

S-043/Stat-151  
Analysis for Clustered and Longitudinal Data  
(Multilevel & Longitudinal Models)

**Unit 2, Lecture 1**  
**Random Intercept Models**

Instructor: Prof. Luke W. Miratrix  
lmiratrix@g.harvard.edu  
Larsen 603

# Today's Goals

---

- ★ Introduce the Random Intercept Model (the simplest of the multilevel models)
- ★ Talk about “partial pooling” where we borrow information from other clusters to learn about individual clusters (and trends across cluster).
- ★ Talk about how these models give better predictions than completely unpooled (e.g., fixed effect) models
- ★ Talk about how these models allow us to assess variation across clusters
- ★ Discuss how to include level 1 covariates in a random intercept model to improve precision and adjust for group differences

# Recall:

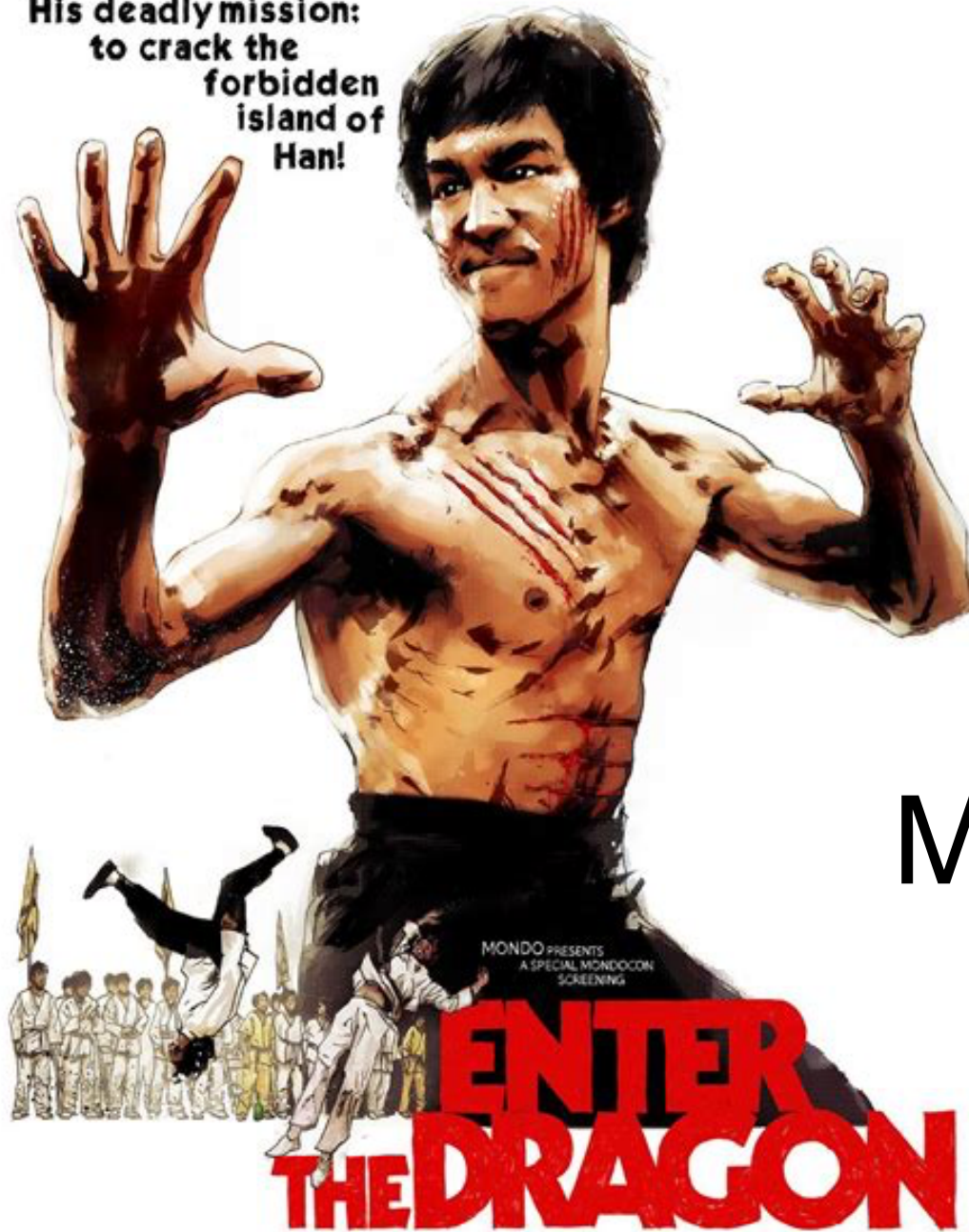
## Aggregation (lecture 1.2) is unsatisfying

We need a better way to analyze multi-level data. Aggregation is cool, but it

1. Is underpowered (school level sampling error gives us problems)
2. Is biased (sampling error in the predictors introduces bias)
3. Makes implausible assumptions (homoscedasticity is almost never going to be correct)

**Admittidly, we can fix 3 with  
econometrics (robust SEs)**

His deadly mission:  
to crack the  
forbidden  
island of  
Han!



# Enter the Multilevel Model

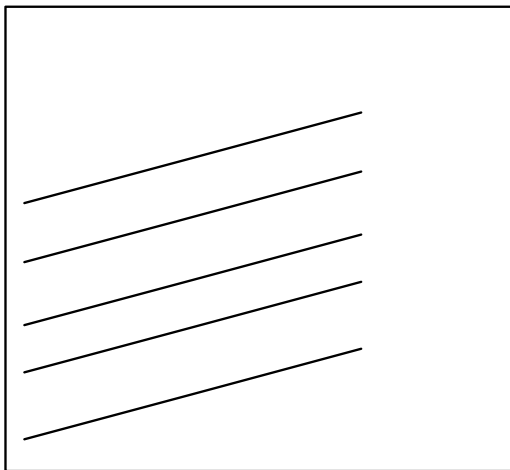
MONDO PRESENTS  
A SPECIAL MONDOCON  
SCREENING  
**ENTER  
THE DRAGON**  
BRUCE LEE IN "ENTER THE DRAGON"  
At AFS CINEMA 6226 Middle Fiskville Rd, AUSTIN, TX 78752 Mondocon  
OCTOBER 22ND 2016 • A 35mm PRESENTATION  
[WWW.MONDOCON.COM](http://WWW.MONDOCON.COM) A MONDO COMMUNICATIONS COMPANY

730268

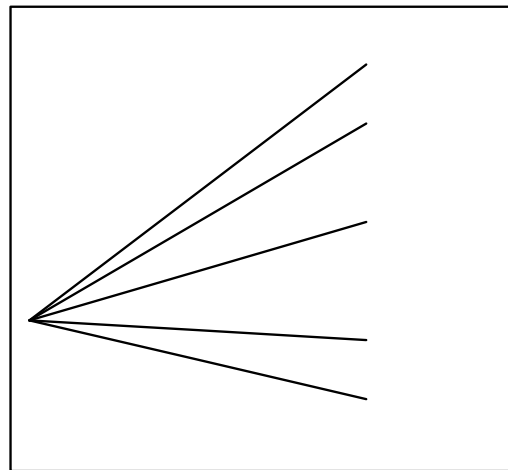
"ENTER THE DRAGON" and all related characters and elements © & TM Warner Bros. Entertainment Inc. ® & TM  
Bruce Lee is a registered trademark of Bruce Lee Enterprises, LLC. The Bruce Lee name, image, likeness and all other indicia are intellectual property of Bruce Lee Enterprises, LLC. All Rights Reserved.

# Multilevel Models are really lots of little models tied together

**Varying intercepts**



**Varying slopes**



**Varying intercepts and slopes**

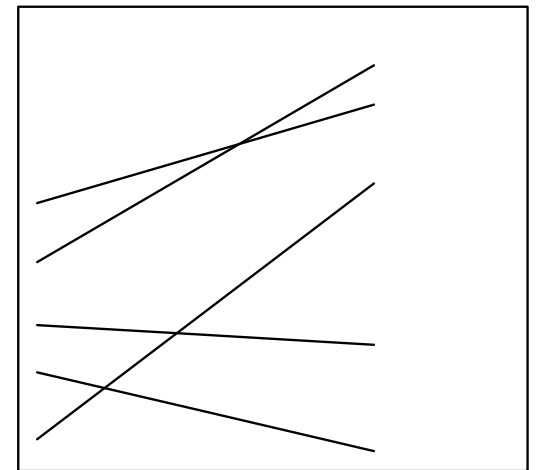
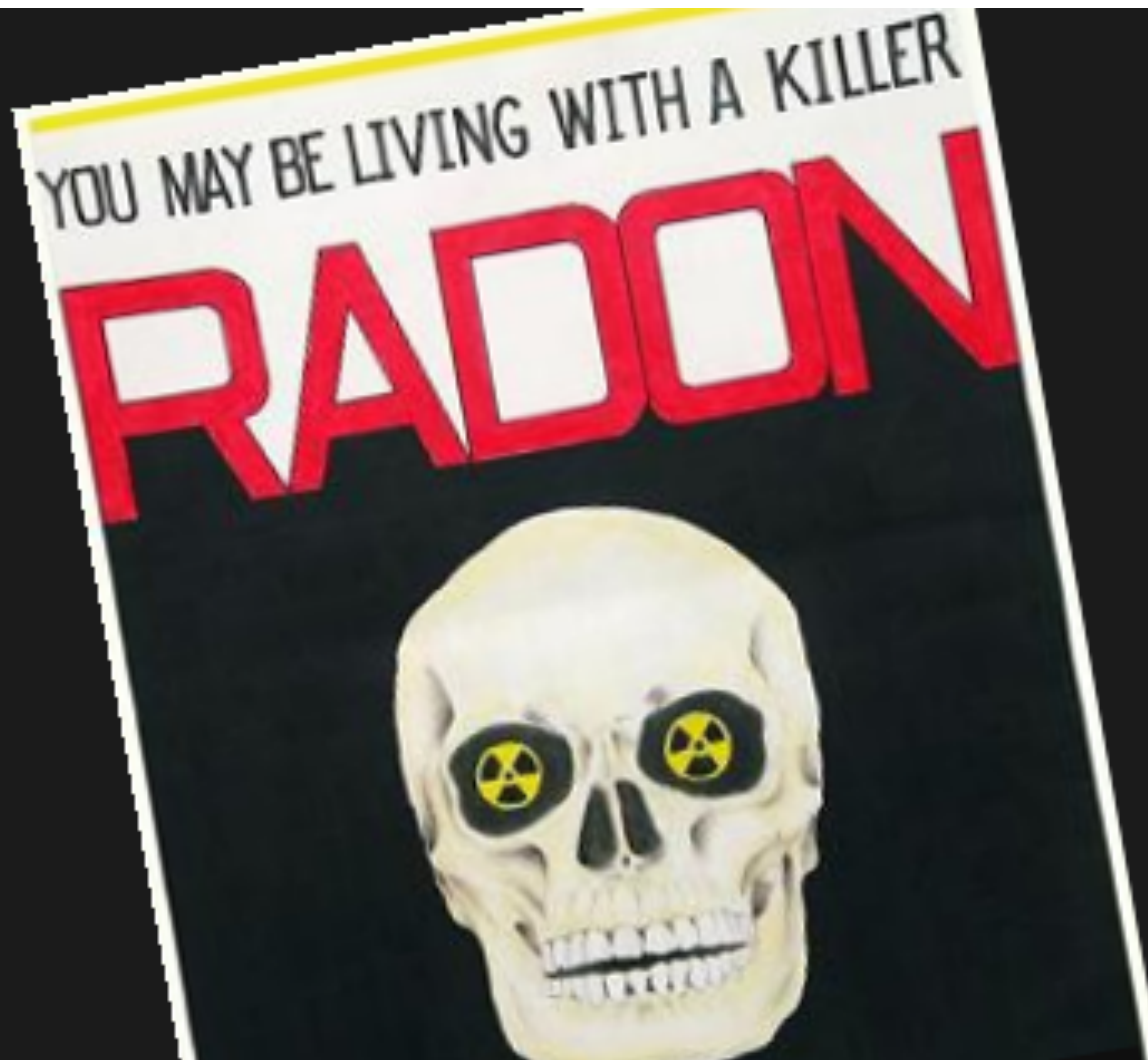


Figure 11.1 *Linear regression models with (a) varying intercepts ( $y = \alpha_j + \beta x$ ), (b) varying slopes ( $y = \alpha + \beta_j x$ ), and (c) both ( $y = \alpha_j + \beta_j x$ ). The varying intercepts correspond to group indicators as regression predictors, and the varying slopes represent interactions between  $x$  and the group indicators.*

A new example (Radon)

(From the Gelman & Hill reading)

A clean and simple example  
(about poison in your home)



# Testing homes for radon?

---

Radon is a radioactive noble (standoffish) gas, *Rn*

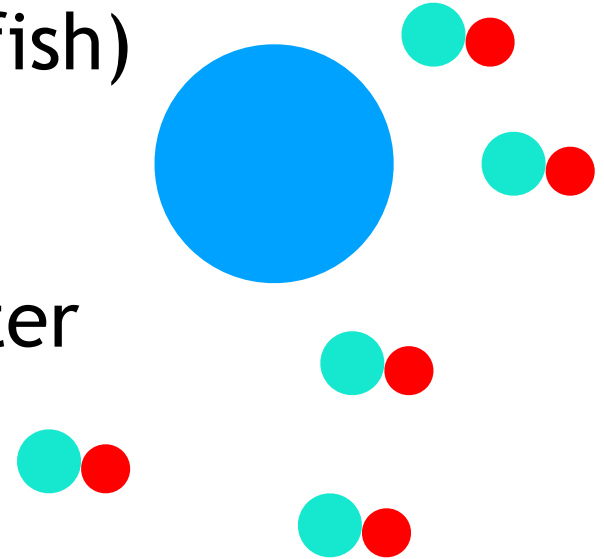
Can enter homes through soil and water

High concentrations can be harmful

Radon can be detected and radon concentrations reduced (sealing and ventilating the basement better), but these are expensive and most homes don't need them

Overarching Question:

How can we use multi-level models to help homeowners decide whether or not to test for radon?





# New Data: Radon levels in counties

---

85 counties in Minnesota

We are interested in country mean levels of log-radon in the houses. (concentrations tend to be similar within a county)

Radon measured either on first floor or basement (if basement exists)

## Research Goals:

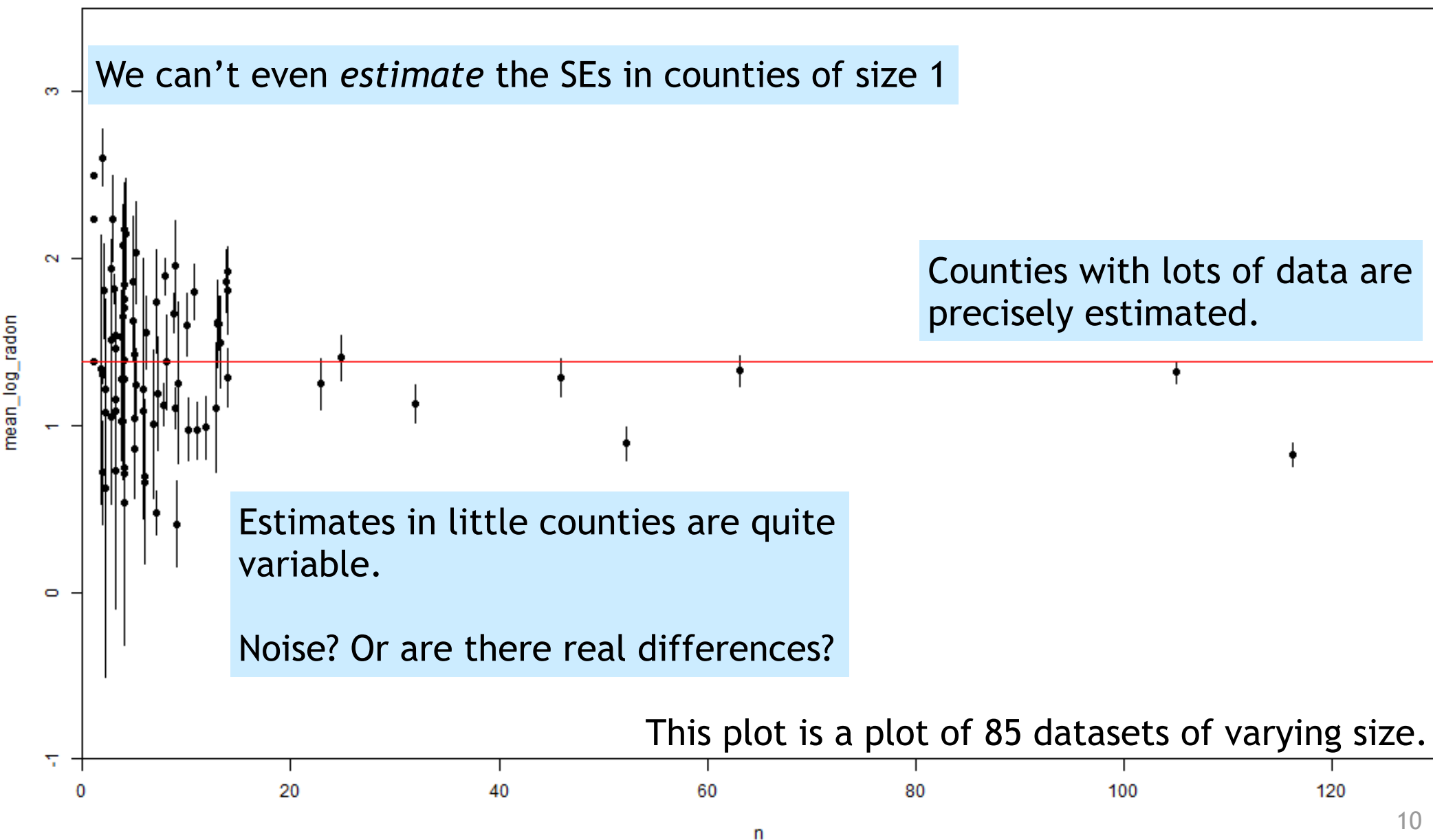
Is there variation in average radon levels across counties? (Are some counties substantially worse off?)

How can we predict radon levels in the different counties?

This may seem new because we're using our model for prediction. That's not as common in the social sciences!

# First first pass: each county by itself

For each county I calculated the mean radon level (and SE of this mean)



# Fitting in one go: the Fixed Effect Regression

---

Notation: if house  $i$  is in county  $j$ , we write  $house_{j[i]}$  or  $house_{ij}$

Specifically,  $house_{j[i]}$  is the (log of the) radon level observed in house  $i$ , which is in county  $j$ .

Our model:

$$house_{j[i]} = \alpha_j + \epsilon_{j[i]}$$

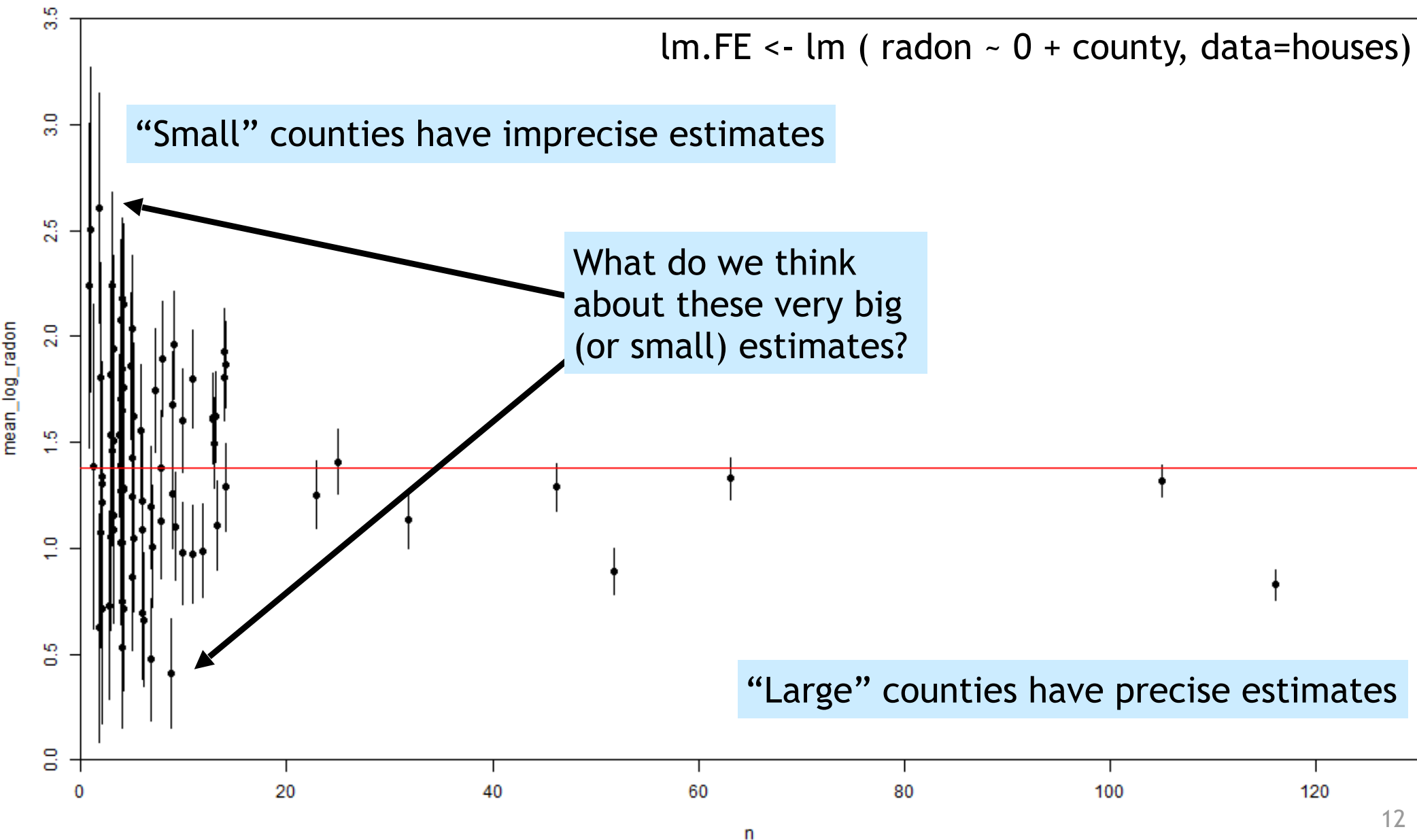
A house's value

County  
mean

Individual  
variability

This is our fixed effects model! We say that the individual county means are *unpooled*, since knowing one mean tells us nothing about any other mean.

# Second pass: estimate mean level for each county-level sample from a fixed-effects model



# So how is this different from just estimating each county by itself?

---

The model:

$$house_i = \alpha_{j[i]} + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

And in what sort of counties will the differences be biggest? (See PollEv question, next.)

# Mini-Recap

---

We just looked at the Radon dataset

We fit a fixed effect regression and looked at all the individual estimates of county-average radon

We are worried that counties with few houses are very noisy, and so our predictions are often way too big or way too small

# Random intercept models (no predictors)



# Thought Experiment: why is it big?

---

Take a county where the average of the houses'  $\log(\text{Radon})$  was 2.1 (this is quite high, most estimates are less than 2, and the average is 1.22). There were 3 houses in our sample.

Why do you think we saw an estimate of 2.1?

What would you guess the true average in the county is? 2.1? More than 2.1? Less than 2.1?

Now say that we see an estimate of 2.1, but this is the average of 30 houses?



# How can we improve our estimates?

---

If a small county has an extreme mean (high or low), some part of that is probably sampling/measurement error

So we want to find a way to “shrink” estimates towards a common mean (pull high means down and low means up), with more shrinkage for small counties and less for big.

# The Random Intercept Model: tying the small worlds together

We have house  $i$  in county  $j[i]$ .

We then have:

One more thing: we assume that  $\epsilon_i$  is independent of  $\alpha_{j[i]}$ .

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

$$\epsilon_i \sim N(0, \sigma_y^2)$$

$$\alpha_j \sim N(\mu, \sigma_\alpha^2)$$

The  $\alpha_j$  are the county means. We assume that these are random draws from a common distribution, hence “random intercepts”

$\sigma_\alpha^2$  is the variance of the county means

$\mu$  is the mean of county means

**Call this a “soft constraint”**

# Concept check

---

This (is supposed to be) a tough one.

We have that

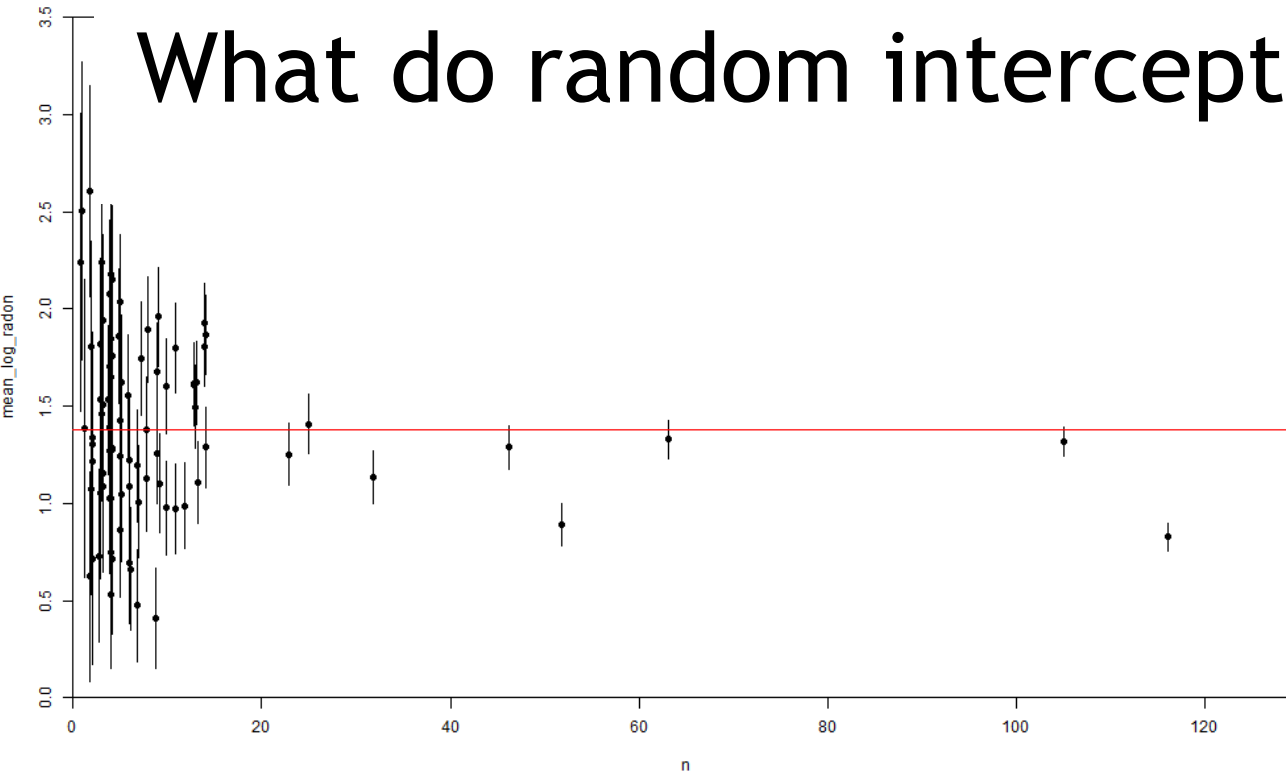
$$\alpha_j \sim N(\mu, \sigma_\alpha^2)$$

Here  $\mu$  is referred to as the “grand mean”; it’s the mean of county means.

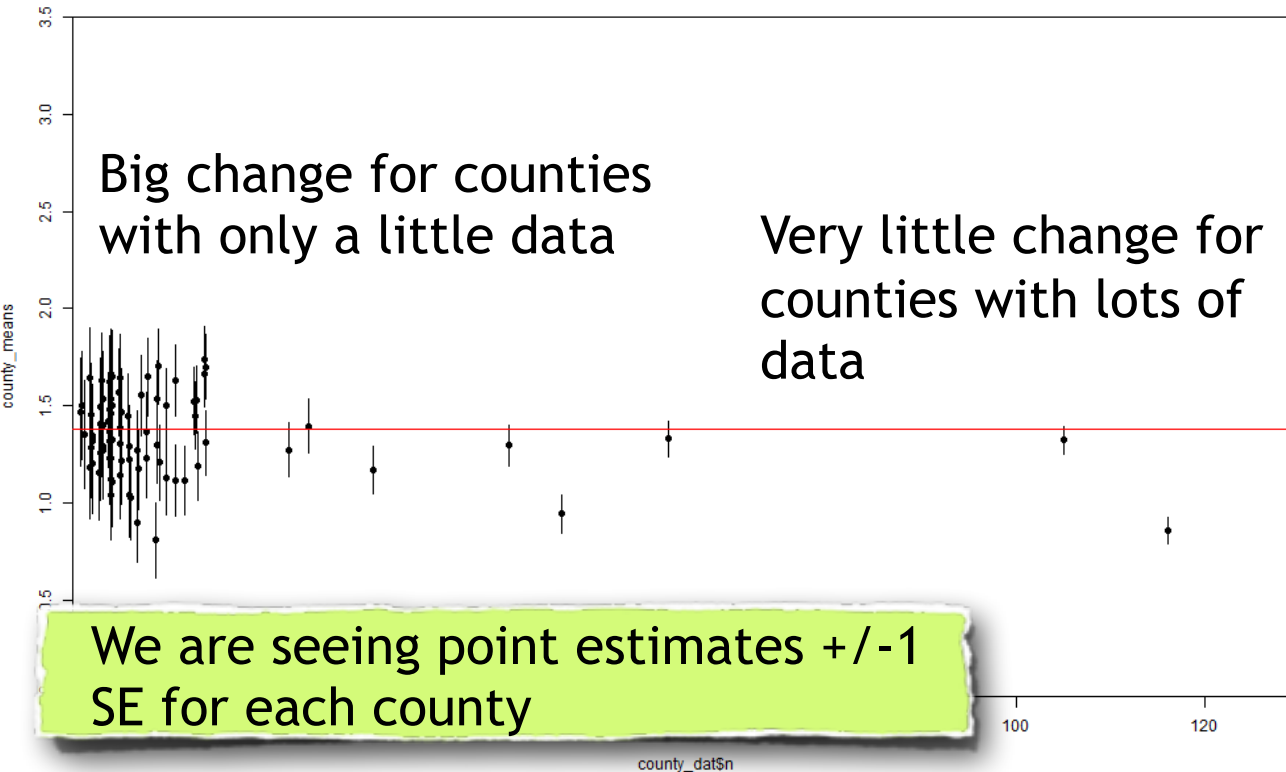
Questions:

- 1) Does this mean (lol) that the average *house* in Minnesota has a value of log\_radon equal to  $\mu$ ?
- 2) Why or why not?

# What do random intercept results look like?



**Unpooled, or fixed effects:**  
county means are estimated separately



**Pooled, or random effects:**  
county means are shrunk towards a grand mean

Far less variability in small counties; these have all been shrunk dramatically toward a common mean.

Also far less uncertainty within counties.

# Revisiting intuition about how estimation works

---

$$house_i = \alpha_{j[i]} + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

$$\alpha_j \sim N(\mu, \sigma_\alpha^2)$$

If the sample mean of county is extreme, that could mean any of the following:

- 1) it really has a large mean ( $\alpha_j$  is actually big), or
- 2) by chance, we have houses with a lot of positive residuals (the  $\epsilon_{i[j]}$  were generally big and positive for this county).

When we fit our model, we balance these two possibilities, shrinking each estimated mean towards the estimated mean of county means ( $\mu$ ); a normal distribution doesn't give many large results.

There will be more shrinking for small counties, and less for big counties. It's easy to imagine that a single house had an unusual value; much harder to imagine that 400 houses did.

# Random Intercepts lead to Partial Pooling

With our model, the county mean estimates are weighted averages of our **best guess given the county** and our **best guess given all the counties**.

Mean in this county, weighted by number of observations and inverse of residual variability

Overall mean, weighted by inverse of county variability

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

**Weights**

**Some weak intuition:**  
**A precision-weighted average**  
**of two estimates.**

# How should we think about the partial pooling equation?

If we have lots of observation per county, we trust the sample mean more

$$\frac{n_j}{\sigma_y^2} \bar{y}_j + \frac{1}{\sigma_a^2} \bar{y}_{all}$$

If there's lots of residual variability, we trust the sample mean less

If the counties are spread out, we do less shrinking

The first term is specific to a single county, county j

This second term is the same for every county

Shrinkage is a weighted average of the county mean and the overall mean

# Empirical Bayes

---

This is known as the “Empirical Bayes” estimate. It’s our best guess at the value of a random intercept if we take the point estimates of the variances ( $\hat{\sigma}_\alpha^2$  and  $\hat{\sigma}_\varepsilon^2$ ) as their true values.

Unlike the county means, these estimates are *biased* towards the grand mean. However, they’re far less variable than the county means.

This is known as a “bias-variance tradeoff”; sometimes we can allow some bias into our estimates in return for (much) less variance.

On the whole, using these predictions will be better, on average, than using the county means. But the estimates will be less variable than reality.



# Model representations: A collapsed version

---

$$house_i = \alpha_{j[i]} + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

$$\alpha_j \sim N(\mu, \sigma_\alpha^2)$$

can be written as

$$house_i = \mu + u_{j[i]} + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

$$u_j \sim N(0, \sigma_\alpha^2)$$

The  $u_j$  is a county offset; the county mean for county  $j$  is equal to  $\mu + u_j$

**This collapsed representation is how R thinks of things.**



# Fitting our model with *lmer*

Include an intercept

```
> library( lme4 )  
> M0 <- lmer (y ~ 1 + (1 | county), data=houses)
```

```
> library( arm )  
> display (M0)
```

Used for the display function

Let that intercept vary by county

```
lmer(formula = y ~ 1 + (1 | county))
```

coef.est	coef.se
1.31	0.05

This is our grand mean/  
mean of county means!

Error terms:

Groups	Name	Std.Dev.
county	(Intercept)	0.31
Residual		0.80

This is the  
variance of the  
county means!

```
---  
number of obs: 919, groups: county, 85  
AIC = 2265.4, DIC = 2251  
deviance = 2255.2
```

“lmer” = Linear mixed effects  
in R

# Comparing the Fixed Effects Model to Random Intercept Model

---

## Fixed effects/unpooled:

- Unbiased estimates of each county

- $k + 1$  parameters (1 per county +  $\sigma^2$ )

- “Simple” to understand and fit

- Can be fit using OLS, a purely geometric/algebraic method:

- Cannot incorporate county-level predictors, and struggles with house-level predictors that don't vary much within a county

## Random effects/partially pooled:

- Biased estimates of each county

- Only 3 parameters ( $\mu$ ,  $\sigma^2_y$ ,  $\sigma^2_\alpha$ )

- Much more complex math and relatively little intuition

- Fit with Maximum Likelihood Estimation (MLE). Find the single set of parameters which maximize the likelihood of the observed data.

- Can include county-level predictors.

We'll cover MLE in the next unit.

# Random intercepts with Level-1 Predictors



# Partial pooling with predictors (G&H 12.3)

---

We know radon varies a lot by where the measurement was taken (basement vs. first floor) because radon sinks.

Some county samples have more basement readings than first floor readings.

Unfair to compare averages. Want to look for differences *adjusting for floor*.

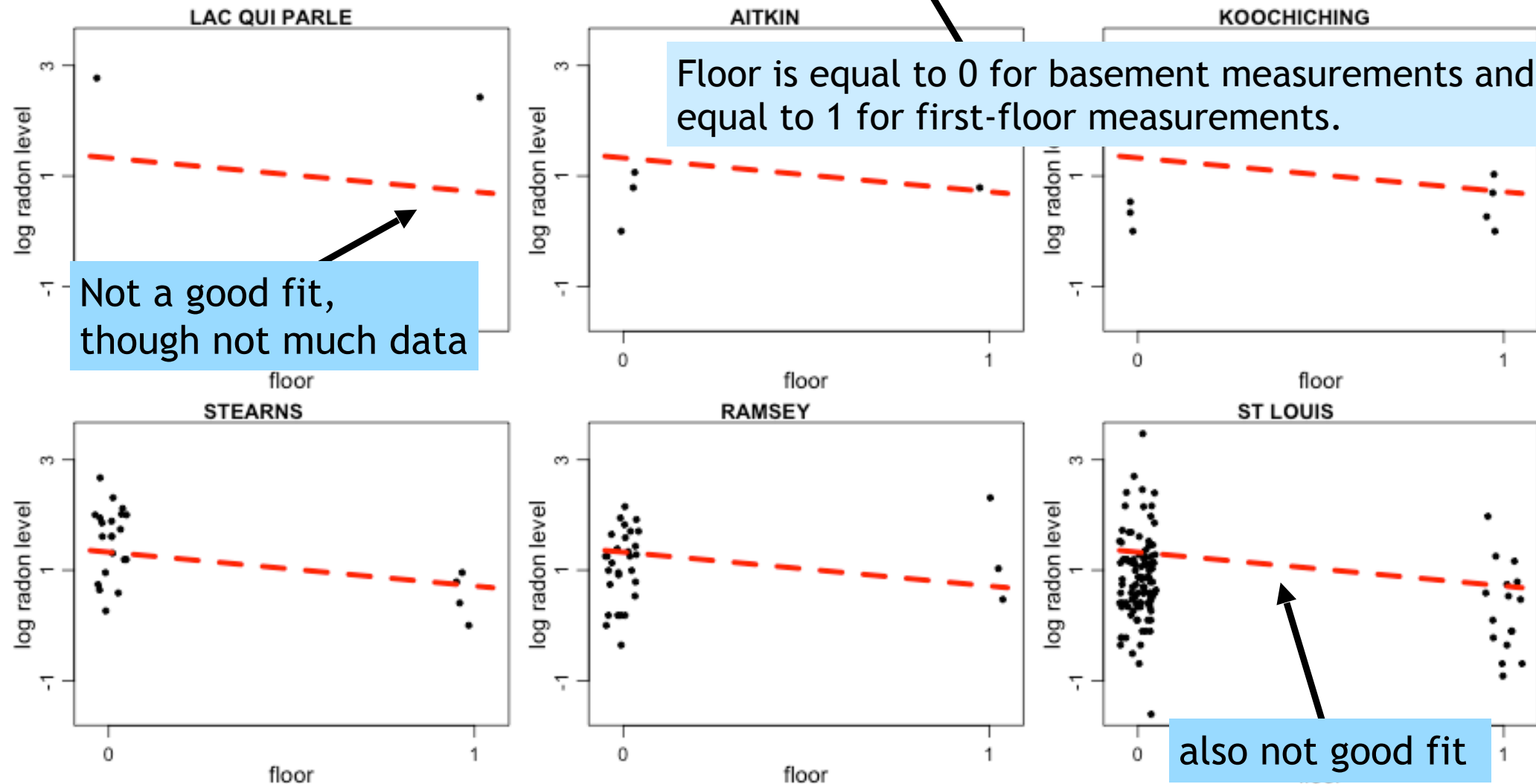
Accounting for floor may also improve *precision* of predictions.

# Complete Pooling of everything

Every county has the same intercept

$$y_i = \alpha + \beta_1 \text{floor}_i + \varepsilon_i$$

Floor is equal to 0 for basement measurements and equal to 1 for first-floor measurements.

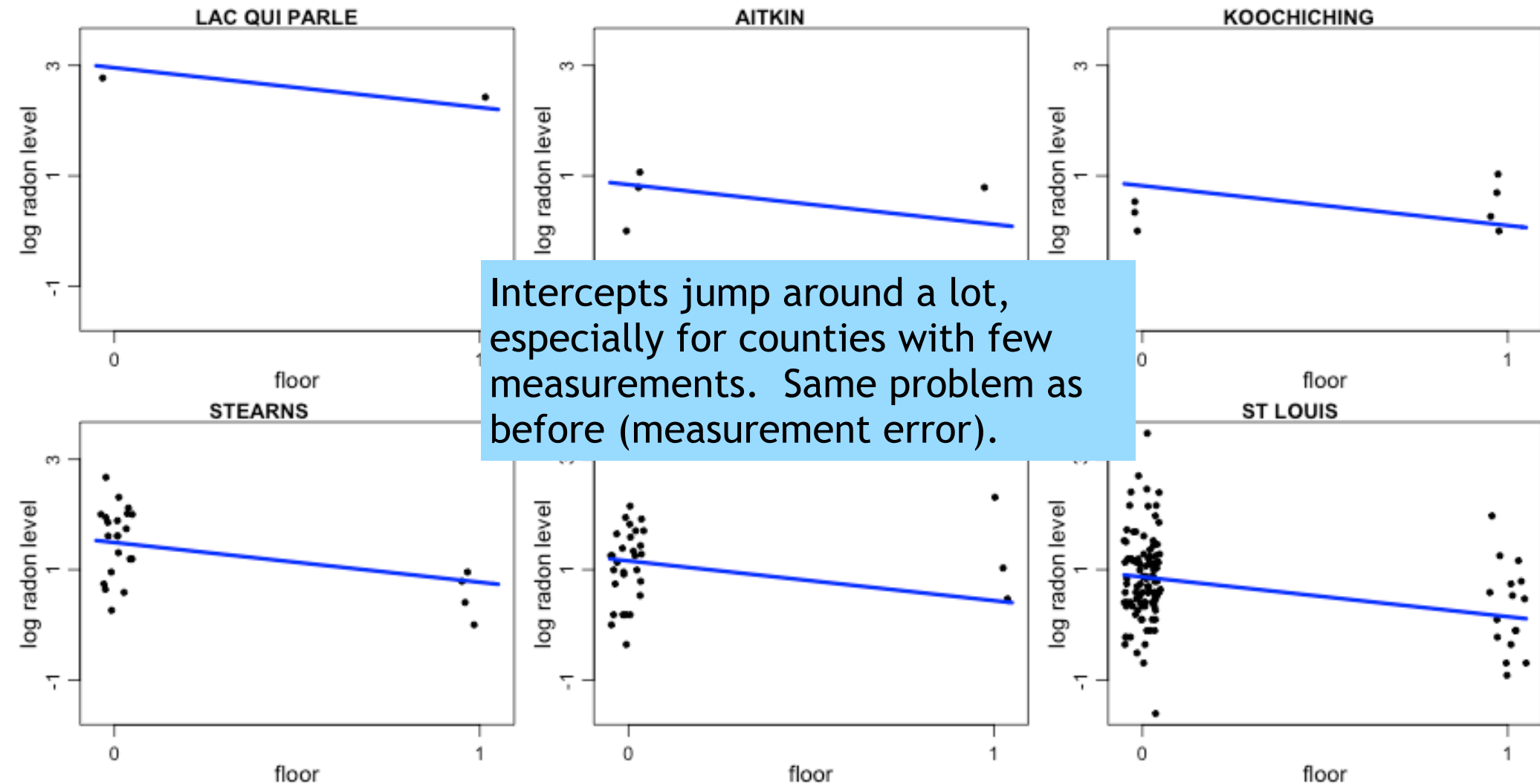


*No Pooling of intercept, just slope*  
(Fixed Effect Regression with additional level 1 (house) covariate)

$$y_i = \alpha_{j[i]} + \beta \text{floor} + \epsilon_i$$

Every county has its own intercept

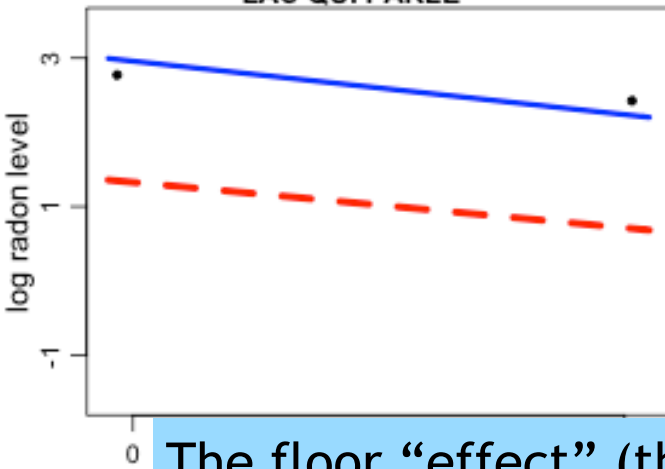
Intercepts jump around a lot, especially for counties with few measurements. Same problem as before (measurement error).



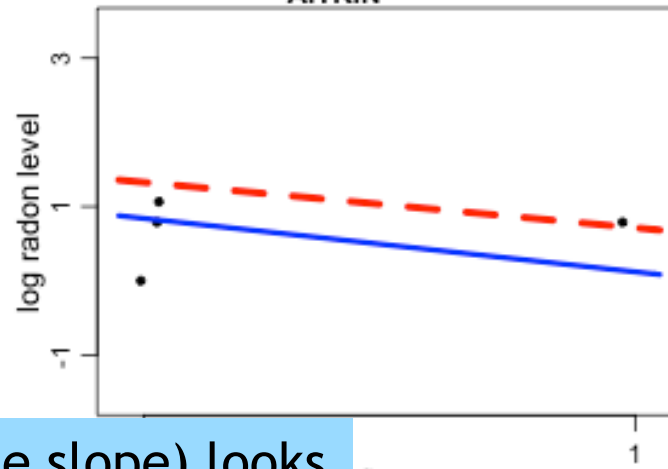
# Complete vs. No Pooling

In this model, what does the intercept represent?

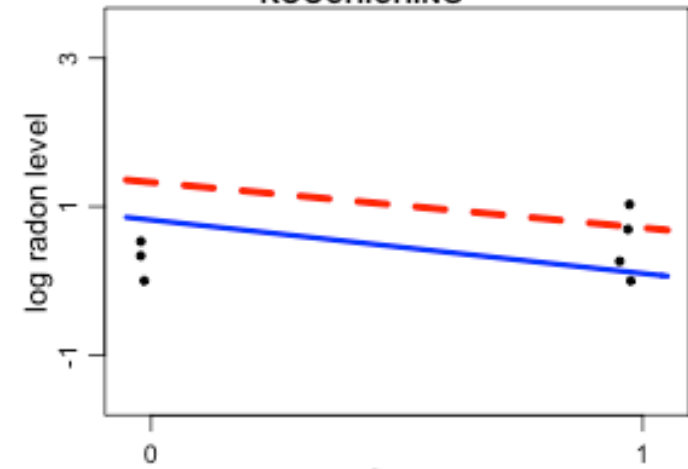
LAC QUI PARLE



AITKIN



KOOCHICHING



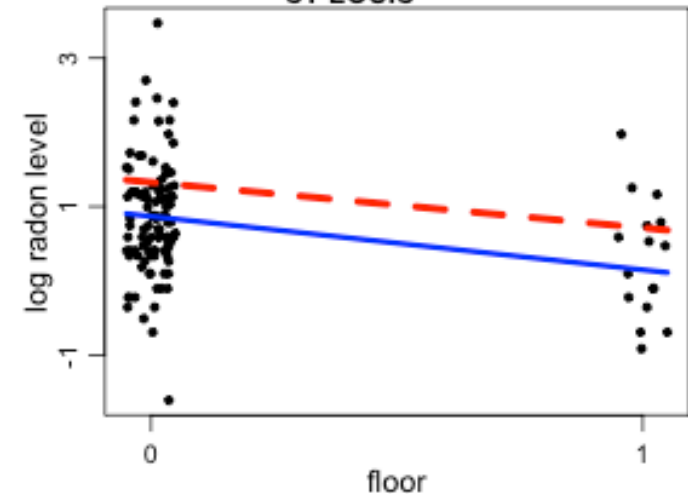
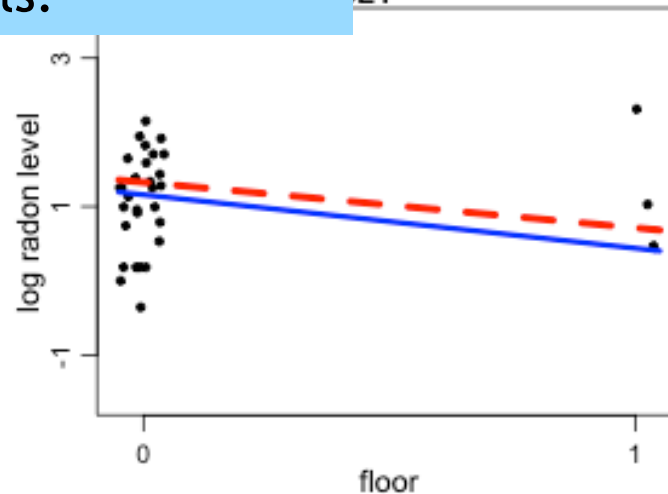
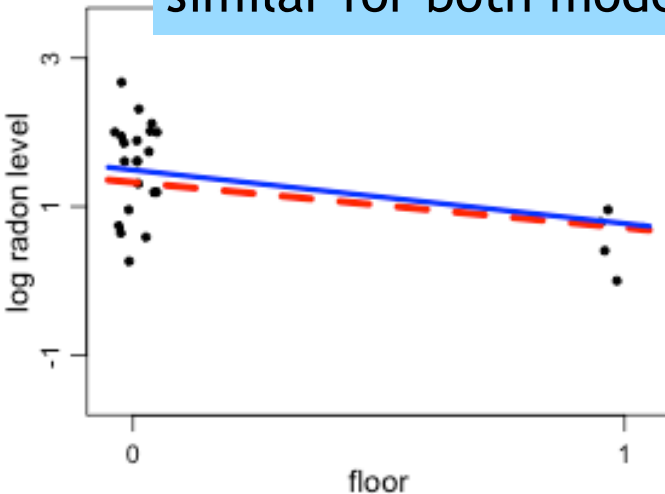
The floor “effect” (the slope) looks similar for both models.

floor

KEY

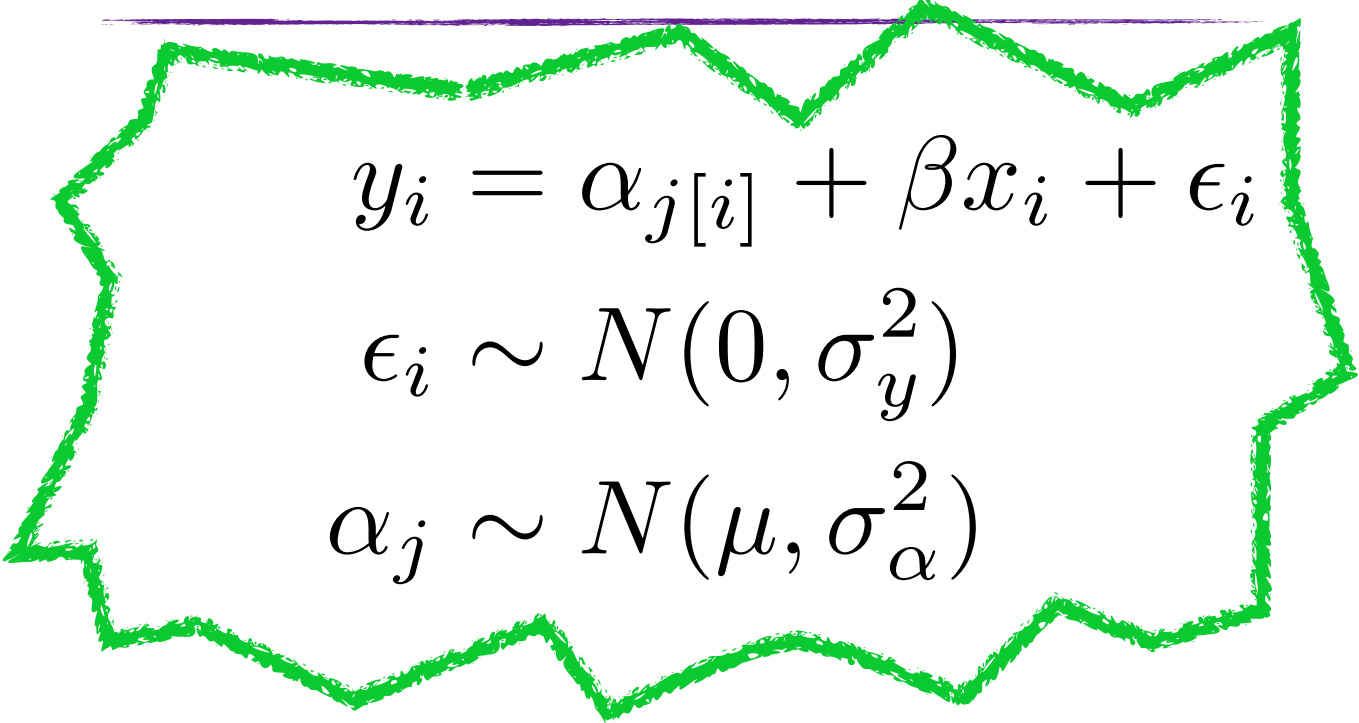
floor

ST LOUIS





# Solution: Tie the intercepts together


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

$$\epsilon_i \sim N(0, \sigma_y^2)$$

$$\alpha_j \sim N(\mu, \sigma_\alpha^2)$$

It is our soft constraint again.

Just like pooling without predictors.

We can then estimate the *parameters* ( $\mu$ ,  $\beta$ ) and *hyperparameters* ( $\sigma_\alpha^2$ ,  $\sigma_y^2$ )

# Our model in expanded form

---

$$\begin{aligned} house_i &= a_{j[i]} + \beta \cdot floor_i + \varepsilon_i \\ &= (\mu_a + u_{j[i]}) + \beta \cdot floor_i + \varepsilon_i \end{aligned}$$

**A normal  
looking  
regression**

$$= \mu_a + \beta \cdot floor_i + (u_{j[i]} + \varepsilon_i)$$

$$\varepsilon_i \sim N(0, \sigma_y^2)$$

$$u_j \sim N(0, \sigma_a^2)$$

**An overall  
residual**

# An approximation for our alpha estimates

---

estimate of  $\alpha_j \approx \frac{\frac{n_j}{\sigma_y^2}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} (\hat{\alpha}_j - \beta \bar{x}_j) + \frac{\frac{1}{\sigma_\alpha^2}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} \mu_\alpha$

**NOT FOR USE**

On left hand side we see classic regression adjustment using the overall beta.

On right hand side we just take the mean of all counties.

Again we see shrinkage: we are taking a weighted average of a county-specific *adjusted* mean and the overall grand mean



# Fitting our multilevel model

Regress log\_radon on floor

```
> mod_cov <- lmer(log_radon ~ 1 + floor + (1|county), data = dat)
```

```
> display(mod_cov)
```

	coef.est	coef.se
(Intercept)	1.49	0.05
floor	-0.66	0.07

Allow the intercept to vary by county

Error terms:

Groups	Name	Std.Dev.
county	(Intercept)	0.32
Residual		0.73

---

number of obs: 919, groups: county, 85

AIC = 2105.5, DIC = 2081.9

deviance = 2089.7

We see estimates for the hyperparameters and parameters.



# Getting the estimated county regression lines

Each county gets its own row.  
We have slopes and intercepts  
for individual regression lines!

```
> names(coef(mod_cov))  
[1] "county"
```

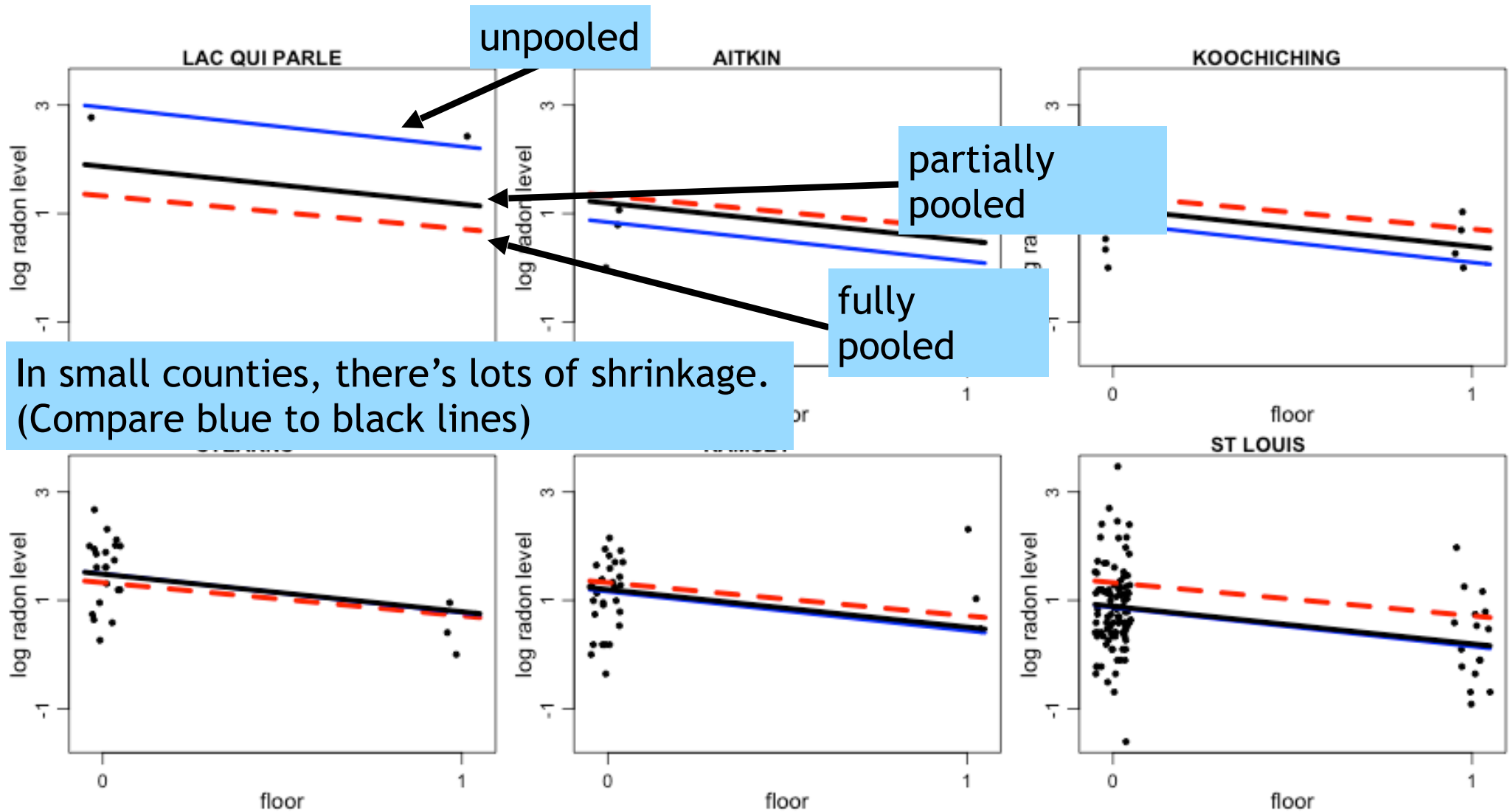
```
> counties <- coef(mod_cov)$county
```

```
> counties
```

	(Intercept)	floor
AITKIN	1.2291388	-0.6628887
ANOKA	0.9811502	-0.6628887
BECKER	1.5066938	-0.6628887
BELTRAMI	1.5377713	-0.6628887
BENTON	1.4732718	-0.6628887
BIG STONE	1.5084495	-0.6628887
...		

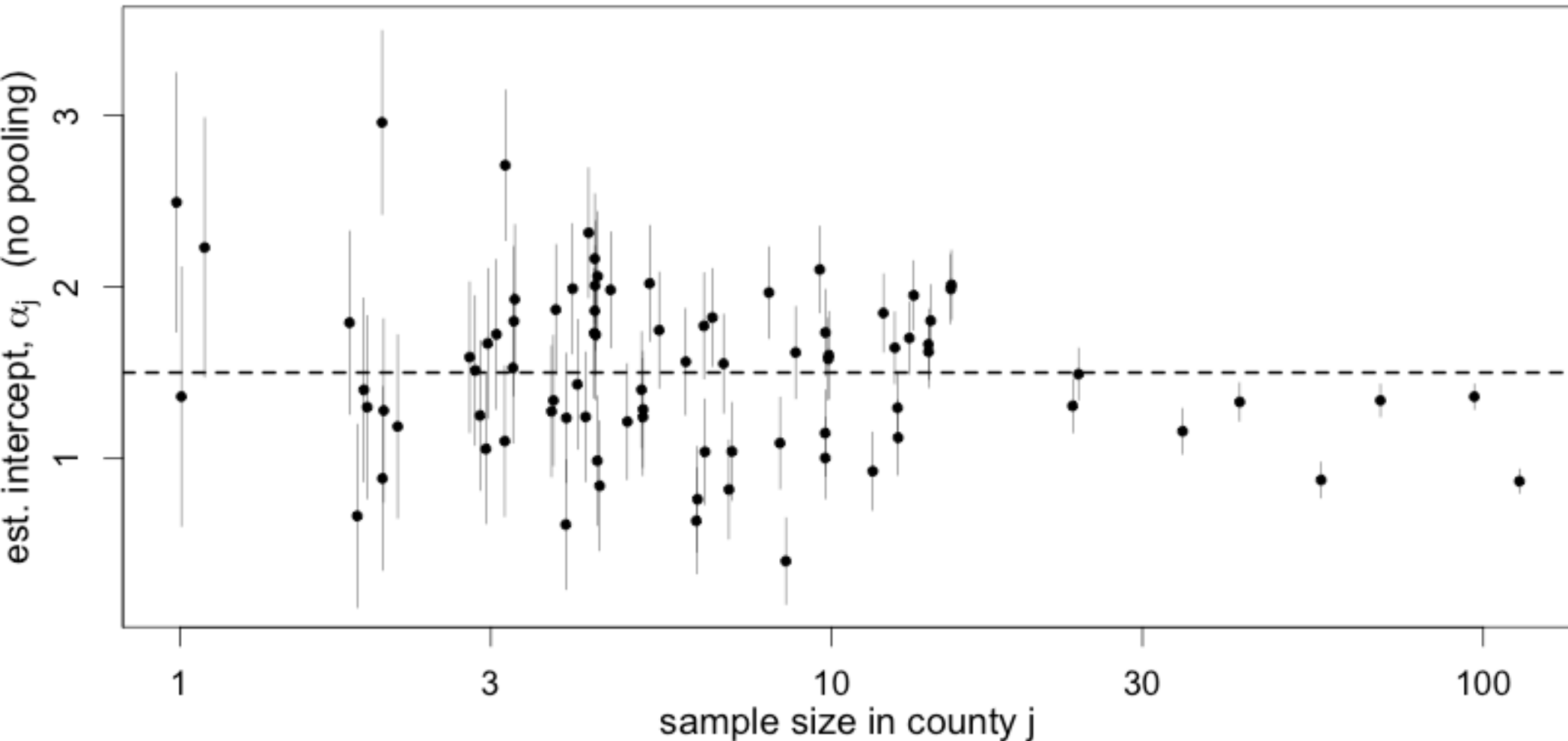
**Warning!**  
coef() gives a LIST of fixed and  
random effects for each grouping  
variable. Hence the "\$county"

# Complete Pooling (red ), No Pooling (blue), and Partial Pooling (black)

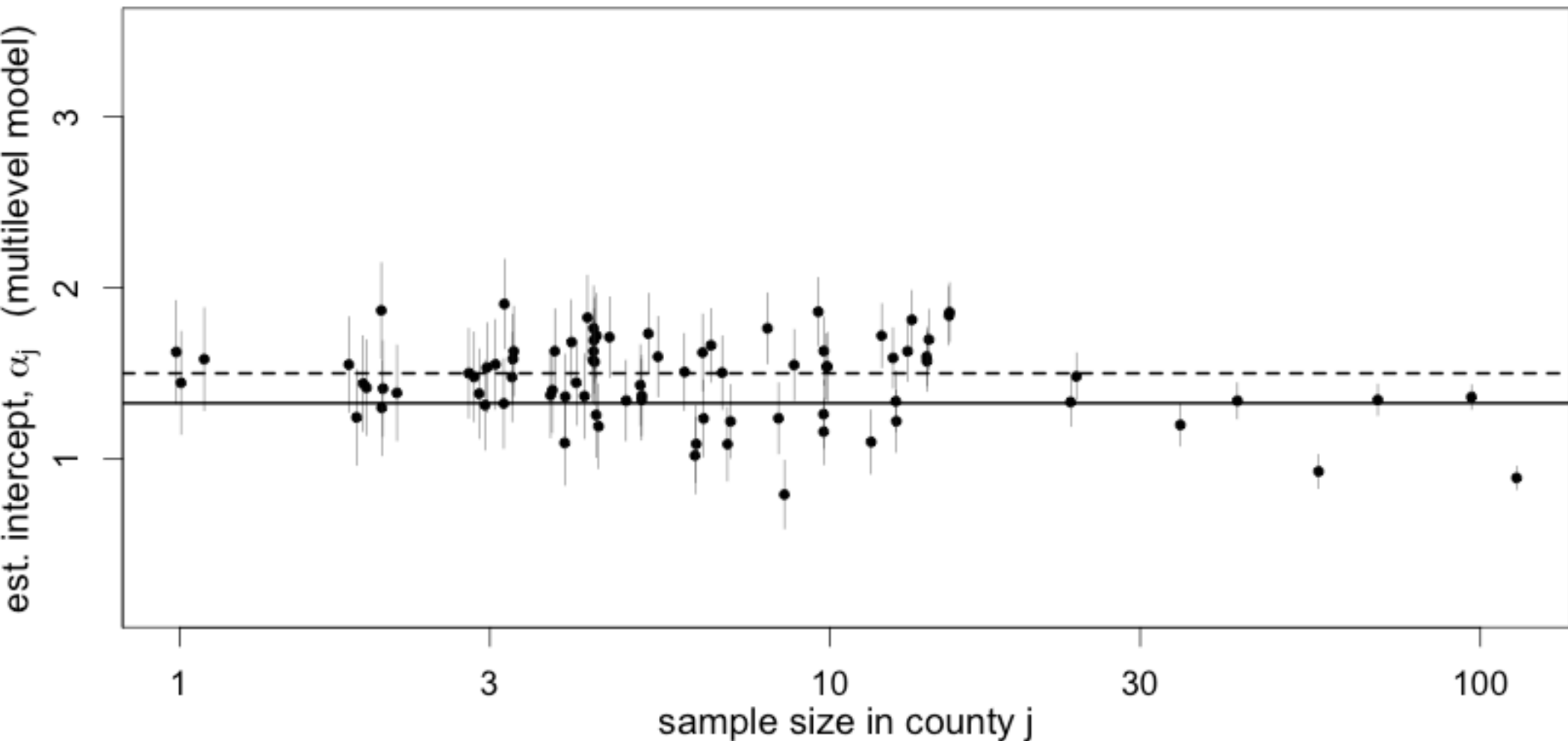


In large counties, there's very little

# Unpooled Intercept Estimates by Sample Size (adjusted by floor)



# Partially Pooled Intercept Estimates by Sample Size (adjusted by floor)





# What to do with a fitted model

---

You can

Describe themes and patterns across your groups (e.g., look at the coefficient of *floor*).

You can

***Attempt to*** describe what is happening with an individual group (e.g., predict a random intercept for a single county).

Describing individuals is **much harder** than describing themes.

# What just happened?



# Recap

---

Check-In  
<http://cs179.org/lec21>

The random intercept model “partially pools” the individual estimates for groups

- ★ This allows us to have better predictions for individual groups
- ★ It also allows us to have estimates of the *true* variation of the groups

We can easily incorporate individual covariates. This does some nice stuff:

- ★ Improves precision of our group mean estimates
- ★ Adjust our group-to-group comparisons to account for differences in the composition of the individuals

We can fit our random intercept model using the lme4 package’s `lmer()` function