BIOS 755: Linear Mixed Models II

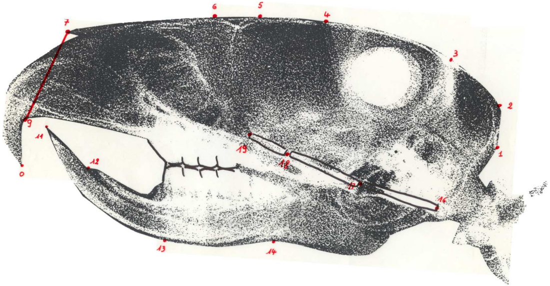
Alexander McLain

Rat Data

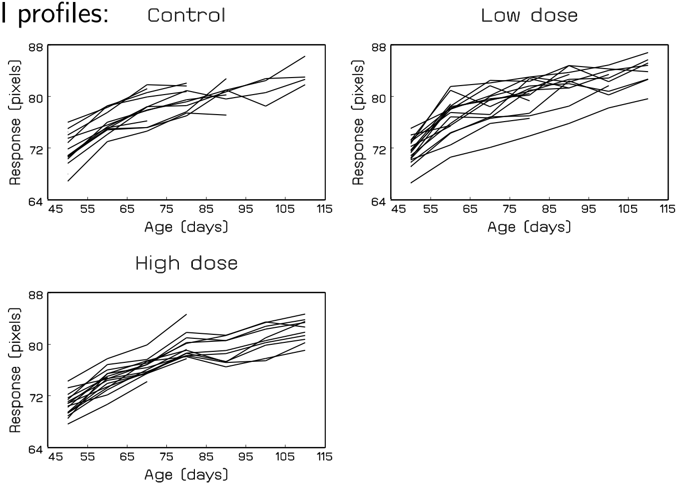
- Randomized experiment in which 50 male Wistar rats are randomized to:
  - Control (15 rats)
  - Low dose of Decapeptyl (18 rats)
  - High dose of Decapeptyl (17 rats)population.
- Question of interest: How does craniofacial growth depend on testosterone production?

Rat Data

- Treatment starts at 45 days; measurements are taken every 10 days from day 50 on.
- The responses are distances (pixels) between well-defined points on x-ray pictures of the skull of each rat:



- We'll consider only one response: the height of the skull.
- Individual profiles:



Models under consideration

- Let's consider the model:

$$Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{1i})t_{ij} + \varepsilon_{ij}$$
$$= \begin{cases} (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + \varepsilon_{ij}, & \text{if low dose} \\ (\beta_0 + b_{0i}) + (\beta_2 + b_{1i})t_{ij} + \varepsilon_{ij}, & \text{if high dose} \\ (\beta_0 + b_{0i}) + (\beta_3 + b_{1i})t_{ij} + \varepsilon_{ij}, & \text{if Control} \end{cases}$$

  Where the covariance of the random effects is

$$\boldsymbol{D} = cov(\boldsymbol{b}) = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$$

Linear Mixed representation

- What are the $\boldsymbol{X}$ and $\boldsymbol{Z}$ from the linear mixed model that corresponds to this model?
- What is the implied mean structure?
- What is the implied variance of Y_{ij}?

Linear Mixed representation

- What is the implied marginal Variance of Y_{ij}?

$$
\begin{aligned}
Var(Y_{ij}) &= \begin{pmatrix} 1 & t_{ij} \end{pmatrix} \boldsymbol{D} \begin{pmatrix} 1 \\ t_{ij} \end{pmatrix} + \sigma^2 \\
&= (d_{11} + \sigma^2) + 2d_{12}t_{ij} + d_{22}t_{ij}^2
\end{aligned}
$$

where

Analysis

- The following model was fitted to the data

$$Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i})t_{ij} + \varepsilon_{ij}$$

- The REML estimates obtained from PROC Mixed are:

Effect	Parameter	REMLE (s.e.)
Intercept	β_0	68.606 (0.325)
Time effects:		
Low dose	β_1	7.503 (0.228)
High dose	β_2	6.877 (0.231)
Control	β_3	7.319 (0.285)
Covariance of $\boldsymbol{b_i}$:		
var(b_{1i})	d_{11}	3.369 (1.123)
var(b_{2i})	d_{22}	0.000 (——)
cov(b_{1i}, b_{2i})	$d_{12} = d_{21}$	0.090 (0.381)
Residual variance:		
var(ε_{ij})	σ^2	1.445 (0.145)
REML log-likelihood	−466.173	

Analysis

- This suggests that the REML likelihood could be further increased by allowing negative estimates for d_{22}.
- In SAS, this can be done by adding the option `nobound` to the `PROC MIXED` statement.
- Results:

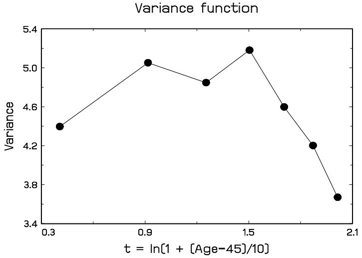
		Parameter restrictions for $\boldsymbol{\alpha}$	
		$d_{ii} \geq 0, \sigma^2 \geq 0$	$d_{ii} \in \mathbb{R}, \sigma^2 \in \mathbb{R}$
Effect	Parameter	REMLE (s.e.)	REMLE (s.e.)
Intercept	β_0	68.606 (0.325)	68.618 (0.313)
Time effects:			
Low dose	β_1	7.503 (0.228)	7.475 (0.198)
High dose	β_2	6.877 (0.231)	6.890 (0.198)
Control	β_3	7.319 (0.285)	7.284 (0.254)
Covariance of $\boldsymbol{b_i}$:			
var(b_{1i})	d_{11}	3.369 (1.123)	2.921 (1.019)
var(b_{2i})	d_{22}	0.000 (——)	−0.287 (0.169)
cov(b_{1i}, b_{2i})	$d_{12} = d_{21}$	0.090 (0.381)	0.462 (0.357)
Residual variance:			
var(ε_{ij})	σ^2	1.445 (0.145)	1.522 (0.165)
REML log-likelihood		−466.173	−465.193

Meaning of a negative variance component

- Fitted variance function

$$
\begin{aligned}
Var(Y_{ij}) &= (\hat{d}_{11} + \hat{\sigma}^2) + 2\hat{d}_{12} t_{ij} + \hat{d}_{22} t_{ij}^2 \\
&= 4.443 + 0.924t - 0.287 t_{ij}^2
\end{aligned}
$$

- The suggested negative curvature in the variance function is supported by the sample variance function:



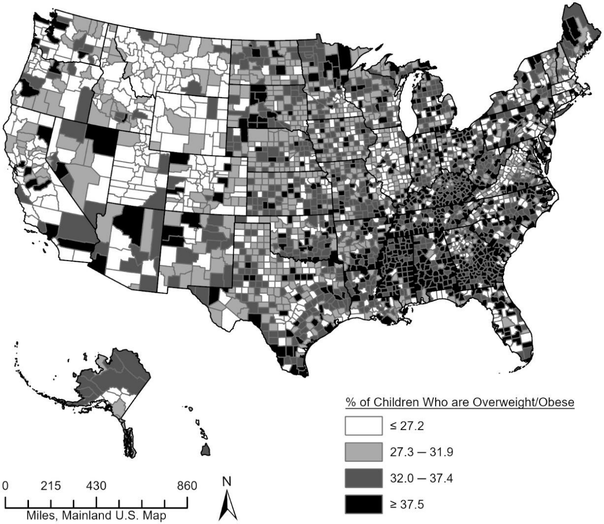
Inference for Linear Mixed Models

Introduction

- In most applications, inference is focused on the fixed effects, $\boldsymbol{\beta}$.
- However, in some studies we may want to predict (or "estimate") subject-specific response profiles.
- Technically, because the $\boldsymbol{b}_i$'s are random, we customarily talk of "predicting" the random effects rather than "estimating" them.
- The predicted random effects can be used to predict values for all levels in the data.

Example: County level disease rates/summaries.

- The National Study of Children's Health (NSCH) gathers data from roughly 50K 2–17 year-old children at each survey.
- The NSCH variables include indicators of ADHD, ASD, and many other conditions.
- It also includes BMI percentile, and other continuous variables.
- A regression model with a random county level intercept was fitted.
- To predict (estimate) county level values, we needed to predict the value of the random effect.


% of Children Who are Overweight/Obese
≤ 27.2
27.3 – 31.9
32.0 – 37.4
≥ 37.5
0 215 430 860
Miles, Mainland U.S. Map
N

Conditional Expectation

- For a RE ANOVA model on $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$ we have

$$\boldsymbol{Y}_{ij} = \mu + b_i + e_{ij}$$

with $b_i \sim N(0, G)$ and $e_i \sim N(0, \sigma^2)$.

- Under this model, the predicted value of the random intercept is

$$E(b_i|Y_i) = \frac{n_i G}{n_i G + \sigma^2}(\bar{Y}_i - \mu)$$

where $Var(b_i) = G$

Best Linear Unbiased Predictor (BLUP)

- When Σ_i is known, the estimator for $\boldsymbol{\beta}$ can be obtained by using ML weighted least square. Then the prediction of $\boldsymbol{b}_i$ is given by

$$\boldsymbol{G}\boldsymbol{Z}_i'\Sigma_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\beta}),$$

where

$$\Sigma_i = \mathrm{var}(\boldsymbol{Y}_i) = \boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}_i' + \boldsymbol{R}_i$$

- This is known as the Best Linear Unbiased Predictor (or BLUP).

Empirical BLUP

- In more general cases we use the Best Linear Unbiased Predictor (or BLUP) to predict the random effects.
- The formula for this is

$$\hat{\boldsymbol{b}}_i = \hat{\boldsymbol{G}}\boldsymbol{Z}_i'\hat{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\beta}),$$

  is often referred to as the "Empirical BLUP" or the "Empirical Bayes" (EB) estimator.
- Furthermore, it can be shown that

$$\text{var}(\hat{\boldsymbol{b}}_i) = \boldsymbol{G}\boldsymbol{Z}_i'\Sigma_i^{-1}\boldsymbol{Z}_i\boldsymbol{G} - \boldsymbol{G}\boldsymbol{Z}_i'\Sigma_i^{-1}\boldsymbol{X}_i\left(\sum_{i=1}^{n}\boldsymbol{X}_i'\Sigma_i^{-1}\boldsymbol{X}_i\right)^{-1}\boldsymbol{X}_i'\Sigma_i^{-1}\boldsymbol{Z}_i\boldsymbol{G}$$

The BLUP Estimation of Individual Mean

- Finally, the ith subject's predicted response profile is,

$$\begin{aligned}\hat{\boldsymbol{Y}}_i &= \boldsymbol{X}_i\hat{\beta} + \boldsymbol{Z}_i\hat{\boldsymbol{b}}_i \\ &= \boldsymbol{X}_i\hat{\beta} + \boldsymbol{Z}_i\hat{\boldsymbol{G}}\boldsymbol{Z}_i'\hat{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\beta}) \\ &= \left(\hat{\boldsymbol{R}}\hat{\Sigma}^{-1}\right)\boldsymbol{X}_i\hat{\beta} + \left(\boldsymbol{I} - \hat{\boldsymbol{R}}_i\hat{\Sigma}_i^{-1}\right)\boldsymbol{Y}_i\end{aligned}$$

- That is, the ith subject's predicted response profile is a weighted combination of the population-averaged mean response profile, $\boldsymbol{X}_i\hat{\beta}$, and the ith subject's observed response profile $\boldsymbol{Y}_i$.

BLUP As a Weighted Average

- When the between-subject variability is greater than the within-subject variability, more weight is given to the ith subject's observed data $\boldsymbol{Y}_i$.
- One formulation of the predicted values can be written as

$$\hat{\boldsymbol{Y}}_i = W_i \boldsymbol{X}_i \hat{\beta}_i^{OLS} + (1 - W_i) \boldsymbol{X}_i \hat{\beta}$$

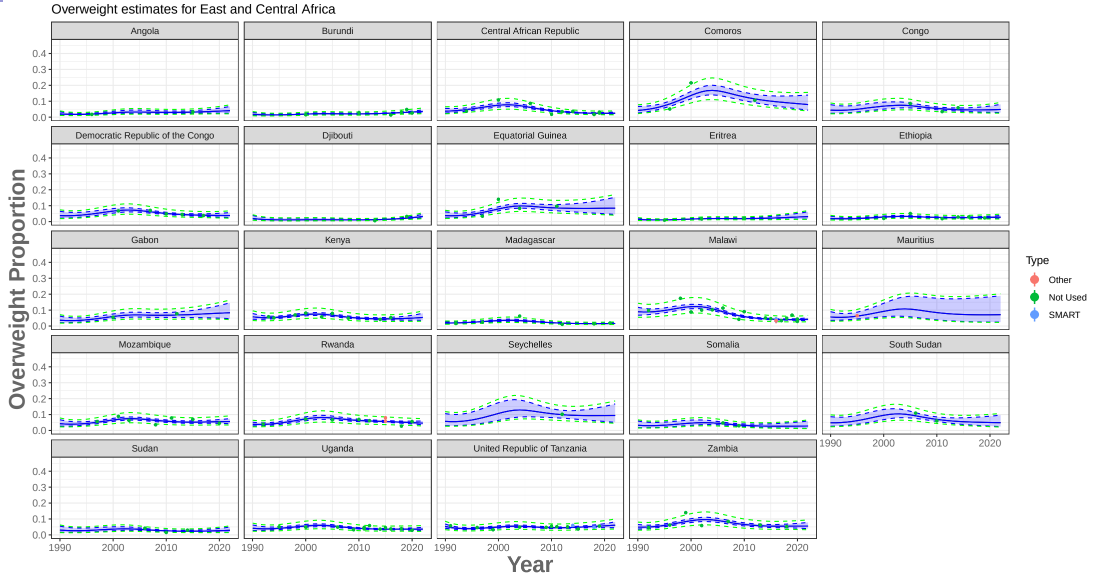
where

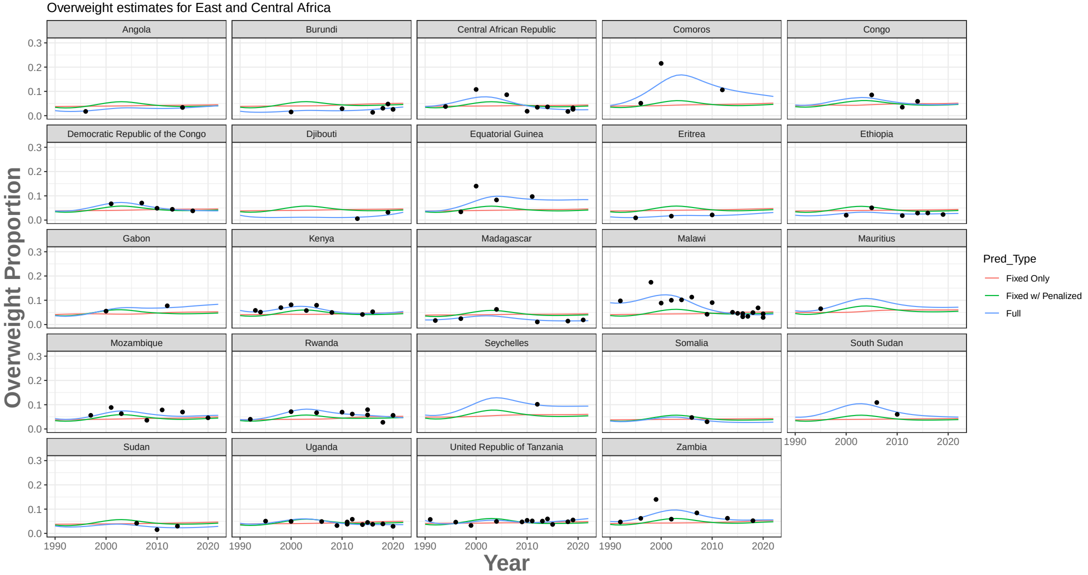
$$W_i = \frac{n_i D_{11}}{n_i D_{11} + \sigma^2}$$

Example: Country level rates of malnutrition.

- Ending malnutrition is a key outcome in the Sustainable Development Goal (SDG)
- The 2nd SDG calls for achieving, by 2025, a reduction of stunting and wasting and halt the rise in overweight in children under 5 years of age.
- Monitoring countries' progress toward the achievement of their SDG targets is an important task, but data sparsity makes monitoring trends challenging.
- The model fitted for this analysis was

$$Y_{ijk} = \beta' \boldsymbol{X}_{ijk} + \boldsymbol{b}_i' \boldsymbol{B}_{ij} + \boldsymbol{b}_{ij}' \boldsymbol{B}_{ijk} + \epsilon_{ij}, \quad (1)$$


Overweight estimates for East and Central Africa
Angola
Burundi
Central African Republic
Comoros
Congo
Democratic Republic of the Congo
Djibouti
Equatorial Guinea
Eritrea
Ethiopia
Gabon
Kenya
Madagascar
Malawi
Mauritius
Mozambique
Rwanda
Seychelles
Somalia
South Sudan
Sudan
Uganda
United Republic of Tanzania
Zambia
Overweight Proportion
Year
0.0
0.1
0.2
0.3
0.4
1990
2000
2010
2020
Type
Other
Not Used
SMART


Overweight estimates for East and Central Africa
Angola
Burundi
Central African Republic
Comoros
Congo
Democratic Republic of the Congo
Djibouti
Equatorial Guinea
Eritrea
Ethiopia
Gabon
Kenya
Madagascar
Malawi
Mauritius
Mozambique
Rwanda
Seychelles
Somalia
South Sudan
Sudan
Uganda
United Republic of Tanzania
Zambia
Overweight Proportion
0.0
0.1
0.2
0.3
Year
1990
2000
2010
2020
Pred_Type
Fixed Only
Fixed w/ Penalized
Full