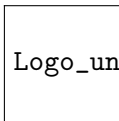


HMMA 307 : Advanced Linear Modeling

Linear mixed models with LM and REML

Delage Cindy

Université de Montpellier



Logo_universit_montpellier.png

Table of Contents

- 1 Introduction to the linear mixed models
- 2 Beta estimation
- 3 Estimation of the parameters variances
- 4 Conclusion

Table of Contents

- 1 Introduction to the linear mixed models
- 2 Beta estimation
- 3 Estimation of the parameters variances
- 4 Conclusion

Introduction to the linear mixed models

Remark

Fixed effect : can be generalised

Random effect : sample-specific

Model

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

$$\begin{pmatrix} u \\ \epsilon \end{pmatrix} \sim (\mathcal{N}(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}))$$

X and Z are design matrix, β is the fix effect vector and u is the random effect vector.

Meaning of this project

Definition

- ML : maximum likelihood regression
- REML : restrained maximum likelihood regression

Goals

- Compare ML and REML
- Application with Python
- explain the differences

Table of Contents

- 1 Introduction to the linear mixed models
- 2 Beta estimation**
- 3 Estimation of the parameters variances
- 4 Conclusion

Manual estimation

Beta

For linear mixed model:

$$\hat{\beta} = (X^{\top} V^{-1} X)^{-1} X^{\top} V^{-1} Y$$

V is the variance of Y , $V = ZGZ^{\top} + R$

Manual estimation : Using these formulas in Python, we obtain the following parameters :

$$\beta = \begin{pmatrix} 70.18571429 \\ 5.71428571 \\ 0.91428571 \end{pmatrix}$$

ML estimation

Mixed Linear Model Regression Results

```
=====
Model:                               MixedLM           Dependent Variable:      pres
No. Observations:                     21                Method:              ML
No. Groups:                           7                Scale:                2.3045
Min. group size:                      3                Log-Likelihood:       -52.8621
Max. group size:                      3                Converged:            No
Mean group size:                      3.0
=====
```

```
-----
                                Coef.  Std.Err.   z    P>|z|  [0.025  0.975]
-----+-----
Intercept                      70.186    1.036  67.718  0.000  68.154  72.217
metal[T.i]                     5.714    1.865   3.064  0.002   2.059   9.369
metal[T.n]                     0.914    1.379   0.663  0.507  -1.789   3.618
Group Var                      5.215
Group x df.metal[T.i] Cov      2.267    2.974
df.metal[T.i] Var              19.736
Group x df.metal[T.n] Cov      0.363    2.380
df.metal[T.i] x df.metal[T.n] Cov 12.379    1.622
df.metal[T.n] Var              8.712
=====
```


REML estimation

Mixed Linear Model Regression Results

```
=====
Model:                               MixedLM           Dependent Variable:      pres
No. Observations:                     21              Method:                REML
No. Groups:                           7              Scale:                  0.9865
Min. group size:                      3              Log-Likelihood:         -49.3760
Max. group size:                      3              Converged:              No
Mean group size:                      3.0
=====
```

```
-----
                                Coef.  Std.Err.   z    P>|z|  [0.025  0.975]
-----+-----
Intercept                      70.186    1.124  62.434  0.000  67.982  72.389
metal[T.i]                     5.714    2.053   2.784  0.005   1.691   9.738
metal[T.n]                     0.914    1.592   0.574  0.566  -2.206   4.035
Group Var                      7.860
Group x df.metal[T.i] Cov      -2.288    5.234
df.metal[T.i] Var              27.526
Group x df.metal[T.n] Cov      -3.292    3.036
df.metal[T.i] x df.metal[T.n] Cov 18.563
df.metal[T.n] Var              15.769
=====
```

Table of Contents

- 1 Introduction to the linear mixed models
- 2 Beta estimation
- 3 Estimation of the parameters variances**
- 4 Conclusion

Variances estimated by ML

Likelihood

$$f_Y(y_1, \dots, y_n) = \frac{e^{\frac{-1}{2}(Y-X\beta)^T V^{-1}(Y-X\beta)}}{\sqrt{2\pi^n |V|}}$$

Optimizing the log likelihood with python gives :

$$\begin{pmatrix} \sigma_Y^2 \\ \sigma_{residuals}^2 \end{pmatrix} = \begin{pmatrix} 9.812383 \\ 8.889932 \end{pmatrix}$$

Variances estimated by REML

Log likelihood

$$-2 \log(\beta, Y) = \log(|V|) + \log(|X^T V^{-1} X|) + (Y - X\beta)^T V^{-1} (Y - X\beta) + \text{Cste}$$

Optimisation

Optimizing the log likelihood with python gives :

$$\begin{pmatrix} \sigma_Y^2 \\ \sigma_{\text{residuals}}^2 \end{pmatrix} = \begin{pmatrix} 11.44780323 \\ 10.3715852 \end{pmatrix}$$

Reason of the differences

Main differences

- Variance of the parameters
- Confidence intervals regression

ML has a bias

$$\mathbb{E}[\hat{\sigma}^2] = \sigma^2 - \frac{\sigma^2}{N}$$

Table of Contents

- 1 Introduction to the linear mixed models
- 2 Beta estimation
- 3 Estimation of the parameters variances
- 4 Conclusion**

Conclusion

Conclusion

REML corrects the bias of ML by adding a term in the log likelihood. The confidence intervals is larger, so are the variances calculated.