36-617: Multilevel and Hierarchical Models

Multilevel Models – The Basics

Brian Junker
132E Baker Hall
brian@stat.cmu.edu

Announcements

- Quiz 06 out today @5pm; due Tues @7pm as usual.
- HW 06 due Weds 1159pm (new std due date!)
    - HW 07 will be out tonight or tomorrow
- Final Report Project
    - Instead of another midterm or final exam
    - Will consist of a data analysis written up as an IDMRAD paper (variation of IMRAD format)
    - Assigned in pieces; first piece will be next week or the following week

Outline

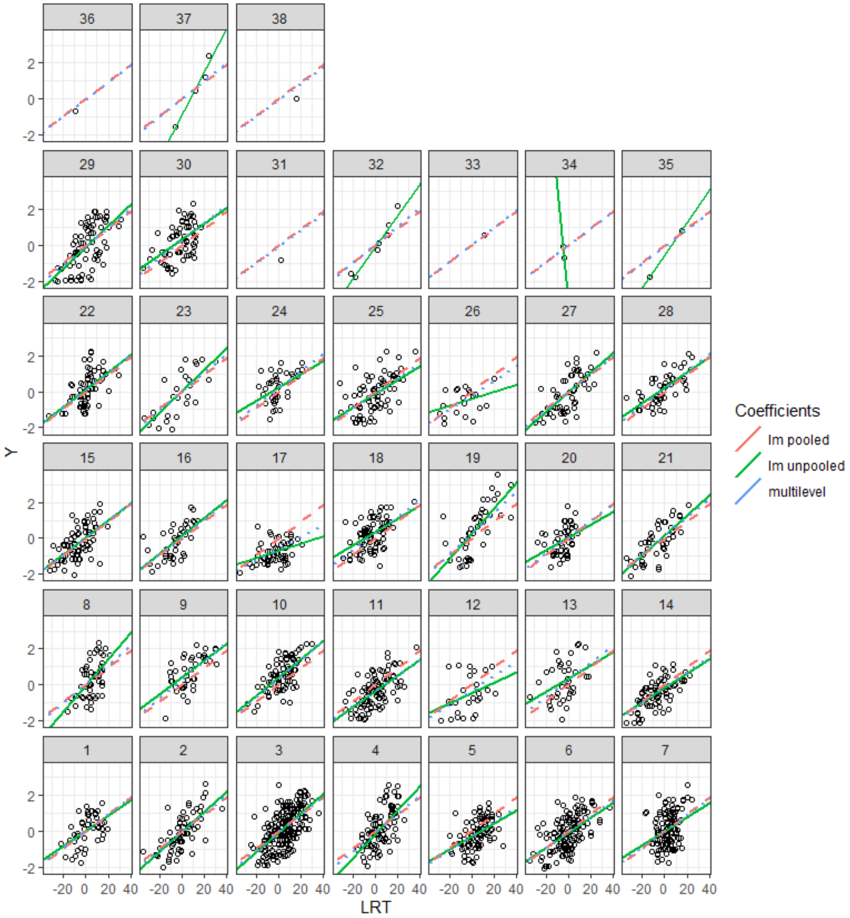
- Multilevel & Hierarchical Models
- Notation for Data in Groups
  - Level 1: Modeling Individual Observations
  - Level 2: Modeling the Groups
- Example: Minnesota Radon Levels, Part 1
  - Totally pooled vs totally unpooled
  - Fitting the random intercept model
  - An MLM phenomenon: Shrinkage
- Final Report Project
  - Will be assigned in “pieces” (nothing yet!)
  - Today we will talk about IMRAD & IDMRAD

Multilevel & Hierarchical Models

- *Most common*: linear regression and generalized linear regression (logistic regression) models
- *Next most common*: hierarchical and multilevel models
- Situations…
  - ❑ Clustered sampling
  - ❑ Grouped experimental trials
    - multicenter clinical trials in medicine
    - group-randomized trials in education
  - ❑ Growth curves and random coefficient models

Multilevel & Hierarchical Models

- Samples of students in 38 schools in London
  - LRT = reading pre-test
  - Y = reading post-test
- Red: fit lm() to all students, ignoring school
- Green: fit lm() to each school individually
- Blue: multilevel model



Multilevel & Hierarchical Models

- Traditional linear regression can either
  - Ignore the clumps completely and fit a single model to all the data
  - Treat each clump completely separately but fail to share information across clumps when some clumps "need help"
  - *Both of these are examples of "Fixed Effects"*
- Multilevel models (MLM's) allow
  - treating clumps separately, ***and***
  - sharing information across clumps to make better estimates
  - *These are examples of "Random Effects"*
- Strong connection between MLM's and Bayesian modeling

Multilevel & Hierarchical Models

All of the following refer to approximately the same class of models:

- These models emphasize connections with linear regression and generalized least squares (GLS):
  - Mixed Effects Models, or just Mixed Models
    - Both fixed and random effects
  - Variance Components Models
- These models emphasize the data generation process (& they are almost Bayesian):
  - Multilevel Models
  - Hierarchical Linear Models

Data for Multilevel Models

- Data that comes to us in clumps[1] of observations that are more like each other within a clump than between clumps, e.g.:
  - Classrooms within schools or schools within a city
  - States or geographic areas within a nation
  - Election precincts within a larger election
  - Answers given by the same student on a test
- Useful when a different linear regression should be fitted within each clump, but there is not enough information to separately estimate all clumps, e.g.:
  - Deducing state opinions from a national opinion survey
  - Fitting separate regressions to compare schools in London – some schools are represented by only 1 or 2 students!

---

[1] More commonly, clumps are called “clusters” or “groups”…

Notation for Data in Groups

$i = 1, \ldots n$ observations

$j = 1, \ldots J$ groups

outcomes

individual-level predictors

group-level predictors

	i	j	y_i	x_{i1}	x_{i2}	$\cdots$	u_{j1}	u_{j2}	$\cdots$
n_1 observations in group 1	1	1	60000	32	66		0	11	
	$\vdots$								
n_2 observations in group 2	33	2	7000	35	64		1	9	
	34	2	53000	55	73		1	9	
	35	2	10000	55	63		1	9	
	$\vdots$								
$\vdots$	123	3	50000	99	63		1	8	
	$\vdots$								
n_J observations in group J	244	J	21000	40	68		0	12	
	245	J	8000	8	64		0	12	
	246	J	3000	55	66		0	12	
	$\vdots$								

$n = n_1 + n_2 + \cdots + n_J$

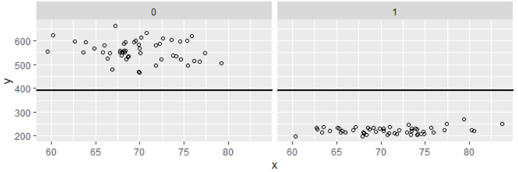
Given i, you can determine j, so $j = j[i]$ (but not vice-versa!)

Level 1 – Modeling Individual Observations

- Here are some models you could fit with lm()…

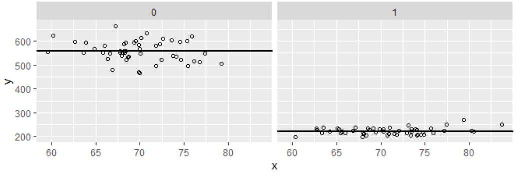
$$y_i = \beta_0 + \varepsilon_i$$

Intercept-only, ignore groups



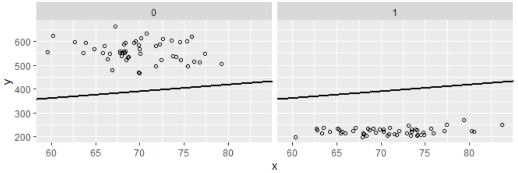
$$y_i = \beta_{0j[i]} + \varepsilon_i$$

Different intercept in each group



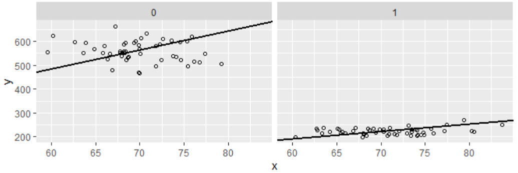
$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i$$

Individual level covariates, ignore groups



$$y_i = \beta_{0j[i]} + \beta_{1j[i]} x_{i1} + \varepsilon_i$$

Individual covariates, different regression in each group



$$y_i = \beta_{0j[i]} + \beta_{1j[i]} x_{i1} + \beta_3 u_{j1} + \varepsilon_i$$

Individual and group level covariates, different regression each group

```
> coef(lm(y ~ gp*x - x - 1 + u))
     gp0      gp1        u    gp0:x    gp1:x 
6.619303 6.362586       NA 3.094159 7.979893 
```

Level 2 – Modeling the Groups

- *Ordinary linear models* can ignore the groups *(totally pooled)*

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \ , \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

or build completely separate models for them *(totally unpooled)*

$$y_i = \beta_{0j[i]} + \beta_{1j[i]} x_{i1} + \varepsilon_i \ , \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

- *Multilevel models* treat the models in different groups as related: their coefficients are sampled from a common distribution *(partial pooling)*:

Level 1

$$\left\{ \; y_i = \alpha_{0j[i]} + \alpha_{1j[i]} x_{i1} + \varepsilon_i \ , \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2) \right.$$

Level 2

$$\left\{ \begin{array}{lcl} \alpha_{0j} & = & \beta_0 + \eta_{0j} \ , \quad \eta_{0j} \overset{iid}{\sim} N(0, \tau_0^2) \\ \alpha_{1j} & = & \beta_1 + \eta_{1j} \ , \quad \eta_{1j} \overset{iid}{\sim} N(0, \tau_1^2) \end{array} \right.$$

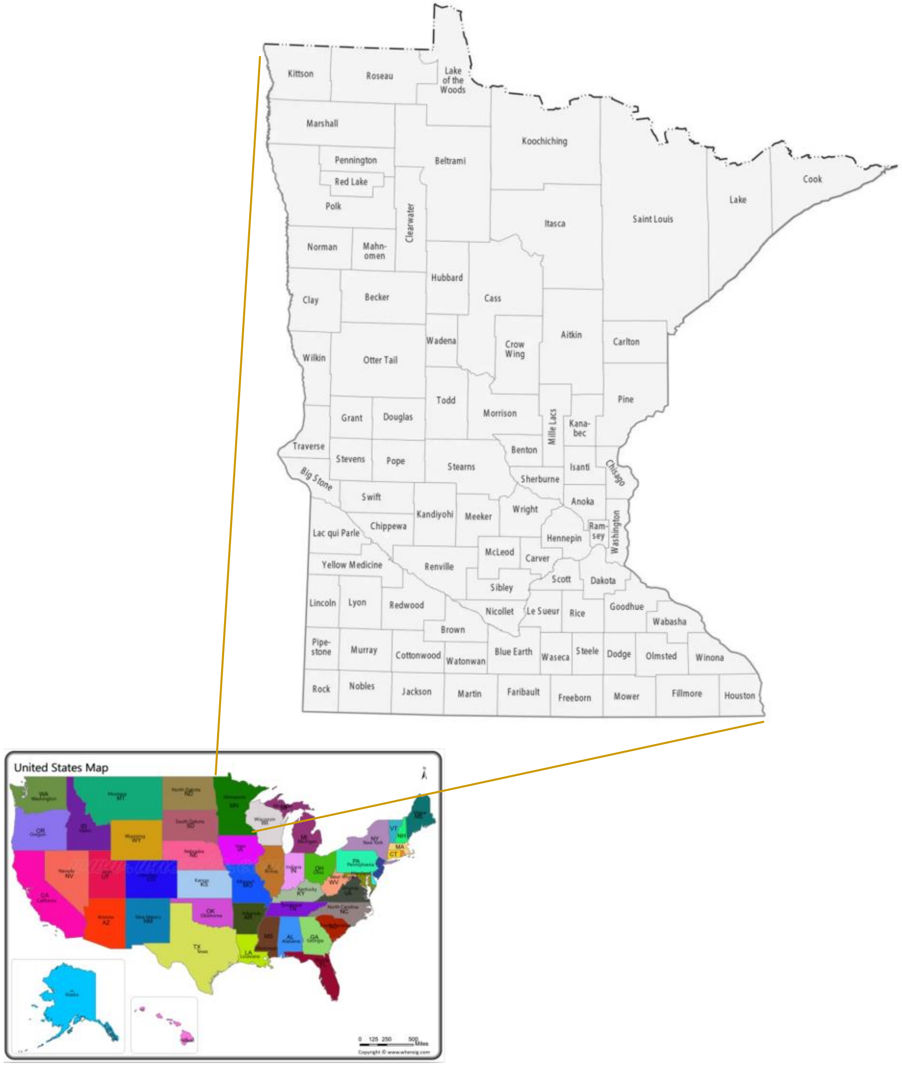
α	"alpha"	β	"beta"	σ	"sigma"
ε	"epsilon"	η	"eta"	τ	"tau"

Minnesota Radon Levels, Part 1

- Each observation in the data set is a house

  Individual-level (house-level) variables:
  - radon, log(radon)
  - floor = 0 if measurement was made in basement;
    = 1 if measurement on first floor
- Houses are grouped into counties

  Group-level (county-level) variables:
  - county.name & county number
  - uranium & log(uranium) – measurement of uranium in the soil in each county
- We want to predict radon levels from the other variables

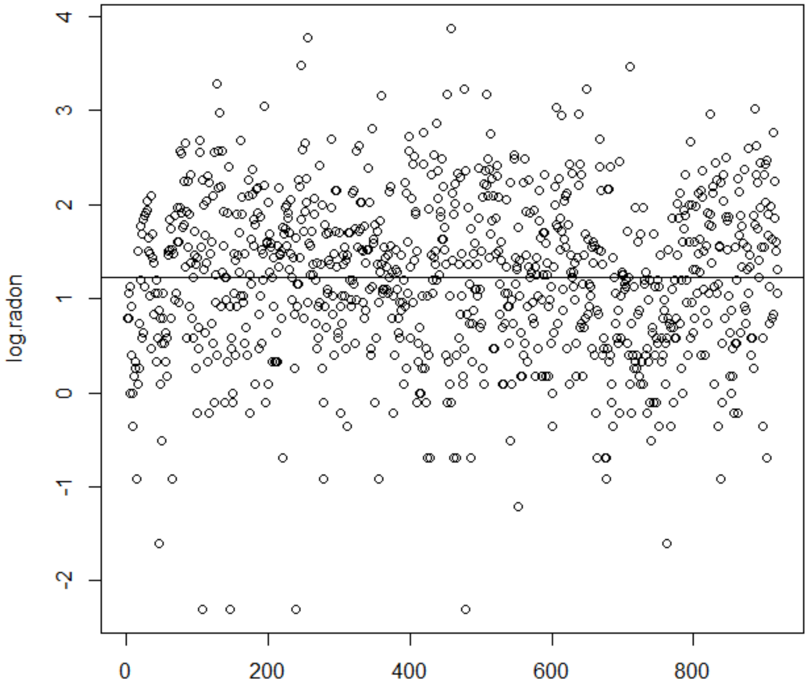


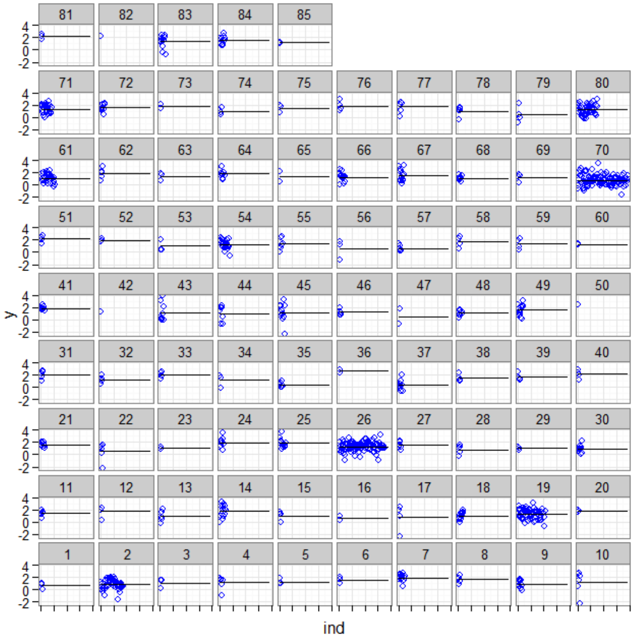
Data from Gelman & Hill (2007).

Mn Radon Levels, Raw Data

```
> mn.radon <- read.table("mn-radon.txt")
> head(mn.radon)
    radon log.radon floor           county.name county  uranium log.uranium
  1   2.2 0.7884574     1 AITKIN                     1 0.502054  -0.6890476
  2   2.2 0.7884574     0 AITKIN                     1 0.502054  -0.6890476
  3   2.9 1.0647107     0 AITKIN                     1 0.502054  -0.6890476
  4   1.0 0.0000000     0 AITKIN                     1 0.502054  -0.6890476
  5   3.1 1.1314021     0 ANOKA                      2 0.428565  -0.8473129
  6   2.5 0.9162907     0 ANOKA                      2 0.428565  -0.8473129
> tail(mn.radon)
    radon log.radon floor           county.name county  uranium log.uranium
914   9.5  2.251292     0 WRIGHT                    84 0.913909 -0.09002427
915   6.4  1.856298     0 WRIGHT                    84 0.913909 -0.09002427
916   4.5  1.504077     0 WRIGHT                    84 0.913909 -0.09002427
917   5.0  1.609438     0 WRIGHT                    84 0.913909 -0.09002427
918   3.7  1.308333     0 YELLOW MEDICINE           85 1.426590  0.35528698
919   2.9  1.064711     0 YELLOW MEDICINE           85 1.426590  0.35528698
>
```

Totally pooled vs *totally unpooled* log(radon) intercept-only models





$y_i = \beta_0 + \epsilon_i$
i = house,
no attention paid to county

$y_i = \beta_{j[i]} + \epsilon_i$
i = house,
$j[i]$ = county that house i is in

Looking at the coefficients from fitting separate (unpooled) models

```
> cties <- as.factor(county)
> contrasts(cties) <- contr.sum(85)
> summary(lm.0 <- lm(y ~ 1))
             Est     SE     t value Pr(>|t|)
(Intercept)  1.22    0.03   43.51   <2e-16 ***

> summary(lm.1 <- lm(y ~ cties))
             Est    SE    t value Pr(>|t|)
(Intercept)  1.34   0.04  32.01   < 2e-16 ***
cties1      -0.68   0.40  -1.72 0.09 .
cties2      -0.51   0.12  -4.36 1.49e-05 ***
cties3      -0.30   0.46  -0.65 0.52
cties4      -0.20   0.30  -0.67 0.50
cties5      -0.09   0.39  -0.23 0.82
cties6       0.17   0.46   0.37 0.71
cties7       0.57   0.21   2.63 0.01 **
cties8       0.29   0.40   0.72 0.47
cties9      -0.41   0.25  -1.63 0.10
cties10     -0.14   0.32  -0.43 0.67
cties11      0.06   0.36   0.16 0.87
cties12      0.39   0.40   0.98 0.33
cties13     -0.30   0.32  -0.94 0.35
cties14      0.44   0.21   2.04 0.04 *
```

```
cties15     -0.37    0.40   -0.92 0.36
cties16     -0.68    0.56   -1.21 0.23
  .            .       .       .    .
  .            .       .       .    .
  .            .       .       .    .
cties70     -0.58    0.08   -6.82 1.80e-11 ***
cties71      0.03    0.16    0.20 0.84
cties72      0.24    0.25    0.93 0.35
cties73      0.45    0.56    0.80 0.42
cties74     -0.36    0.40   -0.90 0.37
cties75      0.14    0.46    0.31 0.76
cties76      0.48    0.40    1.22 0.22
cties77      0.38    0.30    1.25 0.21
cties78     -0.35    0.36   -0.97 0.33
cties79     -0.91    0.40   -2.29 0.02 *
cties80     -0.09    0.12   -0.75 0.45
cties81      0.89    0.46    1.94 0.05 .
cties82      0.89    0.79    1.12 0.26
cties83      0.11    0.22    0.51 0.61
cties84      0.25    0.22    1.11 0.27
> print(c(cties85 = -sum(coef(lm.1)[-1])))
cties85
-0.1571159
```

Problems with totally-pooled vs totally-unpooled

- *Totally-pooled*: It looks like there is some pattern to the county means, so this “over-smooths” (forces all the counties to be the same)
- *Totally-unpooled*: Although the counties have some variation in means, there may not be very much!

Having different means is better than totally-pooled model…

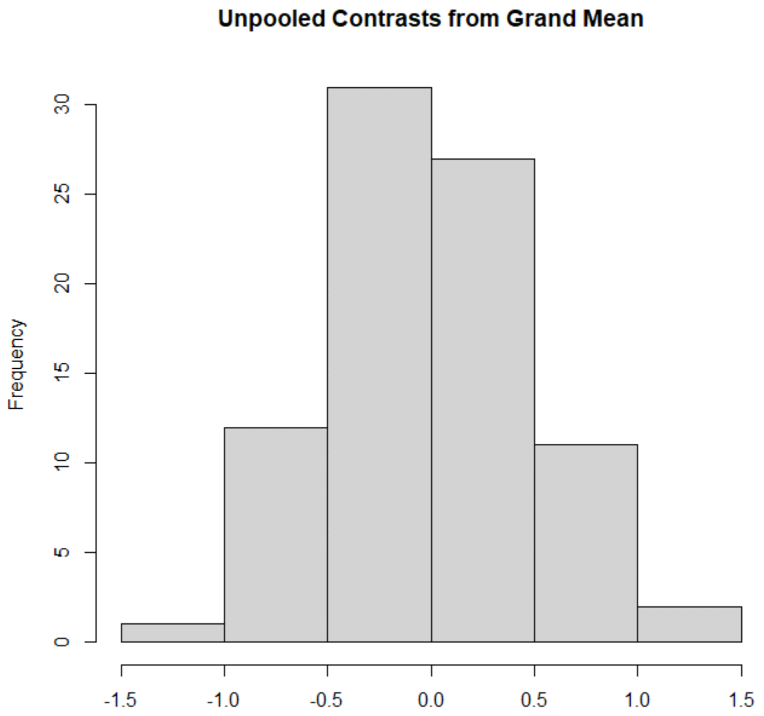
```
cties <- as.factor(county)
contrasts(cties) <- contr.sum(85)
lm.1 <- lm(y ~ cties)
anova(lm.1)
#            Df Sum Sq Mean Sq F value    Pr(>F)
# cties      84 136.89 1.62960  2.5567 1.736e-11 ***
# Residuals 834 531.57 0.63738
```

,…but very few county means are different from overall mean!

```
length(unique(county))
# [1] 85
sum(coef(summary(lm.1))[,4]<0.05)
# [1] 15
15/85
# [1] 0.1764706
```

Motivating the Multilevel Model…

```
> hist(coef(lm.1)[-1],xlab="",
+       main="Unpooled Contrasts from Grand Mean")
```



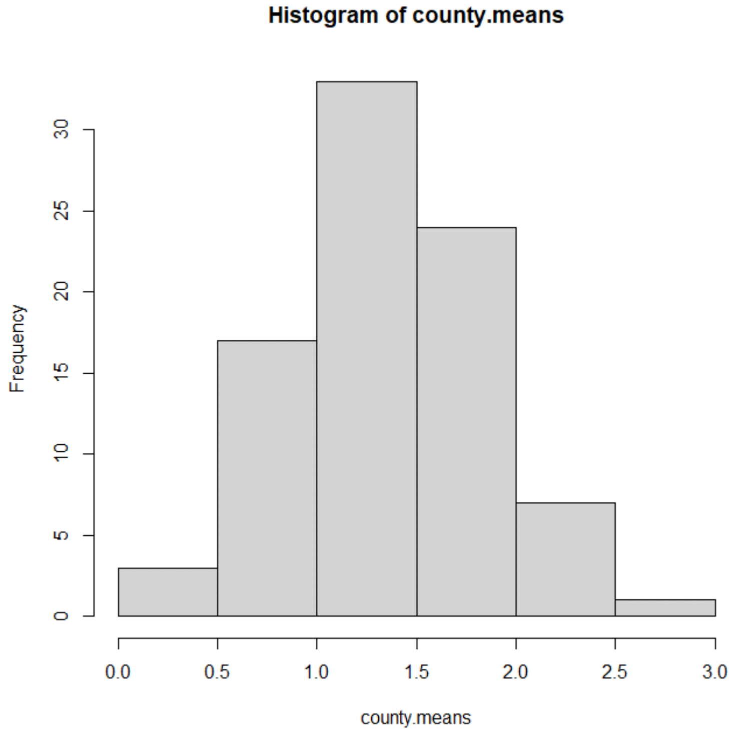
- The coefficients are nearly normally distributed!
- Suggests that we modify our lm() models…

A compromise between totally-pooled and totally-unpooled

- The 85 county means look rather “normal”, so why not model them that way?

$$
\begin{aligned}
y_i &= \alpha_{j[i]} + \epsilon_i,\ \epsilon_i \overset{iid}{\sim} N(0,\sigma^2) \\
\alpha_j &= \beta_0 + \eta_j,\ \eta_j \overset{iid}{\sim} N(0,\tau^2)
\end{aligned}
$$

- Sometimes called a “***random intercept***” model



Same data as on prev. slide, but now centered at $\hat{\beta}_0$, instead of at zero.

Fitting the random-intercept model

$$y_i = \alpha_{j[i]} + \epsilon_i, \ \epsilon_i \overset{iid}{\sim} N(0, \sigma^2) \qquad (1)$$

$$\alpha_j = \beta_0 + \eta_j, \ \eta_j \overset{iid}{\sim} N(0, \tau^2) \qquad (2)$$

Multilevel model (both equations 1 and 2)

```
library(lme4)
random.intercept.model <-
   lmer( y ~ 1 + ( 1 | county.name ) )
summary(lmer.intercept.only)
# Random effects:
#  Groups      Name                Var    SD
#  county.name (Intercept)        0.096 0.310
#  Residual                       0.637 0.798
# Numb. of obs: 919, grps: county.name, 85
#
# Fixed effects:
#             Estimate      SE   t value
# (Intercept)     1.31    0.05     26.84
```

$\hat{\sigma}^2$ $\hat{\tau}^2$

$\hat{\beta}_0$

Unpooled fixed effects (equation 1 only)

```
cties <- as.factor(county)
contrasts(cties) <- contr.sum(85)
lm.1 <- lm(y ~ cties)
summary(lm.1)
# Coefficients:
#                 Est     SE       t Pr(>|t|)
# (Intercept)    1.34   0.04   32.01  < 2e-16 ***
# cties1        -0.68   0.40   -1.72 0.085374 .
# cties2        -0.51   0.11   -4.36 1.49e-05 ***
# cties3        -0.30   0.46   -0.65 0.518720
# […]
# Residual std err:  0.7984 on 834 df
```

We'll get a proper introduction to library(lme4) and lmer() shortly…

Random-intercept model: Where are the intercepts?

```
> summary(random.intercept.model)
Random effects:
 Groups      Name        Variance Std.Dev
 county.name (Intercept) 0.095813 0.30954
 Residual                0.636621 0.79789
Numb. of obs: 919, grps: county.name, 85

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.31257    0.04891   26.84
> fixef(random.intercept.model)
(Intercept)
   1.312574
> ranef(random.intercept.model)
$county.name

                      (Intercept)
AITKIN               -0.245071104
ANOKA                -0.425038053
BECKER               -0.082191868
BELTRAMI             -0.088030506
BENTON               -0.022598796
BIG STONE             0.062346490
BLUE EARTH            0.404629013
[….]
```

```
> summary(lm.1)

Call:
lm(formula = y ~ cties)


Coefficients:
             Estimate Std. Error t value
    Pr(>|t|)
(Intercept)  1.343638   0.041980  32.006
cties1      -0.683231   0.396682  -1.722
cties2      -0.510388   0.117180  -4.356
cties3      -0.295300   0.457408  -0.646
cties4      -0.202652   0.301120  -0.673
cties5      -0.091202   0.396682  -0.230
cties6       0.169372   0.457408   0.370
cties7       0.565589   0.214984   2.631
[…]
```

Random effects – draws from N(0,τ^2)

Fixed effects – estimates of regression coefficients

An MLM phenomenon: Shrinkage

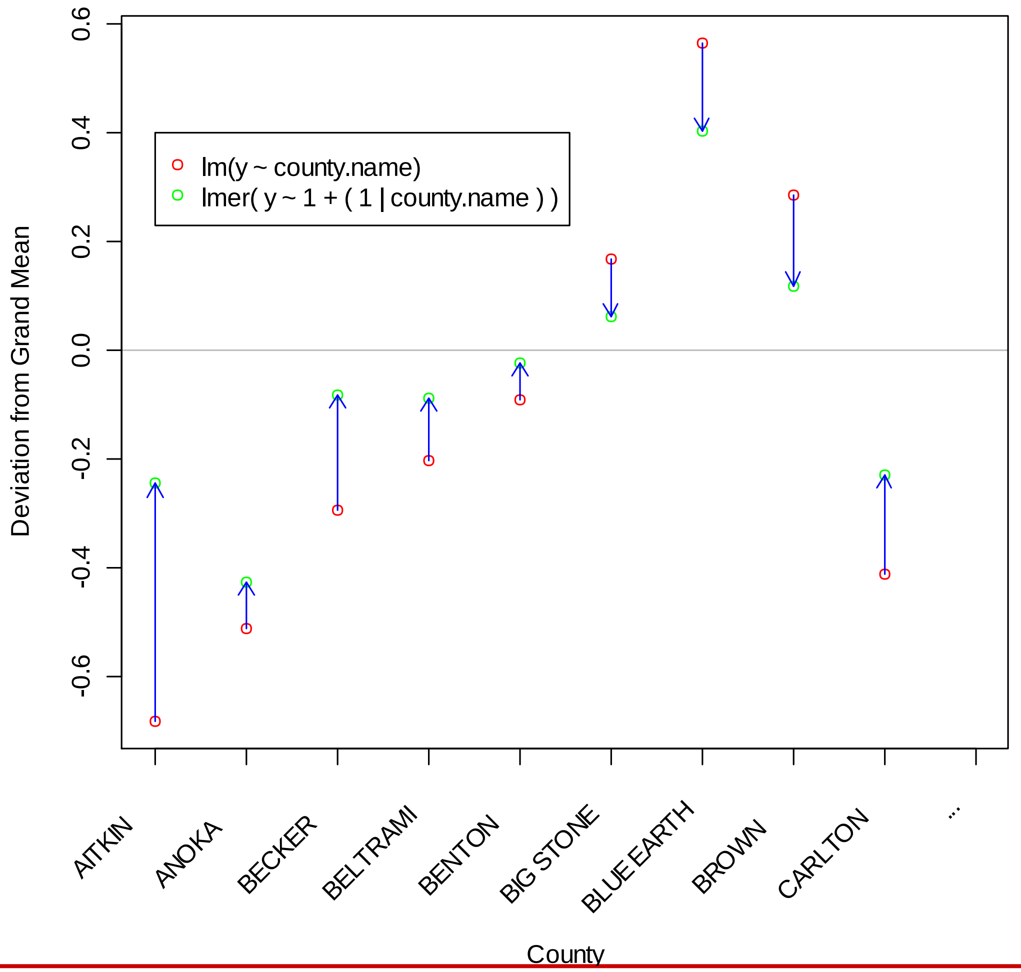
For a county j with n_j houses:

- $\bar{y}_j$= county mean for county j
- $\bar{y}_{\text{all}}$ = grand mean (all of MN)
- $\widehat{\sigma^2}$= variance of y_i's within county j
- $\widehat{\tau^2}$= variance of $\bar{y}_j$'s across MN

Then we will show later in the course that

$$\hat{\alpha}_j \approx \frac{\frac{n_j}{\widehat{\sigma^2}}\bar{y}_j + \frac{1}{\widehat{\tau^2}}\bar{y}_{\text{all}}}{\frac{n_j}{\widehat{\sigma^2}} + \frac{1}{\widehat{\tau^2}}}$$

$$= \left(\frac{\widehat{\tau^2}}{\widehat{\tau^2} + \widehat{\sigma^2}/n_j}\right)\bar{y}_j + \left(\frac{\widehat{\sigma^2}/n_j}{\widehat{\tau^2} + \widehat{\sigma^2}/n_j}\right)\bar{y}_{\text{all}}$$



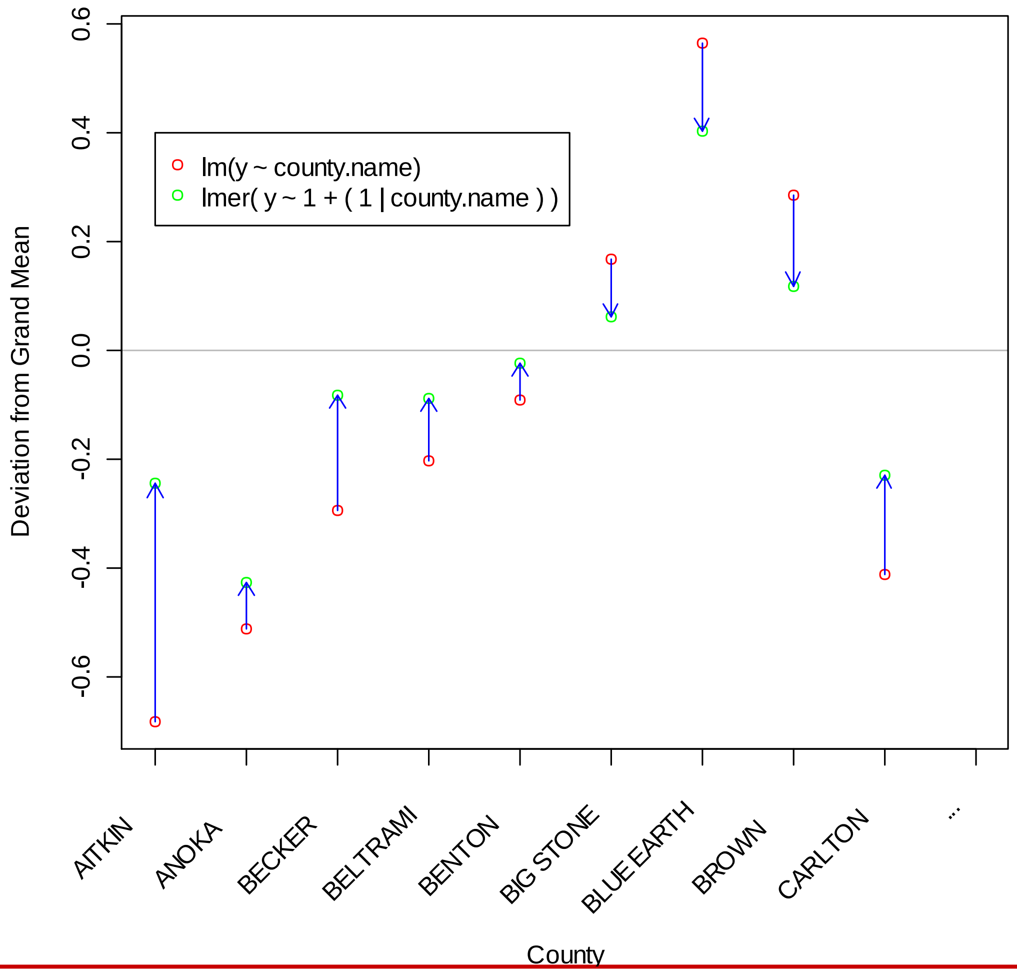
An MLM phenomenon: Shrinkage

The fitted multilevel model underpredicts high obs's and overpredicts low ones.

The distribution assumptions underlying lmer() "smooth out" extreme observations!

Multi-level models provide more smoothing/shrinkage to groups with smaller sample sizes (since there is less evidence that their values should be different from "grand mean".)

This is what lets MLMs "borrow" from all the data to get better estimates of small groups



Final Report Project

- Will be assigned in several pieces (nothing yet!)
- Will involve
  - Extended data analysis (like the midterm)
  - Writing a report using the IDMRAD format
    - Your detailed data analyses will be a technical appendix for this report
- You will
  - Write a rough draft,
  - Review each others’ rough drafts,
  - Submit a final draft in the last week of classes

IMRAD – A canonical way to organize empirical papers & reports

Abstract	Summarize I, M, R and D of paper
(I)introduction	Why would anyone want to read this paper? What questions will be addressed?
(M)ethods	What did you do, to address these questions?
(R)esults	What did you find?
(a)nd	
(D)iscussion	What does it all mean? Typically: answer questions, discuss generalizations & limitations

More information on IMRAD…

- *How prevalent are IMRAD papers? Very…*
  Sollaci et al. (2004). The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. *J Med Libr Assoc 92(3)*, 364—367.

- *Quick advice on IMRAD contents…*
  Aggarwal (2004). *IMRAD: What goes into each section?* (slides). http://www.jpgmonline.com/documents/author/24/2_Aggarwal_10.pdf

From IMRAD to IDMRAD…

Abstract	Summarize I, D, M, R and D of paper
(I)introduction	Why would anyone want to read this paper? What questions will be addressed?
(D)ata	What dataset was used for this study? Typically: Variable definitions, sample size, quick summaries and initial descriptive EDA
(M)ethods	What did you do, to address these questions?
(R)esults	What did you find?
(a)nd	
(D)iscussion	What does it all mean? Typically: answer questions, discuss generalizations & limitations
Technical Appendix	Technical details of carrying out the (M)ethods

The Technical Appendix

- Most statistics papers are based on lots of technical analysis.
- Most readers of the main paper won't want to see all the details, but some (me!) will want to know that you handled the details well.
- A technical appendix is a good place to collect together the analyses that contributed to the main paper, **in the order they will be presented in the paper.**
    - *NOT the order in which you did the analyses!!*
- Don't include lots of analyses not mentioned in the paper.
    - *The paper can and should cite sections of the appendix* to show reader where the details are, for the interested reader.
- Do include text and comments in the appendix explaining why you did the analyses you did.

Summary

- Multilevel & Hierarchical Models
- Notation for Data in Groups
  - Level 1: Modeling Individual Observations
  - Level 2: Modeling the Groups
- Example: Minnesota Radon Levels, Part 1
  - Totally pooled vs totally unpooled
  - Fitting the random intercept model
  - An MLM phenomenon: Shrinkage
- Final Report Project
  - Will be assigned in “pieces” (nothing yet!)
  - Today we will talk about IMRAD & IDMRAD