36-617: Applied Linear Models

Intro to Multi-level Models, II

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

Announcements

- Quiz on Sheather 10.1 today
- HW09 due this Weds
- HW10 due next Weds
- Projects:
  - I still owe you grades on Project 01
  - Project 02 out now (HW10, #2: tech appx!)
- Project 02 Schedule:
  - **Wed Nov 17:** Draft Technical Appendix with HW 10.
  - **Wed Nov 24 (or earlier):** Full IDMRAD paper first draft.
  - **Wed Dec 1:** Peer reviews due.
  - **Fri Dec 10 (or earlier):** Full IDMRAD paper final draft!

Plan for rest of semester

- M Nov 8 – intro to MLM's, continued
- W Nov 10 – residuals for MLM's
- M Nov 15 – estimation and model selection
- W Nov 17 – shrinkage, crash course on Bayes
- M Nov 22 – catch-up, or multilevel glm's
- M Nov 29 – ?? Maybe spline smoothing
- W Dec 1 – ?? Maybe spline smoothing

Outline

- "Shrinkage" in multilevel models
- lmer() notation, variance components models, and multi-level models
- Fixed & random effects; multiple effects

(Note: We will talk about how estimation is done - ML, REML and Empirical Bayes - later. For now we focus on fitting the model & interpreting the fit.)

Fitting the random-intercept model

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

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

Multilevel model (both equations 1 and 2)

```
library(lme4)
lmer.intercept.only <-
   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.unpooled.contrast.from.grand.mean <-
   lm(y ~ cties)
summary(lm.unpooled.contrast.from.grand.mean)
# 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
```

Random-intercept model: Where are the intercepts?

```
> summary(lmer.intercept.only)
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(lmer.intercept.only)
(Intercept)
   1.312574
> ranef(lmer.intercept.only)
$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.unpooled.con-
   trast.from.grand.mean)

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

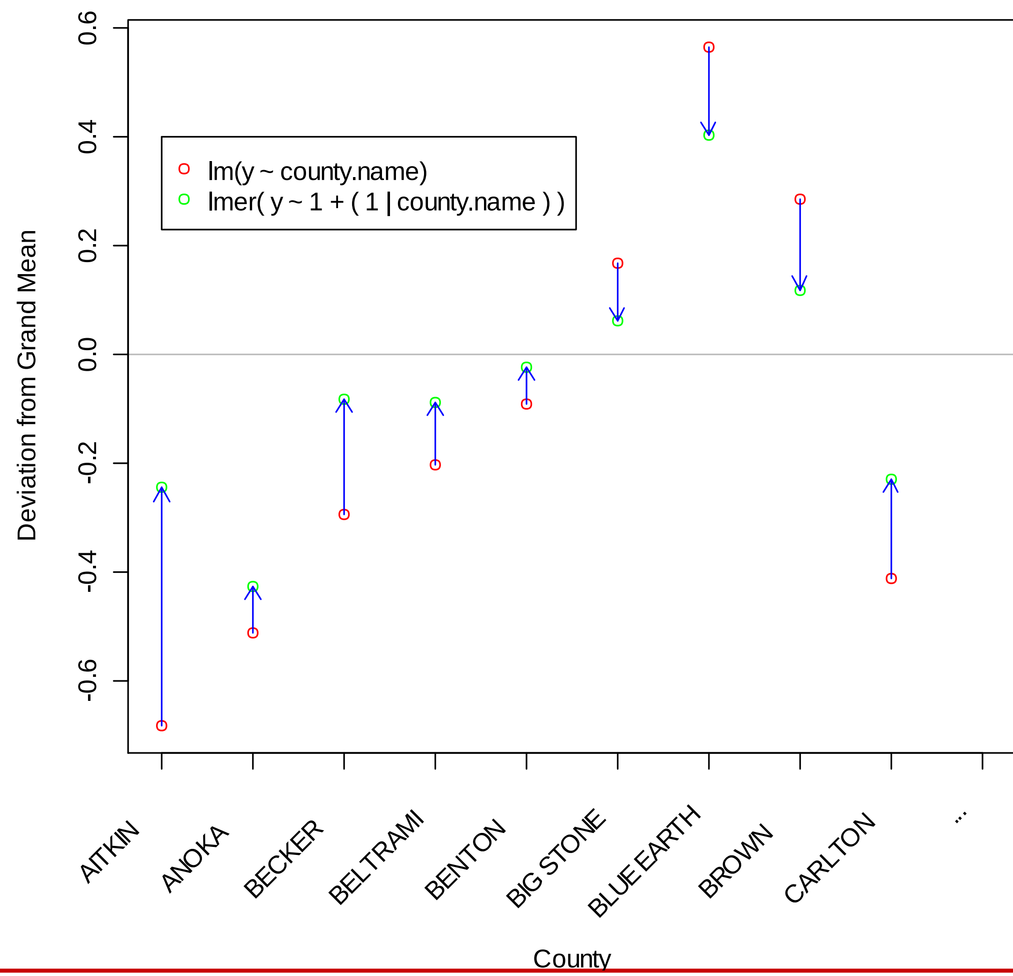
An MLM phenomenon: Shrinkage

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".)

We'll talk about ***why*** when we talk about estimation, later on…



lmer() notation…

- Multilevel model:

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

- Variance components model:

$$
y_i = \beta_0 + \beta_1 u_{j[i]} + \alpha_1 x_i + \eta_{j[i]} + \epsilon_i
$$

- lmer() model:

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

$$y_i = \beta_0 + \beta_1 u_{j[i]} + \alpha_1 x_i + \eta_{j[i]} + \epsilon_i$$

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

- Some authors say "varying effects" instead of "random effects"…

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

- Intercept-only random-intercept model

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

- Random-intercept model w / individual-level predictors

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

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

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

Intercept-only random-intercept model

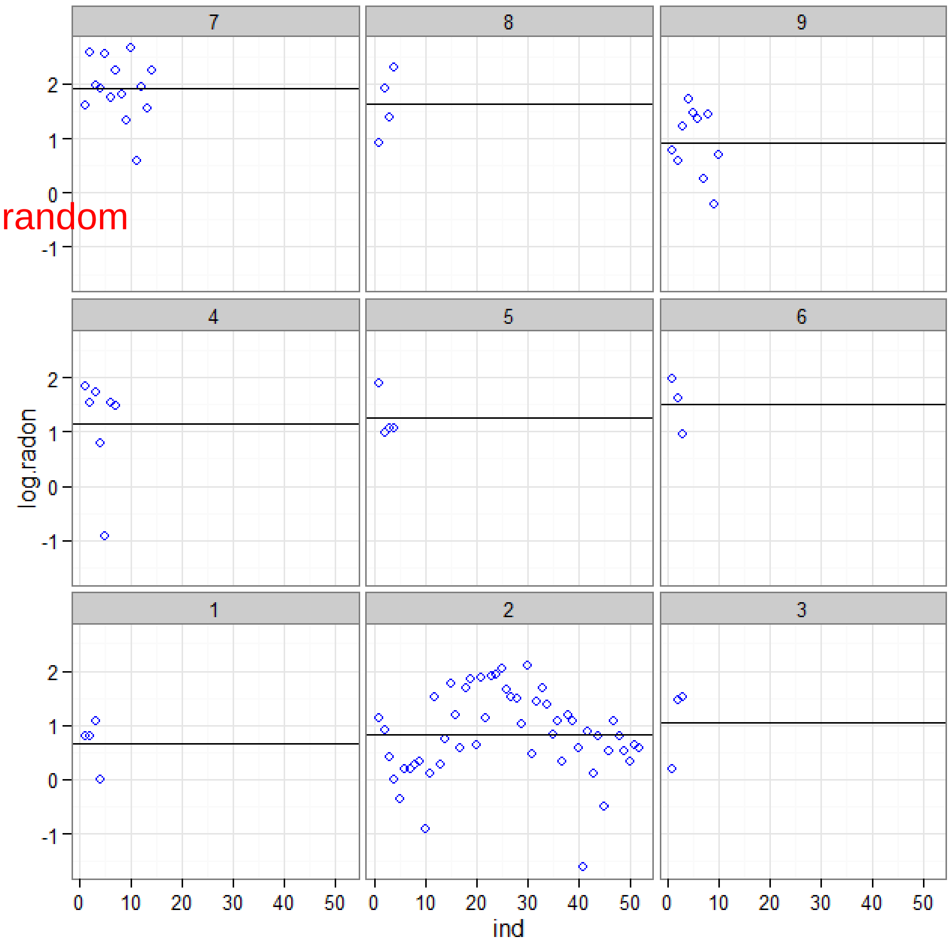
$$y_i = \alpha_{j[i]} + \epsilon_i$$
$$\alpha_j = \beta_0 + \eta_j$$

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
> county.sample <- mn.radon$county %in%
   1:9
> subset <- mn.radon[county.sample,]
> facet_data <-
   split(subset,subset$county)
> g <- ggplot(subset,
+    aes(x=ind, y=log.radon)) +
+  facet_wrap( ~ county,as.table=F) +
+  geom_point(pch=1,color="blue")
> for (j in 1:9)
+  g <- g +
     geom_hline(data=facet_data[[j]],
+      aes(yintercept=mean(log.radon)),
+      color="black")
> plot(g)
```

Fixed effect

SD of random effect

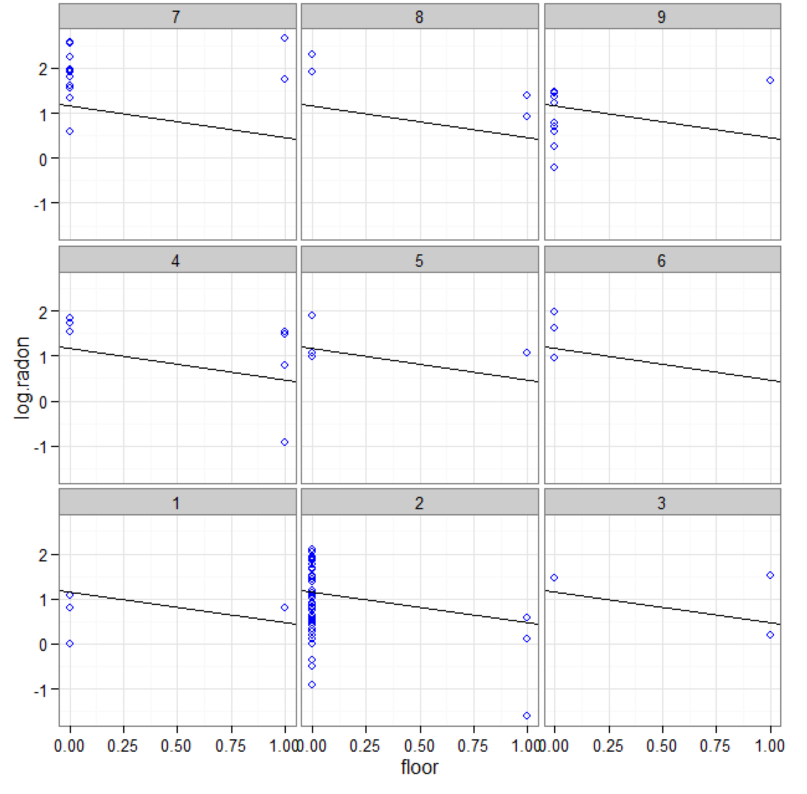


Random-intercept model with individual predictors

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

$$\alpha_{0j} = \beta_0 + \eta_j$$

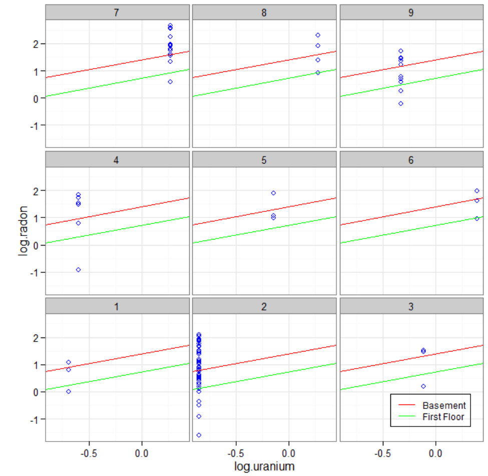
```
> y <- log.radon
> x <- floor
> M1 <-
    lmer(y ~ x + (1 | county) )
> display(M1)
 coef.est coef.se
(Intercept)  1.46     0.05
x           -0.69     0.07
Error terms:
 Groups   Name        Std.Dev.
 county   (Intercept) 0.33
 Residual             0.76
number of obs: 919, groups:
   county, 85
AIC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

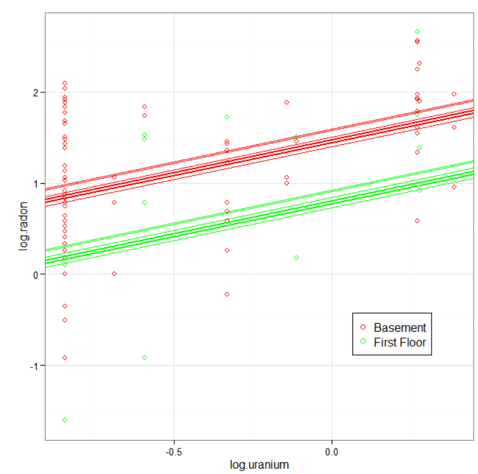


Random-intercept model w / individual & group level predictors

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

```
> 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
```





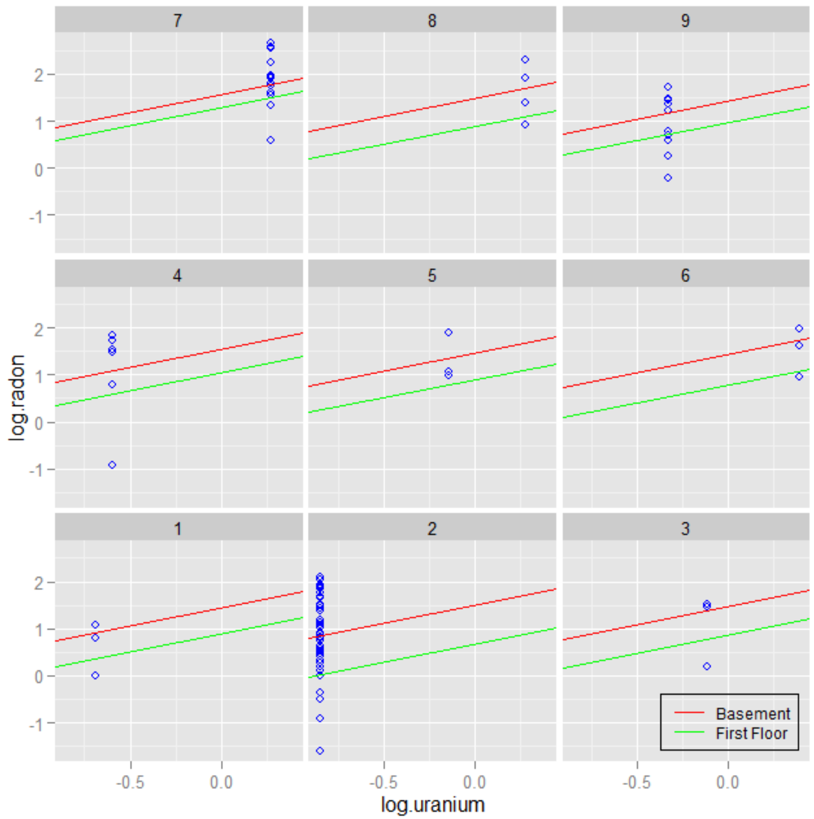
One last Radon Model: Random Intercept, Random Slope, Gp Predictor

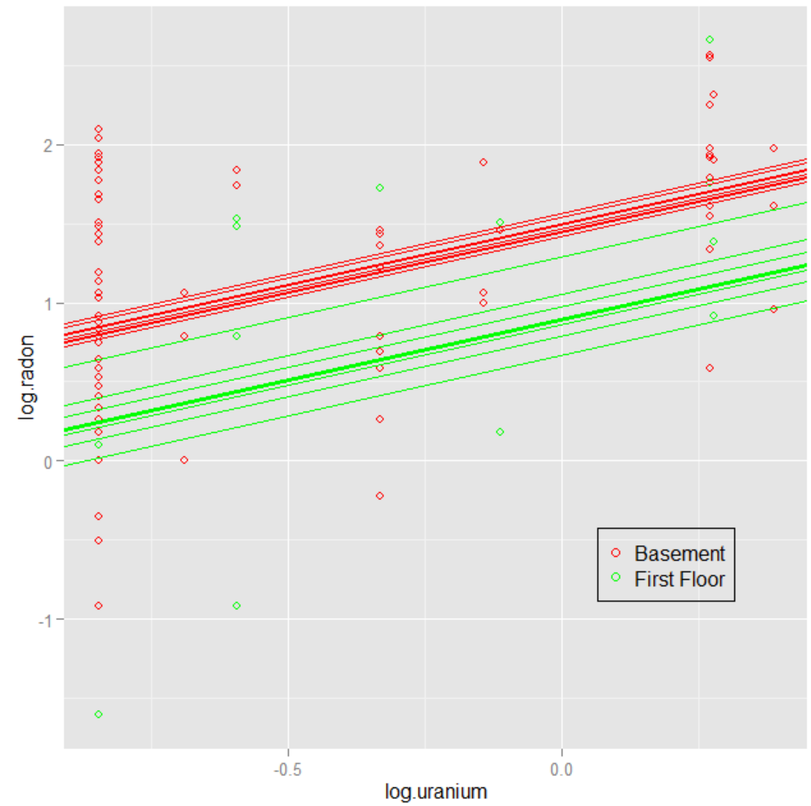
$$
\begin{aligned}
y_i &= \alpha_{0j[i]} + \alpha_{1j[i]} x_i + \epsilon_i \\
\alpha_{0j} &= \beta_{00} + \beta_{01} u_j + \eta_{0j} \\
\alpha_{1j} &= \beta_{10} + \eta_{1j} \\
& \epsilon_i \overset{iid}{\sim} N(0, \sigma^2) \\
& \eta_{0j} \overset{iid}{\sim} N(0, \tau_0^2) \\
& \eta_{1j} \overset{iid}{\sim}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

```
> 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.:

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

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

Multiple Random Effects

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

```
> 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
```

Summary

- "Shrinkage" in multilevel models
- lmer() notation, variance components models, and multi-level models
- Fixed & random effects; multiple effects

(Note: We will talk about how estimation is done - ML, REML and Empirical Bayes - later. For now we focus on fitting the model & interpreting the fit.)