STAT3015/7030:
Generalised Linear Modelling
Multilevel Models - basics

Bronwyn Loong

Semester 2 2014

References

Ch 12 - Gelman and Hill

Ch 8.8 - Faraway

Motivation

- Would you rather get surgery at a hospital with 4 patients and 4 successes, or one with 100 patients and 95 success?
- Would you rather live in a neighbourhood with 1000 households and 4 burglary events in the past month, or a neighbourhood with 2000 households and 12 burglarly events in the past month?

What do the following types of data have in common?

A community sample of 1282 women provided survey data on walking for leisure and transport; educational level; enjoyment of walking; social support for physical activity; and sporting/recreational club membership. These data were linked with objective environmental data on the density of public open space and walking tracks in the womens local neighborhood.

What do the following types of data have in common?

There is widespread consensus that weak health systems hamper the effective provision of HIV/AIDS services. The aim of this study was to examine the effect of an HIV/AIDS program on six government-run general clinics in Kampala. Longitudinal information on the delivery of health services was collected at each clinic. Monthly changes in the volume of HIV and non-HIV services were analyzed. We also have background information on the overall efficacy of each clinic.

What do the following types of data have in common?

With interest in racial biases in police stops, in this article we analyze data from 125,000 pedestrian stops by the New York Police Department over a 15 month period (data such as age, race, gender of suspect, precinct of stop). We have further information on each stop by police precinct (geography, arrest rate, racial composition of precinct).

Features of Multilevel Data

- Data are structured as observations within groups
- Coefficient estimates in models may vary by group
- Account for variation within groups and across groups
- Account for dependencies in data due to grouping beyond what can be explained by group level predictors

Data for Analysis

We are interested in estimating the distribution of radon in each of the 85 counties in Minnesota. Radon measurements are taken on individual houses within counties, and the number of houses sampled differs across counties.

- Response: y = logarithm of the radon measurement for each house
- $x = 0$ if radon measured in basement, 1 if first floor (radon comes from underground)
- u = logarithm of measurement of soil uranium at the county level
- county = name of county

Data for Analysis

There are 919 observations in the total data set, representing 85 counties in Minnesota. The first 6 rows of the data are shown below.

```
          y x          u county
1 0.7884574 1 -0.6890476      1
2 0.7884574 0 -0.6890476      1
3 1.0647107 0 -0.6890476      1
4 0.0000000 0 -0.6890476      1
5 1.1314021 0 -0.8473129      2
6 0.9162907 0 -0.8473129      2
```

How would you analyse the data?

What goes wrong with the following approaches?

- Individual level regression: $y \sim x + u$; $y \sim x + u + factor(county)$
- Group level regression: include individual predictors as county averages

Notation

- Groups j = 1,...,J
- n_j= number of observations in group j
- $j[i]$ codes group membership, e.g. if $j[173] = 15$ then the 173rd observation is in county 15.
- Standard deviations σ_y are data-level errors and σ_α group-level errors

Multilevel Model: no predictors

$$y_i \sim N(\alpha_{j[i]}, \sigma_y^2) \; i = 1, .., n$$

$$\alpha_{j[i]} \sim N(\mu_\alpha, \sigma_\alpha^2) \; j = 1, .., J$$

What will our estimated $\hat{\alpha}_j$'s look like?

$$\hat{\alpha}_{j[i]} \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

Multilevel Model: no predictors

$$\hat{\alpha}_{j[i]} \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

What happens to our estimate of $\hat{\alpha}_j$ as

- n_j gets large?
- n_j gets small?
- σ_α^2 gets large relative to σ_y^2?
- σ_α^2 gets small relative to σ_y^2?

Example: baseball data

Illustration: Morris' Baseball Data

i	player	y_i	θ_i	i	player	y_i	θ_i
1	Clemente	.400	.346	10	Swoboda	.244	.230
2	F. Robinson	.378	.298	11	Unser	.222	.264
3	F. Howard	.356	.276	12	Williams	.222	.256
4	Johnstone	.333	.222	13	Scott	.222	.303
5	Berry	.311	.273	14	Petrocelli	.222	.264
6	Spencer	.311	.270	15	E. Rodriguez	.222	.226
7	Kessinger	.289	.263	16	Campaneris	.200	.285
8	L. Alvarado	.267	.210	17	Munson	.178	.316
9	Santo	.244	.269	18	Alvis	.156	.200

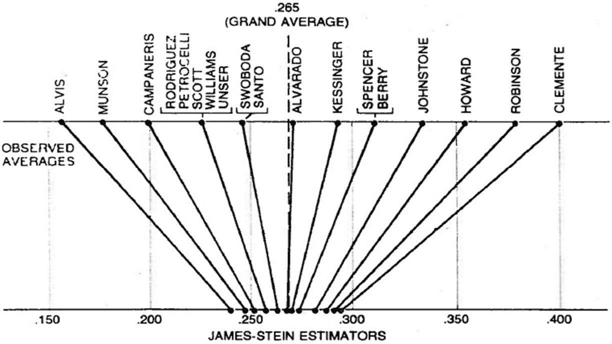
For players i=1,...,18

- y_i = batting average after first 45 at bats in 1970,
- θ_i= true 1970 batting ability (pretend the final 1970 averages measure this)

Example: baseball data

- How should I estimate Roberto Clementes season long batting average on the basis of his first 45 at bats?
- What is the maximum likelihood estimate of this value?
- Why is the multilevel approach better, intuitively?
- How will my answers differ?
- How does this relate to the concept of regression toward the mean?

Example: baseball data

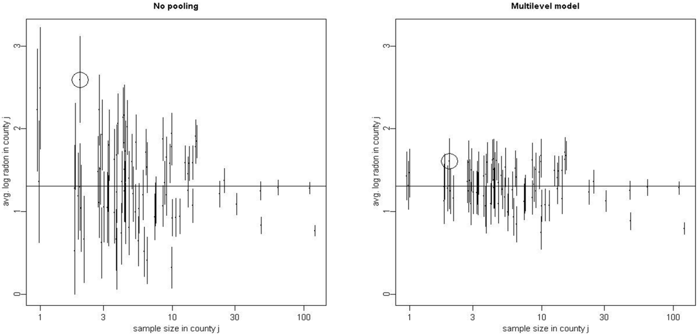


Multilevel model: pooling

A multilevel model is a compromise between two extremes.

- Complete pooling: all groups are identical (exclude group level predictors from the usual regression model)
- No pooling: all groups are different (fit separate model for each group)
- As $\sigma_\alpha^2 \to \infty$, $\hat{\alpha}_j \to \bar{y}_j$ (no pooling)
- If $\sigma_\alpha^2 = 0$, $\hat{\alpha}_j = \bar{y}_{all}$ (complete pooling)

Multilevel model vs pooling



Complete Pooling: Individual Level Predictors

```
> model1<-lm(formula = y ~ x)
> summary(model1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.32674    0.02972  44.640   <2e-16 ***
x           -0.61339    0.07284  -8.421   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.8226 on 917 degrees of freedom
Multiple R-squared: 0.07178, Adjusted R-squared: 0.07077
F-statistic: 70.91 on 1 and 917 DF,  p-value: < 2.2e-16
```

No Pooling: Individual Level Predictors

```
model2<-lm(formula = y ~ x + factor(county) - 1)
summary(model2)
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
x                -0.72054    0.07352  -9.800  < 2e-16 ***
factor(county)1   0.84054    0.37866   2.220 0.026701 *
factor(county)2   0.87482    0.10498   8.333 3.23e-16 ***
..........................................................
factor(county)85  1.18652    0.53487   2.218 0.026801 *


Residual standard error: 0.7564 on 833 degrees of freedom
Multiple R-squared: 0.7671, Adjusted R-squared: 0.7431
F-statistic: 31.91 on 86 and 833 DF,  p-value: < 2.2e-16
```

Multilevel Model: Individual Level Predictors

$$y_i \sim N(\alpha_{j[i]} + \beta x_i, \sigma_y^2), \ i = 1, ..., n$$

$$\alpha_{j[i]} \sim N(\mu_\alpha, \sigma_\alpha^2), \ j = 1, ..., J$$

What happends as

- $\sigma_\alpha^2 \to \infty$?
- $\sigma_\alpha^2 \to 0$?

To fit these models in R, 'lmer' function in library(arm).

Multilevel Model: Individual Level Predictors

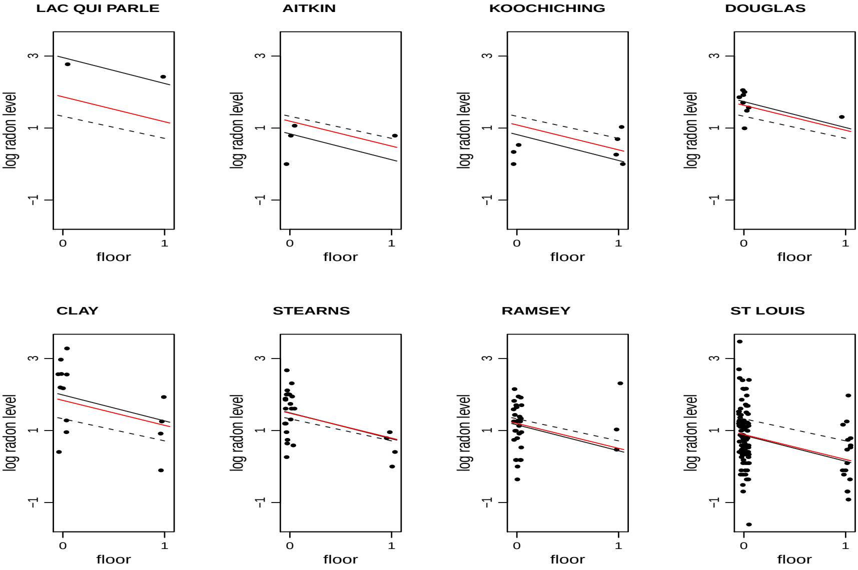
```
> model3<-lmer(formula = y ~ x + (1 | county))
> summary(model3)

Random effects:
 Groups   Name        Variance Std.Dev.
 county   (Intercept) 0.10773  0.32822
 Residual             0.57091  0.75559
Number of obs: 919, groups: county, 85

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.46159    0.05157  28.340
x           -0.69299    0.07043  -9.839

Correlation of Fixed Effects:
  (Intr)
x -0.288
```

Interpreting Error Term Estimates

$\hat{\sigma}_y$=0.76, $\hat{\sigma}_\alpha$=0.33

- $\hat{\sigma}_y^2/\hat{\sigma}_\alpha^2 = 0.76^2/0.33^2 = 5.3$. The standard deviation of average radon levels between counties is the same as the standard deviation of the average of 5.3 measurements within a county.
- Intraclass correlation: $\frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_\alpha^2+\hat{\sigma}_y^2} = \frac{0.33^2}{0.33^2+0.76^2} = 0.16$.This value ranges from 0 (grouping conveys no information) to 1 (all group members identical).

Pooling



LAC QUI PARLE
AITKIN
KOOCHICHING
DOUGLAS
CLAY
STEARNS
RAMSEY
ST LOUIS
log radon level
floor
3
1
−1
0
1

Useful lmer functions

```
> coef (model3)
$county
   (Intercept)          x
1    1.1915015 -0.6929905
2    0.9276037 -0.6929905
.......
85   1.3862299 -0.6929905
> fixef (model3)
(Intercept)          x
  1.4615940  -0.6929905
```

Useful lmer functions

```
> ranef (model3)
$county
   (Intercept)
1  -0.27009244
2  -0.53399029
.....
85 -0.07536403

> se.ranef (model3)
$county
 [1,] 0.24778450
 [2,] 0.09982720
.....
[85,] 0.27967312
```

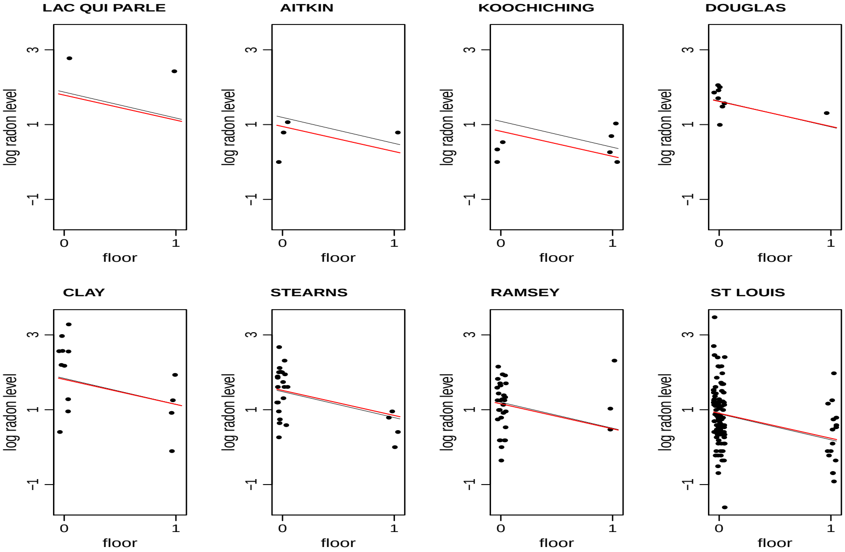
Multilevel model:Group level Predictors

$$y_i \sim N(\alpha_{j[i]} + \beta x_i, \sigma_y^2), \ i = 1, ..., n$$

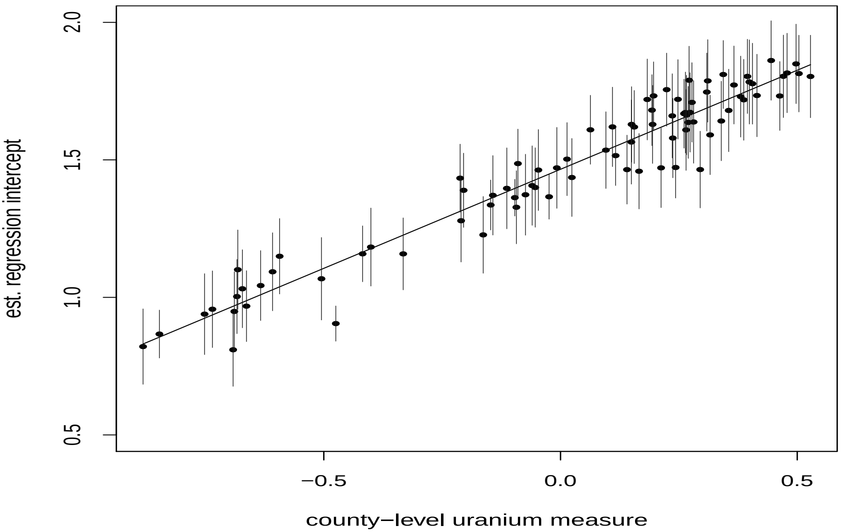
$$\alpha_{j[i]} \sim N(\gamma_0 + \gamma_1 u_j, \sigma_\alpha^2), \ j = 1, ..., J$$

```
 model4<-lmer(formula = y ~ x + u.full + (1 | county))
> display(model4)
lmer(formula = y ~ x + u.full + (1 | county))
            coef.est coef.se
(Intercept)  1.47     0.04
x           -0.67     0.07
u            0.72     0.09

Error terms:
 Groups   Name        Std.Dev.
 county   (Intercept) 0.16
 Residual             0.76
```

Multilevel model:Group level Predictors



Multilevel model:Group level Model



Interpreting Error Term Estimates with Group Level Predictor

$\hat{\sigma}_y = 0.76$, $\hat{\sigma}_\alpha = 0.16$

- Why is $\hat{\sigma}_\alpha = 0.16$ so much smaller than in the model without a group level predictor?
- Why is $\hat{\sigma}_y = 0.76$ the same as in the model without a group level predictor?
- $\hat{\sigma}_y^2/\hat{\sigma}_\alpha^2 = 0.76^2/0.16^2 = 22.56$. The county level model is as good as 23 observations within any county.

Other Notes

- Multilevel modeling is most important when it is close to complete pooling, at least for some of the groups
- Using group level predictors makes partial pooling more effective
- Multilevel models can be extended in many ways: varying slope models, glms, multiple levels, non-nested group structures, etc.
- If there are a small number of groups (e.g. < 5) then it will be difficult to estimate group level variation

When should you use multilevel models?

- *How many groups?* If J is small, it will be difficult to estimate between group variation, and multilevel models wont offer much improvement, but will not be worse either.
- *How many observations per group?* Even one or two observations per group is acceptable.
- *When is multilevel modeling most effective?* When some of the groups are close to complete pooling. In that case σ_α is relatively small, and groups can borrow information from each other.
- In general, there is no penalty to use a multilevel model over a classical model, but large potential gains.

Varying Intercepts and Slopes

What is new about the following model? What are the pros/cons of fitting this type of model? Do you think this will provide an important improvement over earlier models for the radon data?

$$y_i \sim N(\alpha_{j[i]} + \beta_{j[i]} x_i, \sigma_y^2),\ \ i = 1, ..., n$$

$$\begin{pmatrix} \alpha_{j[i]} \\ \beta_{j[i]} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix}\right),\ j = 1, ..., J$$

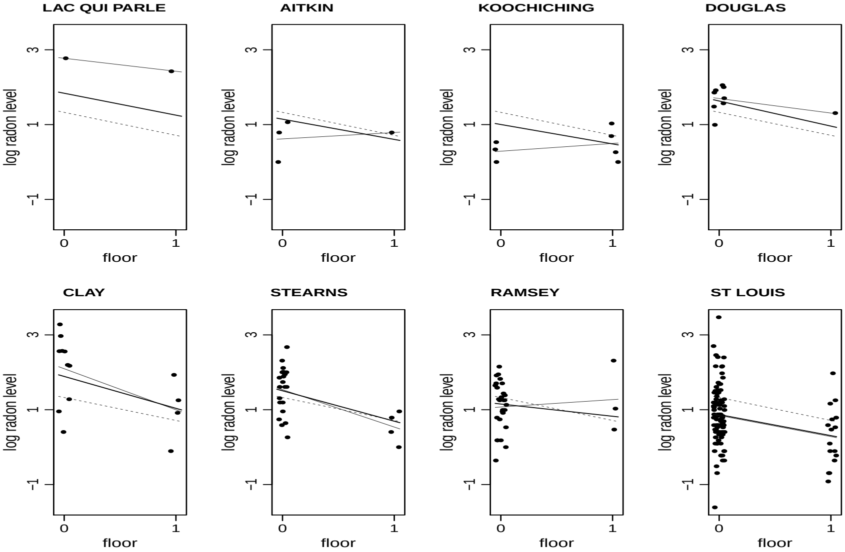
Varying Intercepts and Slopes

```
> M3<-lmer(y~x + (1 + x | county))
> display(M3)
lmer(formula = y ~ x + (1 + x | county))
            coef.est coef.se
(Intercept)  1.46     0.05
x           -0.68     0.09

Error terms:
 Groups   Name        Std.Dev. Corr
 county   (Intercept) 0.35
          x           0.34     -0.34
 Residual             0.75
---
number of obs: 919, groups: county, 85
AIC = 2180.3, DIC = 2153.9
deviance = 2161.1
```

Varying Intercepts and Slopes

```
> coef(M3)
$county
   (Intercept)          x
1    1.1445240 -0.5406161
2    0.9333816 -0.7708545
------
85   1.3787927 -0.6531793
> ranef(M3)
$county
    (Intercept)            x
1  -0.318245970  0.140484571
2  -0.529388410 -0.089753841
------
85 -0.083977357  0.027921419
```

Varying Intercepts and Slopes



Varying Slopes with Group Level Predictors

$$y_i \sim N(\alpha_{j[i]} + \beta_{j[i]} x_i, \sigma_y^2), \;\; i = 1, ..., n$$

$$\begin{pmatrix} \alpha_{j[i]} \\ \beta_{j[i]} \end{pmatrix} \sim N\left(\begin{pmatrix} \gamma_0^\alpha + \gamma_1^\alpha u_j \\ \gamma_0^\beta + \gamma_1^\beta u_j \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right), \;\; j = 1, ..., J$$

Varying Slopes with Group Level Predictors

```
lmer(formula = y ~ x + u.full + x:u.full + (1 + x | county)
            coef.est coef.se
(Intercept)  1.47     0.04
x           -0.67     0.08
u.full       0.81     0.09
x:u.full    -0.42     0.23

Error terms:
 Groups   Name        Std.Dev. Corr
 county   (Intercept) 0.12
          x           0.31     0.41
 Residual             0.75
---
number of obs: 919, groups: county, 85
AIC = 2142.6, DIC = 2102.1
deviance = 2114.3
```

Varying Slopes with Group Level Predictors