(mdf.summary())
```

Comparaison des RMSE

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

Method	RMSE
LMM	29.410624
GLM	47.448898
OLS	47.448898

- On remarque que l'erreur quadratique moyenne du **LMM** est la plus petite.

LMM Résultat

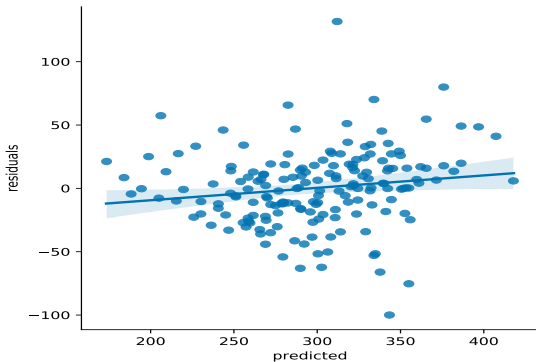


Figure: Représentation du temps de réaction en fonctions des jours

- Le temps de réactions étant mieux répartis en fonction des jours.

Conclusion

- ▶ Les modèles linéaires mixtes sont utilisés pour tenir compte de la non-indépendance des données
- ▶ le modèle fourni généralement un meilleur ajustement et explique plus de variation dans les données.