

Statistical Analysis of Repeated Measurements Data

Dimitris Rizopoulos

Department of Biostatistics, Erasmus University Medical Center

`d.rizopoulos@erasmusmc.nl`

April 18 – 22, 2016

Contents

| | | |
|----------|--|-----------|
| 1 | Motivating Data Sets | 1 |
| 1.1 | Motivating Longitudinal Studies | 2 |
| 1.2 | Features of Longitudinal Data | 14 |
| 1.3 | Review of Key Points | 26 |
| | | |
| 2 | Marginal Models for Continuous Data | 27 |
| 2.1 | Simple Methods | 28 |
| 2.2 | Review of Linear Regression | 37 |
| 2.3 | Marginal Models | 46 |

| | |
|--|-----|
| 2.4 Interpretation | 52 |
| 2.5 Estimation | 63 |
| 2.6 Fitting Marginal Models in R | 69 |
| 2.7 Covariance Matrix | 73 |
| 2.8 Model Building | 84 |
| 2.9 Hypothesis Testing | 87 |
| 2.10 Confidence Intervals | 111 |
| 2.11 Design Considerations - Sample Size | 113 |
| 2.12 Residuals | 118 |
| 2.13 Review of Key Points | 134 |

| | | |
|----------|---|------------|
| 3 | The Linear Mixed Effects Model | 136 |
| 3.1 | The Linear Mixed Model | 137 |
| 3.2 | Interpretation | 143 |
| 3.3 | Hierarchical vs Marginal | 151 |
| 3.4 | Estimation | 161 |
| 3.5 | Mixed-Effects Models in R | 171 |
| 3.6 | Nested and Crossed Random Effects* | 179 |
| 3.7 | Mixed Models with Correlated Errors | 190 |
| 3.8 | Time-Varying Covariates* | 196 |
| 3.9 | Model Building | 206 |
| 3.10 | Hypothesis Testing | 209 |

| | |
|-------------------------------------|-----|
| 3.11 Residuals | 210 |
| 3.12 Review of Key Points | 220 |

4 Marginal Models for Discrete Data 223

| | |
|---|-----|
| 4.1 Review of Generalized Linear Models | 224 |
| 4.2 Generalized Estimating Equations | 237 |
| 4.3 Interpretation | 245 |
| 4.4 Generalized Estimating Equations in R | 252 |
| 4.5 Working Correlation Matrix | 255 |
| 4.6 Hypothesis Testing | 266 |
| 4.7 Review of Key Points | 275 |

| | | |
|----------|--|------------|
| 5 | Mixed Models for Discrete Data | 277 |
| 5.1 | Generalized Linear Mixed Models | 278 |
| 5.2 | Interpretation | 285 |
| 5.3 | Estimation | 304 |
| 5.4 | GLMMs in R | 306 |
| 5.5 | Model Building | 308 |
| 5.6 | Hypothesis Testing | 310 |
| 5.7 | Review of Key Points | 312 |
| | | |
| 6 | Statistical Analysis with Incomplete Grouped Data | 315 |
| 6.1 | Missing Data in Longitudinal Studies | 316 |

| | |
|---|-----|
| 6.2 Missing Data Mechanisms | 321 |
| 6.3 Analysis with Incomplete Data | 336 |
| 6.4 Summary | 358 |
| 6.5 Review of Key Points | 360 |

7 Closing 362

| | |
|----------------------------------|-----|
| 7.1 Concluding Remarks | 363 |
|----------------------------------|-----|

Practicals 367

| | |
|---|-----|
| Practical 1: Marginal Models Continuous | 368 |
| Practical 2: Mixed Models Continuous | 378 |
| Practical 3: Marginal Models Discrete | 387 |

| | |
|--|-----|
| Practical 4: Mixed Models Discrete | 395 |
|--|-----|

What is this Course About

Grouped data arise in a wide range of disciplines

- Typical examples of grouped data
 - ▷ *repeated measurements*: measuring the same outcome multiple times on the same sample unit (e.g., biomarkers in patients)
 - ▷ *multilevel data*: outcomes measured on sample units that are organized in different levels (e.g., patients in medical centers or students in schools)

What is this Course About (cont'd)

- Statistical analysis of grouped data
 - ▷ Features of grouped data
 - ▷ describe their distribution
 - ▷ inference using suitable regression models

Learning Objectives

- Goals: After this course participants will be able to
 - ▷ identify settings in which a repeated measurements model is required,
 - ▷ construct and fit an appropriate model to the data at hand, and
 - ▷ correctly interpret the results
- Even though the course will be primarily explanatory
 - ▷ sufficient mathematical detail will be provided in order participants to obtain a clear view on the different modeling approaches, and how they should be used in practice

Agenda

- **Chapter 1:** Motivating Data Sets

- ▷ Data sets that we will use throughout the course
- ▷ General repeated measurements settings
- ▷ Research questions

- **Chapter 2:** Marginal Models for Continuous Data

- ▷ Features of repeated measurements data
- ▷ Naive approaches
- ▷ Review linear regression
- ▷ Marginal models

Agenda (cont'd)

- **Chapter 3:** The Linear Mixed Effects Model
 - ▷ Intuition behind mixed models
 - ▷ nested and cross random effects
- **Chapter 4:** Marginal Models for Discrete Data
 - ▷ Review generalized linear models
 - ▷ Generalized estimating equations

Agenda (cont'd)

- **Chapter 5:** Mixed Models for Discrete Data

- ▷ Generalized linear mixed effects models
- ▷ approximations of the integrand & integral
- ▷ interpretation of parameters

- **Chapter 6:** Statistical Analysis with Incomplete Grouped Data

- ▷ Problems with incomplete data
- ▷ Missing data mechanisms
- ▷ Valid inferential approaches

Structure of the Course & Material

- Lectures & software practicals using R
- Material:
 - ▷ Course Notes
 - ▷ R code in soft format
- Within the course notes there are several examples of R syntax – these are denoted by the symbol 'R> '

Software Requirements

- The up-to-date versions of R and Rstudio; downloadable from
 - ▷ <http://cran.r-project.org/>
 - ▷ <http://www.rstudio.com/>
- Additional required packages
 - ▷ **nlme, lme4, MCMCglmm, geepack,**
 - ▷ **MASS, lattice, shiny, corrplot**

Software Requirements

- Up-to-date versions of these packages and their dependencies can be installed using the command

```
install.packages(c("shiny", "nlme", "lattice", "lme4",  
                  "MCMCglmm", "geepack", "MASS", "corrplot"),  
                dependencies = TRUE)
```

- Up-to-date version of a modern web browser, e.g.,
 - ▷ Mozilla Firefox (<https://www.mozilla.org/firefox/>)
 - ▷ Google Chrome (<http://www.google.com/chrome/>)

Software Requirements

- A [shiny](#) web app that replicates all analyses in the course including also some additional illustrations
- The app is available on GitHub and can be invoked using the following two-step procedure (assuming internet connection is available)

1. Start R

2. Run the command

```
shiny::runGitHub("Repeated_Measurements", "drizopoulos")
```

this will open a new web browser window (or tab) with the app

- Note: in order the app to be functional you should **not** close R

References

- Some texts in longitudinal data analysis
 - ▷ Demidenko, E. (2004). *Mixed Models: Theory and Applications*. New York: John Wiley & Sons.
 - ▷ Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. (2002). *Analysis of Longitudinal Data*, 2nd edition. New York: Oxford University Press.
 - ▷ Galecki, A. and Burzykowski, T. (2013). *Linear Mixed-Effects Models Using R*. New York: Springer-Verlag.
 - ▷ Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.
 - ▷ Fitzmaurice, G., Laird, N., and Ware, J. (2011). *Applied Longitudinal Analysis*, 2nd Ed. Hoboken: John Wiley & Sons.
 - ▷ Hand, D. and Crowder, M. (1995). *Practical Longitudinal Data Analysis*. London: Chapman & Hall.

References (cont'd)

- Some texts in longitudinal data analysis
 - ▷ Hedeker, D. and Gibbons, R. (2006). *Longitudinal Data Analysis*. New York: John Wiley & Sons.
 - ▷ Lindsey, J. (1993). *Models for Repeated Measurements*. Oxford: Oxford University Press.
 - ▷ Pinheiro, J. and Bates, D. (2000). *Mixed Effects Models in S and S-plus*. New York: Springer-Verlag.
 - ▷ Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.

Use of Statistical Models

... the megalomaniacal strategy of fitting a grand unified model, supposedly capable of answering any conceivable question that might be posed, is, in our view, dangerous, unnecessary and counterproductive.

Drum and McCullach (1993, *Statistical Science* **8**, 300–301)