

Linear Regression Models

P8111

- HW 4 notes
 - no skewed CV
 - outlier removal
 -

Lecture 23

Jeff Goldsmith
April 19, 2016



THE DEPARTMENT OF
BIostatISTICS



Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Today's Lecture

- Multilevel models ✓
 - ▶ Hierarchical / nested models
 - ▶ Crossed designs ✓
- Bayesian methods

Longitudinal data

- We observe data y_{ij}, x_{ij} for subjects $i = 1, \dots, I$ at visits $j = 1, \dots, J_i$
- Overall, we pose the model

$$\underline{y = X\beta + \epsilon}$$

where $\underline{\text{Var}(\epsilon) = \sigma^2 V}$ and

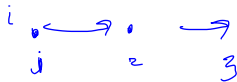
$v(\rho)$

$$V = \begin{bmatrix} V_1 & 0 & \dots & 0 \\ 0 & V_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & V_I \end{bmatrix}$$

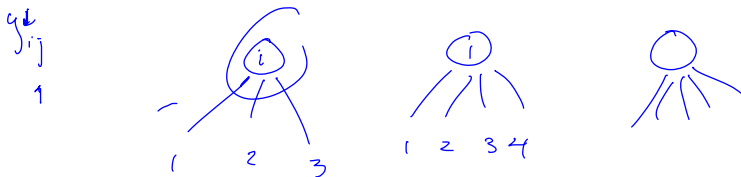
Longitudinal data

- Extended cross-sectional models to allow repeated subject observations
- Repeated observations had a time element
- One basic approach was random effects

Multilevel models



- Multilevel models are a 'more general' class of models
- Repeated observations don't necessarily have to be taken in time
- Examples of two-level models include students in a class, members in a family, patients in a hospital, etc



Two-level model

The repeated observations structure we developed for longitudinal data helps for two-level models. Specifically for repeated observations j within clusters i , we could write

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b_i + \epsilon_{ij}$$

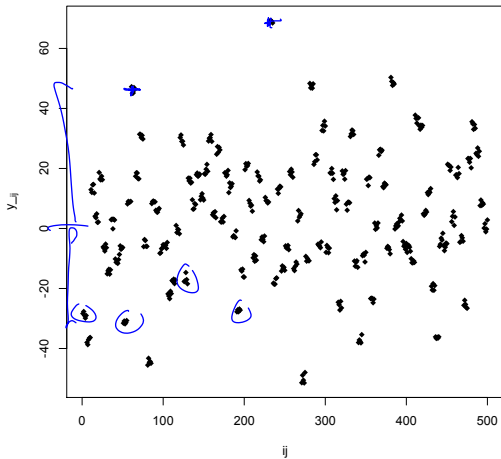
with

- $b_i \sim N[0, \tau^2]$
- $\epsilon \sim N[0, \nu^2]$

Intuition, estimation, induced correlation, interpretation – all of these were established for LDA and transfer here

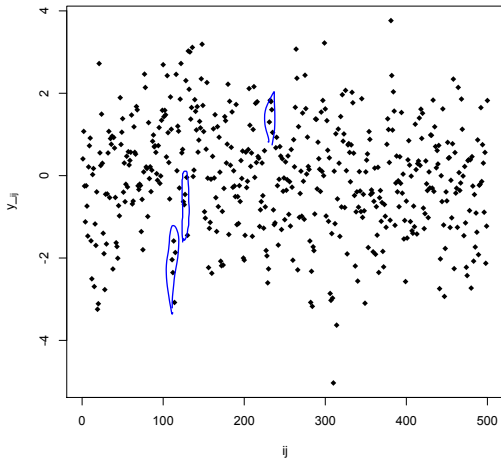
Example I

$$\tau^2 \Rightarrow \sigma^2 \quad / \quad ICC \approx 1$$



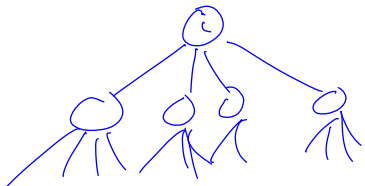
Example II

$$\gamma^2 \ll \nu^2 \quad / \quad \text{ICC} \approx 0$$



Three level model

- Sometimes, the data have a more complex nested structure
- Each cluster is part of a larger cluster
- Examples include students in classes in universities, members in families in towns, patients in hospitals in regions



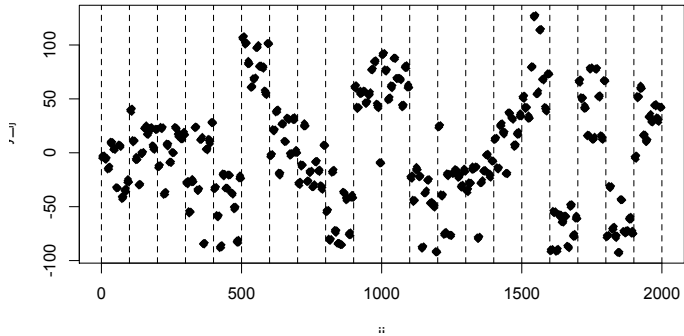
ICCs

This model gives two levels of correlation (observations within clusters, clusters within super clusters), and therefore a couple of ICCs:

- $cov(y_{ijk}, y_{ijk'}) =$

- $cov(y_{ijk}, y_{ij'k}) =$

Example



Example

Nested model

```
> nested.mod = lmer(yij ~ (1 | L1) + (1 | L2))
> summary(nested.mod)
Linear mixed model fit by REML ['lmerMod']
Formula: yij ~ (1 | L1) + (1 | L2)
```

REML criterion at convergence: 7464.337

Random effects:

Groups	Name	Variance	Std.Dev.
L2	(Intercept)	527.003	22.957
L1	(Intercept)	2137.453	46.233
	Residual	1.004	1.002

Number of obs: 2000, groups: L2, 200; L1, 20

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	0.7712	10.4646	0.074

Crossed designs

- Alternatively to nested (hierarchical) models, sometimes there is a crossed design
- Each subject is observed under multiple “treatments”, so there are both subject and treatment effects
- For example, each student is graded in multiple classes; each patient is assayed for multiple genes

Crossed designs

For a crossed model (with subjects i and treatments j), we can write

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b_i + b_j + \epsilon_{ij}$$

with

- $b_i \sim \text{N} \left[0, \tau_{(1)}^2 \right]$
- $b_j \sim \text{N} \left[0, \tau_{(2)}^2 \right]$
- $\epsilon \sim \text{N} \left[0, \nu^2 \right]$

Here there is covariance within subjects across treatments, and within treatments across subjects.

ICCs

Here there is covariance within subjects across treatments, and within treatments across subjects.

- $cov(y_{ij}, y_{ij'}) =$

- $cov(y_{i'j}, y_{ij}) =$

LDA and MLM

- Estimation works basically the same for these models as for random intercept models
- Intuition is the same as well – you want to borrow strength for one subject from the population of other subjects
- Interpretation of fixed effects is *marginal*; interpretation of random effects is *conditional*
- Using randomness both decreases the number of parameters and induces correlation structures

Bayesian methods

Longitudinal data analysis and multilevel models are a good place to start “thinking Bayesian”

- Even though they’re frequentist, they include randomness at subject levels
- The idea of “shrinking toward a population mean” or “borrowing strength” is a pretty Bayesian concept
- Even writing down random effect distributions is reminiscent of defining prior distributions

Basic Bayes

LDA and MLM are fairly advanced topics, so we'll start with a simpler example

- Suppose I gather data y_i and want to learn about $E(y)$
- Suppose even more I think I already know *something* about $E(y)$
- I might write down something about what I want to learn and what I think I know

Basic Bayes

What do I think I know?

- $y_i | \mu \sim \text{N} [\mu, \sigma_y^2]$
- $\mu \sim \text{N} [\mu_0, \sigma_0^2]$

What do I want to learn?

- $\mu | y_i \sim ???$

Basic Bayes

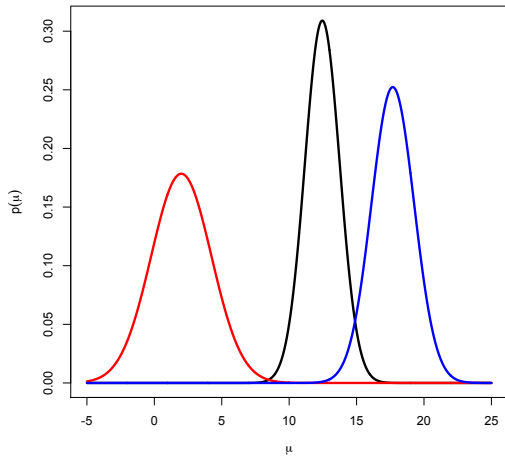
Luckily, this is all related through Bayes' formula:

$$p(\mu|y_i) \propto p(y_i|\mu)p(\mu)$$

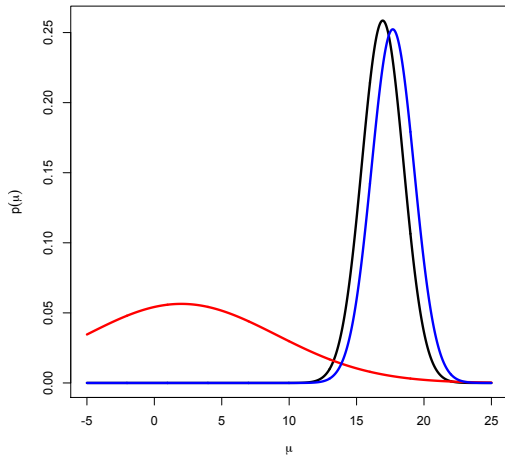
- For the Normal likelihood with a Normal prior for the mean, the posterior is also Normal:

$$\mu|y_i \sim \text{N} \left[\frac{\sigma_\mu^2}{\frac{\sigma_y^2}{n} + \sigma_\mu^2} \bar{y} + \frac{\frac{\sigma_y^2}{n}}{\frac{\sigma_y^2}{n} + \sigma_\mu^2} \mu_0, \frac{\frac{\sigma_y^2}{n} \sigma_\mu^2}{\frac{\sigma_y^2}{n} + \sigma_\mu^2} \right]$$

Effect of Prior



Effect of Prior



Bayesian regression

How can we pose the regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

with $\boldsymbol{\epsilon} \sim \text{N}[0, I_n]$ in a Bayesian framework?

- By making distributional assumptions about the $\boldsymbol{\beta}$
- Normal priors seemed to work well in the past ...
- Try $\boldsymbol{\beta} \sim \text{N}\left[0, \sigma_{\boldsymbol{\beta}}^2 I_p\right]$ where p includes the intercept

Bayesian regression

We want to obtain the posterior

$$p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) \propto p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{X})p(\boldsymbol{\beta})$$

Bayesian regression

Can show that

$$[\boldsymbol{\beta} | \mathbf{y}, \mathbf{X}] \sim \mathbf{N} [\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p]$$

where

$$\boldsymbol{\Sigma}_p = \left(\frac{1}{\sigma_\epsilon^2} \mathbf{X}^T \mathbf{X} + \frac{1}{\sigma_\beta^2} \mathbf{I} \right)^{-1}$$

and

$$\boldsymbol{\mu}_p = \boldsymbol{\Sigma}_p \left(\frac{1}{\sigma_\epsilon^2} \mathbf{X}^t \mathbf{y} \right)$$

So, about the variances

- Throughout all of this we have implicitly conditioned on the variances σ_ϵ^2 and σ_β^2
- Doesn't affect any of our calculations – the terms involving μ don't overlap with terms involving σ_ϵ^2 or σ_β^2
- σ_β^2 is often treated as fixed; σ_ϵ^2

The full posterior

- Need $[\beta, \sigma_\epsilon^2 | \mathbf{y}, \mathbf{X}]$
- “Intractable” problem
- Just as good: sample from the posterior

Sampling from the posterior

- aka where Bayes gets really weird
- You can draw a sample from the posterior even if you can't write down exactly what it is
- That sample is your basis for inference
 - ▶ Posterior sample average is your estimate
 - ▶ Quantiles on the posterior sample define your credible interval
- Sample describes the *joint distribution* of all model parameters

Some notes on this business

- Joint distributions are often worth the trouble
- Bayesian methods were really controversial for a long time, but are at least less controversial now
- The introduction of “prior knowledge” happens even in frequentist methods, although it is often not explicitly acknowledged

Today's big ideas

- Nested and crossed random effects models
 - Bayesian stuff
-