

# 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)

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

- $X$ : Matrice ( $N \times p$ )
- $\beta$ : Les coefficients de régression pour chaque variable indépendante du modèle
- *epsilon*: Contient les erreurs (résidus) du modèle
- $Z$ : Matrice qui contient les valeurs observées pour chaque individu  $N$  pour chaque covariable  $q$  des effets aléatoires.
- $u$  est un vecteur ( $q \times 1$ ) qui contient les effets aléatoires des  $q$  covariables de  $Z$

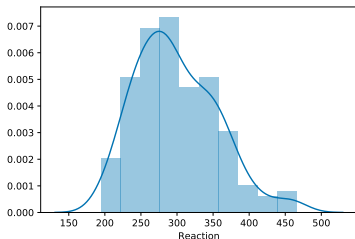
## Exemple: sleepstudy

|   | 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     |

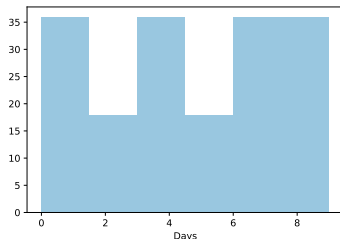
► **180 individus** × **3 variables**

- **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

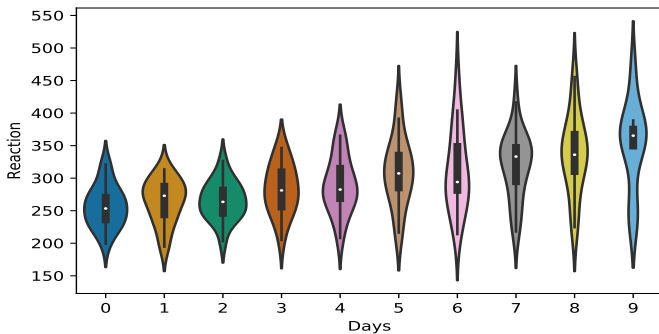


Figure: violins-plots de nos données, jours /réactions.

- 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

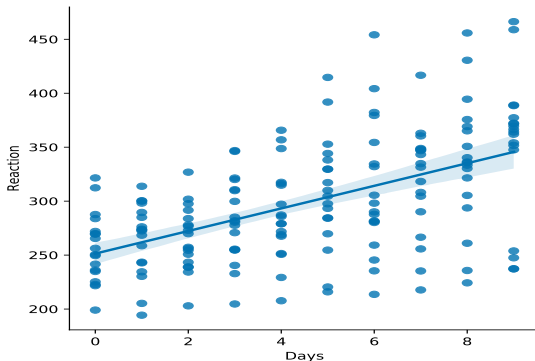


Figure: Représentation du temps de réaction en fonctions des jours

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

# Application

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



# 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())
```

# Application

## OLS Regression Results

|           | Coef.    | Std.Err. | z      | $P >  z $ | [0.0250.975]    |
|-----------|----------|----------|--------|-----------|-----------------|
| Intercept | 251.4051 | 6.610    | 38.033 | 0.000     | 238.361 264.449 |
| Days      | 10.4673  | 1.238    | 8.454  | 0.000     | 8.024 12.911    |

## Generalized Linear Model Regression Results

|           | Coef.    | Std.Err. | z      | $P >  z $ | [0.0250.975]    |
|-----------|----------|----------|--------|-----------|-----------------|
| Intercept | 251.4051 | 6.610    | 38.033 | 0.000     | 238.449 264.361 |
| Days      | 10.4673  | 1.238    | 8.454  | 0.000     | 8.040 12.894    |

## Mixed Linear Model Regression Results

|           | Coef.    | Std.Err. | z      | $P >  z $ | [0.0250.975]    |
|-----------|----------|----------|--------|-----------|-----------------|
| Intercept | 251.405  | 9.747    | 25.794 | 0.000     | 232.302 270.508 |
| Days      | 10.467   | 0.804    | 13.015 | 0.000     | 8.891 12.044    |
| Group Var | 1378.176 | 17.156   |        |           |                 |

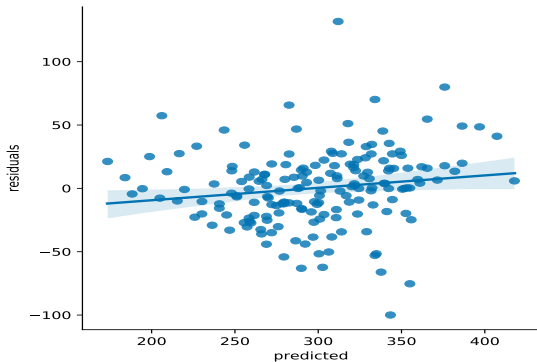
## Comparaison des RMSE

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

| Method | RMSE      |
|--------|-----------|
| LMM    | 29.410624 |
| GLM    | 47.448898 |
| OLS    | 47.448898 |

- On remarque que l'erreur quadratique moyenne du **MML** 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.