From Classroom to Clinic: Applying Linear Mixed Models to Understand Real-World Medical Research Data

Joshua J. Cook, M.S., Syed Ahzaz H. Shah, B.S., Jacob Hernandez, B.S., Sara Basili, M.S., Achraf Cohen, Ph.D.

University of West Florida (UWF) - STA 6257 (Advanced Statistical Modeling)

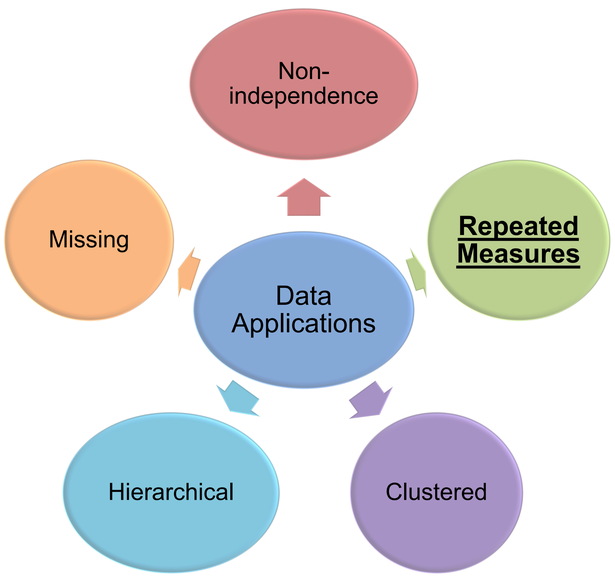
Department of Mathematics and Statistics
UNIVERSITY of WEST FLORIDA

INTRODUCTION

Defining a **linear mixed model (LMM):**

"advanced statistical tools designed to analyze data that exhibit complex structures, such as hierarchical organization, repeated measures, and random effects"

-IBM



What are mixed effects?

- **Fixed effects** – the variable of interest and controlled variables (*i.e., those that are directly)*
- **Random effects** –measured random variability between individuals, clusters, or hierarchies (*i.e., those that are not typically measured directly)*

MATHEMATICAL FOUNDATIONS

Mathematical Foundations: Linear Algebra

Simple **LMMs** can be **defined** by:

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

- **Y** is the response vector.
- **X** is the design matrix for **fixed effects.**
- **β** is the vector of **fixed effect**s
- **Z** is the design matrix for **random effects**.
- ***u*** is the vector of **random effects**
- **ϵ** is the vector of **residual errors.**

An Example:

- 2-level longitudinal structure
- 100 students (*random effect*)
- 10 test scores (*dependent variable – fixed effect*)
- Associated study time each (independent variable - *fixed effect*)
- N = 1000 (students * tests)
- J = 10 (scores per student)
- P = 2 (random intercept + fixed effect)
- This LMM can be **defined** by:

$$Y_{1000\times1} = X_{1000\times2}\,\beta_{2\times1} + Z_{1000\times10}\,u_{10\times1} + \epsilon_{1000\times1}$$

Level 1 (Time):

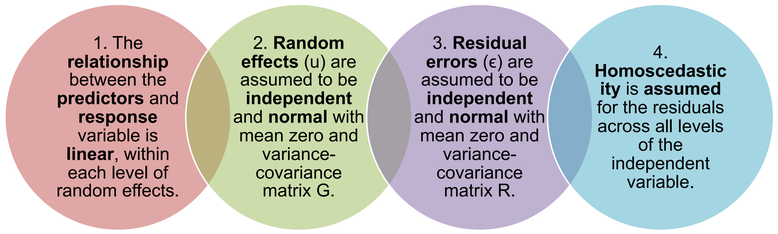
$$Y_{ti} = \beta_{0j} + \beta_{1j}\cdot \mathrm{StudyTime}_{ti} + e_{ti}$$

Level 2 (Student):

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

- γ_{00} is the grand intercept mean.
- u_{0j} is the deviation of the j_{th} **group**

ASSUMPTIONS



THE CAPSTONE PROJECT

The Dataset:

- **Longitudinal**, retrospective study.
- Impact of BMI on **airway resistance** and reactance in children with Sickle Cell Disease (C-SCD) and asthma (C-Asthma).
- Same patients (n=85), **multiple (repeated) measurements** over time (6 visits total).
- 14 measured variables (*fixed effects*)
- 2 grouping variables (*random effects*)
- 4 of the fixed effects are *dependent variables* representing airway resistance.
- **Data required winsorization to satisfy LMM assumption of normality. All other assumptions were met.**

The Final Model:

$$\begin{aligned}Y_{ijk} = {} & (\beta_0 + u_{0j} + u_{1j}\times Observation\_number2_{ijk} + \ldots + u_{5j}\times Observation\_number6_{ijk}) \\ & + \beta_1(BMI_{ijk}) + \beta_2(AsthmaYes_{ijk}) + \beta_3(ICSYes_{ijk}) + \beta_4(LABA_{ijk}) \\ & + \beta_5(Gender_{ijk}) + \beta_6(Age\_months_{ijk}) + \beta_7(Height\_cm_{ijk}) + \beta_8(Weight\_Kg_{ijk}) \\ & + \beta_9(Group_{ijk}) + \epsilon_{ijk}\end{aligned}$$

- **Fixed effects:** Diagnosis (C-SCD or C-Asthma) and all demographic and co-morbidity variables.

- **Random effects:** Subject and observation number.

- Adding *observation_number* increased AIC, but these effects were **kept in the final model due to the goal of the research project** (predicting airway resistance).
- All **assumptions satisfied.**
- Akaike Information Criterion **(AIC) = 1801.60.**
- **Smaller Mean Squared Error (41.95 v. 117.1 and 270.60) and Mean Absolute Error (5.07 v. 8.48 and 12.15)** than previous models.

CONCLUSION AND ADDITIONAL RESOURCES

Conclusion:

- **LMM's are versatile tools** for modeling complex relations with multiple effects (fixed and random), as well as missing and non-independent data.
- For the given capstone dataset, the **generated LMM can reliably predict** measures of airway resistance and reactance given demographic and co-morbidity data.
- This model can be reliably used for **both children with Sickle Cell Disease and those with asthma.**

View our FULL literature, report, and slides for the capstone project:

REFERENCES

1. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015), "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, 67, 1–48. https://doi.org/10.18637/jss.v067.i01.
2. Galecki, A. T., Kathleen B. Welch (2014), *Linear Mixed Models: A Practical Guide Using Statistical Software, Second Edition*, New York: Chapman; Hall/CRC. https://doi.org/10.1201/b17198.