36-463/663: Multilevel and Hierarchical Models

Multilevel Models – The Basics II

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

Announcements

- Quiz 06 (the short answer part) will be graded shortly.
- HW 06 due tonight 1159pm (new std due date!)
  - HW 07 is out; due next Weds
- Final Report Project
  - Assigned in pieces; first piece will be next week or the following week
- Reading for next week:
  - Finish G&H Ch 13 and read G&H Ch 14 (in week09 folder)

Outline

- Three Ways to Write The Model
- Brief intro to library(lme4) and lmer()
- lmer() notation / modeling language
- Minnesota Radon Levels, Part 2
  - Predicting county means with log(uranium)
  - Intercept-only random-interecept model
  - Random-intercept model with level 1 predictors
  - Random-intercept model with level 1 and level 2 predictors
- More than one random effect
  - Random intercept, random slope
  - Correlation between random effects

There is a lot of good code in mn-radon.r, that you will want to use as the course progresses.

Different ways to write the random-intercept model

- Multi-level Model (emphasize regression)

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

- Variance Components Model (substitute for α_j)

$$y_i = \beta_0 + \eta_{j[i]} + \epsilon_i, \qquad \epsilon_i \overset{iid}{\sim} N(0,\sigma^2)$$
$$\eta_j \overset{iid}{\sim} N(0,\tau^2)$$

- Hierarchical Model (emphasize distributions)

Level 2: $\alpha_j \overset{iid}{\sim} N(\beta_0, \tau^2)$

Level 1: $y_i \overset{indep}{\sim} N(\alpha_{j[i]}, \sigma^2)$

Multi-level Model (a.k.a. Hierarchical Linear Model)

- Emphasize Regression Structure

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

Level 2 $\left\{ \alpha_j \;=\; \beta_0 + \eta_j, \;\; \eta_j \overset{iid}{\sim} N(0, \tau^2) \right.$

- Easy to use intuitions from lm() at each “level” of the model, to build and evaluate models

Variance Components Model

- Emphasize Error Structure

$$y_i = \beta_0 + \eta_{j[i]} + \epsilon_i, \qquad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

$$\eta_j \overset{iid}{\sim} N(0, \tau^2)$$

- Errors from different sources
  - η_j from groups/counties (j); Var(county level) = τ^2
  - ε_i from individual houses (i); Var(arbitrary house) = $\tau^2 + \sigma^2$
  - If $j[i] \neq j[i']$: $\mathrm{Cov}(y_i, y_{i'}) = 0$;
  - If $j[i] = j[i']$: $\mathrm{Cov}(y_i, y_{i'}) = \tau^2$, $\mathrm{Cor}(y_i, y_{i'}) = \tau^2/(\tau^2 + \sigma^2)$

Intra-class correlation (ICC)

- $\mathrm{Var}(\bar{y}_j) = \mathrm{Var}(\beta_0 + \eta_j + \frac{1}{n_j} \sum_{\text{all } i \in \text{ county } j} \epsilon_i) = \tau^2 + \sigma^2/n_j$;
  - The average is a ***reliable*** measure of county levels if σ^2/n_j is much: smaller than τ^2:

$$\frac{\text{Var (county level)}}{\text{Var (average of houses in county)}} = \frac{\tau^2}{\tau^2 + \sigma^2/n_j}$$

reliability

Hierarchical (Bayes) Model

- Emphasize Distribution Structure

$$\text{Level 2:} \quad \alpha_j \overset{iid}{\sim} N(\beta_0, \tau^2)$$

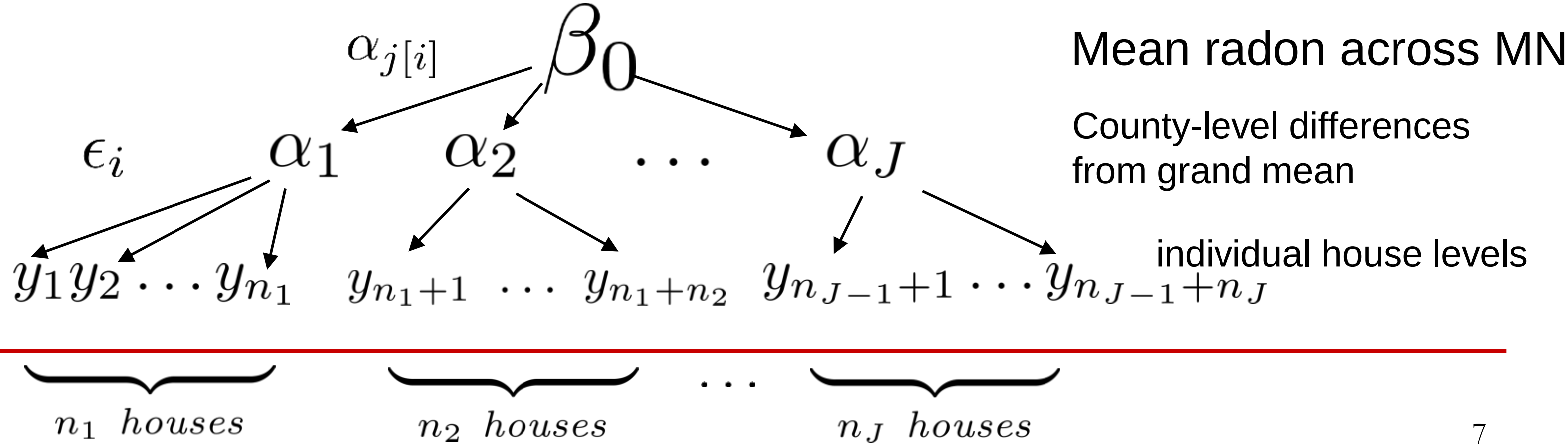
$$\text{Level 1:} \quad y_i \overset{indep}{\sim} N(\alpha_{j[i]}, \sigma^2)$$

- Emphasize Bayesian point of view

$$\text{Prior:} \quad \alpha_j \overset{iid}{\sim} N(\beta_0, \tau^2)$$

$$\text{Likelihood:} \quad y_i \overset{indep}{\sim} N(\alpha_{j[i]}, \sigma^2)$$

- Emphasize two-stage (multistage) sampling



A brief introduction to lmer()…

- library(lme4)*
- Main functions
  - lmer() for multilevel models that generalize lm()
  - glmer()** for multilevel models that generalize glm()
- Helper functions
  - summary() – like for lm()
  - fixef() – gets the β's
  - ranef() – gets the η's
  - anova(), AIC(), BIC(), etc. – we'll talk about in detail later…
- Additional libraries
  - library(arm) – display() and other helper functions for lmer()
  - library(lmerTest), library(cAIC4), ... – model comparison/selection
  - Other libraries as needed…

*Not library(Matrix) as in G&H
**Not lmer() again as in G&H

lmer() notation follows the variance components formulation…

- Multilevel model:

$$
\begin{aligned}
y_i &= \alpha_{0j[i]} + \beta_2 x_i + \epsilon_i \\
\alpha_{0j} &= \beta_0 + \beta_1 u_j + \eta_j
\end{aligned}
$$

- Variance components model:

$$
y_i = \beta_0 + \beta_1 u_{j[i]} + \beta_2 x_i + \eta_{j[i]} + \epsilon_i
$$

- lmer() model:

```
lmer(y ~ 1 + u + x + (1 | group)) + ϵi
```

Fixed Effects, Random Effects

- Fixed effects are considered to be "fixed but unknown" and we try to estimate them, e.g. with lm(), or the non-parenthesis terms in lmer()
- Random effects are considered to be draws from a distribution (not fixed)

$$y_i = \beta_0 + \beta_1 u_{j[i]} + \beta_2 x_i + \eta_{j[i]} + \epsilon_i$$

```
lmer(y  ~        u    +  x    +  (1 | group))
```

lmer() notation examples…

Formula	Alternative	Meaning
`(1 | g)`	`1 + (1 | g)`	Random intercept with fixed mean.
`0 + offset(o) + (1 | g)`	`-1 + offset(o) + (1 | g)`	Random intercept with *a priori* means.
`(1 | g1/g2)`	`(1 | g1) + (1 | g1:g2)`	Intercept varying among `g1` and `g2` within `g1`.
`(1 | g1) + (1 | g2)`	`1 + (1 | g1) + (1 | g2)`	Intercept varying among `g1` and `g2`.
`x + (x | g)`	`1 + x + (1 + x | g)`	Correlated random intercept and slope.
`x + (x || g)`	`1 + x + (1 | g) + (0 + x | g)`	Uncorrelated random intercept and slope.

Table 2: Examples of the right-hand-sides of mixed-effects model formulas. The names of grouping factors are denoted `g`, `g1`, and `g2`, and covariates and *a priori* known offsets as `x` and `o`.

From: Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67(1),* 1–48. https://doi.org/10.18637/jss.v067.i01

Try it with the random intercept model…

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

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

$$\Rightarrow y_i = \beta_0 + \eta_{j[i]} + \varepsilon_i$$ Variance Components Model

```
lmer(y ~ 1 + (1 | county.name)) + ε
```

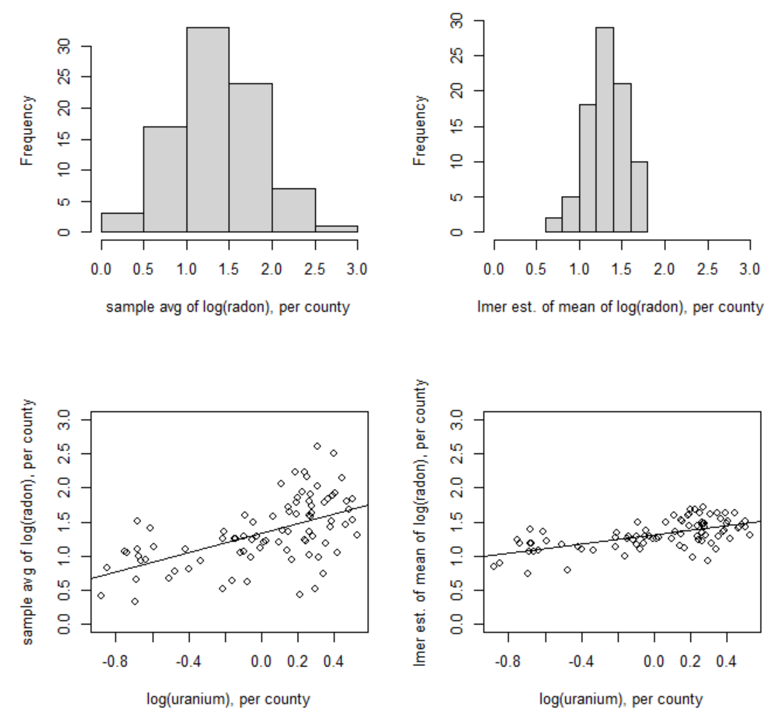
R…

Minnesota Radon Levels, Part 2

```
> str(mn.radon)
'data.frame':   919 obs. of  7 variables:
 $ radon       : radon level in each house
 $ log.radon   : log(radon)
 $ floor       : where measured: 0 = basement; 1 = 1st floor
 $ county.name:  name of county
 $ county      : numerical index of county
 $ uranium     : average uranium content of soil, per county
 $ log.uranium: log(uranium)

> attach(mn.radon)
> y <- log.radon
> x <- floor
> u <- sapply(split(log.uranium,county), function(x){x[1]})
> detach()
> rand.int.model <- lmer(y ~ 1 + (1|county.name))
```

Is log(uranium) useful in the model?

```
> par(mfrow=c(2,2))
> county.means <- sapply(split(log.radon,
+ county), mean)
> hist(county.means,nclass=6,xlim=c(0,3),
+ main="", xlab="avg log(radon), per county")
> ajhat <- fixef(ran.int.model) +
+ ranef(rand.int.model)$county.name[,1]
> hist(ajhat,nclass=6,xlim=c(0,3), main="",
+ xlab="lmer est. of mean log(radon) ")
> plot(county.means ~ u, ylim=c(0,3),
+ xlab="log(uranium), per county",
+ ylab="avg log(radon), per county")
> abline(lm(county.means ~ u))
> plot(ajhat ~ u, ylim=c(0,3),
+ xlab="log(uranium), per county",
+ ylab="lmer est. of mean log(radon) ")
> abline(lm(ajhat ~ u))
```



`> summary(lm(county.means ~ u))`

	Est	SE	t-val	p-val	
(Int)	1.33	0.04	30.34	< 2e-16	***
u	0.71	0.12	6.20	2.06e-08	***

`> summary(lm(ajhat ~ u))`

	Est	S	t-val	p-val	
(Int)	1.31	0.02	70.97	< 2e-16	***
u	0.33	0.05	6.95	7.61e-10	***

Suggests:

- Instead of

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

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

we could try to fit

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

$$\alpha_j = \beta_0 + \beta_1 u_j + \eta_j, \eta_j \overset{iid}{\sim} N(0, \tau^2)$$

U_j = log(uranium$_j$)

- Variance components form:

$$y_i = \beta_0 + \beta_1 u_{j[i]} + \eta_{j[i]} + \epsilon_i$$

- lmer model:

```
lmer( y ~ 1 + log.uranium + (1|county.name))
```

Fitting this model to the radon data...

```
> summary(lmer.intercepts.depend.on.log.ur-
    anium)
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ 1 + log.uranium +
    (1 | county.name)

REML criterion at convergence: 2219.794

Random effects:
 Groups      Name        Variance Std.Dev.
 county.name (Intercept) 0.01406  0.1186
 Residual                0.64037  0.8002
Number of obs: 919, groups: county.name, 85

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.33305    0.03397   39.24
log.uranium  0.71912    0.08777    8.19

Correlation of Fixed Effects:
            (Intr)
log.uranium 0.197
```

$\tau^2 = \text{Var}(\eta_j)$

$\sigma^2 = \text{Var}(\varepsilon_i)$

β_0

β_1

```
> fixef(lmer.intercepts.depend.on.log.ur-
    anium)
(Intercept) log.uranium
  1.3330508   0.7191188
> ranef(lmer.intercepts.depend.on.log.ur-
    anium)
$county.name
                        (Intercept)
AITKIN                -0.0142971713
ANOKA                  0.0583741025
BECKER                -0.0125490841
BELTRAMI               0.0312484900
BENTON                 0.0017869830
BIG STONE             -0.0060780289
BLUE EARTH             0.0895241245
BROWN                  0.0078003746
CARLTON               -0.0293551573
CARVER                -0.0230826914
CASS                   0.0499879229
CHIPPEWA               0.0161734868
CHISAGO                0.0272838175
CLAY                   0.0475401692
[…]
```

Estimates of the η_j's themselves

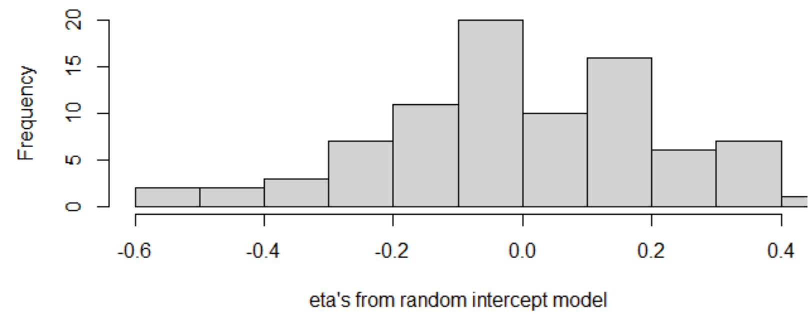
We can see that this improves the fit of the model by looking at the η's or the τ's & σ's

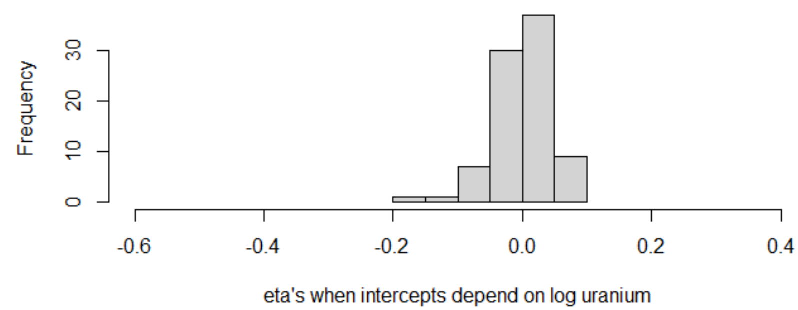
```
> hist(ranef(random.intercept.model)$
+ county.name[,1], xlim=c(-.6,.4),main="",
+ xlab="eta's from rand int model")
> hist(ranef(lmer.ints.depend.on.log.u)$
+ county.name[,1], xlim=c(-.6,.4),main="",
+ xlab="eta's when ints depend on log u")
> VarCorr(rand.int.model)
```

Groups	Name	Std.Dev.	
county.name	(Int)	0.30954	← $\hat{\tau}$
Residual		0.79789	← $\hat{\sigma}$

```
> VarCorr(lmer.ints.depend.on.log.u)
```

Groups	Name	Std.Dev.	
county.name	(Int)	0.11855	← $\hat{\tau}$
Residual		0.80023	← $\hat{\sigma}$





There are lots of different models we could fit… Here are some examples.

- Intercept-only random-intercept model

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

- Random-intercept model w / individual-level predictors

$$\begin{aligned} y_i &= \alpha_{0j[i]} + \alpha_1 x_i + \epsilon_i, \ \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2) \\ \alpha_{0j} &= \beta_0 + \eta_j, \qquad \eta_j \stackrel{iid}{\sim} N(0, \tau^2) \end{aligned}$$

- Random-intercept model w / individual & group level predictors

$$\begin{aligned} y_i &= \alpha_{0j[i]} + \alpha_1 x_i + \epsilon_i, \ \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2) \\ \alpha_{0j} &= \beta_0 + \beta_1 u_j + \eta_j, \qquad \eta_j \stackrel{iid}{\sim} N(0, \tau^2) \end{aligned}$$

Intercept-only random-intercept model

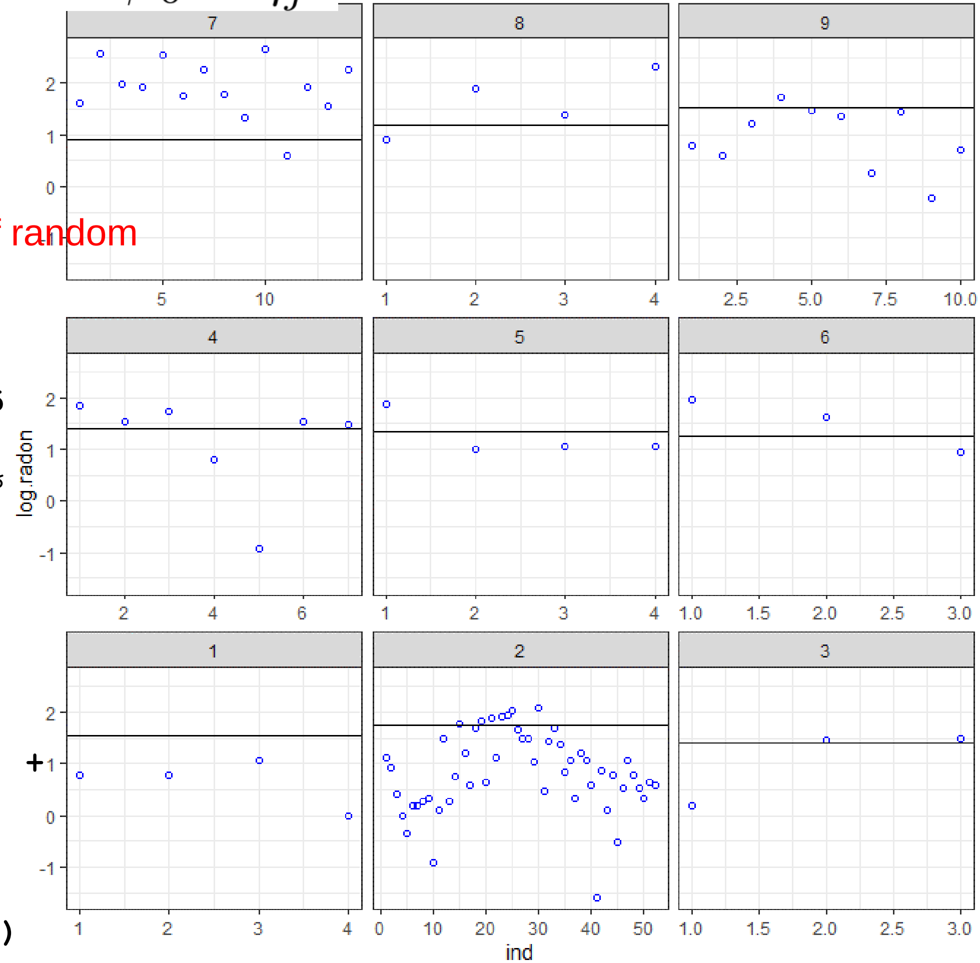
$$y_i = \alpha_{j[i]} + \epsilon_i$$
$$\alpha_j = \beta_0 + \eta_j$$

display() is a briefer version of summary() from library(arm)

```
> y <- log.radon
> M0 <- lmer(y ~ 1 + (1 | county) )
> display(M0)
coef.est  coef.se
    1.31     0.05
Error terms:
 Groups   Name        Std.Dev.
 county   (Intercept) 0.31
 Residual             0.80
number of obs: 919, groups: county, 85
AIC = 2265, DIC = 2251, deviance = 2255
> plot.counties <- 1:9
> county.sample <- mn.radon$county %in%
   plot.counties
> subset <- mn.radon[county.sample,]
> alpha0 <- coef(M0)$county[,1]
> params <- data.frame(plot.counties,
   alpha0[plot.counties],slopes=0)
> names(params) <- c("county","alpha0",
   "slopes")
> ggplot(subset,aes(x=ind,y=log.radon)) +
    facet_wrap( ~ county, as.table=F,
   scales="free_x") +
    geom_point(pch=1,color="blue") +
   geom_abline(data=params,
   aes(intercept=alpha0, slope=slopes))
```

Fixed effect

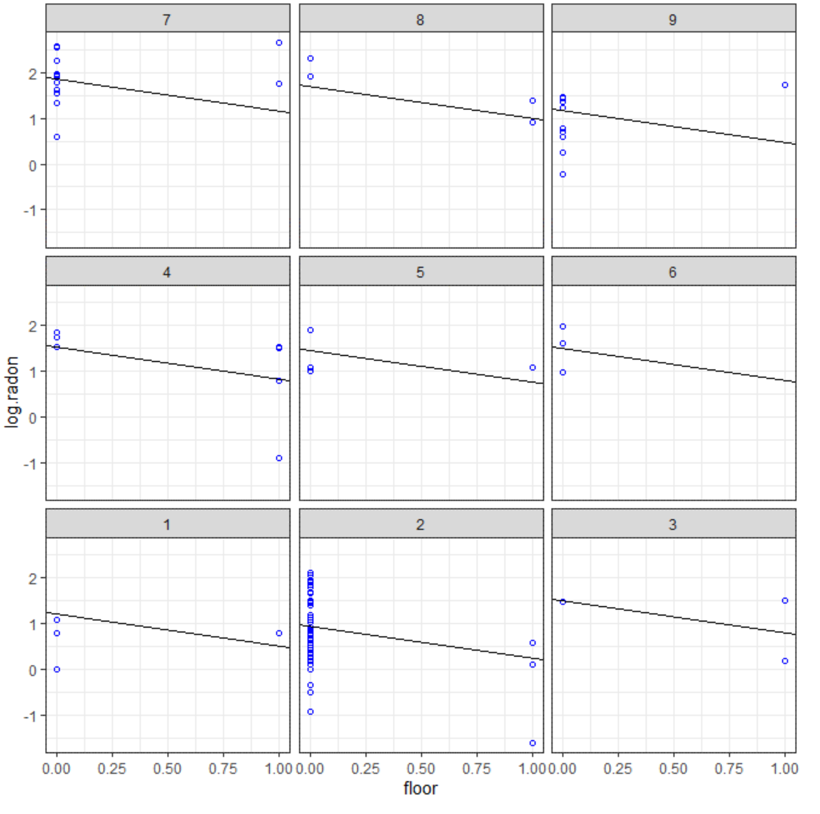
SD of random effect



Random-intercept model with individual predictors

$$y_i = \alpha_{0j[i]} + \alpha_1 x_i + \epsilon_i$$
$$\alpha_{0j} = \beta_0 + \eta_j$$

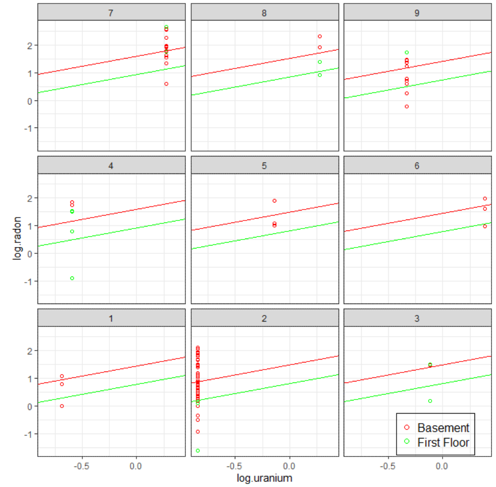
```
> y <- log.radon
> x <- floor
> M1 <-
    lmer(y ~ x + (1 | county) )
> display(M1)
 coef.est coef.se
(Intercept)  1.46     0.05
x           -0.69     0.07
Error terms:
 Groups   Name        Std.Dev.
 county   (Intercept) 0.33
 Residual             0.76
number of obs: 919, groups:
    county, 85
AIC = 2179.3, DIC = 2156
deviance = 2163.7
```

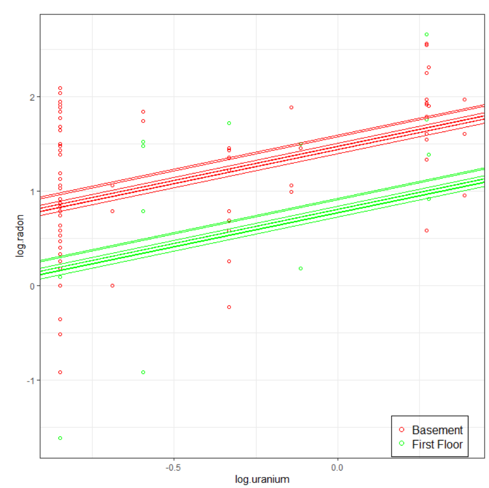


Random-intercept model w / individual & group level predictors

$$y_i = \alpha_{0j[i]} + \alpha_1 x_i + \epsilon_i$$
$$\alpha_{0j} = \beta_0 + \beta_1 u_j + \eta_j$$

```
> y <- log.radon
> x <- floor
> u.full <- log.uranium
> M2 <- lmer(y ~ x + u.full + (1 | county))
> display(M2)

coef.est coef.se
(Intercept)  1.47     0.04
x           -0.67     0.07
u.full       0.72     0.09

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

number of obs: 919, groups: county, 85
AIC = 2144.2, DIC = 2111.5
deviance = 2122.9
```





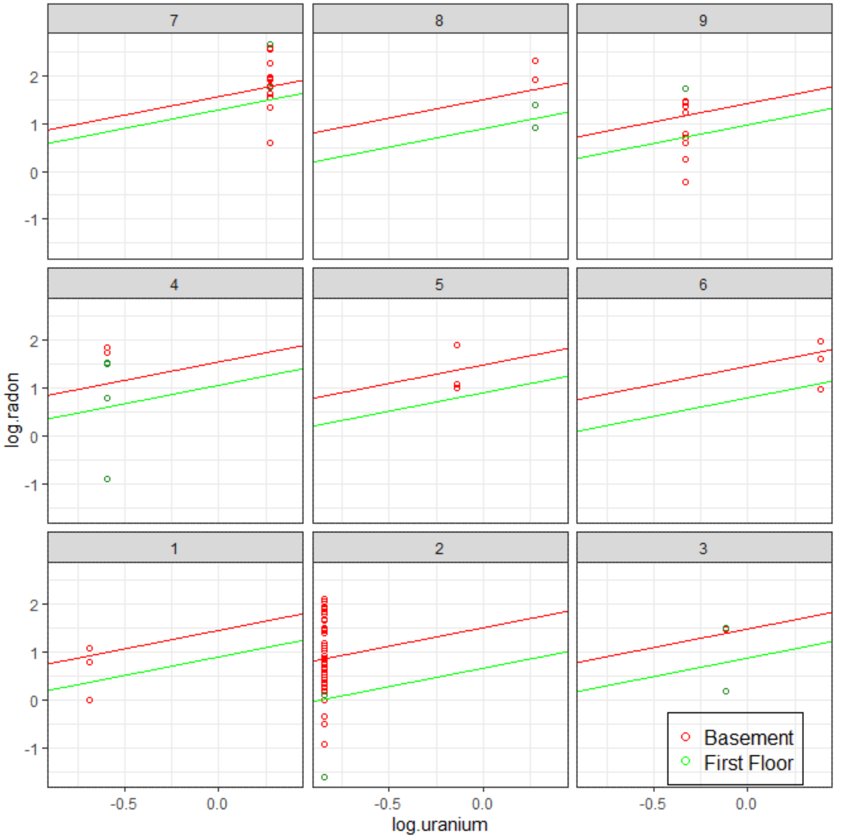
One last Radon Model: Random Intercept, Random Slope, Gp Predictor

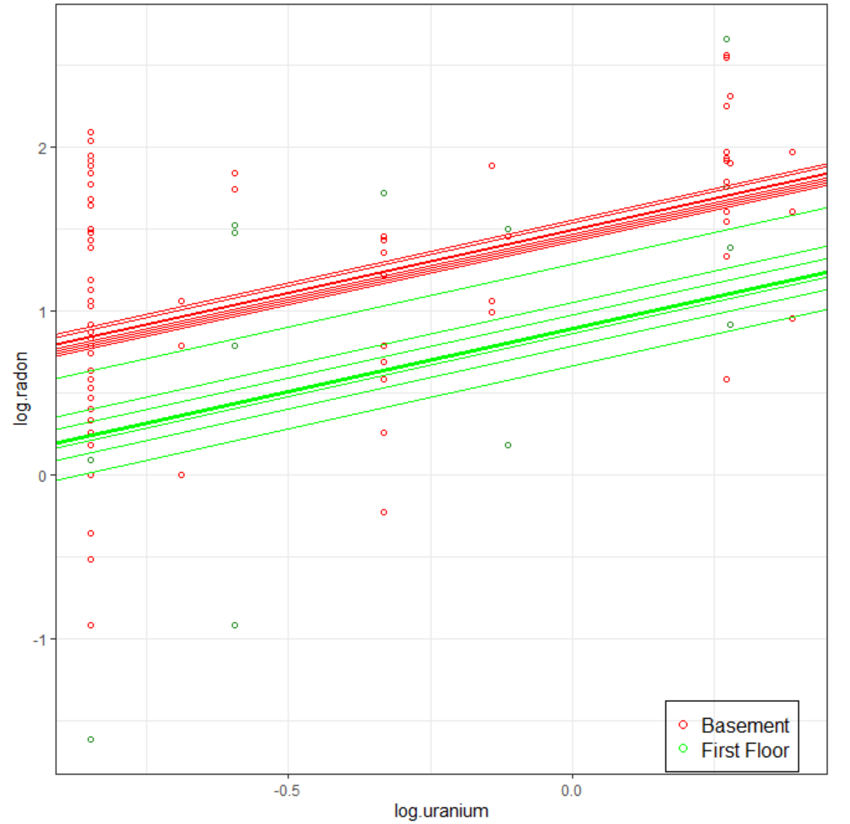
$$\begin{aligned} y_i &= \alpha_{0j[i]} + \alpha_{1j[i]} x_i + \epsilon_i \\ \alpha_{0j} &= \beta_{00} + \beta_{01} u_j + \eta_{0j} \\ \alpha_{1j} &= \beta_{10} + \eta_{1j} \\ &\epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2) \\ &\eta_{0j} \stackrel{iid}{\sim} N(0, \tau_0^2) \\ &\eta_{1j} \stackrel{iid}{\sim} N(0, \tau_1^2) \end{aligned}$$

```
> y <- log.radon
> x <- floor
> u.full <- log.uranium

> M3 <- lmer(y ~
  x + u.full + (1 + x | county) )
```

```
> display(M3)
            coef.est  coef.se
(Intercept)  1.46      0.04
x           -0.64      0.09
u.full       0.77      0.09

Error terms:
 Groups   Name        SD   Corr
 county   (Intcpt) 0.13
          x        0.36  0.21
 Residual          0.75

number of obs: 919, groups:
   county, 85
AIC = 2142.6, DIC = 2106.7
deviance = 2117.7
```

One last Radon Model: Random Intercept, Random Slope, Gp Predictor





More on Multiple Random Effects

$$
\begin{aligned}
y_i &= \alpha_{0j[i]} + \alpha_{1j[i]} x_i + \epsilon_i & \epsilon_i &\overset{iid}{\sim} N(0, \sigma^2) \\
\alpha_{0j} &= \beta_{00} + \beta_{01} u_j + \eta_{0j} & \eta_{0j} &\overset{iid}{\sim} N(0, \tau_0^2) \\
\alpha_{1j} &= \beta_{10} + \eta_{1j} & \eta_{1j} &\overset{iid}{\sim} N(0, \tau_1^2)
\end{aligned}
$$

- We always model ϵ_i as "independent of everything" because it is the "unexplainable variation"
- η_{0j} and η_{1j} might be dependent on each other!
  - Perhaps radon levels are relatively high in some counties (α_{0j} large) but dissipate to a relatively constant level above ground (α_{1j} large too).
  - Suggests η_{0j} and η_{1j} might be correlated.

Multiple Random Effects

- Thus we often do (and lmer() does) model the correlation between random effects, e.g.:

$$
\begin{aligned}
y_i &= \alpha_{0j[i]} + \alpha_{1j[i]} x_i + \epsilon_i \qquad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2) \\
\alpha_{0j} &= \beta_{00} + \beta_{01} u_j + \eta_{0j} \\
\alpha_{1j} &= \beta_{10} + \eta_{1j}
\end{aligned}
$$

$$
\begin{pmatrix} \eta_{0j} \\ \eta_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \rho_{01}\tau_0\tau_1 \\ \rho_{01}\tau_1\tau_0 & \tau_1^2 \end{pmatrix} \right)
$$

Multiple Random Effects

$$
\begin{aligned}
y_i &= \alpha_{0j[i]} + \alpha_{1j[i]} x_i + \epsilon_i \\
\alpha_{0j} &= \beta_{00} + \beta_{01} u_j + \eta_{0j} \\
\alpha_{1j} &= \beta_{10} + \eta_{1j} \\
\epsilon_i &\stackrel{iid}{\sim} N(0, \sigma^2) \\
\eta_{0j} &\stackrel{iid}{\sim} N(0, \tau_0^2) \\
\eta_{1j} &\stackrel{iid}{\sim} N(0, \tau_1^2) \\
\mathrm{Cor}(\eta_{0j}, \eta_{1j}) &= \rho
\end{aligned}
$$

```
> y <- log.radon
> x <- floor
> u.full <- log.uranium

> M3 <- lmer(y ~
  x + u.full + (1 + x | county) )
```

```
> display(M3)
            coef.est  coef.se
(Intercept)  1.46      0.04
x           -0.64      0.09
u.full       0.77      0.09

Error terms:
 Groups   Name        SD   Corr
 county   (Intcpt)  0.13
          x         0.36  0.21
 Residual           0.75

number of obs: 919, groups:
   county, 85
AIC = 2142.6, DIC = 2106.7
deviance = 2117.7
```

Correlation between random effects

- In general it is better to leave the correlation in the model, than to take it out.
  - We are sometimes interested in centering one or more predictors for multilevel models, just as we were for ordinary linear models (for interpretation, better numerical behavior, etc.)
  - MLM's with estimated correlation are invariant (except for obvious main effects shifts) to centering and other linear shifts of the X's
  - MLM's without the correlation do not have this invariance property

Summary

- Three Ways to Write The Model
- Brief intro to library(lme4) and lmer()
- lmer() notation / modeling language
- Minnesota Radon Levels, Part 2
  - Predicting county means with log(uranium)
  - Intercept-only random-interecept model
  - Random-intercept model with level 1 predictors
  - Random-intercept model with level 1 and level 2 predictors
- More than one random effect
  - Random intercept, random slope
  - Correlation between random effects