Linear Regression Models
P8111

Lecture 21

Jeff Goldsmith
April 12, 2016

THE DEPARTMENT OF BIOSTATISTICS
Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Today's Lecture

- LDA review
- Model interpretation
- Model estimation
- Example (pig data!)

Recall the setting

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

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

where $\mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2 V$ and

$\sigma^2 V_i$

$\underbrace{\boxed{V_i}}_{J_i \times J_i}$

Uniform = exchangeable

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

General ideas

$y_{ij} | b_i = \beta_0 + b_i + \beta_1 x_{ij1} + \epsilon_{ij}$

- We discussed two approaches: $Var(y_{ij})$; $cov(y_{ij}, y_{ij'}) \neq 0$
    - Random (mixed) effects models introduce *random* subject effects and assume uncorrelated errors
    - Marginal models include only fixed effects and assume correlated subject-level errors $y_{ij} = \beta_0 + \beta_1 x_{ij1} + e_{ij}$;
- In some cases these introduce equivalent correlation structures; an example is that a random intercept model induces uniform within-subject correlations $Var(\epsilon_i) = V_i$
- Marginal models separate the mean structure and correlation; random effect models induce correlation by introducing random quantities into the mean

Some comparisons

- Marginal models may be less immediately obvious or interpretable
- However, marginal models can be more robust to misspecification through robust SE estimates

Marginal model

The marginal model formulation is

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

where

$$E(y_{ij} \mid x_{ij1} = 1) - E(y_{ij} \mid x_{ij1} = 0) = \beta_1$$

- $\underline{\boldsymbol{\epsilon} \sim \mathrm{N}\left[0, \sigma^2 V\right]}$
- This approach focuses on the *marginal* distribution of $\boldsymbol{y}$, rather than on a subject-level *conditional* distribution
- Coefficients have a marginal interpretation – compare subjects based only on covariate values
- Interpretation is analogous to a cross-sectional model

Remember GLS

Given the model

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

where $\boldsymbol{\epsilon} \sim N(0, \sigma^2 V)$ with V known, we are essentially assuming

$$\boldsymbol{y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 V)$$

Using MLE, we find that $\hat{\boldsymbol{\beta}}_{GLS} = (\boldsymbol{X}^T V^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T V^{-1}$

$L(\beta | y) \propto \exp \ldots \ldots$

Estimation – marginal model

- If we can use MLE when V is known, maybe we can use MLE to estimate V as well
- Our log likelihood function is

$$\begin{aligned} l(\boldsymbol{\beta}, \sigma^2, V; \boldsymbol{y}, \boldsymbol{X}) &= -\frac{1}{2}\Big[n \log(\sigma^2) + \log(|V|) \\ &\quad + \frac{1}{\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T V^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\Big] \end{aligned}$$

- Using profile likelihood, we find that for any V_0

$$\hat{\boldsymbol{\beta}}(V_0) = (\boldsymbol{X}^T V_0^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T V_0^{-1}$$

Estimation – marginal model

$L(V, \sigma^2 | y, x)$

- Estimation of V and σ is done through restricted maximum likelihood
  - Standard MLE produces biased variance estimates; REML adjusts for the number of fixed effects components that are estimated
- Often V is structured parametrically to ease estimation and computation
- We won't worry about how this is done

$V(\hat{\rho})$

Random intercept model

$$E(y_{ij} \mid \boxed{b_i}, x_{ij}) = \beta_0 + b_i + \beta_1 x_{ij}$$

$$E(y_{ij} \mid b_i, x_{ij} = 1) - E(y_{ij} \mid b_i, x_{ij} = 0) = \beta_1$$

A random intercept model with one covariate is given by

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

where

$$E(y_{ij} \mid b_i = 0) = \beta_0 + \beta_1 x_{ij}$$

- $b_i \sim \mathrm{N}\left[0, \tau^2\right]$
- $\epsilon_{ij} \sim \mathrm{N}\left[0, \nu^2\right]$

$$\mathrm{Var}(y_{ij} \mid b_i) = \nu^2 I$$

$$E(y \mid b_i) = \beta_0 + b_i + \beta_1 x_{ij}$$

Random intercept model

More compactly, we write

Z

$\sum_i J_i \times I$

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

where

- $b \sim \mathrm{N}\left[0, \boxed{\tau^2} I_I\right]$
- $\epsilon \sim \mathrm{N}\left[0, \nu^2 I_n\right]$

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1J_1} \\ \vdots \\ y_{IJ_I} \end{bmatrix} = \begin{bmatrix} x_{111} & \cdots & x_{11p} \\ x_{1J_11} & \cdots & x_{1J_1p} \\ \vdots & & \vdots \\ x_{IJ_I1} & \cdots & x_{IJ_Ip} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} +$$

$$y = X\beta$$

$$+ Zb$$

$$\begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} J_1 \\ \\ \\ \\ J_2 \\ \\ \vdots \\ \\ J_I \end{array} & \begin{bmatrix} 1 & 0 & \cdots \\ 1 & 0 & \\ 1 & 0 & \\ 1 & 0 & \\ 0 & 1 & \\ 0 & 1 & \\ 0 & \vdots & \vdots \\ \vdots & 0 & \vdots \\ \vdots & 0 & \vdots \\ 0 & 0 & \vdots \end{bmatrix} \end{array} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_I \end{bmatrix}$$

Random intercept model

In the model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\epsilon}$$

why might we assume $\boxed{\boldsymbol{b}}$ are random rather than estimating them as fixed effects?

- Interpretation
- "Penalization" / shrinkage / Borrowing Strength
- # Parameters

Random intercept model

In the model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\epsilon}$$

why might we assume $\boldsymbol{b}$ are fixed rather than considering them to be random?

- You don't want penalization
- "Random" interp doesn't make sense

(mostly don't do this)

Random intercept model interpretation

Random effect models have a conditional interpretation.

- Mean conditions on subject effects: $\mathrm{E}(y_{ij}|x_{ij}, \boldsymbol{\beta}, b_i)$
- To derive a marginal (averaged across subjects) mean, one can use iterated expectations

$$E(y|x) = E\left(E_{|b}(y|x,b)\right) = E_b(X\beta + Zb) = X\beta$$

- (For an identity link, the coefficients also have a marginal interpretation – this is not true for generalized RE models)

$$E\left(g^{-1}(X\beta + Zb)\right) \neq g^{-1}\left(E(X\beta + Zb)\right)$$

Estimation – random effect model

Still done using MLE, but now we include random effects; our model is

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\epsilon}$$

where

- $\boldsymbol{b} \sim \mathrm{N}\left[0, \tau^2 I_I\right]$
- $\boldsymbol{\epsilon} \sim \mathrm{N}\left[0, \nu^2 I_n\right]$

Estimation – random effect model

$$L(\beta, b ; y) \propto p(y \mid \beta, b) \cdot p(b)$$

$$\propto \exp\left\{ \frac{-1}{2\tau^2} (y - X\beta - Zb)^T (y - X\beta - Zb) \right\}$$

$$\cdot \exp\left\{ \frac{-1}{2\nu^2} b^T b \right\}$$

$$\gamma = \begin{bmatrix} \beta \\ b \end{bmatrix}$$

$$C = \begin{bmatrix} X & Z \end{bmatrix}$$

$$R = \begin{array}{c|c} & p \\ p & \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & I \end{array}\right] \\ & I \end{array}$$

$$\underline{y - X\beta - Zb} = y - C\gamma$$

$$b^T b = \gamma^T R \gamma$$

Estimation – random effect model

$$\propto \exp\left\{ \frac{-1}{2\tau^2} \underline{(y - C\gamma)^T (y - C\gamma)} \right\}$$

$$\exp\left\{ \frac{-1}{2\nu^2} \gamma^T \underline{\underline{R}} \gamma \right\}$$

$$= \arg_\gamma \left[\underline{(y - C\gamma)^T (y - C\gamma) + \underbrace{\lambda}_{} \gamma^T R \gamma} \right]$$

$\approx$ Ridge

Estimation – BLUPs

$$\hat{\beta}_{\text{Ridge}} = \left(X^T X + \lambda P\right)^{-1} X^T y$$

- Our estimate for fixed and random effects are

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{b}} \end{bmatrix} = \left(\boldsymbol{C}^T\boldsymbol{C} + \boxed{\frac{\nu^2}{\tau^2}}R\right)^{-1}\boldsymbol{C}^T\boldsymbol{y}$$

- These are referred to as "BLUPs" (the "P" is for "predictions")

~ BLU from GM

- The variances ν^2 and τ^2 are estimated via REML

BLUPs

BLUPs are a lot like BLUEs

- They are the best linear unbiased predictors when one has both fixed and random effects
- We derived them using Normal distributions, but even without distributional assumptions these are BLUP
- The Normal distributions help with the assumptions are satisfied, in that one can get distribution-based inference

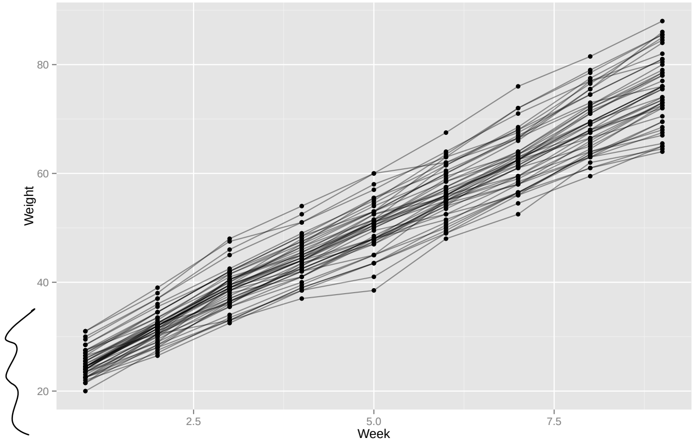
Penalized spline regression

Note that BLUPs look an awful lot like penalized regression estimates

- The "mixed model framework" is commonly used for penalized spline estimation
- The "tuning parameter" is a ratio of variances, and can be estimated via REML (rather than CV)
- This approach provides a method for inference

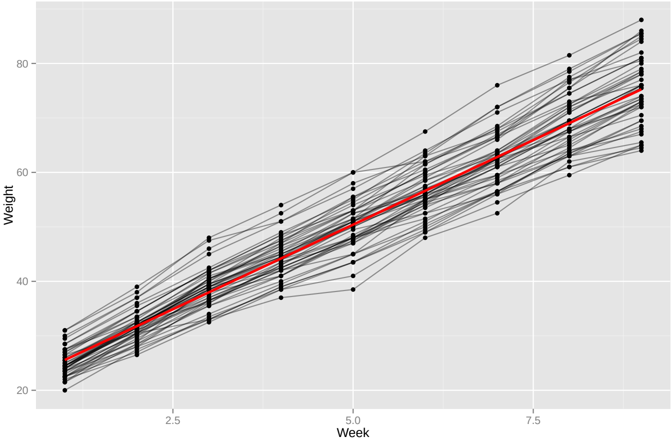
Penal $\neq$ MM, at least conceptually.

Pig data



Pig data

OLS fit for pig data



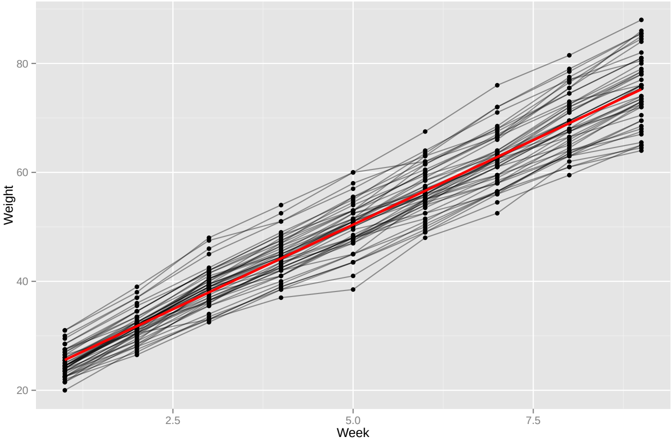
Pig data

OLS code

```
>   lin.mod = lm(weight ~ num.weeks, data = pig.weights)
>   summary(lin.mod)

Residuals:
     Min       1Q   Median       3Q      Max
-11.9051  -2.5348  -0.1952   2.5949  13.1751

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 19.35561    0.46054   42.03   <2e-16 ***
num.weeks    6.20990    0.08184   75.88   <2e-16 ***

Residual standard error: 4.392 on 430 degrees of freedom
```

Pig data

Marginal model fit for pig data



Pig data

"Generalized" Est Equations

$y = X\beta + \epsilon, \quad \epsilon \sim (0, \sigma^2 V)$

Marginal model code

V_i

```
> library(gee)
> marg.mod = gee(weight ~ num.weeks, id = id.num, corstr = "exchangeable",
               data = pig.weights)
> summary(marg.mod)

Model:
 Link:                      Identity
 Variance to Mean Relation: Gaussian
 Correlation Structure:     Exchangeable

Summary of Residuals:
       Min         1Q      Median         3Q        Max
-11.9050926  -2.5347801  -0.1951968   2.5949074  13.1751157
```

exch = truth   ? truth

```
Coefficients:
             Estimate Naive S.E.   Naive z Robust S.E. Robust z
(Intercept) 19.355613  0.5983680  32.34734  0.39963854 48.43280
num.weeks    6.209896  0.0393321 157.88366  0.09107443 68.18485
```

Misspecification ??

```
Working Correlation
          [,1]      [,2]      [,3]      [,4] ...
 [1,] 1.0000000 0.7690313 0.7690313 0.7690313 ...
 [2,] 0.7690313 1.0000000 0.7690313 0.7690313 ...
 [3,] 0.7690313 0.7690313 1.0000000 0.7690313 ...
 [4,] 0.7690313 0.7690313 0.7690313 1.0000000 ...
 ...
```

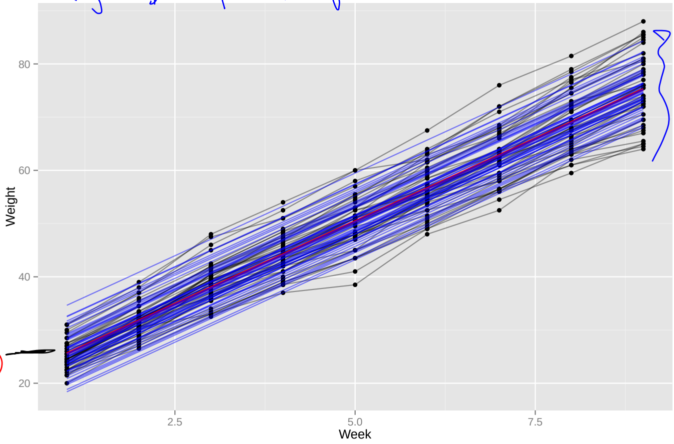
$\rightarrow \rho$

$$\begin{pmatrix} 1 & \rho & \rho & \rho \\ & 1 & \rho & \\ \rho & \rho & 1 & 1 \\ \rho & \rho & \rho & 1 \\ & & & \ddots \end{pmatrix}$$

Pig data

Mixed effect model fit for pig data

$E(y \mid b_i) = \beta_0 + \beta_1 x_{ij} + b_i$



b_i

$E(y \mid b=0)$

Pig data

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

Mixed effect model code

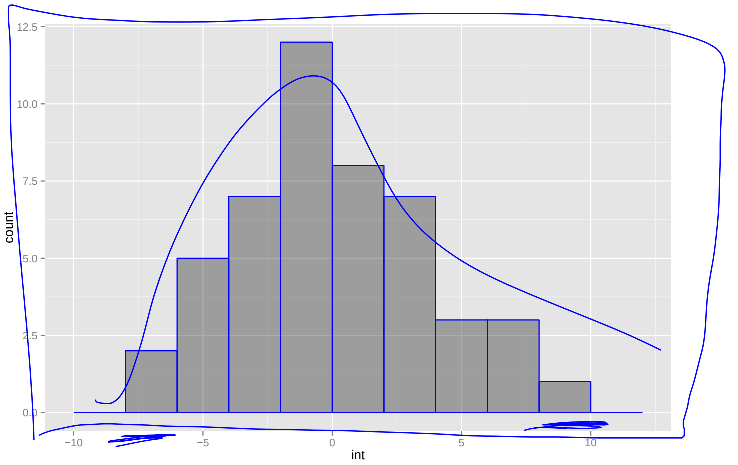
= intercept

```
> library(lme4)
> ranef.mod = lmer(weight ~ (1 | id.num) + num.weeks, data = pig.weights)
> summary(ranef.mod)
Linear mixed model fit by REML
Formula: weight ~ (1 | id.num) + num.weeks
   Data: pig.weights
  AIC  BIC logLik deviance REMLdev
 2042 2058  -1017     2030    2034
Random effects:
 Groups   Name        Variance Std.Dev.
 id.num   (Intercept) 15.1418  3.8913
 Residual              4.3947  2.0964
Number of obs: 432, groups: id.num, 48

Fixed effects:
            Estimate Std. Error t value
(Intercept) 19.35561    0.60311   32.09
num.weeks    6.20990    0.03906  158.97

>   (15.1418) / (15.1418 + 4.3947)
[1] 0.7750518
```

ICC

Pig data

Histogram of estimated random intercepts



Today's big ideas

- Estimation in LDA
- Example + code

Potential reading on mixed effects models – Semiparametric Regression (Ruppert, Wand, Carroll) Ch 4