Advanced Data Analysis

Semiparametric Mixed Models

Spinal Bone Mineral Density Data

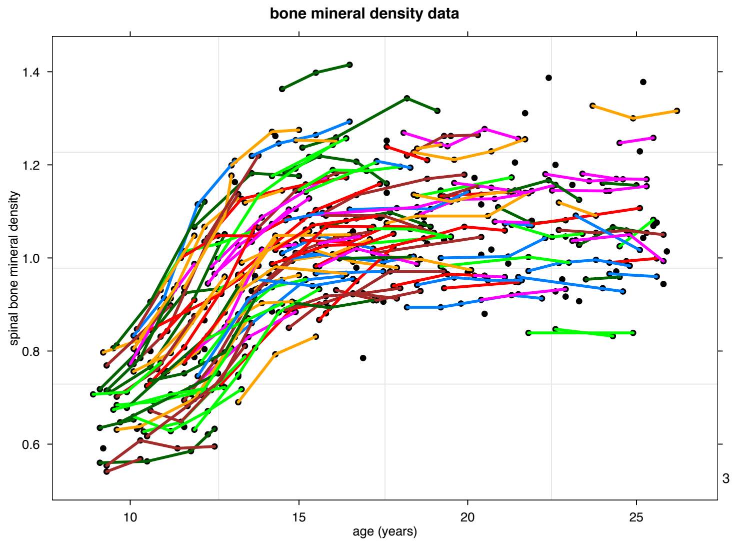
The following slide shows

longitudinal or repeated measures

data on

spinal bone mineral density (SBMD)

for 230 girls and young women.



Double Subscript Notation

Let

m = number of subjects

n_i = number of measurements on subject i $(1 \leq i \leq m)$

Then

y_{ij} = response for jth measurement on subject i.

Double Subscript Notation for Example

$\texttt{SBMD}_{ij}$ = jth spinal bone mineral density for subject i

$\texttt{age}_{ij}$ = age corresponding to $\texttt{SBMD}_{ij}$ measurement

Semiparametric Mixed Model

$$\texttt{SBMD}_{ij} = U_i + f(\texttt{age}_{ij}) + \varepsilon_{ij}$$

$U_i \overset{\text{ind.}}{\sim} N(0, \sigma_u^2)$ is a random subject intercept

Treating Longitudinal and Smoothing Together

$$\boldsymbol{Z} = \left[\begin{array}{l|l} \text{random} & \text{spline} \\ \text{intercept} & \text{basis} \\ \text{indicators} & \text{functions} \end{array} \right]$$

$$\text{Cov}(\boldsymbol{u}) = \begin{bmatrix} \sigma_U^2 \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 \boldsymbol{I} \end{bmatrix}$$

σ_U^2 = controls between-subject variability

σ_u^2 = controls amount of smoothing in estimation of mean

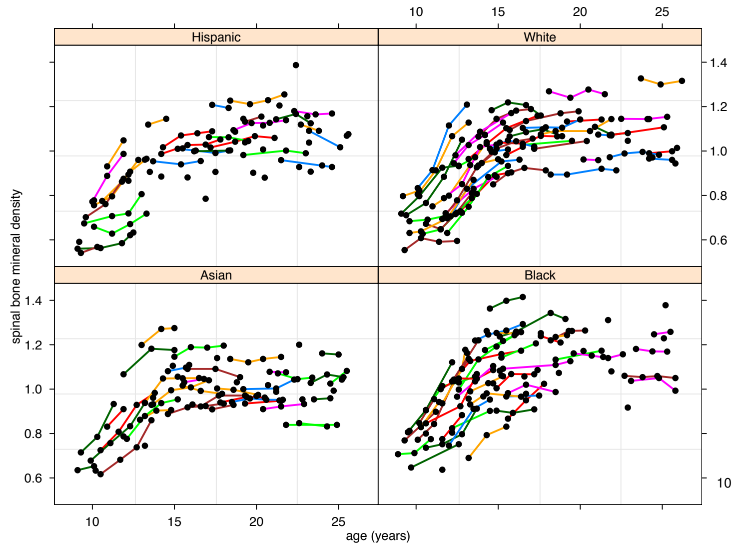
Explicit Representation of $\boldsymbol{Z}$

$$\boldsymbol{Z} = \begin{bmatrix} 1 & \cdots & 0 & (\texttt{age}_{11} - \kappa_1)_+ & \cdots & (\texttt{age}_{11} - \kappa_K)_+ \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cdots & 0 & (\texttt{age}_{1n_1} - \kappa_1)_+ & \cdots & (\texttt{age}_{1n_1} - \kappa_K)_+ \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & (\texttt{age}_{m1} - \kappa_1)_+ & \cdots & (\texttt{age}_{m1} - \kappa_K)_+ \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & (\texttt{age}_{mn_m} - \kappa_1)_+ & \cdots & (\texttt{age}_{mn_m} - \kappa_K)_+ \end{bmatrix}$$

Additive Model Extension

The SBMD data also have a variable for

ethnic background



Additive Model Extension

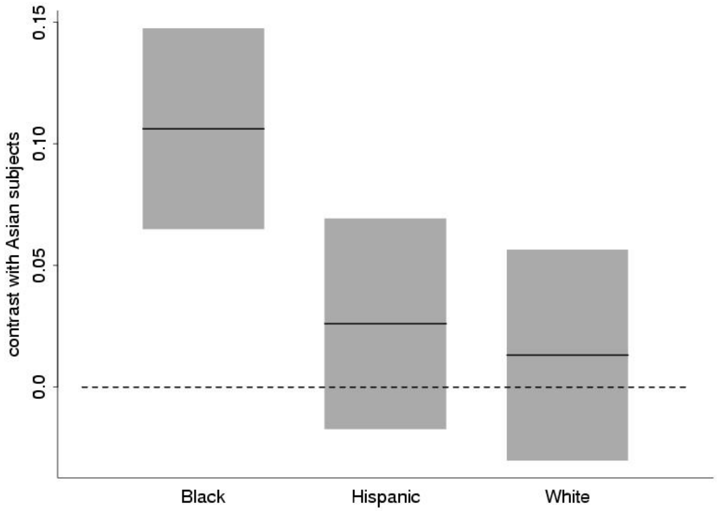
With the Asian ethnicity as a baseline we have

$$\begin{aligned}\texttt{SBMD}_{ij} &= U_i + f(\texttt{age}_{ij}) + \beta_1\,\texttt{black}_i + \beta_2\,\texttt{hispanic}_i \\ &\quad + \beta_3\,\texttt{white}_i + \varepsilon_{ij},\end{aligned}$$

Effect of Ethnicity

From the mixed model output:

variable	Value	Approx. Std. Error	z ratio
black	0.1062	0.02066	5.141
hispanic	0.0260	0.02164	1.203
white	0.0131	0.02165	0.6069

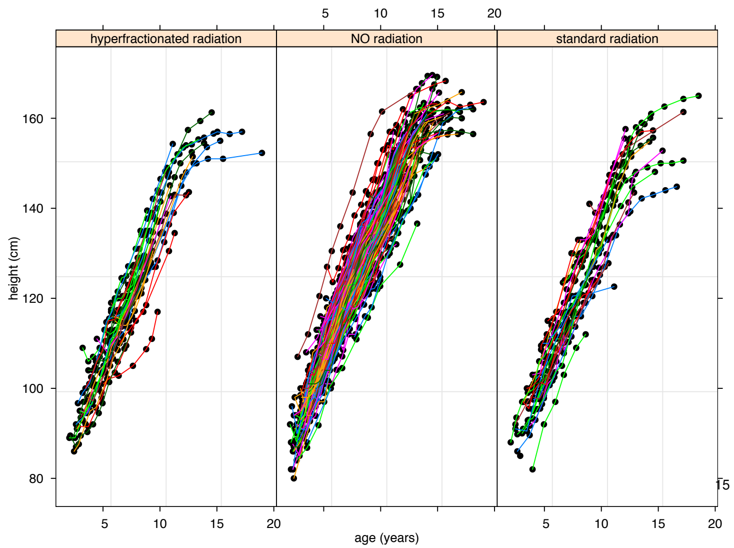


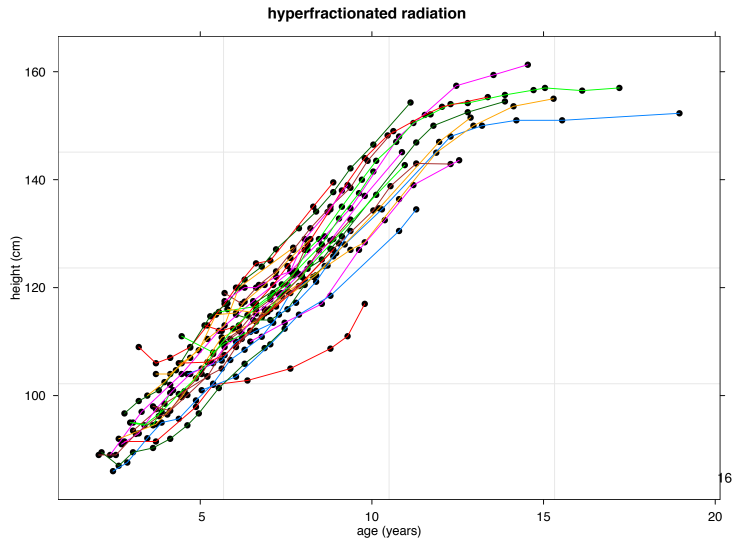
Acute Lymphoblastic Leukeamia Data Set

Cancer clinical trials at Dana Farber Cancer Institute, Boston, USA.

Growth curves for 197 girls undergoing three treatments for acute lymphoblastic leukeamia

What are long-term effects of treatment on growth?





Subject Specific Curves Extension

$$y_{ij} = U_i + f(x_{ij}) + \varepsilon_{ij}$$

$$\Downarrow$$

$$y_{ij} = g_i(x_{ij}) + f(x_{ij}) + \varepsilon_{ij}$$

Subject Specific Curves Model

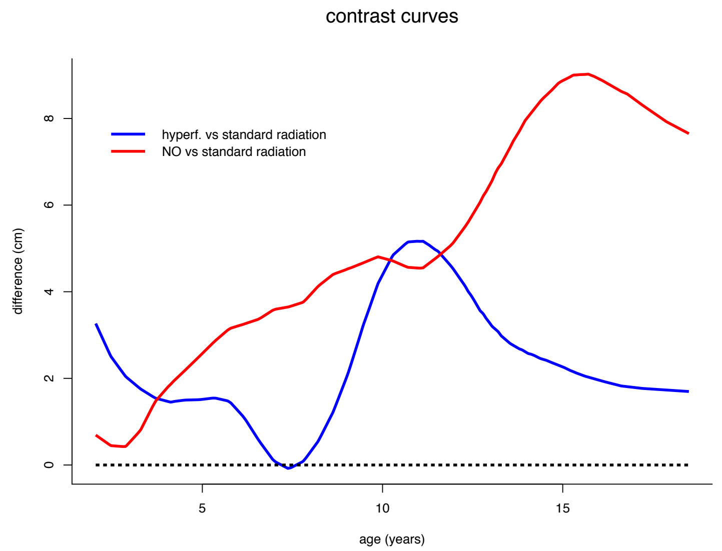
$\boldsymbol{X\beta}$: global linear component and treatment indicators

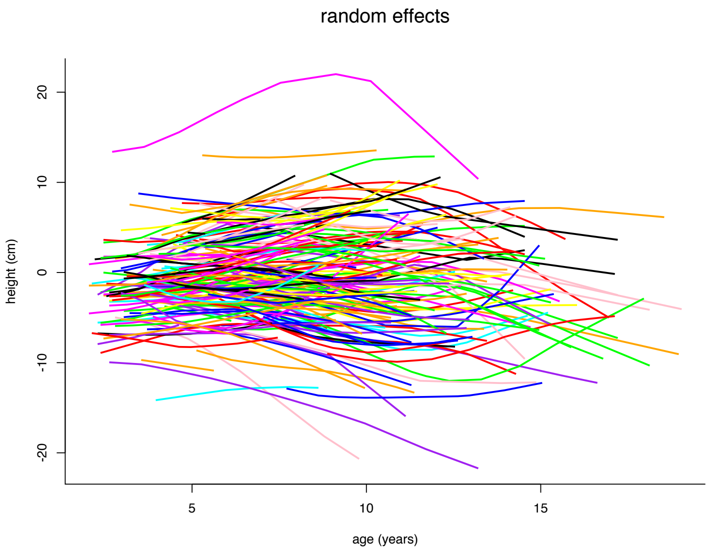
$\boldsymbol{Zu}$:

- spline basis functions for global non-parametric component with variance component σ^2_{glob}.
- individual linear effects with 2×2 covariance matrix $\boldsymbol{\Sigma}$.
- spline basis functions for subject specific deviations with single variance component σ^2_{subj}.

Distribution of $\boldsymbol{u}$

$$\boldsymbol{u} \sim N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\sigma^2_{\text{glob}}\boldsymbol{I} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \underset{1\le i\le m}{\text{blockdiag}}(\boldsymbol{\Sigma}) & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \sigma^2_{\text{subj}}\boldsymbol{I}\end{bmatrix}\right)$$





Subject Specific Curves Code

The following paper contains

R code

for illustration of subject specific curve analyses:

Durban, M., Harezlak, J., Wand, M.P. and Carroll, R.J. (2005). Simple fitting of subject-specific curves for longitudinal data. Statistics in Medicine, 24, 1153–1167.

R Demonstration

In this R demonstration we illustrate extend

subject-specific curve fitting via
linear mixed model software

to obtain a contrast curve for mean difference in heights between the radiation and non-radiation groups.

The code for this demonstration is in the script growthLeukCont.Rs

Main Points

- Many longitudinal applications benefit from semiparametric regression.
- Mixed models are the standard vehicle for longitudinal data analysis.
- Penalised splines with mixed model representations allow for a seamless fusion between semiparametric regression and longitudinal data analysis.