# 36-463-663: Hierarchical Linear Models

Intro to Multi-level Models, II

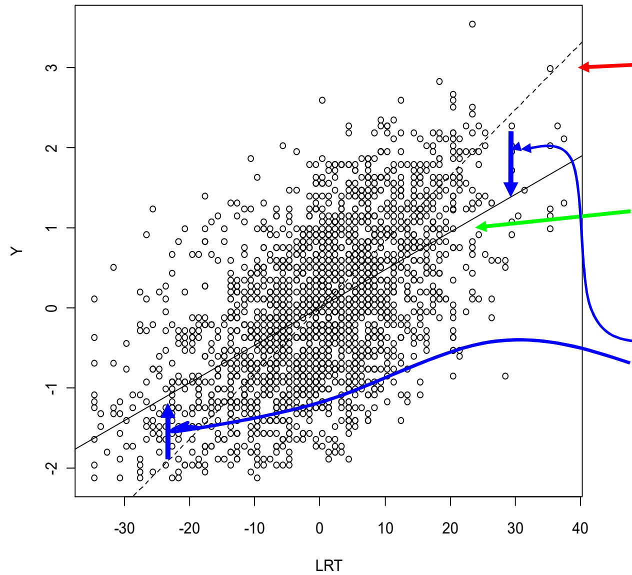
Brian Junker
132E Baker Hall
brian@stat.cmu.edu



## Outline

- Regression to the Mean & "Shrinkage"
- lmer() notation, variance components models, and multi-level models
- Fixed effects, random effects, varying effects
- Multiple random effects
- Read: Ch 13
  - Skip material on inverse-Wishart distribution (pp 286-287) for now – premature for us!
- Note on library(ggplot2) vs library(lattice):
  - The plots in the slides are made with an older version of ggplot2. The current version seems fussier in some ways.
  - In the R handout (online) I have also included commands for xyplot (from library "lattice") which produce similar plots to the plots in the slides.

# Regression to the Mean




$$\frac{y - mean(y)}{SD(y)} = \frac{x - mean(x)}{SD(x)}$$

$$\frac{y - mean(y)}{SD(y)} = r \cdot \frac{x - mean(x)}{SD(x)}$$

The fitted regression line underpredicts high obs's and overpredicts low ones.

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

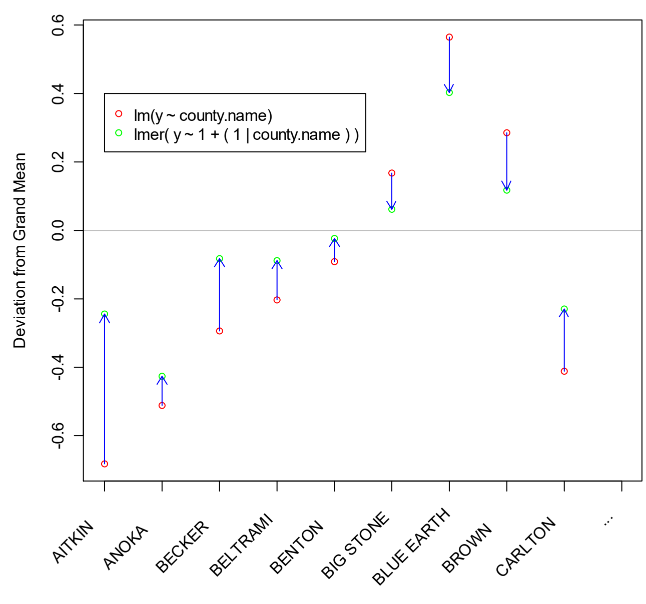


# Shrinkage in MLM's

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".)



# 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))
```

- This is less obvious in Bayesian modeling so G&H avoid the terms, or just say "varying effects"

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

- Intercept-only random-intercept* model

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

- Random-intercept* model w / individual-level predictors

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

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

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

*G&H say "varying intercept" rather than "random intercept"



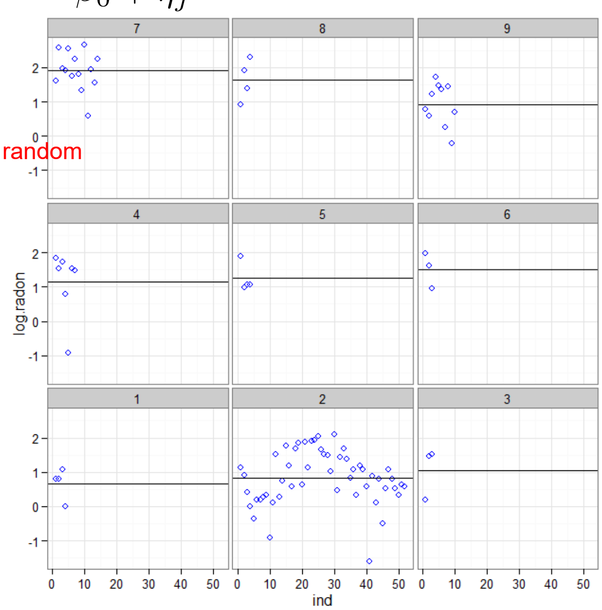
# Intercept-only random-intercept* model

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

```
> 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

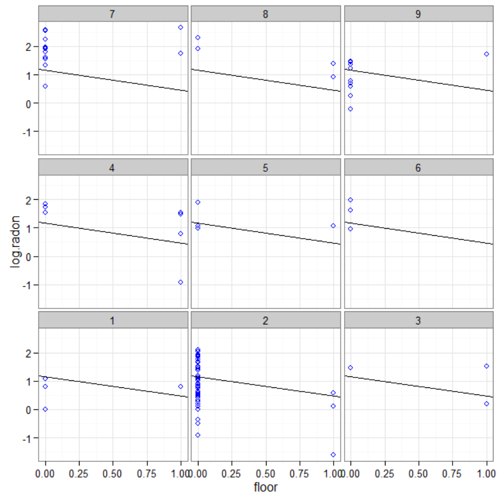
SD of random effect



# Random-intercept* model with individual predictors

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

```
> 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
```

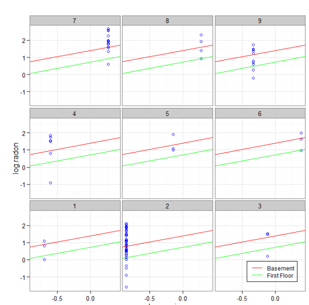


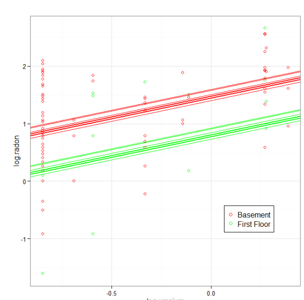



# Random-intercept* model w / individual & group level predictors

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

```
> 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

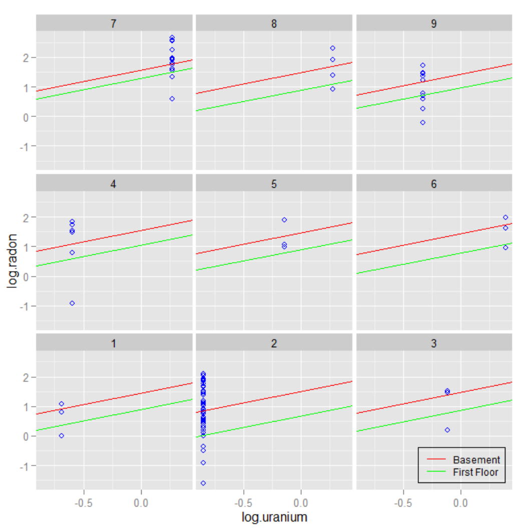
$$
\begin{aligned}
y_i &= \alpha_{0j[i]} + \alpha_{1j[i]} x_i + \epsilon_i \\
\alpha_{0j} &= \beta_{00} + \beta_{01} u_j + \eta_{0j} \\
\alpha_{1j} &= \beta_{10} + \eta_{1j} \\
& \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2) \\
& \eta_{0j} \stackrel{iid}{\sim} N(0, \tau_0^2) \\
& \eta_{1j} \stackrel{iid}{\sim}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

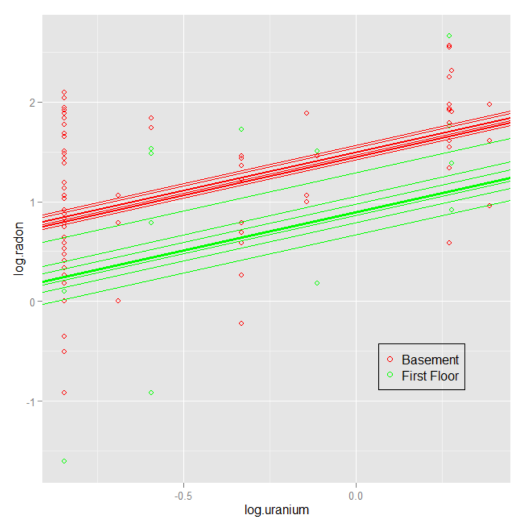
```
> y <- log.radon
> x <- floor
> u.full <- log.uranium

> 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 $\epsilon_i$ as "independent of everything" because it is the "unexplainable variation"
- $\eta_{0j}$ and $\eta_{1j}$ might be dependent on each other!
  - If radon levels are relatively high in one county ($\alpha_{0j}$ large) then there might not be much room for basement-to-first-floor differences ($\alpha_{1j}$ small).
  - Suggests $\eta_{0j}$ and $\eta_{1j}$ might be negatively correlated



## Multiple Random Effects

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

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

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

# Multiple Random Effects

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

```
> 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
```



# We can also force $\rho$=0…

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

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

> M3i <- lmer(y ~ x + u.full +
  (1 | county) + (0 + 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)  0.14
          x         0.37  0.00
 Residual           0.75

number of obs: 919, groups:
   county, 85
AIC = 2140.8, DIC = 2106.8
deviance = 2117.8
```

# Summary

- Regression to the Mean & “Shrinkage”
- lmer() notation, variance components models, and multi-level models
- Fixed effects, random effects, varying effects
- Multiple random effects
- Read: Ch 13
  - Skip material on inverse-Wishart distribution (pp 286-287) for now – premature for us!