Linear Model with Random Intercept

LM->General Linear Model

- General Linear Model is Linear Regression model with correlated errors.
- For GLM case, GEE is used.

Linear Regression Model

Model: $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$

$$\boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

Estimation: MLE

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{y})$$

$$\boldsymbol{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$$

Variance Estimate: REML

$$\hat{\sigma}^2 = \frac{1}{n-p}(\boldsymbol{y} - \widehat{\boldsymbol{y}})^T(\boldsymbol{y} - \widehat{\boldsymbol{y}})$$

Special Case

General Linear Model

Model: $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$

$$\boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{V})$$

Estimation: GLS

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y})$$

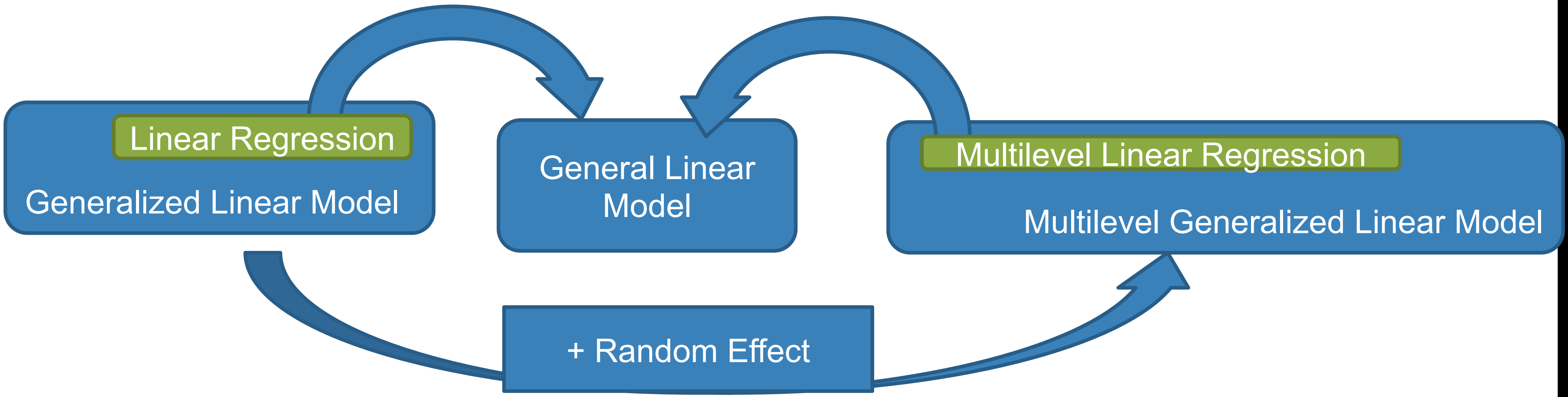
$$\boldsymbol{Var}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$$

Variance Estimate: Sandwich Estimator

Dealing within group correlation

- The BIG model:
  - Repeated Measurement GLM -> GEE

- The Smaller model:
  - Multilevel regression



Repeated measurement normal linear model

- The usual model

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

- The difference is

$$\epsilon = N(\mathbf{0}, \mathbf{V})$$

- Therefore

$$\mathbf{y} \sim N(\mu = \mathbf{X}\beta, \mathbf{V})$$

- where

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \cdots & \mathbf{0} \\ \vdots & & \ddots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V}_N \end{bmatrix}$$

$$\mathbf{y} = \begin{pmatrix} \left.\begin{matrix} y_{11} \\ y_{12} \\ \cdots \\ y_{1K} \end{matrix}\right\} \text{1st individual} \\ \left.\begin{matrix} y_{21} \\ y_{22} \\ \cdots \\ y_{2K} \end{matrix}\right\} \text{2nd individual} \\ y_{31} \\ \left.\begin{matrix} \cdots \\ y_{Nk} \end{matrix}\right\} \text{Nth individ} \end{pmatrix}$$

Different forms of V_j

- Equicorrelation

$$\mathbf{V}_j = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & & \ddots & \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

Random Intercept

- Distance based

$\rho_{jk}=f(t_j,t_k)$

$$\mathbf{V}_j = \sigma^2 \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & & \ddots & \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{bmatrix}$$

Random Slope

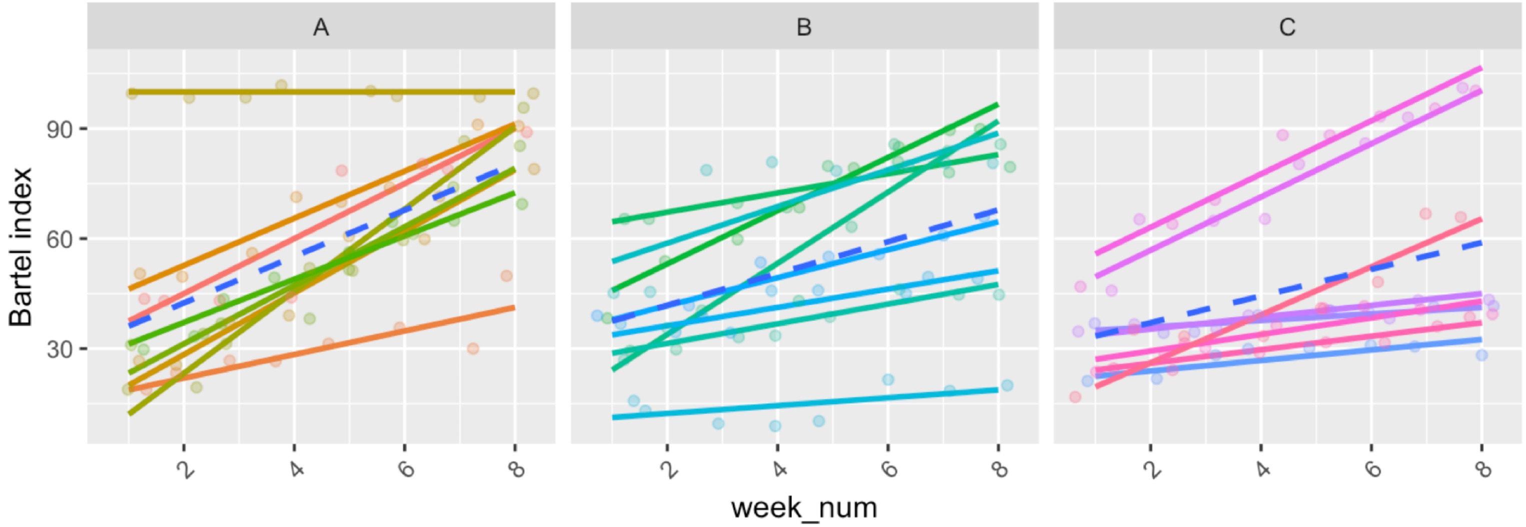
- Unstructured

Is correlation a bad thing?

- The big model a linear regression model that needs “correction”.
- Since it violates the assumption of uncorrelated errors, the standard error estimates will be off if we don’t account for this information.
- It makes sense and
- It’s important if testing is the primary goal of the analysis.
- Under this framework, however, I cannot get rid of the feeling that the correlation is treated as something bad.

Embracing the variability

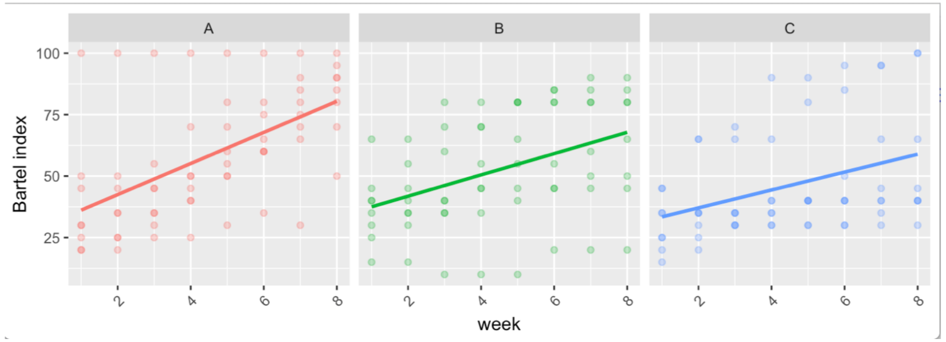
- No one looks exactly like the group average



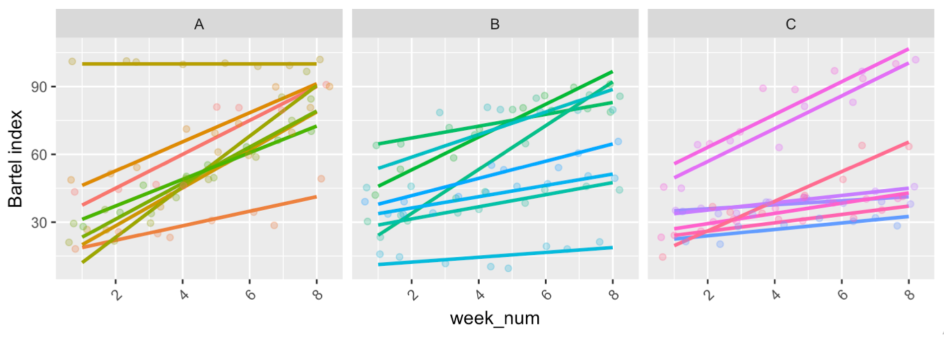
- However, we also believe there’s shared characteristics within groups.
- How can we let the individual data speak for itself while trying to asses the characteristics of the groups?

What lies between the two extremes?

Overall trend and everyone deviates from it.



Everyone is unique.



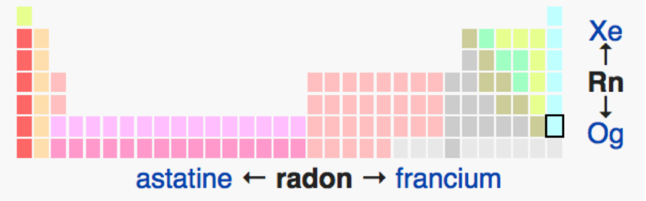
?

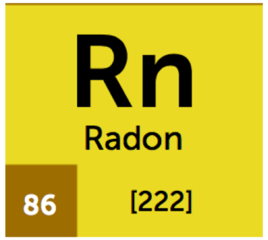
Complete – ? – No pooling

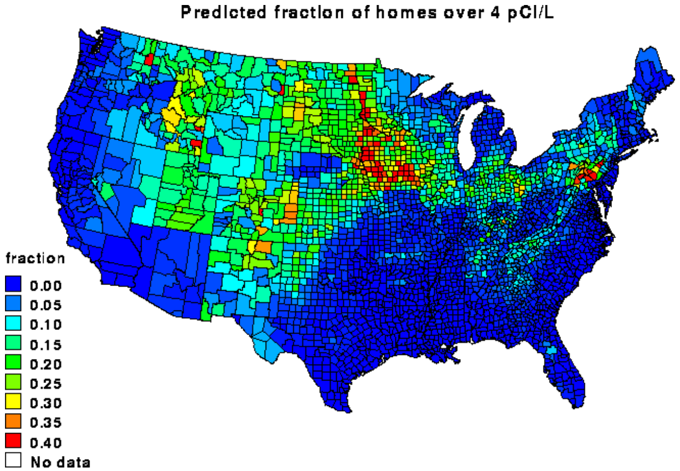
Random Intercept

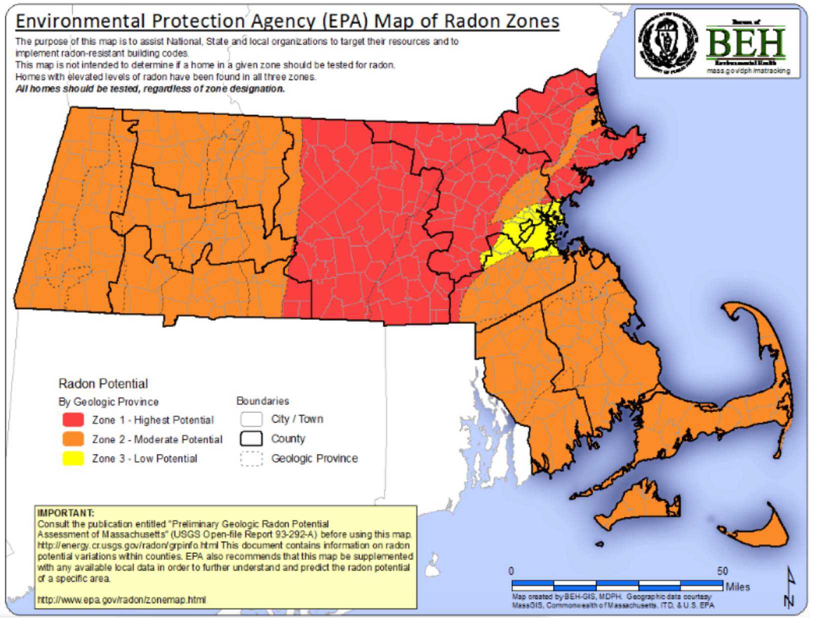
Home radon measurement and remediation

- Radon: a naturally occurring radioactive gas
  - cause lung cancer in high concentrations and
  - 1000s of lung cancer deaths per year in US.
- Radon levels distribution in U.S. homes varies
  - some houses have dangerously high concentrations.
- Environmental Protection Agency
  - radon measurements in 80,000 houses (random)







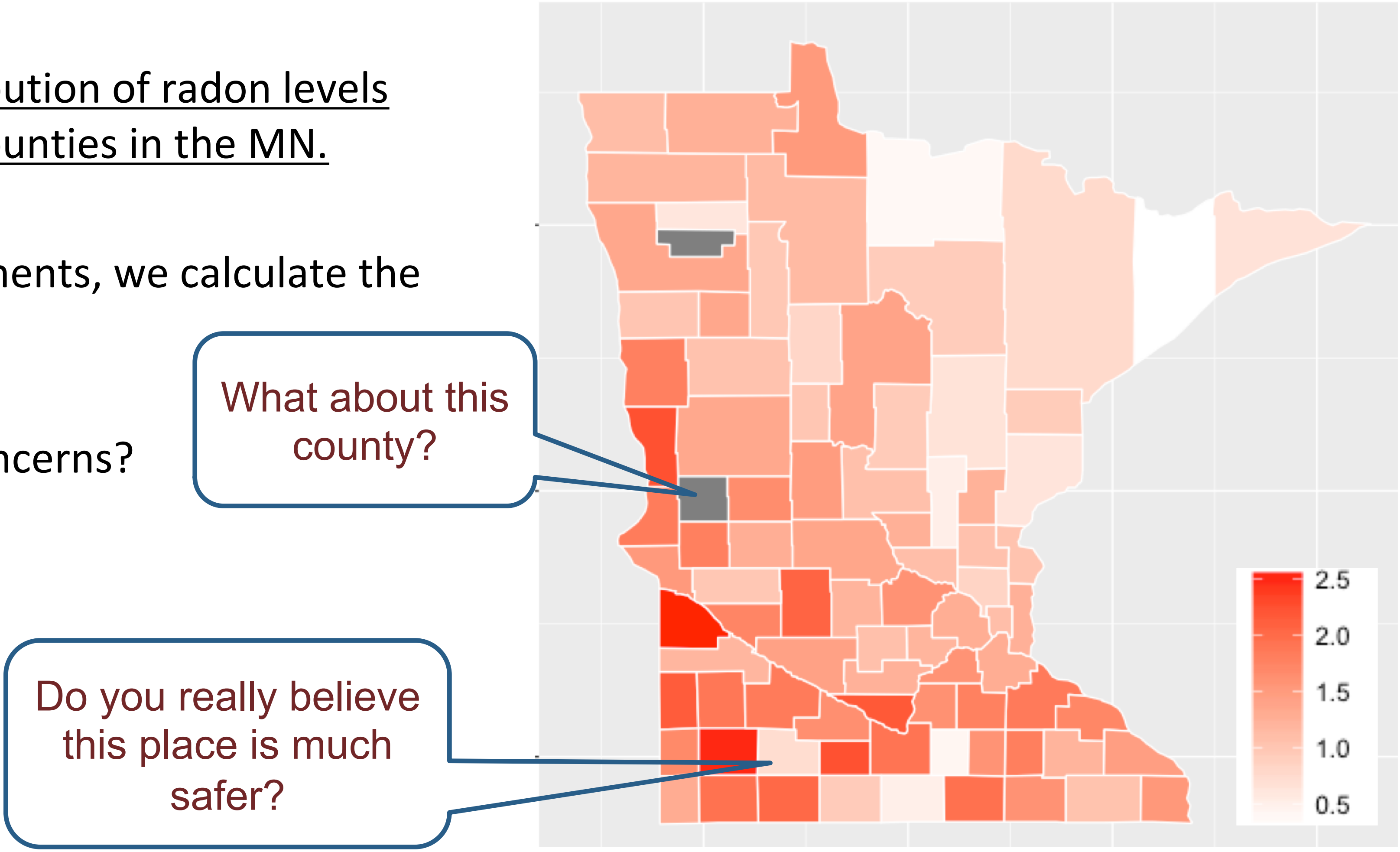


https://en.wikipedia.org/wiki/Radon

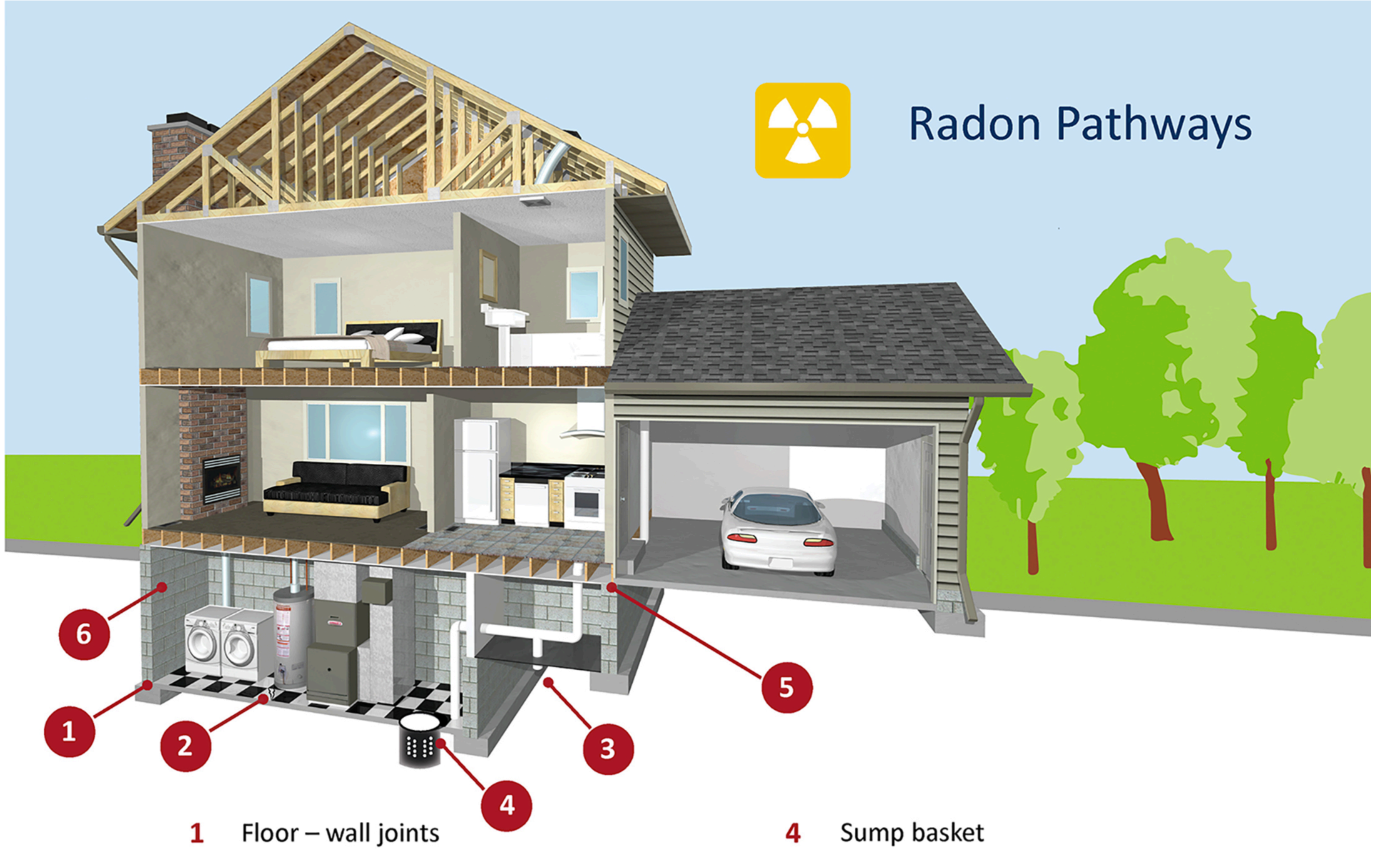
Radon Levels in MN

- Goal:
  - Estimate the distribution of radon levels in each of the 85 counties in the MN.
- That's easy!
  - We have measurements, we calculate the average, done.
  - Right?
- What are some of the concerns?

Map of average log radon level by county



Common radon pathways



1 Floor – wall joints

2 Pores and cracks in concrete blocks or slabs

3 Exposed soil, such as in a crawl space

4 Sump basket

5 Open tops of block walls

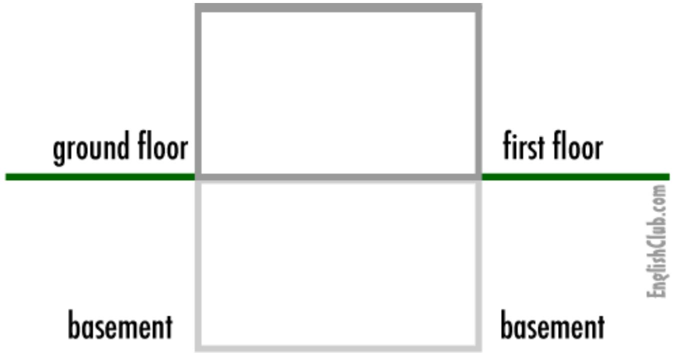
6 Mortar joints

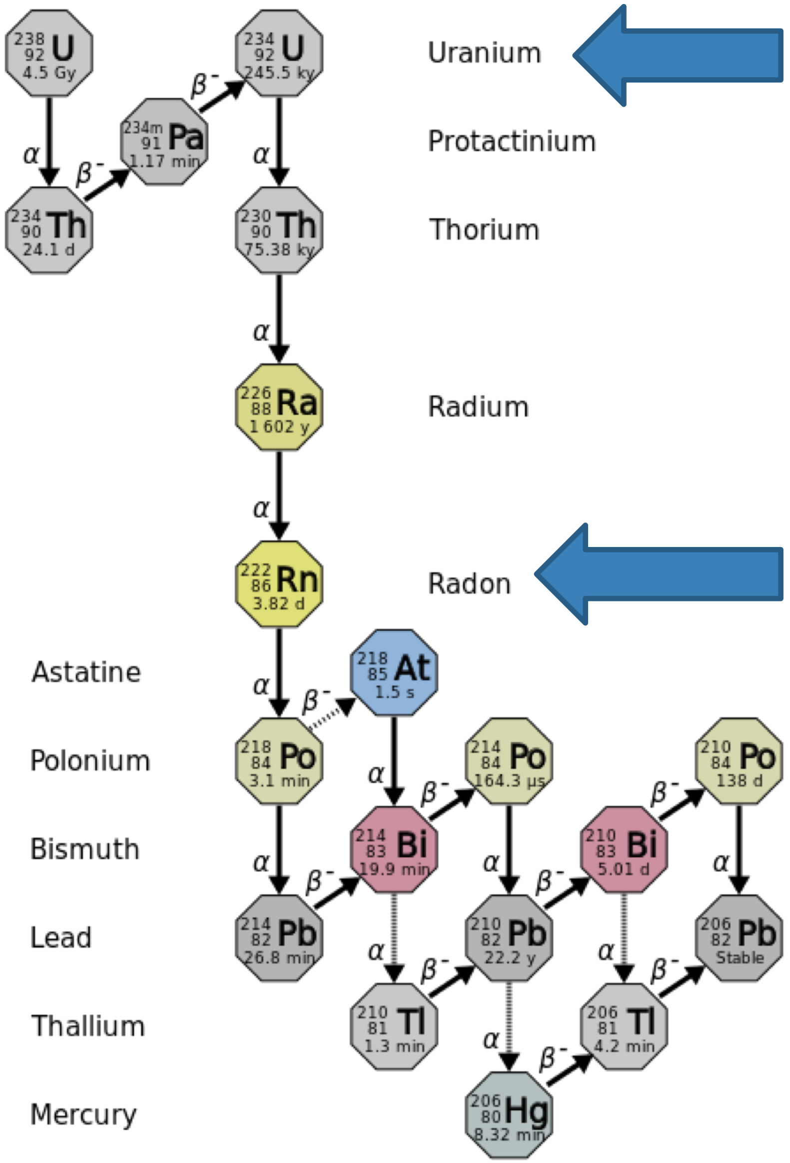
http://www.health.state.mn.us/divs/eh/indoorair/radon/

Other information

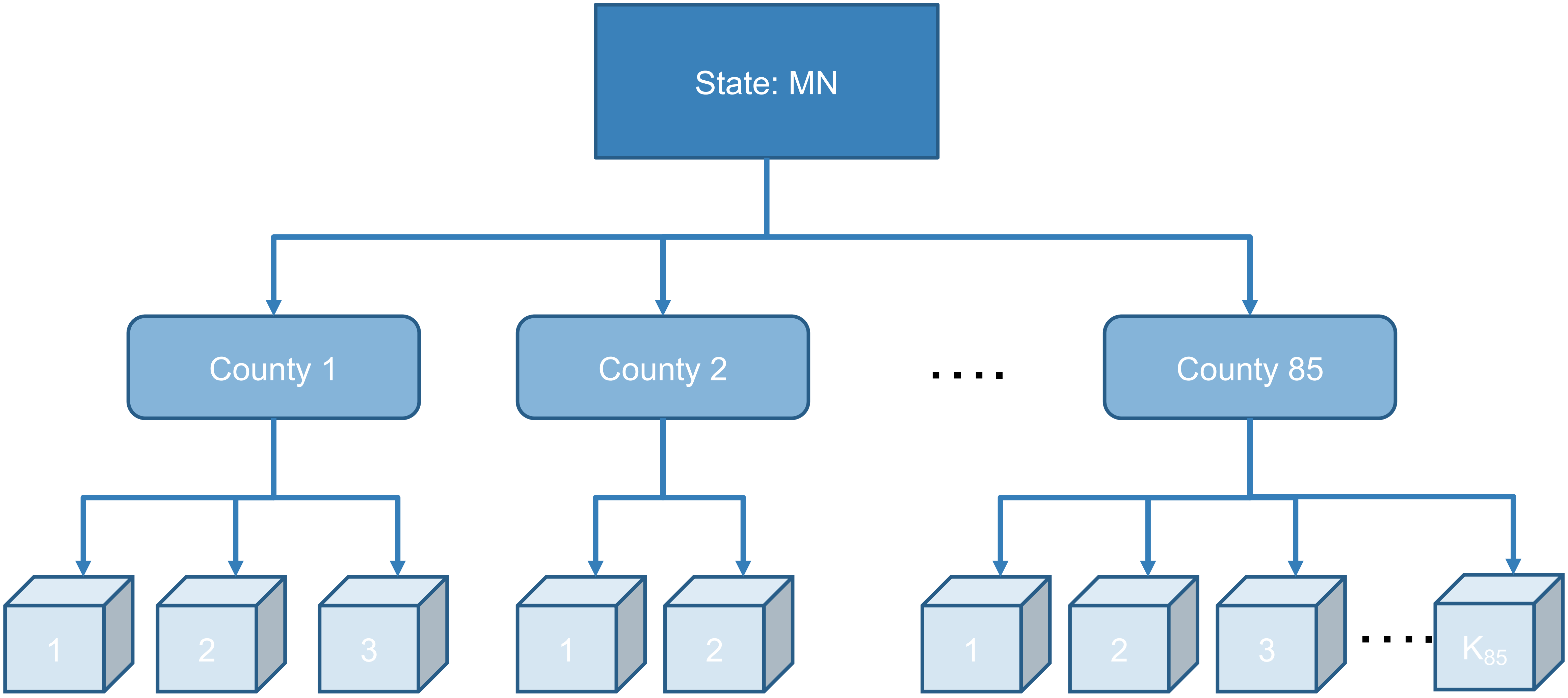
- Soil uranium level measurement per county
  - Radon happens during the decay process of Uranium.
  - You have a measurement of soil uranium per county.

- The floor of measurement per house
  - either basement or first floor;
  - radon comes from underground and can enter more easily when a house is built into the ground.

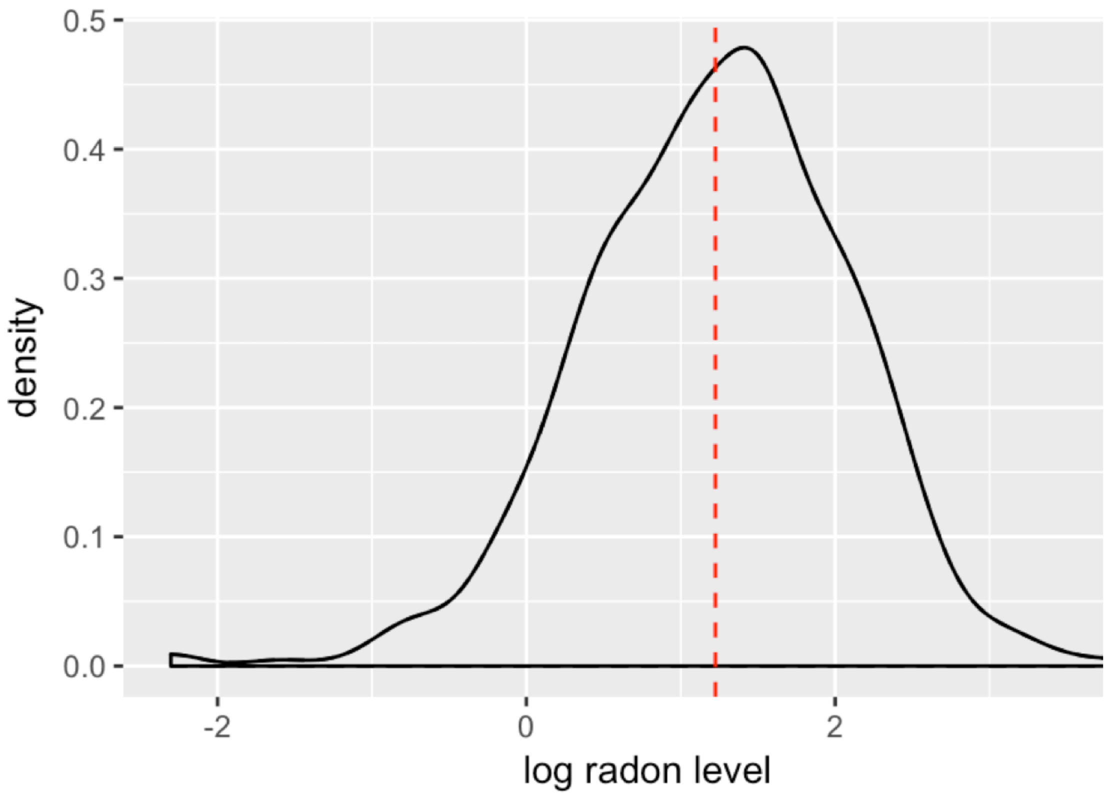


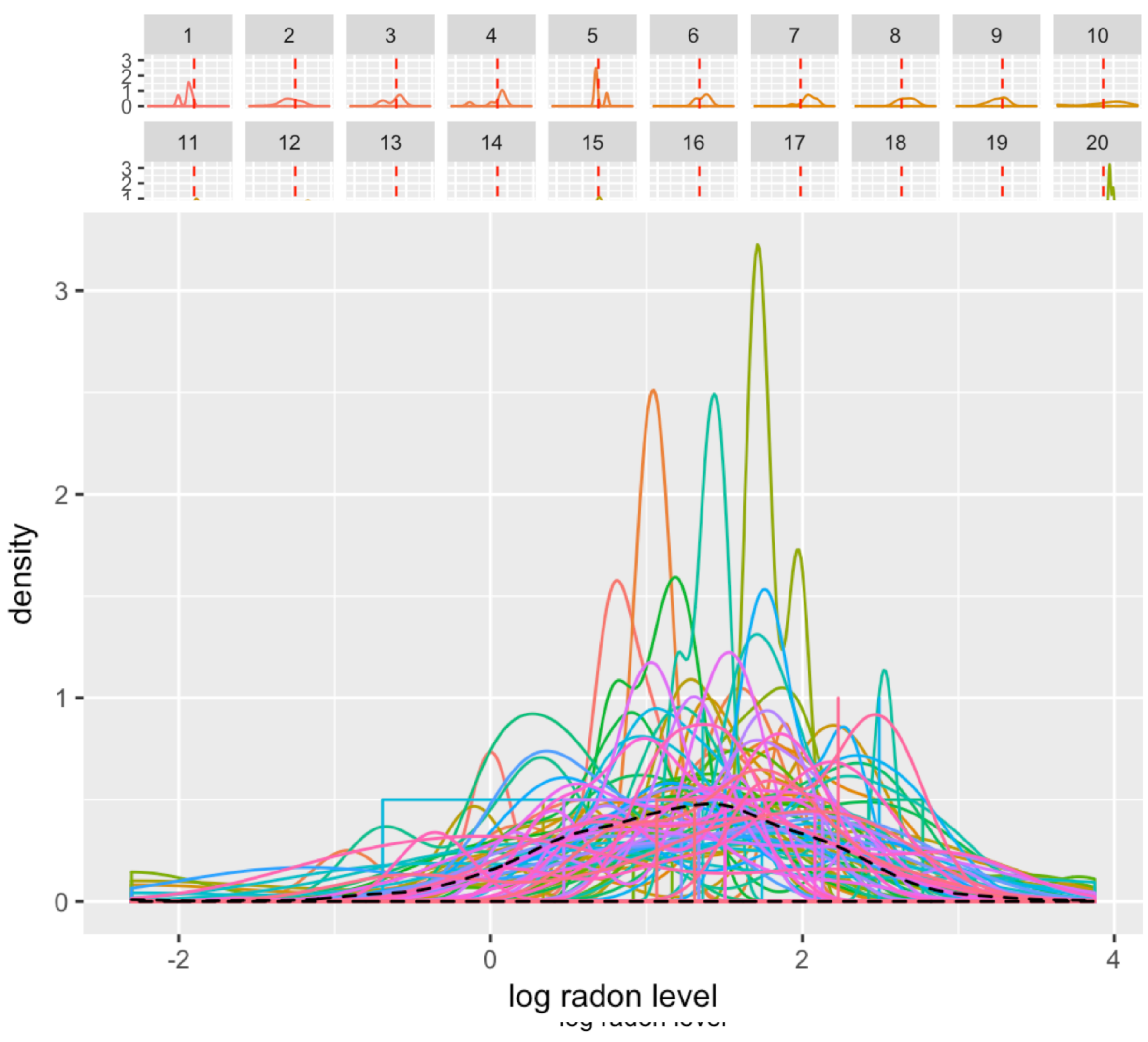


Data Structure

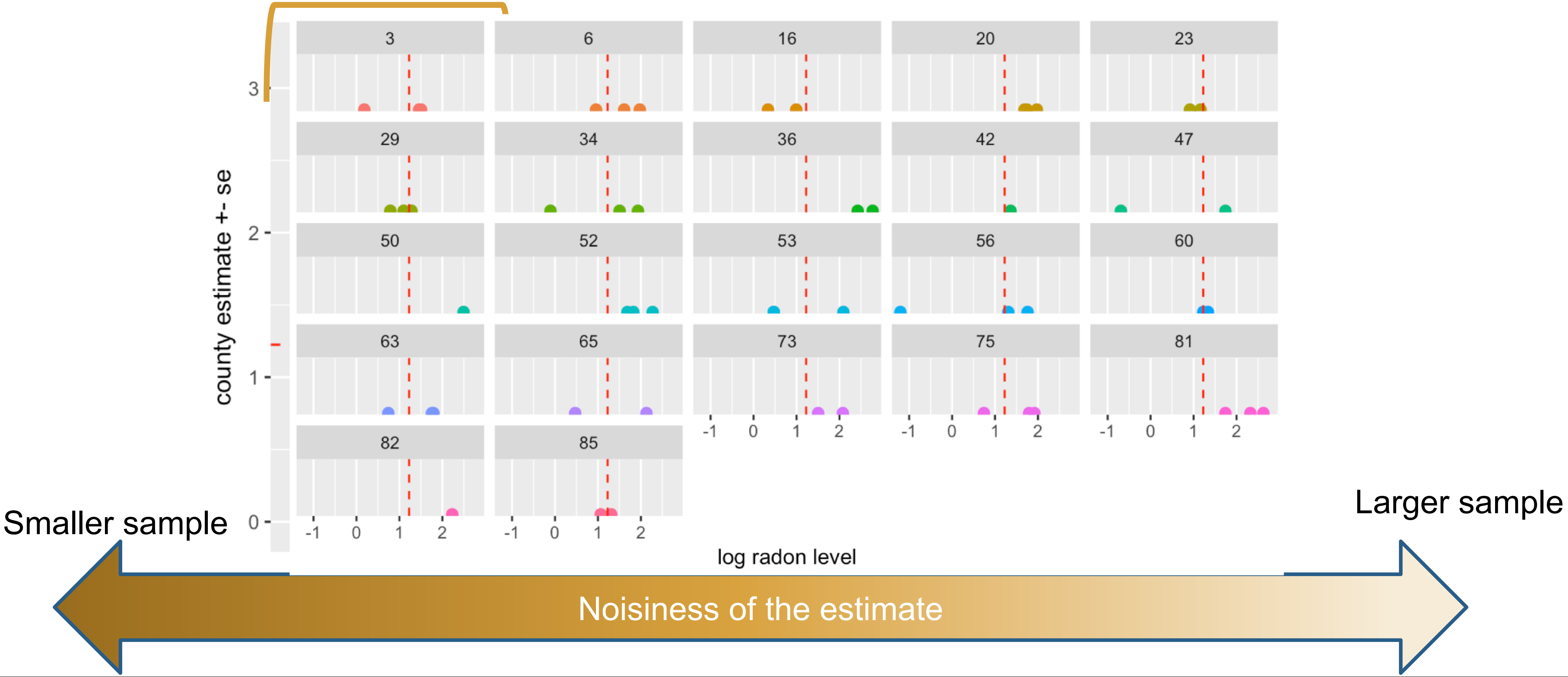


Distribution of radon levels of the houses within each of 85 counties in Minnesota



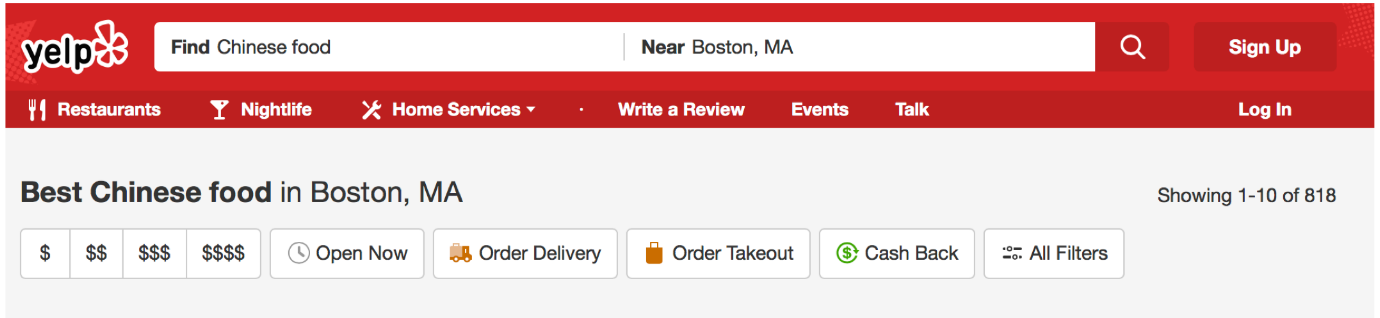


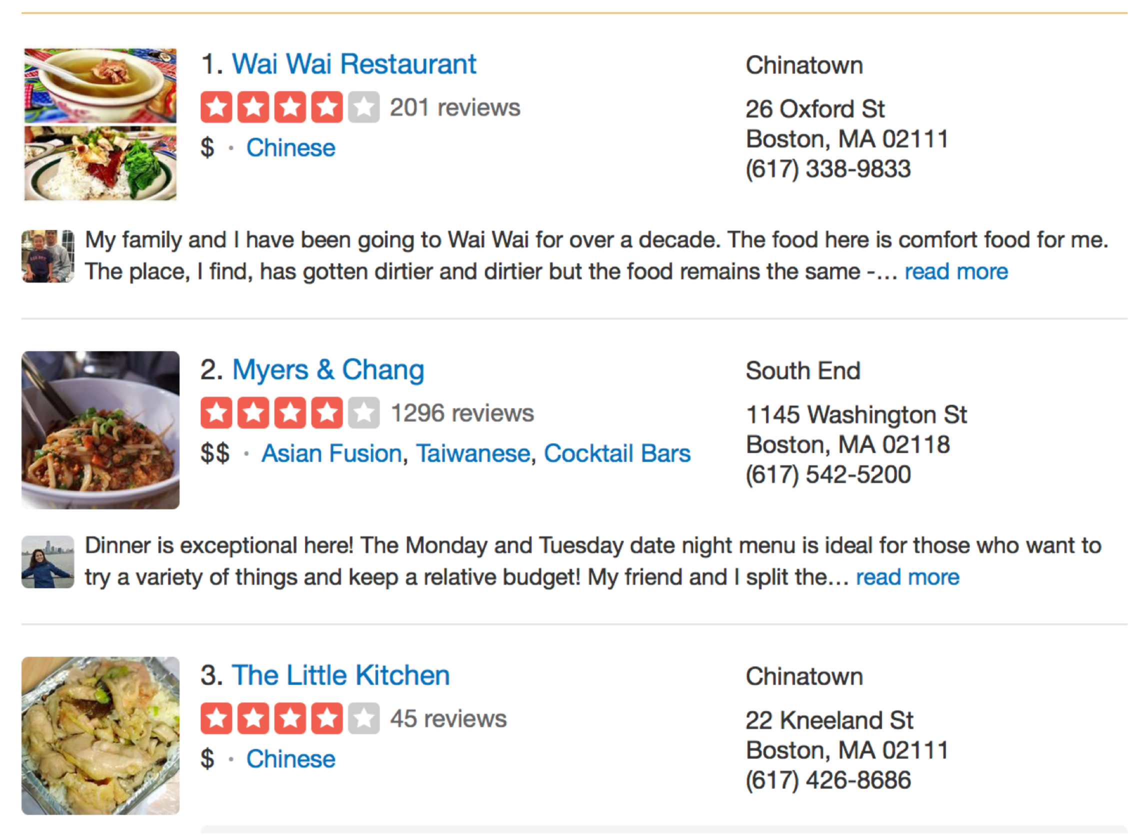
No pooling estimate vs Complete pooling

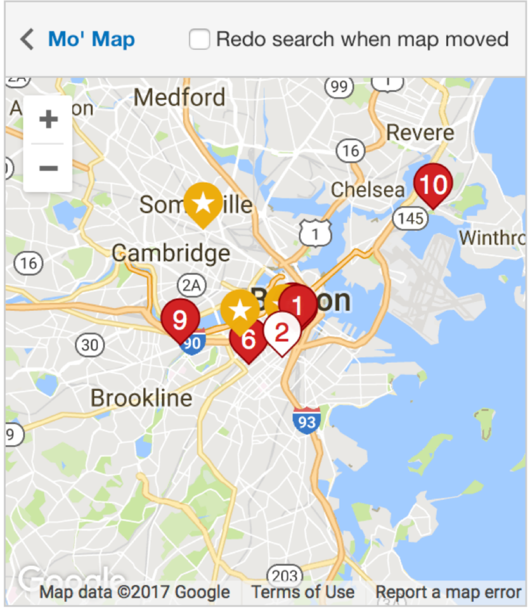


Big picture

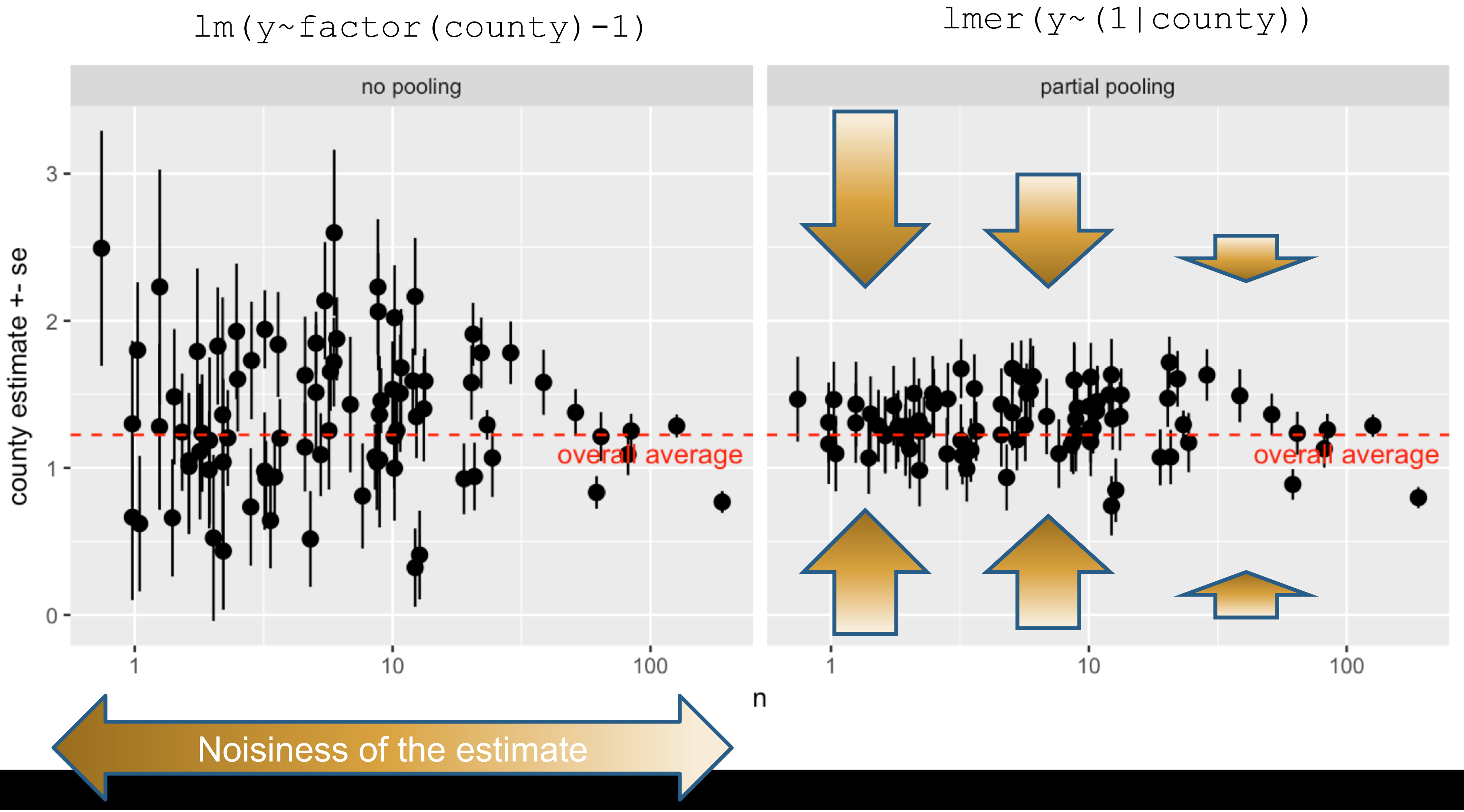
- Which review do you trust more?







Partial Pooling: Less certainty more shrinkage



j[i] is a compact way to express the structure

- intercept $\alpha_{j[i]}$
- slope $\beta_{j[i]}$

- Say we have 3 cities with coefficients for each city j.
  - $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3$
- An individual i in one of 3 cities.
- If an individual 2 comes from city 1 then j[2]=1

```
> i<-1:6
> j<-c(2,1,1,2,3,3)
> j[2]
[1] 1
```

Individual i	Group j
1	2
2	1
3	1
4	2
5	3
6	3

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

$$y_1 = \alpha_2 + \beta_2 x_1 + \varepsilon_1$$
$$y_2 = \alpha_1 + \beta_1 x_2 + \varepsilon_2$$
$$y_3 = \alpha_1 + \beta_1 x_3 + \varepsilon_3$$
$$y_4 = \alpha_2 + \beta_2 x_4 + \varepsilon_4$$
$$y_5 = \alpha_3 + \beta_3 x_5 + \varepsilon_5$$
$$y_6 = \alpha_3 + \beta_3 x_6 + \varepsilon_6$$

The models

`lm(y~county-1)`

No Pooling

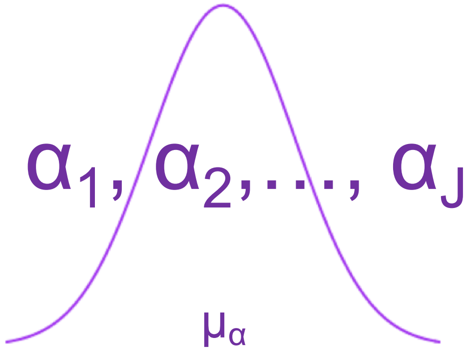
$$y_i = \alpha_{j[i]} + \underline{\varepsilon}_i$$
$$\underline{\varepsilon}_i \sim N(0,\ \sigma^2{}_y)$$

$\alpha_1, \alpha_2, \ldots, \alpha_J$

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

Partial Pooling

$$y_i = \underline{\alpha}_{j[i]} + \underline{\varepsilon}_i$$
$$\underline{\varepsilon}_i \sim N(0,\ \sigma^2{}_y)$$
$$\underline{\alpha}_j \sim N(\mu_\alpha, \sigma^2{}_\alpha)$$



$\alpha_1, \alpha_2, \ldots, \alpha_J$

`lm(y~1)`

Complete Pooling

$$y_i = \alpha + \underline{\varepsilon}_i$$
$$\underline{\varepsilon}_i \sim N(0,\ \sigma^2{}_y)$$

α

Partial pooled estimate

No Pooling

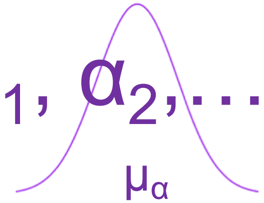
$$y_i = \alpha_{j[i]} + \underline{\varepsilon}_i$$
$$\underline{\varepsilon}_i \sim N(0, \sigma^2_y)$$

$$\alpha_1, \alpha_2, \ldots, \alpha_J$$

Partial Pooling

$$y_i = \underline{\alpha}_{j[i]} + \varepsilon_i$$
$$\underline{\varepsilon}_i \sim N(0, \sigma^2_y)$$
$$\underline{\alpha}_j \sim N(\mu_\alpha, \sigma^2_\alpha)$$

$$\alpha_1, \alpha_2, \ldots, \alpha_J$$



Complete Pooling

$$y_i = \alpha + \underline{\varepsilon}_i$$
$$\underline{\varepsilon}_i \sim N(0, \sigma^2_y)$$

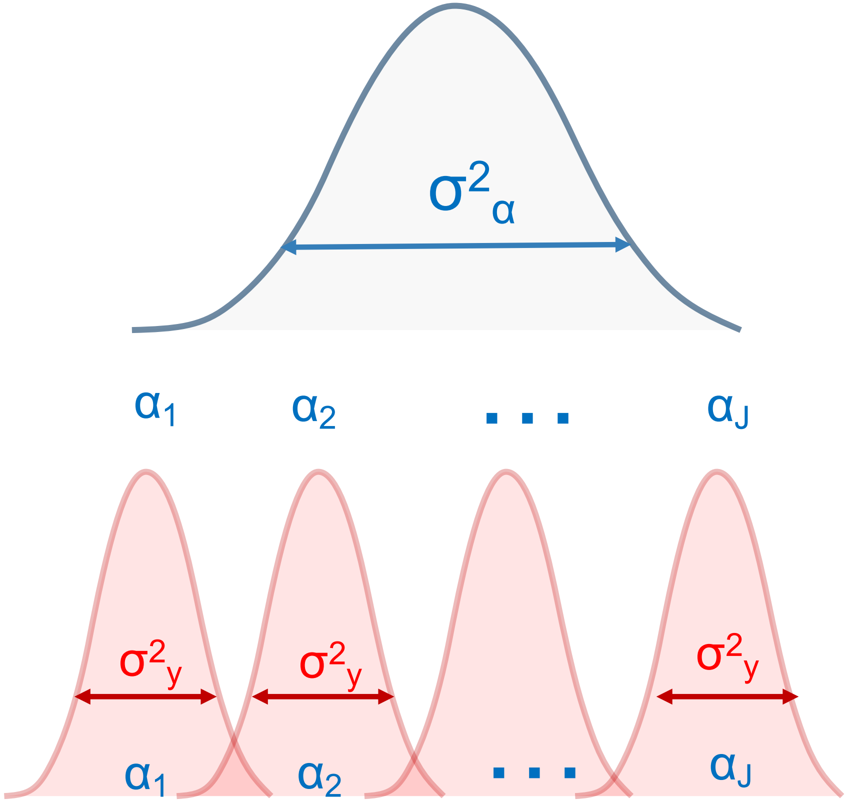
$$\alpha$$

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

Difference in the error terms

- σ^2_α: the variance among the average log radon levels of the different counties
- σ^2_y: the within-county variance in log radon measurements

```
lmer(formula = y ~ 1 + (1 | county))
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 = 2265.4, DIC = 2251
deviance = 2255.2
```



Sources of variability and level of shrinkage

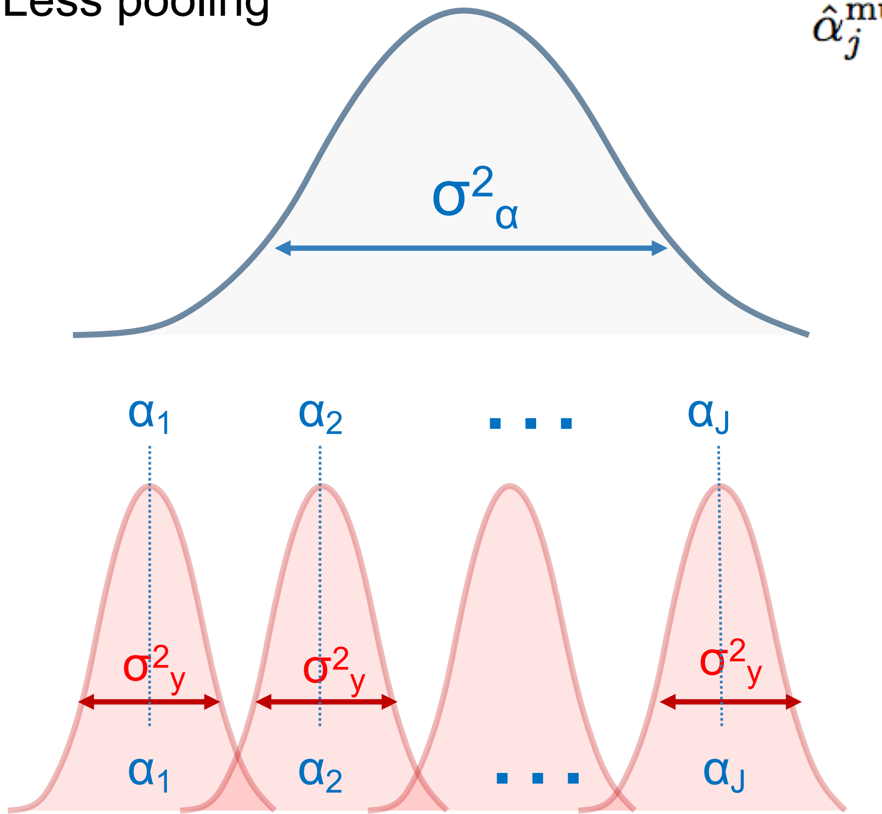
Individual estimate more reliable

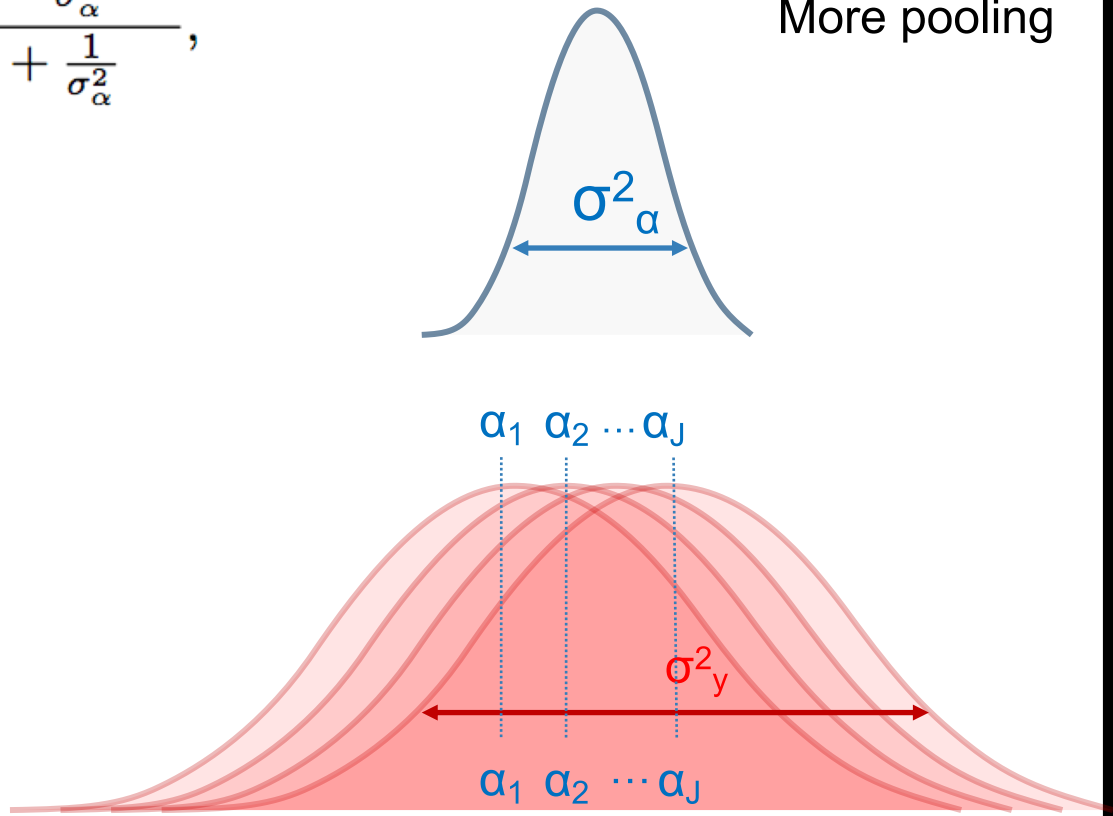
Overall estimate more reliable

Less pooling

More pooling

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





Relation to the repeated measure model.

- WLG, assume we have J groups with K individuals each
- We re write the model as

$y_i = \underline{\alpha}_{j[i]} + \varepsilon_i\ ,\ i = 1,\dots,J*K$
$\underline{\varepsilon}_i \sim N(0,\ \sigma^2_y)$
$\underline{\alpha}_j \sim N(\mu_\alpha,\sigma^2_\alpha)\ ,\ j= 1,\dots,J$

$\Rightarrow$

$y_{jk} = \mu+\underline{\alpha}_j + \underline{\varepsilon}_{jk},\ j= 1,\dots,J;\ k=1,\dots,K$
$\underline{\varepsilon}_{jk}\sim N(0,\sigma^2_y)$
$\underline{\alpha}_j \sim N(0,\sigma^2_\alpha)$

- Where $\underline{\alpha}_j$ and $\underline{\varepsilon}_{jk}$ are independent of one another.
- Then

$$E(y_{jk})= E(\mu+ \underline{\alpha}_j + \underline{\varepsilon}_{jk})= \mu$$

$$Var(y_{jk})= E[(y_{jk}- \mu)^2]= E[(\underline{\alpha}_j + \underline{\varepsilon}_{jk})^2]= \sigma^2_y + \sigma^2_\alpha$$

$$Cov(y_{jk},y_{jm})= E[[(\underline{\alpha}_j + \underline{\varepsilon}_{jk})(\underline{\alpha}_j + \underline{\varepsilon}_{jm})]= \sigma^2_\alpha$$

$$Cov(y_{jk},y_{lm})= E[[(\underline{\alpha}_j + \underline{\varepsilon}_{jk})(\underline{\alpha}_l + \underline{\varepsilon}_{lm})]= 0$$

Random intercept is equicorrelation model

- Since
  - $\mathrm{Var}(y_{jk}) = \sigma^2_y + \sigma^2_\alpha$
  - $\mathrm{Cov}(y_{jk}, y_{jm}) = \sigma^2_\alpha$
  - $\mathrm{Cov}(y_{jk}, y_{lm}) = 0$

Partial Pooling

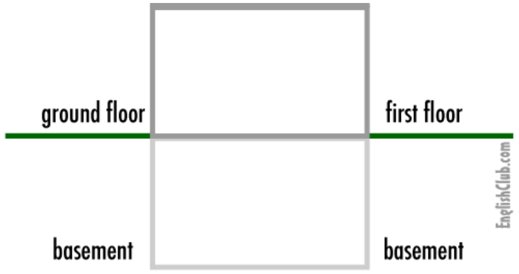
$$y_{jk} = \mu + \underline{\alpha_j} + \underline{\varepsilon_{jk}}$$
$$\underline{\varepsilon_i} \sim N(0, \sigma^2_y)$$
$$\underline{\alpha_j} \sim N(0, \sigma^2_\alpha)$$

- Where $\rho=\sigma^2_\alpha/(\sigma^2_y+\sigma^2_\alpha)$ is called the inter-class correlation coefficient.
  - Describes the proportion of the total variance due to within-cluster variance.

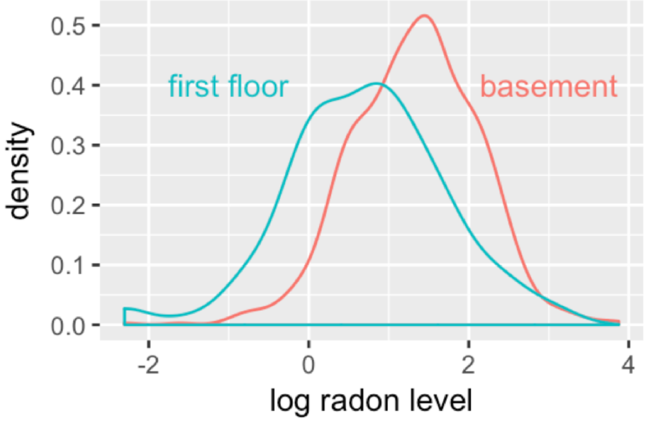
$$\mathbf{V}_j = \begin{bmatrix} \sigma^2_\alpha + \sigma^2_y & \sigma^2_\alpha & \cdots & \sigma^2_\alpha \\ \sigma^2_\alpha & \sigma^2_\alpha + \sigma^2_y & \cdots & \sigma^2_\alpha \\ \vdots & & \ddots & \\ \sigma^2_\alpha & \sigma^2_\alpha & \cdots & \sigma^2_\alpha + \sigma^2_y \end{bmatrix} = \sigma^2_\alpha + \sigma^2_y \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & & \ddots & \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

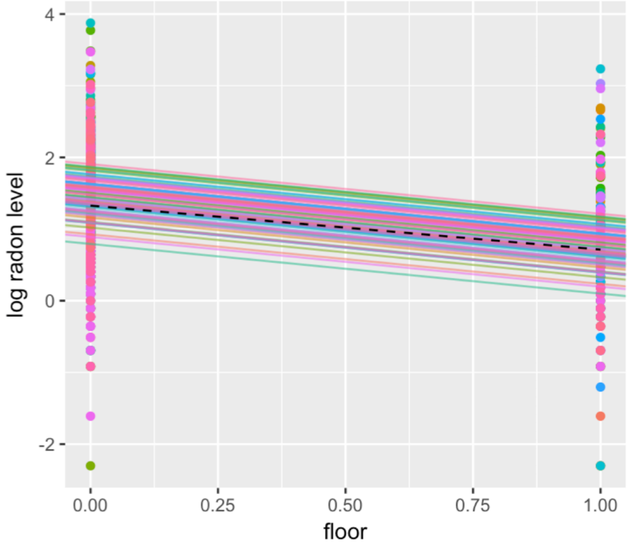
Random Intercept + a slope

- Measurements were taken in the lowest living area of each house,
  - 0: basement
  - 1: first floor coded as 1.



- Complete Pooling (dotted black line)
  - $y_i = \alpha + \beta x_i + \underline{\varepsilon}_i$
  - $\underline{\varepsilon}_i \sim N(0, \sigma^2_y)$
- No Pooling (colored lines)
  - $y_i = \alpha_{j[i]} + \beta x_i + \underline{\varepsilon}_i$
  - $\underline{\varepsilon}_i \sim N(0, \sigma^2_y)$





Adding a predictor floor

Complete Pooling

```
lm(formula = y ~ x)
            coef.est coef.se
(Intercept)  1.33     0.03
x           -0.61     0.07
---
n = 919, k = 2
residual sd = 0.82, R-Squared = 0.07
```

No pooling

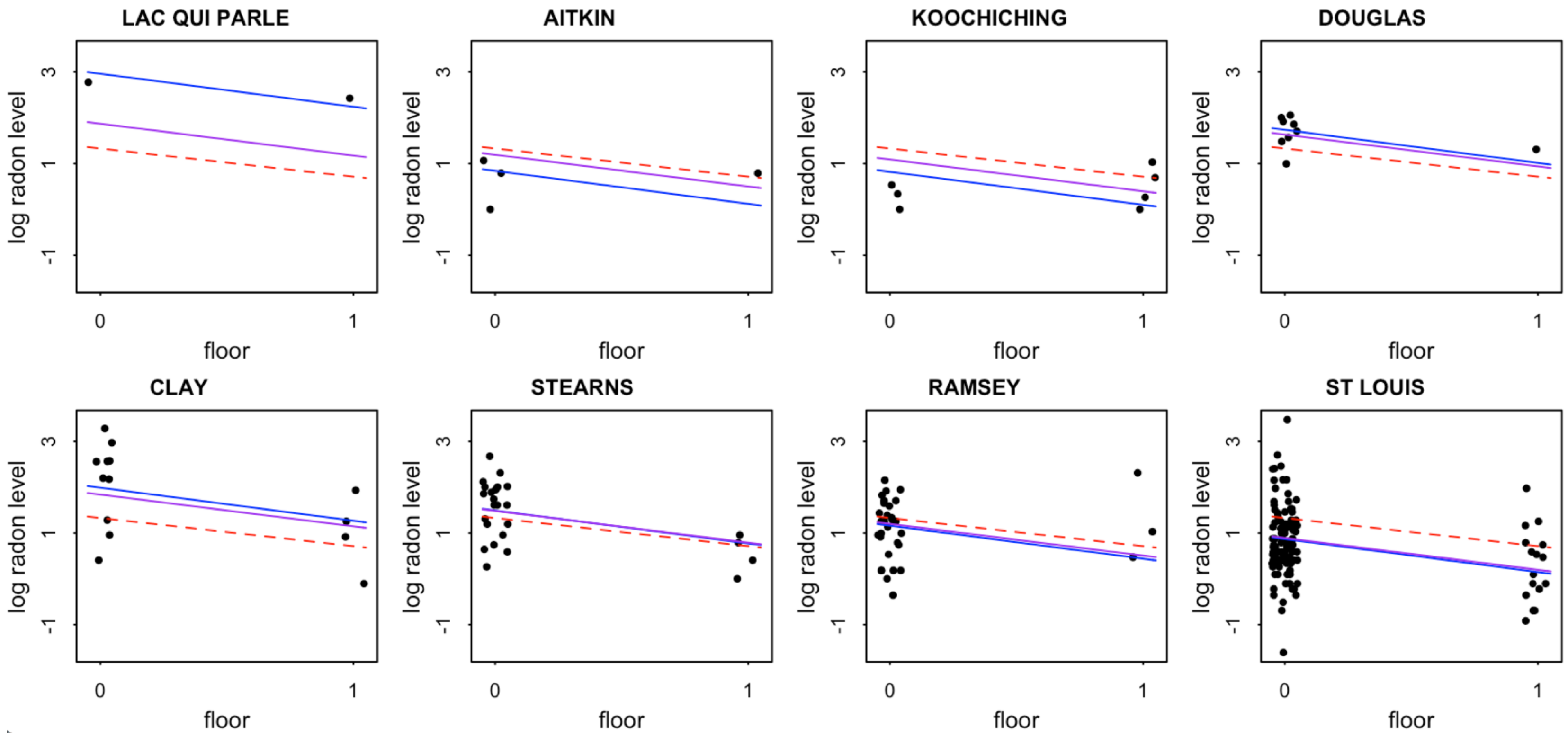
```
lm(formula = y ~ x + factor(county) - 1)
                 coef.est coef.se
x                -0.72     0.07
factor(county)1   0.84     0.38
factor(county)2   0.87     0.10
....
factor(county)84  1.65     0.21
factor(county)85  1.19     0.53
---
n = 919, k = 86
residual sd = 0.76, R-Squared = 0.77
```

Partial Pooling

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

Floor as the predictor

- Red complete pooling, Blue no pooling, purple partial pooling
- The more points in county, closer the Blue and Purple



The models

`lm(y~x+county)`

No Pooling

$y_i = \alpha_{j[i]} + \beta x_i + \underline{\varepsilon}_i$

$\underline{E}_i \sim N(0,\ \sigma^2_y)$

$\alpha_1, \alpha_2, \ldots, \alpha_J$

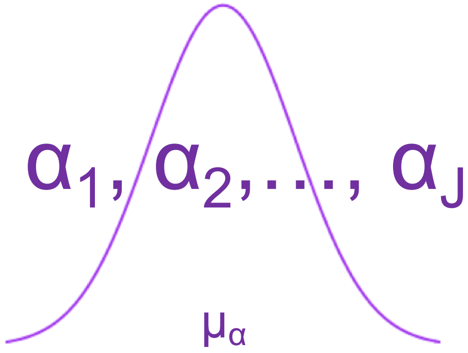
`lmer(y~x+(1|county))`

Partial Pooling

$y_i = \underline{\alpha}_{j[i]} + \beta x_i + \varepsilon_i$

$\underline{\varepsilon}_i \sim N(0,\ \sigma^2_y)$

$\underline{\alpha}_j \sim N(\mu_\alpha, \sigma^2_\alpha)$



$\alpha_1, \alpha_2, \ldots, \alpha_J$

`lm(y~x)`

Complete Pooling

$y_i = \alpha + \beta x_i + \underline{\varepsilon}_i$

$\underline{\varepsilon}i \sim N(0,\ \sigma^2_y)$

α

Interpretation

- Average line for all the counties

  y = 1.46-0.69 x

- Interpretation?
- $\sigma^2_\alpha : \sigma^2_y = 0.33^2 : 0.76^2 = 1 : 5$

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

- $\cong (n_j/5)(1/\sigma^2_\alpha)$ (county level estimate) + $(1/\sigma^2_\alpha)$(overall estimate)

Partial Pooling

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

Adding a predictor floor

Complete Pooling

```
lm(formula = y ~ x)
            coef.est coef.se
(Intercept)  1.33     0.03
x           -0.61     0.07
---
n = 919, k = 2
residual sd = 0.82, R-Squared = 0.07
```

No pooling

```
lm(formula = y ~ x + factor(county) - 1)
                 coef.est coef.se
x                -0.72     0.07
factor(county)1   0.84     0.38
factor(county)2   0.87     0.10
....
factor(county)84  1.65     0.21
factor(county)85  1.19     0.53
---
n = 919, k = 86
residual sd = 0.76, R-Squared = 0.77
```

Partial Pooling

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

Group Level Predictor

Random Intercept but not random slope

`lm(y~x+county)`

No Pooling

$$y_i = \alpha_{j[i]} + \beta x_i + \underline{\varepsilon}_i$$

$$\underline{\varepsilon}_i \sim N(0, \sigma^2_y)$$

$\alpha_1, \alpha_2, \ldots, \alpha_J$

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

Partial Pooling

$$y_i = \underline{\alpha}_{j[i]} + \beta x_i + \varepsilon_i$$

$$\underline{\varepsilon}_i \sim N(0, \sigma^2_y)$$

$$\underline{\alpha}_j \sim N(\mu_\alpha, \sigma^2_\alpha)$$

$\alpha_1, \alpha_2, \ldots, \alpha_J$

μ_α

`lm(y~x)`

Complete Pooling

$$y_i = \alpha + \beta x_i + \underline{\varepsilon}_i$$

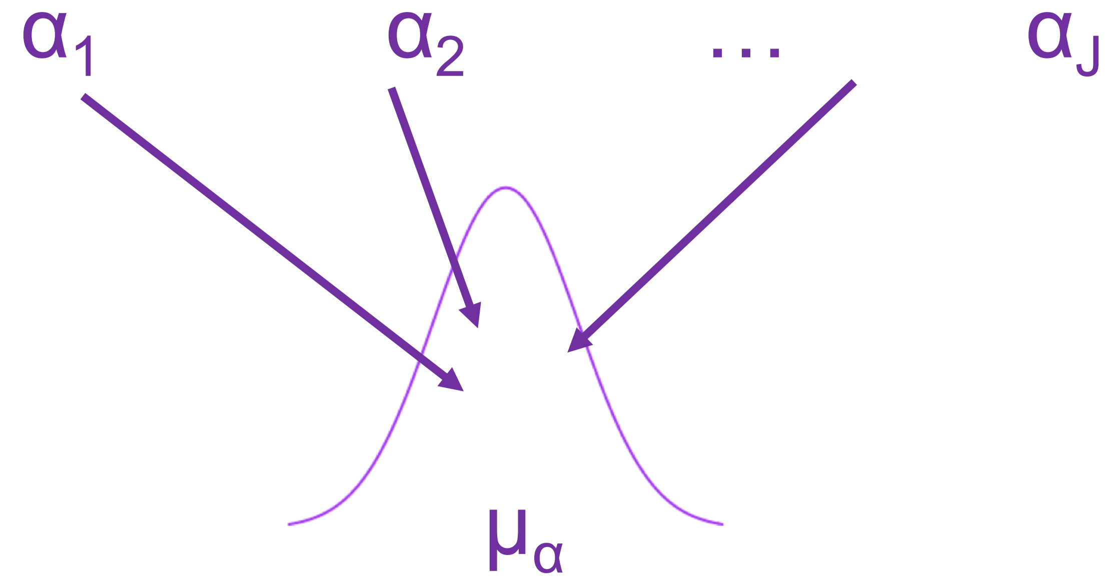
$$\underline{\varepsilon}_i \sim N(0, \sigma^2_y)$$

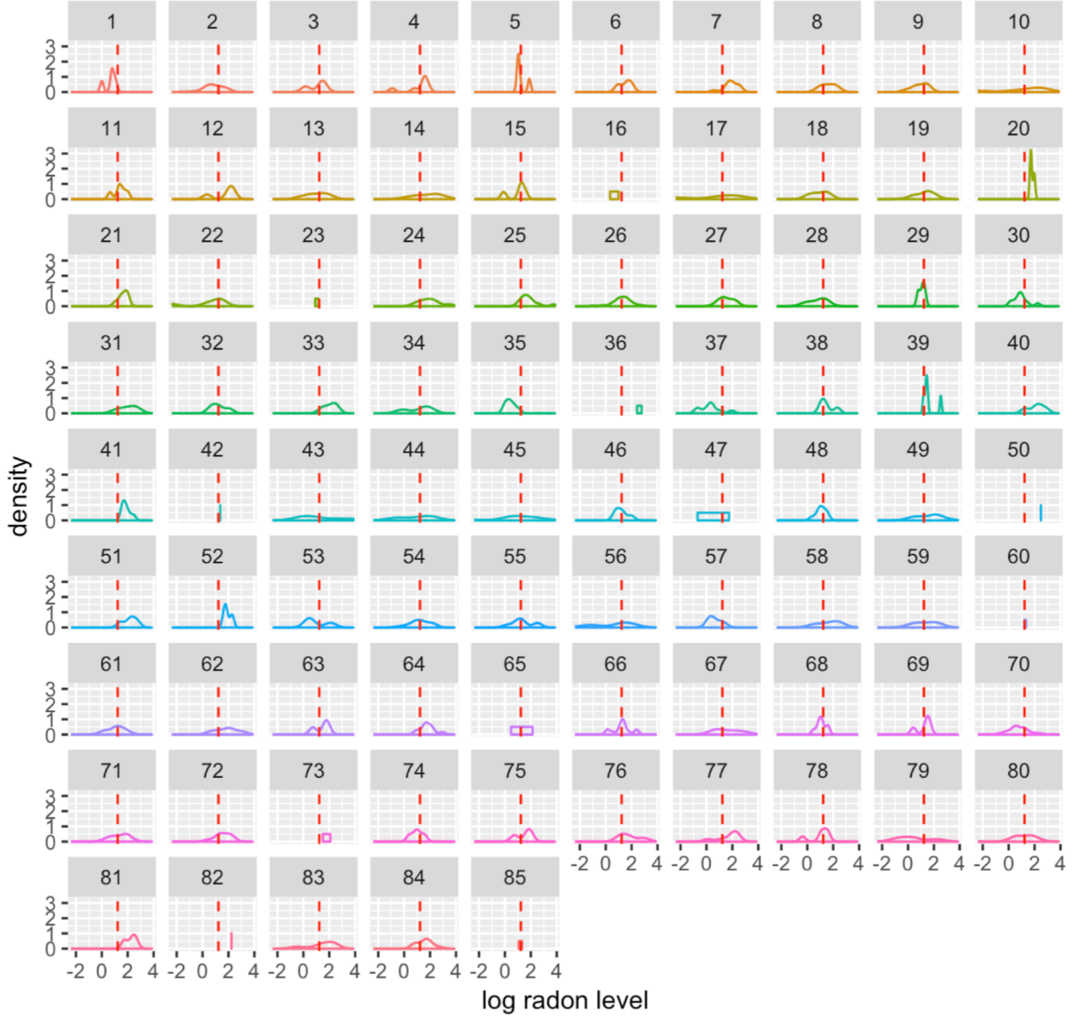
α

Shrinkage toward means

- Estimates are shrunken toward the overall mean

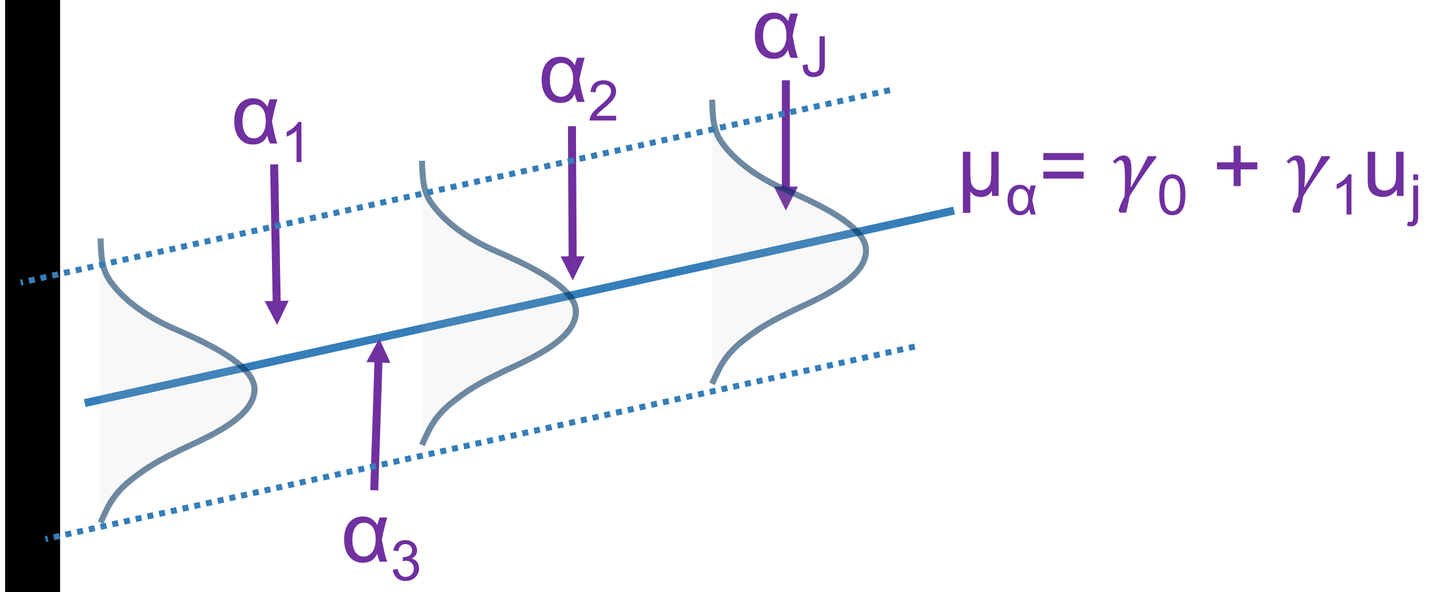
$$\hat{\alpha}_j^{\text{multilevel}} \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{\text{all}}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}},$$





Shrinkage towards conditional means

- Group level predictor makes partial pooling more effective.



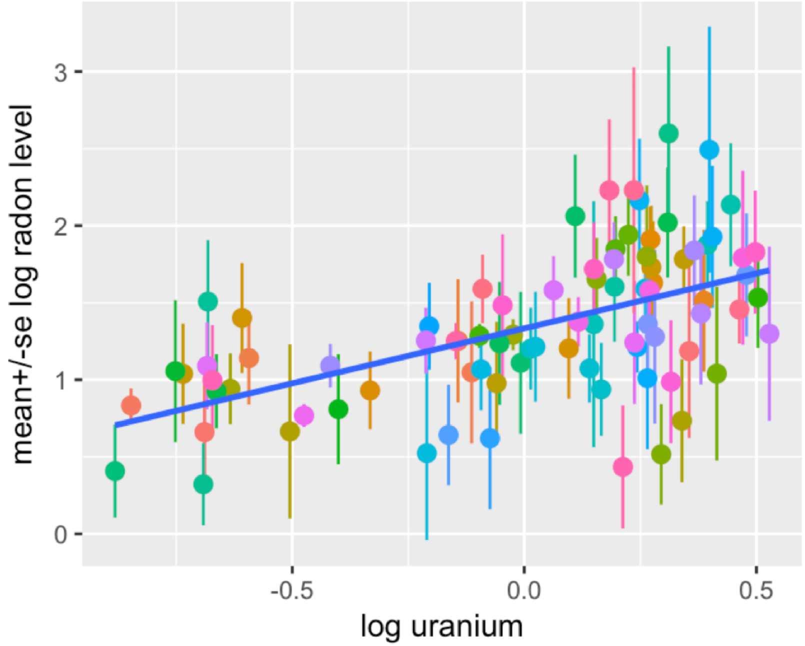
$$\text{estimate of } \alpha_j \approx \frac{\frac{n_j}{\sigma_y^2}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} \cdot (\text{estimate from group } j) + $$
$$+ \frac{\frac{1}{\sigma_\alpha^2}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} \cdot (\text{estimate from regression}).$$

Partial Pooling

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

$\underline{\varepsilon_i} \sim N(0, \sigma^2_y)$

$\underline{\alpha_j} \sim N(\gamma_0 + \gamma_1 u_j, \sigma^2_\alpha)$

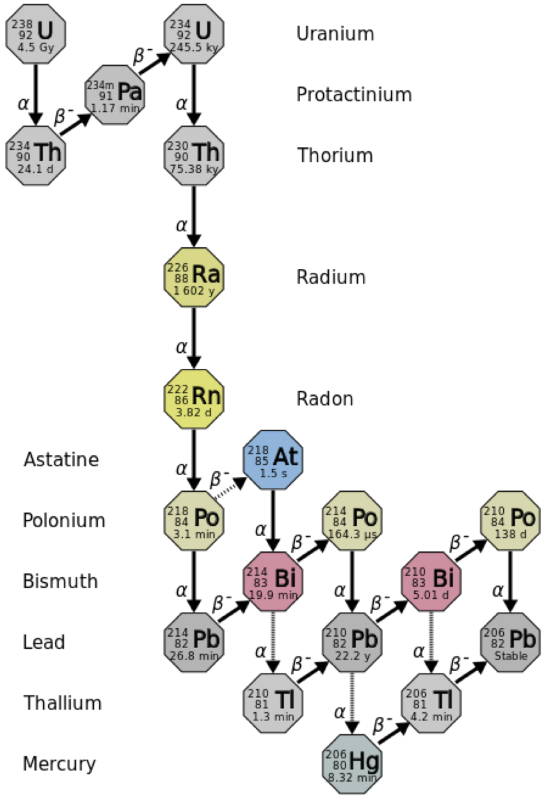


Group-level predictors

- Soil Uranium Measurement.
- Only available at the county level.

Household & County

	county	floor	y	log_uranium
1	1	1	0.7884574	-0.6890476
2	1	0	0.7884574	-0.6890476
3	1	0	1.0647107	-0.6890476
4	1	0	0.0000000	-0.6890476
5	2	0	1.1314021	-0.8473129
6	2	0	0.9162907	-0.8473129



Fit using R

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

```
lmer(formula = y ~ x + u.full + (1 | county))
             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 = 2111.4
deviance = 2122.8
```

Interpretation

- Average line for all the counties

  y = 1.47-0.67 x+0.72u

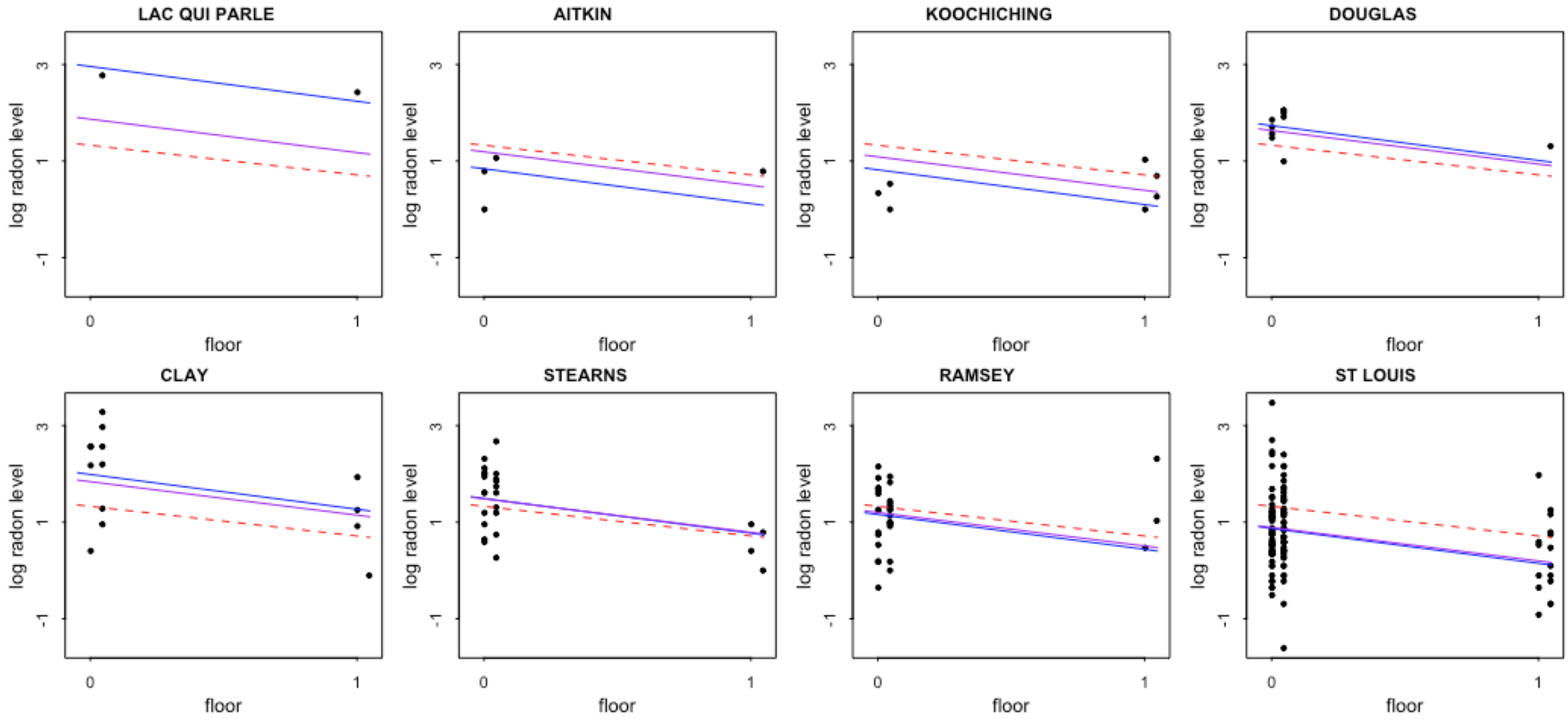
- Interpretation?
- $\sigma^2_\alpha : \sigma^2_y = 0.16^2 : 0.76^2 = 1 : 23$

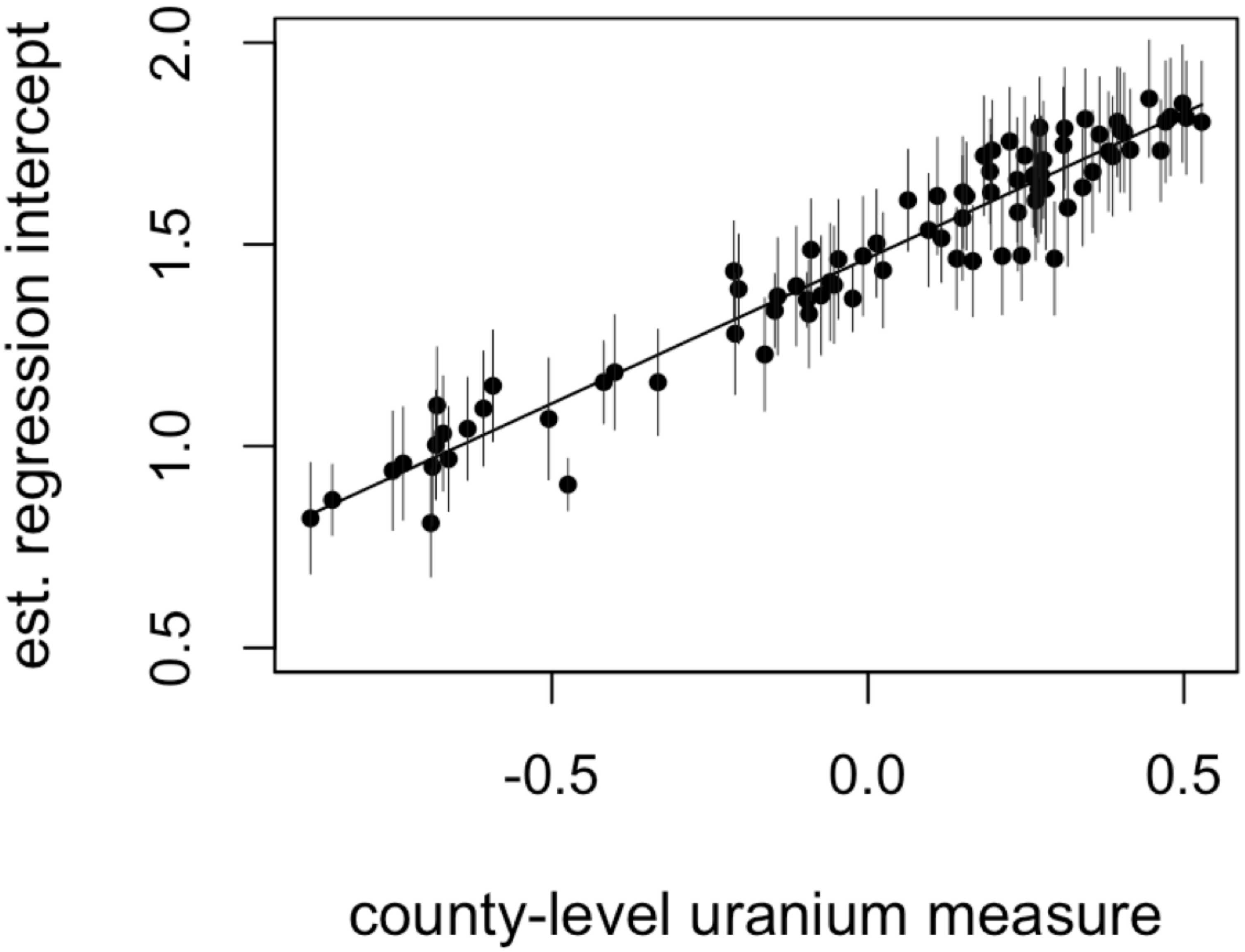
$$\text{estimate of } \alpha_j \approx \frac{\frac{n_j}{\sigma_y^2}}{\frac{n_j}{\sigma_y^2}+\frac{1}{\sigma_\alpha^2}} \cdot (\text{estimate from group } j) + $$
$$+ \frac{\frac{1}{\sigma_\alpha^2}}{\frac{n_j}{\sigma_y^2}+\frac{1}{\sigma_\alpha^2}} \cdot (\text{estimate from regression}).$$

- $\cong (n_j/23)(1/\sigma^2_\alpha)$ (county level estimate j)
  \+ $(1/\sigma^2_\alpha)$(estimate from regression)

Partial Pooling

```
lmer(formula = y ~ x + u.full + (1 |
county))
              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 = 2111.4
deviance = 2122.8
```


LAC QUI PARLE
AITKIN
KOOCHICHING
DOUGLAS
CLAY
STEARNS
RAMSEY
ST LOUIS
log radon level
floor
-1
1
3
0
1


est. regression intercept
county-level uranium measure
2.0
1.5
1.0
0.5
-0.5
0.0
0.5

Different ways to express the model

5 ways to write the varying intercept model

1. Allowing regression coefficients to vary across groups

$$y_i \sim N(\alpha_{j[i]} + \beta x_i, \sigma^2{}_y), \quad \alpha_j = \gamma_0 + \gamma_1 u_j + \eta_j, \quad \eta_j \sim N(0, \sigma^2{}_\alpha)$$

2. Combining separate local regressions

Within county j: $y_i \sim N(\alpha_j + \beta x_i , \sigma^2{}_y)$, for $i = 1, \ldots, n_j$,
$\alpha_j \sim N(\gamma_0 + \gamma_1 u_j , \sigma^2{}_\alpha)$ for $j = 1, \ldots J$

3. Modeling the coefficients of a large regression model

$y_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \sigma^2{}_y)$, for $i = 1, \ldots, n$,
$\beta_j \sim N(\mu_\alpha, \sigma^2{}_\alpha)$ for $j = 4, \ldots J+3$

Note: X is n by J+3 columns
first 3 columns in X are intercept, floor and uranium

4. Regression with multiple error terms

$$y_i \sim N(\mathbf{X}_i\boldsymbol{\beta} + \eta_{j[i]}, \sigma^2{}_y), \eta_j \sim N(0, \sigma^2{}_\alpha)$$

5. Large regression with correlated errors:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{V})$$

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{V_J} \end{bmatrix}$$

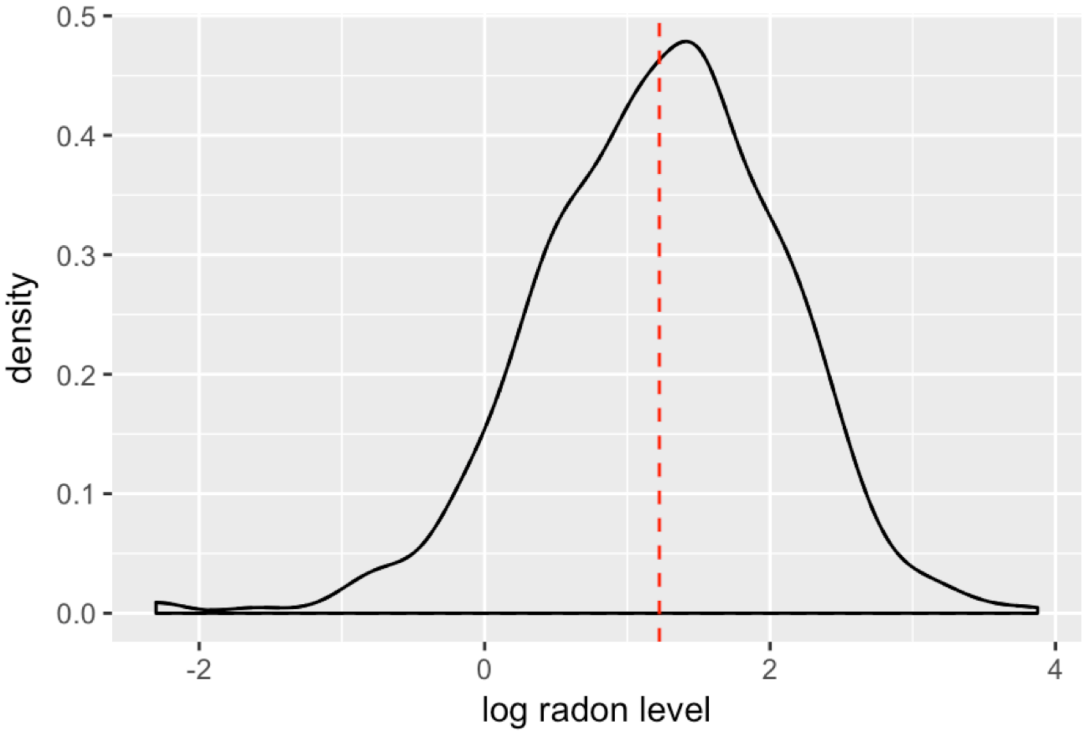
Fixed? Random? Mixed?

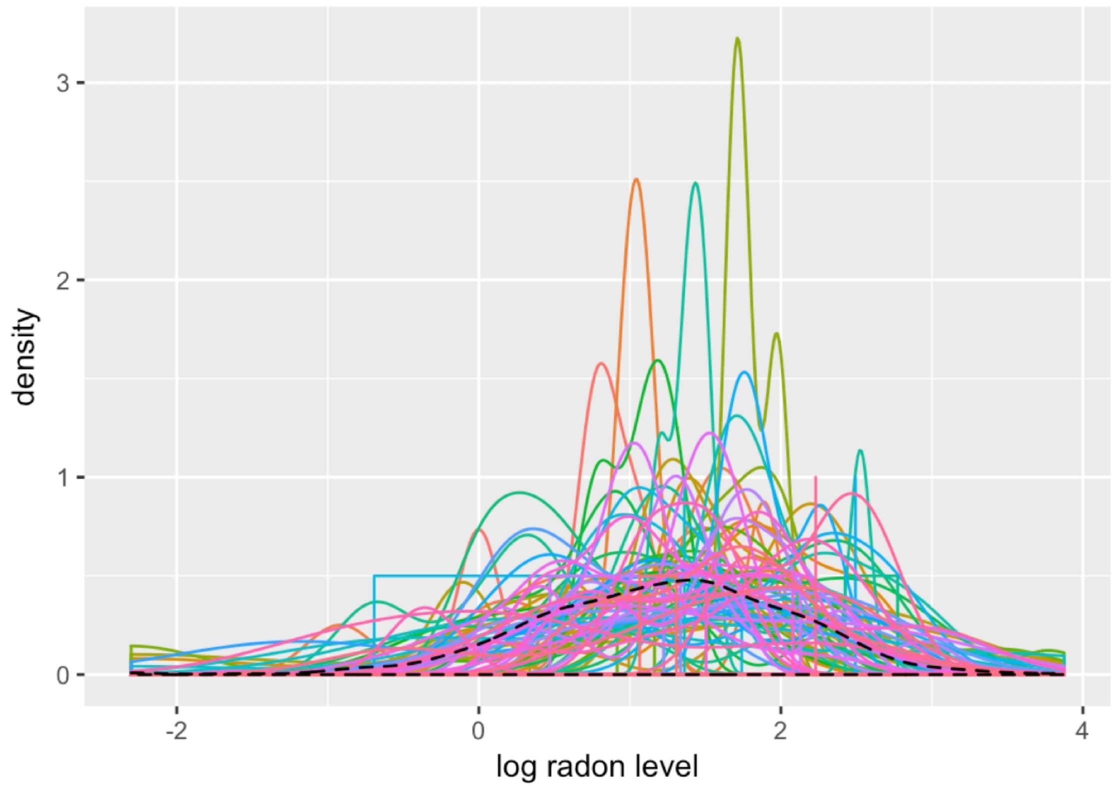
Fixed and random effects

- The varying coefficients in a multilevel model are sometimes called random effects.
- The term fixed effects is used in contrast to random effects.
- A model with both fixed and random effects are called the mixed effect models. (LMM, GLMM, etc)
- There are many confusing suggestions on when to use which:
  - “fixed effects are appropriate if group-level coefficients are of interest, and random effects are appropriate if interest lies in the underlying population”
  - “fixed effects when the groups in the data represent all possible groups, and random effects when the population includes groups not in the data. “

Data Generation

Data generating process





How would you generate a fake radon data?

Without group level predictor

- Write a code to simulate a fake radon data assuming complete pooling

```
# parameters
ncounty <- 85;   nhouse    <- 919
mu_a    <- 1.46; b         <- -0.69
sigma_y <- 0.76;
# data
x       <- rbinom( nhouse,1,0.5)
county  <- sample(1:ncounty, nhouse, TRUE)
# generated fake outcome
ycp     <- rnorm (nhouse, mu_a + b*x , sigma_y)
```

- How would you modify it to accommodate the partial pooling aspect?

```
sigma_a <- 0.33
ac      <- rnorm( ncounty, 0, sigma_a )
ypp     <- rnorm( nhouse, mu_a+ ac[county]  + b*x , sigma_y )
```

How would you generate a fake radon data?

With group level predictor

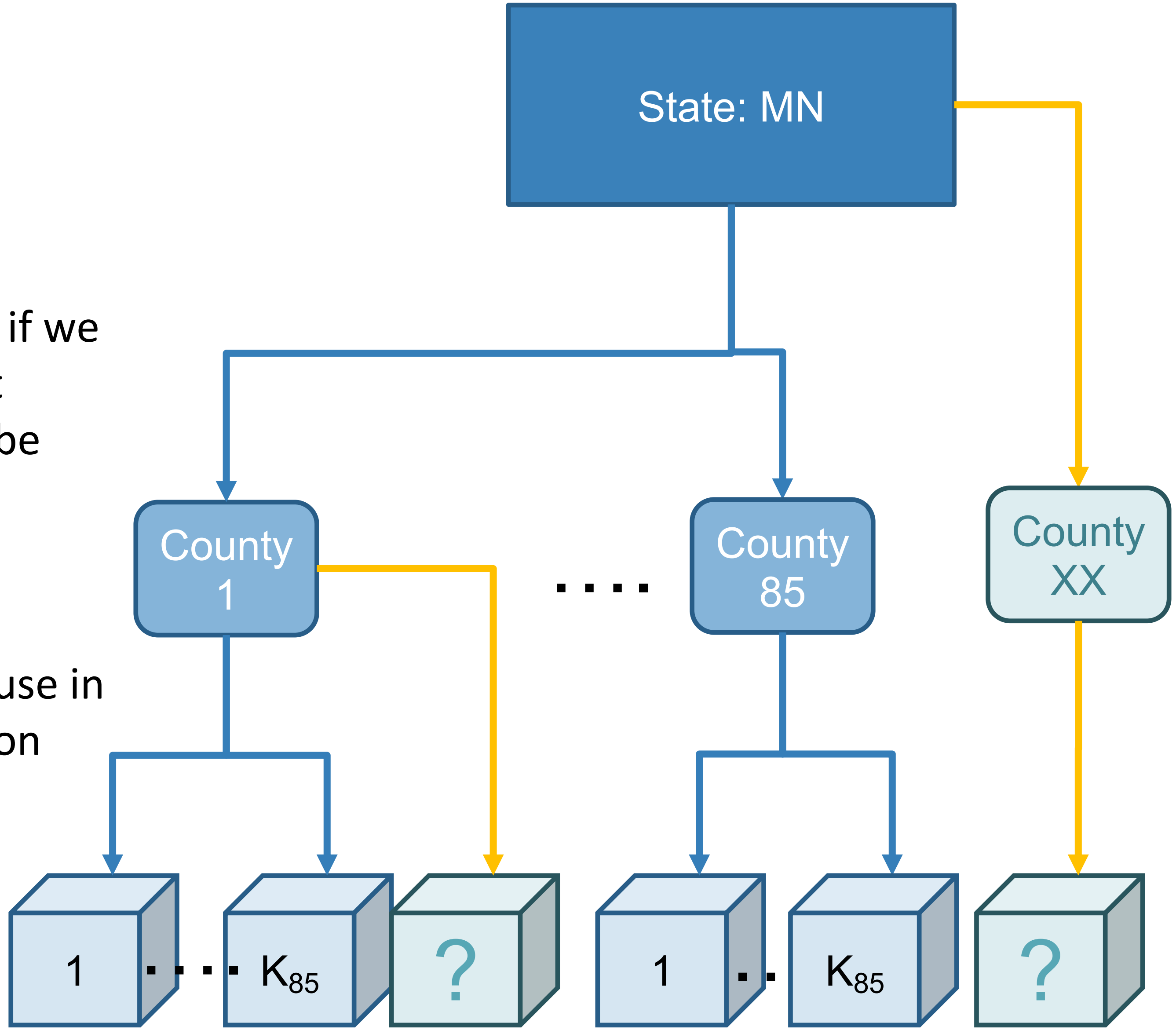
- Write a code to simulate a fake radon data with group level predictor

```
# parameters
ncounty <- 85;   nhouse    <- 919
mu_a    <- 1.46; b         <- -0.69;
sigma_y <- 0.76;

# data
x       <- rbinom( nhouse,1,0.5)
county  <- sample(1:ncounty, nhouse, TRUE)

# generated fake outcome
sigma_a <- 0.16; g <- 0.72
u       <- runif(ncounty,-1,1)
ac      <- rnorm(ncounty, u*g, sigma_a)
ypp     <- rnorm( nhouse, mu_a+ ac[county]  + b*x , sigma_y )
```

Prediction

Prediction:

- A new observation in
- an existing group.
  - For a county that we estimated, if we were to get another house what would the radon measurement be like.
- a new group.
  - For a county that we did not measure, if we were to get a house in that county what would the radon measurement be like?



New observation in an existing group

- New house in county 26 measured on the 1st floor
- $\tilde{y}_i \sim N(\hat{\alpha}_{26}+\hat{\beta}x+\hat{\gamma}_1 u_{26}, \hat{\sigma}^2_y)$

Partial Pooling

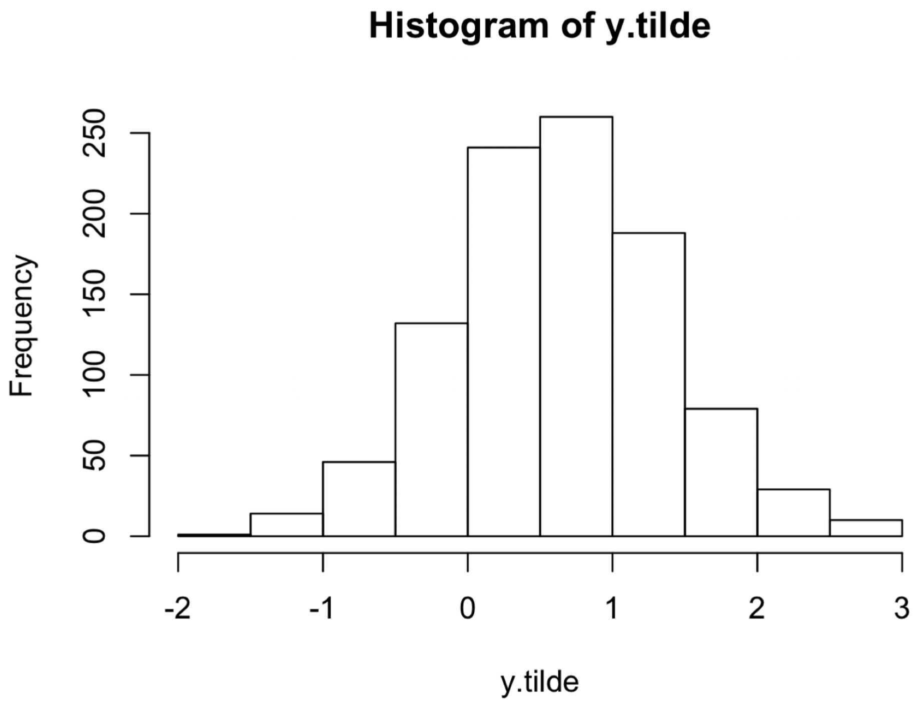
$$\hat{y}_i = \hat{\alpha}_{j[i]} + \hat{\beta}x_i + \hat{\varepsilon}_i$$
$$\hat{\varepsilon}_i \sim N(0, \hat{\sigma}^2_y)$$
$$\hat{\alpha}_j \sim N(\hat{\gamma}_0 + \hat{\gamma}_1 u_j, \hat{\sigma}^2_\alpha)$$

```
> coef.hat <- coef(M2)$county[26,]; coef.hat
   (Intercept)          x    u.full
26    1.432151 -0.6682448 0.7202676
> x.tilde <- 1
> sigma.y.hat <- sigma.hat(M2)$sigma$data; sigma.y.hat
[1] 0.7584393
> y.tilde <- rnorm (n.sims, coef.hat %*% c(1, x.tilde,
u[26]), sigma.y.hat)
```

```
lmer(formula = y ~ x + u.full + (1
| county))
            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
```

New observation in an existing group

- Predictive uncertainty can be assessed using simulation.
  - Histogram
  - percentiles of the sample



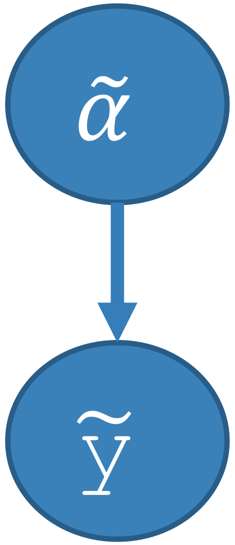
- We are ignoring uncertainties in the estimated parameters α,β, and σ. But these can be added back using `sim`.

```
> quantile (y.tilde, c(.25, .5, .75))
      25%       50%       75%
0.1655713 0.6155451 1.1423393
```

A new observation in a new group.

- For a county that is not in the data we need to
  - first predict the county level intercept.
  - Then predict the radon measurement in that county.

- When there is a uranium measurement in the county we can use that information. But if otherwise we need to supply something.



Partial Pooling

$$\hat{y}_i = \hat{\alpha}_{j[i]} + \hat{\beta} x_i + \hat{\varepsilon}_i$$
$$\hat{\varepsilon}_i \sim N(0, \hat{\sigma}^2_y)$$
$$\hat{\alpha}_j \sim N(\hat{\gamma}_0 + \hat{\gamma}_1 u_j, \hat{\sigma}^2_\alpha)$$

A new observation in a new group.

- For a house in new county with average uranium level

$$\tilde{\alpha} \sim N(\hat{\gamma}_0 + \hat{\gamma}_1 u, \hat{\sigma}^2_\alpha)$$
$$\tilde{y}_i \sim N(\tilde{\alpha} + \hat{\beta} x_i, \hat{\sigma}^2_y)$$

Partial Pooling

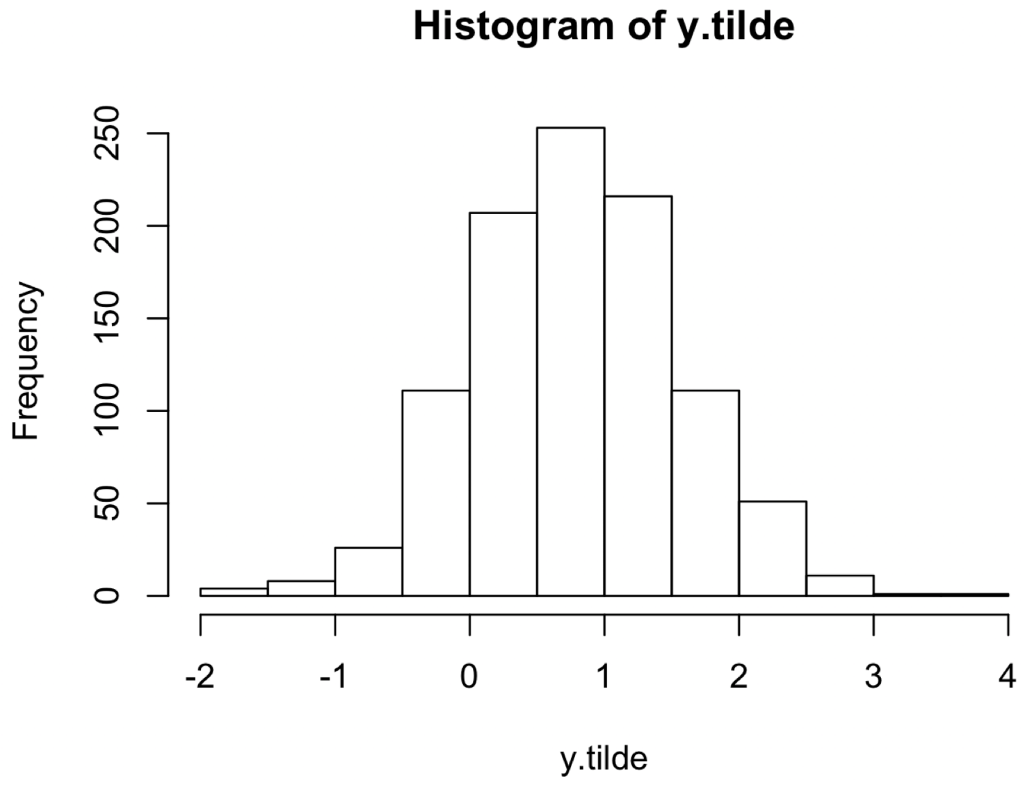
$$\hat{y}_i = \hat{\alpha}_{j[i]} + \hat{\beta} x_i + \hat{\varepsilon}_i$$
$$\hat{\varepsilon}_i \sim N(0, \hat{\sigma}^2_y)$$
$$\hat{\alpha}_j \sim N(\hat{\gamma}_0 + \hat{\gamma}_1 u_j, \hat{\sigma}^2_\alpha)$$

```
> u.tilde <- mean (u)
> g.0.hat <- fixef(M2)["(Intercept)"]
> g.1.hat <- fixef(M2)["u.full"]
> sigma.a.hat <- sigma.hat(M2)$sigma$county
>
> b.hat <- fixef(M2)[2]
> a.tilde <- rnorm (n.sims, g.0.hat + g.1.hat*u.tilde, sigma.a.hat)
> y.tilde <- rnorm (n.sims, a.tilde + b.hat*x.tilde, sigma.y.hat)
```

A new observation in a new group.

- Predictive uncertainty can be assessed using simulation.
  - Histogram
  - percentiles of the sample

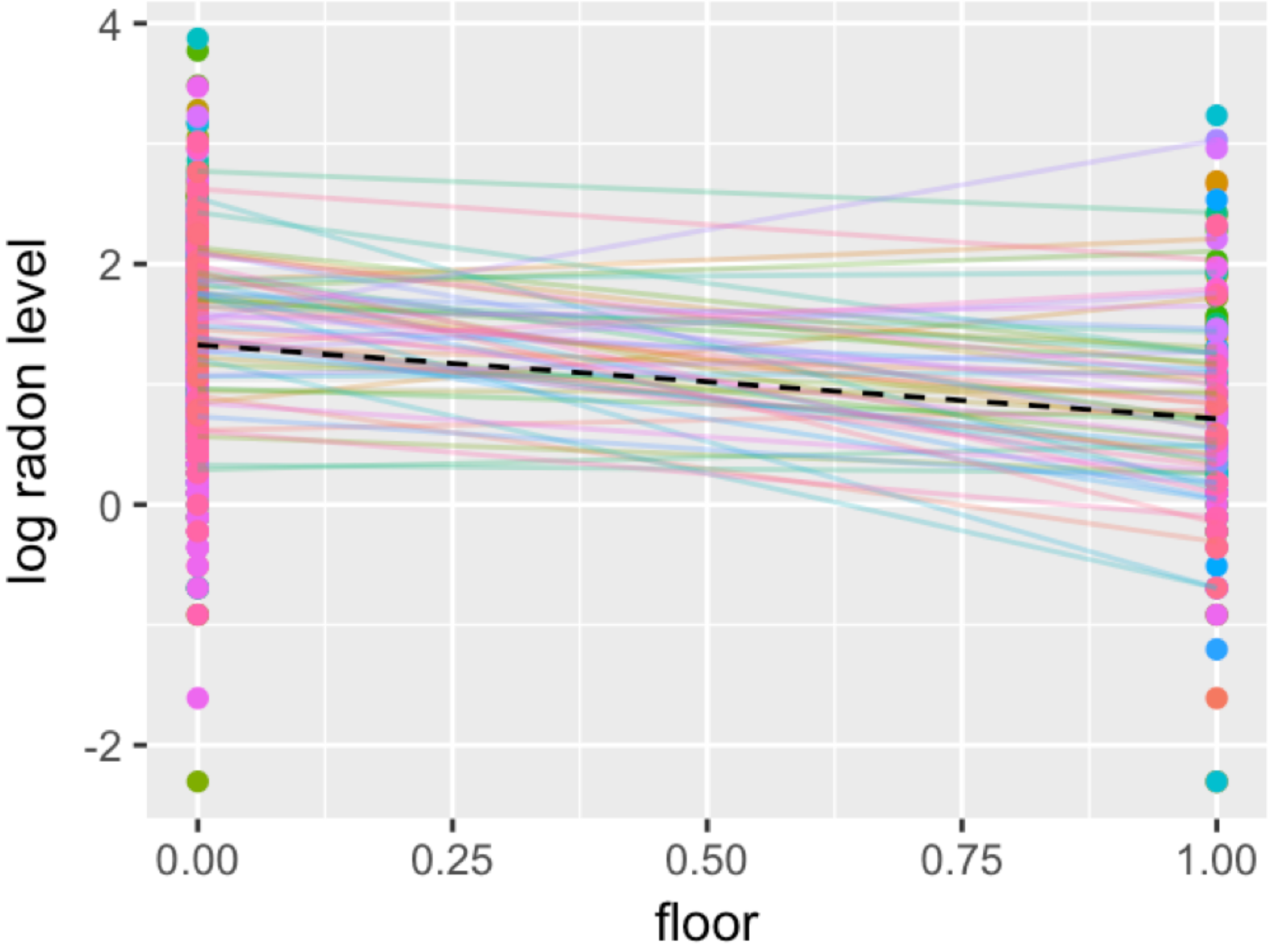
```
> quantile (y.tilde, c(.25, .5, .75))
      25%       50%       75%
0.2768836 0.7712910 1.3036386
```



Random Intercept and Random Slope

Random Intercept + Random slope

- The floor effect is not uniform.
- Complete Pooling
  - $y_i = \alpha + \beta x_i + \underline{\varepsilon}_i$
  - $\underline{\varepsilon}_i \sim N(0, \sigma^2_y)$
- No Pooling
  - $y_i = \alpha_{j[i]} + \beta_{\mathbf{j[i]}} x_i + \underline{\varepsilon}_i$
  - $\underline{\varepsilon}_i \sim N(0, \sigma^2_y)$



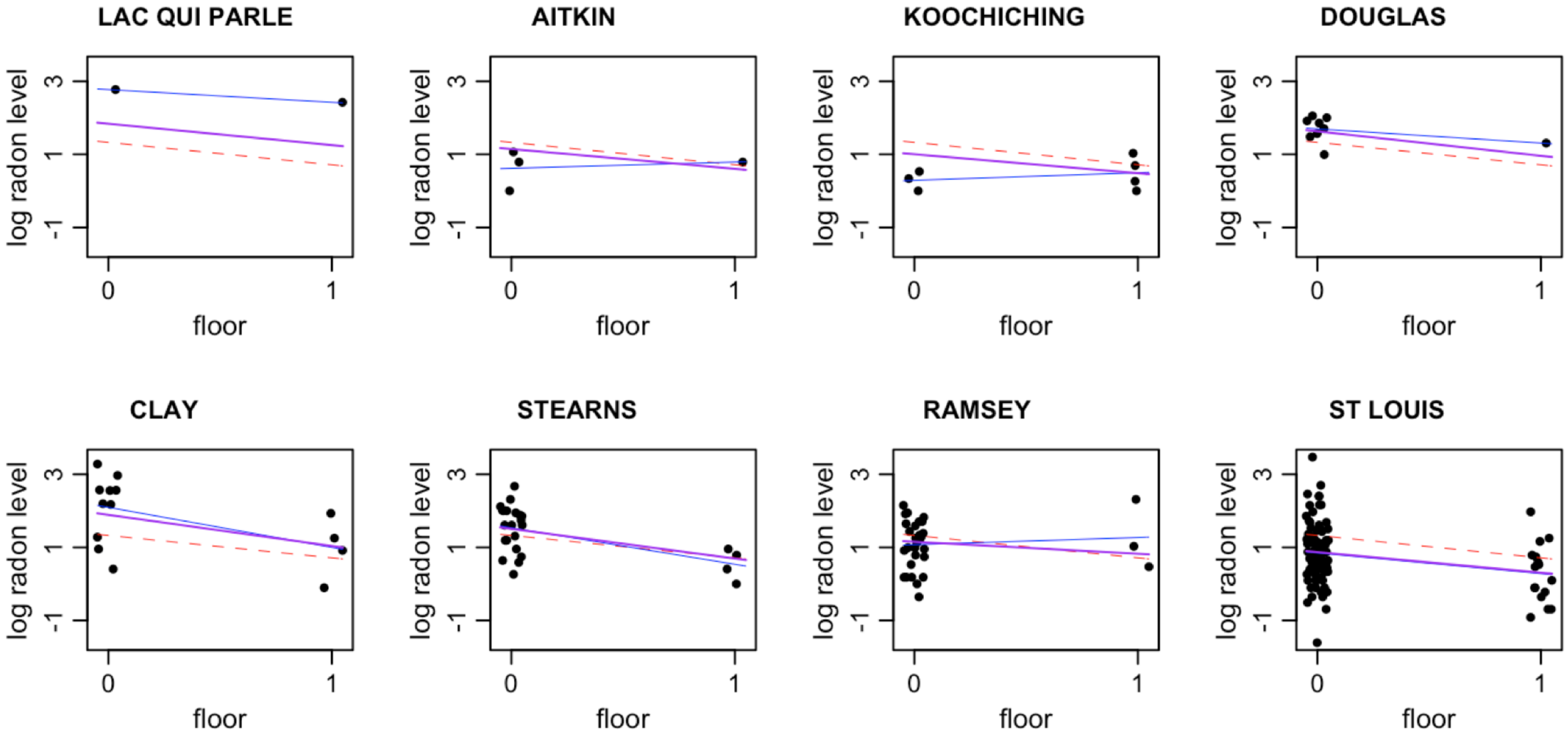
Adding Random Slope

```
lm(formula = y ~ x * factor(county) - 1)
                   coef.est coef.se
x                   0.17     0.86
factor(county)1     0.62     0.43
factor(county)2     0.90     0.11
factor(county)3     1.46     0.75
…
factor(county)85    1.19     0.53
x:factor(county)2  -1.38     0.97
x:factor(county)3  -0.79     1.26
…
x:factor(county)84 -0.99     1.16
---
n = 919, k = 145
residual sd = 0.75, R-Squared = 0.79
```

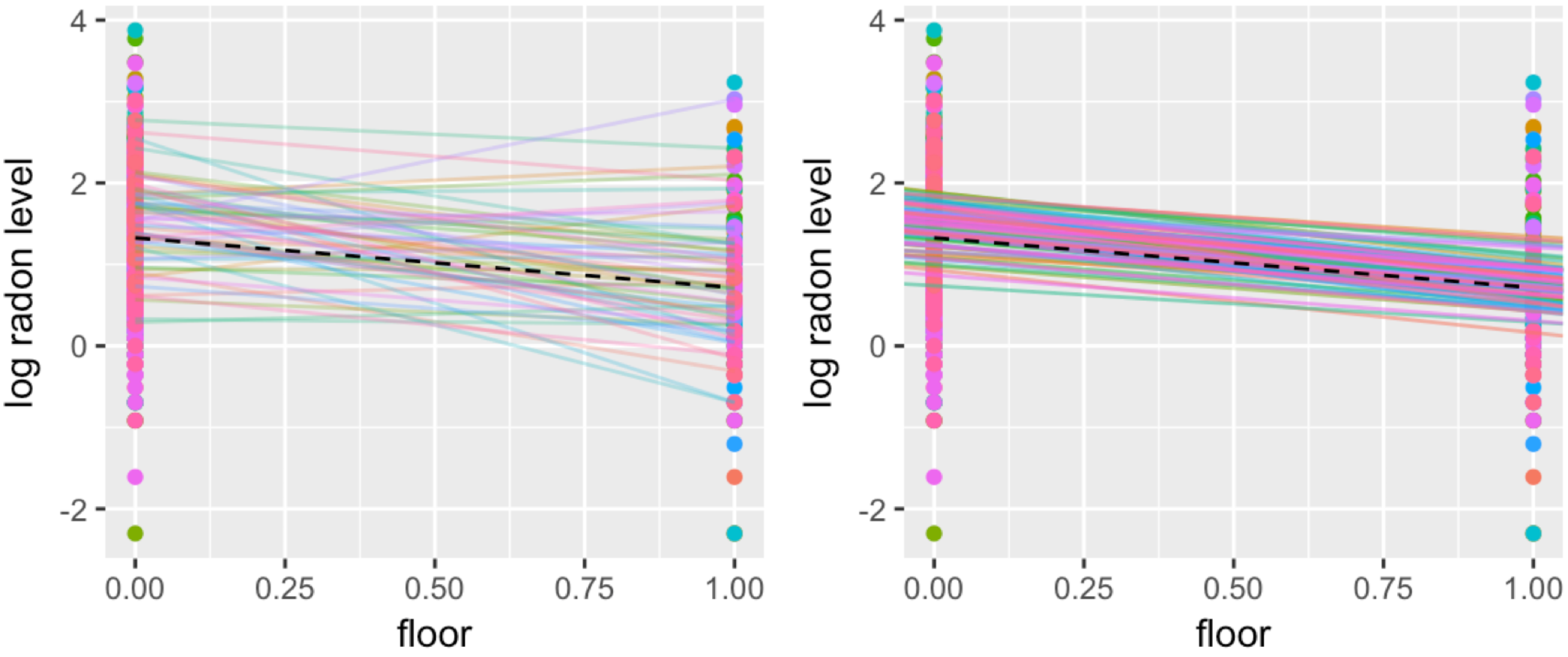
```
lmer(formula = y ~ x + (1 + x | county))
coef.est coef.se
(Intercept)  1.46     0.05
x           -0.68     0.09
Error terms:
Groups   Name        Std.Dev. Corr
county   (Intercept) 0.35
         x           0.34     -0.34
Residual             0.75
---
number of obs: 919, groups: county, 85
AIC = 2180.3, DIC = 2153.9
deviance = 2161.1
```

The no pooled estimate is corrected

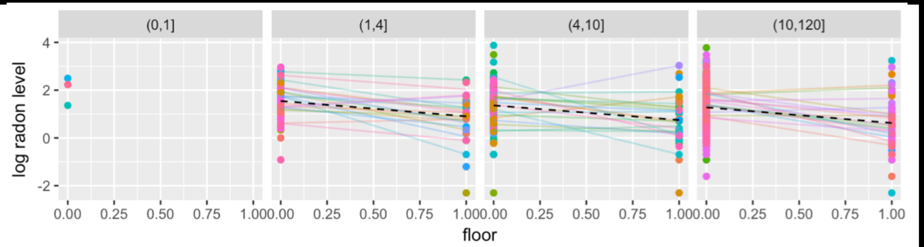
- Red complete pooling, Blue no pooling, purple partial pooling
- The more points in county, closer the Blue and Purple



Overall estimated slopes

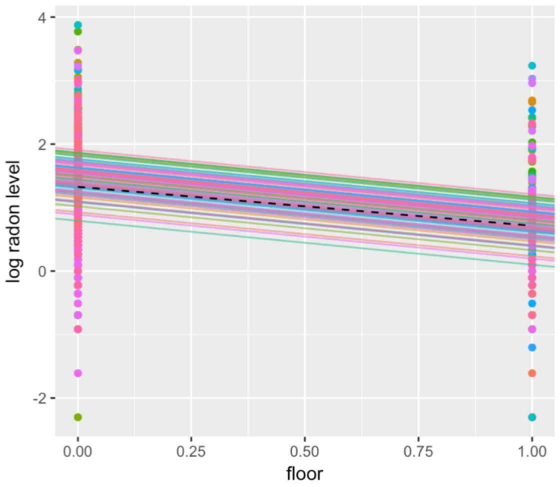


Why is this better?

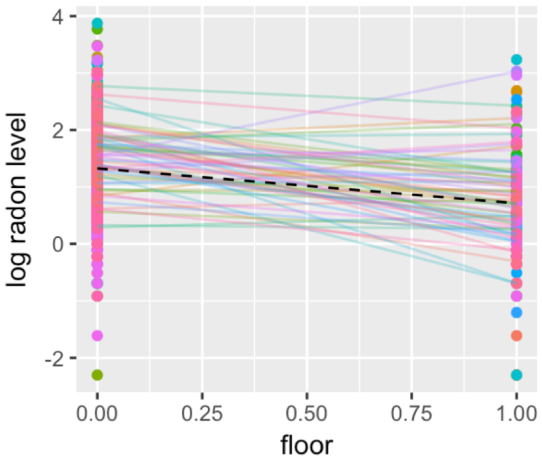


- Sharing information across counties gives you more estimates
  - Out of 85 counties, 25 counties had no slope estimates since they only had 1 observation within the county.
- Letting the data speak so data tells you the model they prefer
  - Result is close to random intercept model.

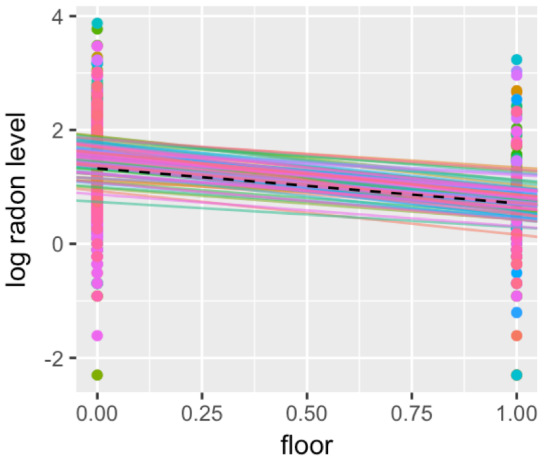
Random intercept
Fixed slope



No Pooling



Random intercept
Random slope



Random **Intercept** and random slope

`lm(y~x*county)`

No Pooling

$$y_i = \alpha_{j[i]} + \beta_{j[i]}\, x_i + \underline{\varepsilon}_i$$

$$\underline{\varepsilon}_i \sim N(0, \sigma^2{}_y)$$

$$\alpha_1, \alpha_2, \ldots, \alpha_J$$

$$\beta_1, \beta_2, \ldots, \beta_J$$

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

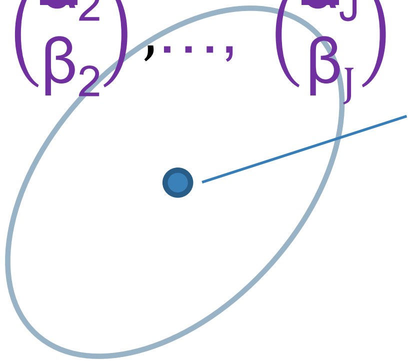
Partial Pooling

$$y_i = \underline{\alpha}_{j[i]} + \beta_{j[i]} x_i + \varepsilon_i$$

$$\underline{\varepsilon}_i \sim N(0, \sigma^2{}_y)$$

$$\begin{pmatrix}\alpha_j \\ \beta_j\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_\alpha \\ \mu_\beta\end{pmatrix}, \begin{pmatrix}\sigma^2{}_\alpha & \rho\sigma^2{}_\alpha\sigma^2{}_\beta \\ \rho\sigma^2{}_\alpha\sigma^2{}_\beta & \sigma^2{}_\beta\end{pmatrix}\right)$$

$$\begin{pmatrix}\alpha_1 \\ \beta_1\end{pmatrix}, \begin{pmatrix}\alpha_2 \\ \beta_2\end{pmatrix}, \ldots, \begin{pmatrix}\alpha_J \\ \beta_J\end{pmatrix}$$



$$\begin{pmatrix}\mu_\alpha \\ \mu_\beta\end{pmatrix}$$

`lm(y~x)`

Complete Pooling

$$y_i = \alpha + \beta x_i + \underline{\varepsilon}_i$$

$$\underline{\varepsilon}_i \sim N(0, \sigma^2{}_y)$$

$$\alpha$$

$$\beta$$

How is this related to the big model?

- WLG, assume we have J groups with K individuals each

$$y_{jk} = b_0 + \alpha_{j[i]} + (b_1 + \beta_{j[i]})t_k + \varepsilon_{jk},\ j= 1,\ldots,J;\ k=1,\ldots,K$$
$$\varepsilon_{jk} \sim N(0,\sigma^2_y)$$
$$\alpha_j \sim N(0,\sigma^2_\alpha)$$
$$\beta_j \sim N(0,\sigma^2_\beta)$$

- Assume α_j, β_j, and ε_{jk} are independent of one another.
- Then

$$E(y_{jk}) = b_0+b_1t_k$$

$$Var(y_{jk}) = \sigma^2_y + t^2_k\sigma^2_\beta+\sigma^2_\alpha$$

$$Cov(y_{jk},y_{jm}) = E[[(\alpha_j + \beta_j t_k + \varepsilon_{jk})(\alpha_j + \beta_j t_m + \varepsilon_{jm})] = \sigma^2_\alpha+t_mt_k\sigma^2_\beta$$

$$Cov(y_{jk},y_{lm}) = E[[(\alpha_j + \beta_j t_k + \varepsilon_{jk})(\alpha_l + \beta_l t_m + \varepsilon_{lm})] = 0$$

Random intercept & slope is distance based model

- Since
  - $\mathrm{Var}(y_{jk}) = \sigma^2_y + t^2_k\sigma^2_{\underline{\beta}} + \sigma^2_\alpha$
  - $\mathrm{Cov}(y_{jk}, y_{jm}) = \sigma^2_\alpha + t_k t_m \sigma^2_{\underline{\beta}}$
  - $\mathrm{Cov}(y_{jk}, y_{lm}) = 0$

Partial Pooling

$$y_{jk} = b_0 + \underline{\alpha}_{j[i]} + (b_1 + \underline{\beta}_{j[i]})t_k + \underline{\varepsilon}_{jk},$$
$$\underline{\varepsilon}_i \sim N(0, \sigma^2_y)$$
$$\underline{\alpha}_j \sim N(0, \sigma^2_\alpha)$$
$$\underline{\beta}_j \sim N(0, \sigma^2_{\underline{\beta}})$$

- Since $\rho_{km}=f(t_k, t_m)$ this is distance based formulation

Adding group level predictor

- We can also add group level predictor for the random slope

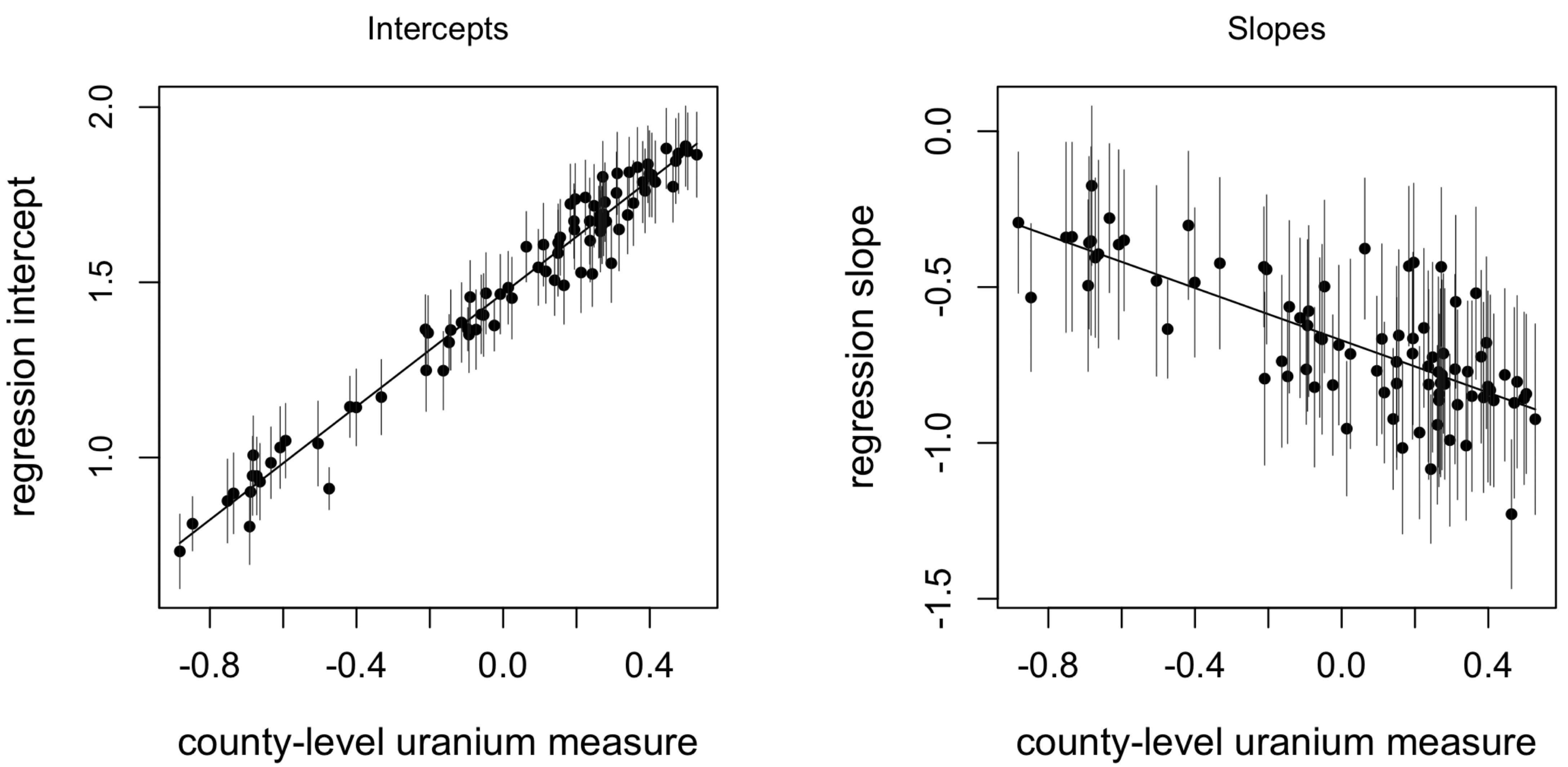
```
lmer(formula = y ~ x + u.full + x:u.full + (1 + x | county))
```

$$y_i = \underline{\alpha}_{j[i]} + \beta x_{j[i]} + \varepsilon_i$$

$$\underline{\varepsilon}_i \sim N(0, \sigma^2_y)$$

$$\begin{pmatrix}\alpha_j\\ \beta_j\end{pmatrix} \sim N\left(\begin{pmatrix}\gamma_0^{\alpha} + \gamma_1^{\alpha} u_j\\ \gamma_0^{\beta} + \gamma_1^{\beta} u_j\end{pmatrix}, \begin{pmatrix}\sigma^2_{\alpha} & \rho\sigma^2_{\alpha}\sigma^2_{\beta}\\ \rho\sigma^2_{\alpha}\sigma^2_{\beta} & \sigma^2_{\beta}\end{pmatrix}\right)$$

```
            coef.est  coef.se
(Intercept)  1.47      0.04
x           -0.67      0.08
u.full       0.81      0.09
x:u.full    -0.42      0.23

Error terms:
 Groups   Name        Std.Dev. Corr
 county   (Intercept) 0.12
          x           0.31     0.41
 Residual             0.75
---
number of obs: 919, groups: county, 85
AIC = 2142.6, DIC = 2101.9
deviance = 2114.2
```



Intercepts
Slopes
regression intercept
regression slope
2.0
1.5
1.0
0.0
-0.5
-1.0
-1.5
-0.8
-0.4
0.0
0.4
county-level uranium measure
county-level uranium measure

Estimation

Multilevel Normal Linear Regression Model

- Likelihood:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{U}\gamma + \epsilon \qquad \epsilon \sim N(\mathbf{0}, \sigma_y^2 \boldsymbol{I})$$

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_J \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_J \end{pmatrix}, \mathbf{U} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2 & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{U}_J \end{pmatrix}, \gamma = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_J \end{pmatrix}, \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_J \end{pmatrix}$$

- Second Level model:

$$\gamma \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\Gamma)$$

$$\boldsymbol{\Sigma}_\Gamma = \begin{pmatrix} \boldsymbol{\Sigma}_\gamma & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\gamma & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Sigma}_\gamma \end{pmatrix}$$

Marginal model MNLR

- If you integrate out the random effects from the conditional model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{U}\gamma + \epsilon$$

- Marginally

$$Var(\mathbf{y}) = \mathbf{U}\boldsymbol{\Sigma}_\Gamma\mathbf{U}^T + \sigma_y^2\mathbf{I}$$

- Which is the same as the “big model” with

$$\mathbf{y} \sim N\left(\mathbf{X}\beta, \mathbf{V} = \mathbf{U}\boldsymbol{\Sigma}_\Gamma\mathbf{U}^T + \sigma_y^2\mathbf{I}\right)$$

- We can use the same technique we used for GLS

Best Linear Unbiased Prediction (BLUP)

- After getting $\widehat{\boldsymbol{\beta}}$ and $\widehat{V}$ we want prediction $\widetilde{\boldsymbol{\gamma}}$ of random effect $\boldsymbol{\gamma}$.
- Because $\boldsymbol{\gamma}$ is random effect we call it *prediction* instead of estimation.
- A prediction $\widetilde{\boldsymbol{\gamma}}$ is called BLUP of $\boldsymbol{\gamma}$ if
  - $E(\widetilde{\boldsymbol{\gamma}}) = 0$ and $E(\mathbf{a}^T \widetilde{\boldsymbol{\gamma}} - \mathbf{a}^T \widetilde{\boldsymbol{\gamma}})^2$ is minimized for all $\mathbf{a}^T$.

Best Linear Unbiased Prediction (BLUP)

- The joint distribution of y and $\boldsymbol{\gamma}$

$$\begin{pmatrix} \mathbf{y} \\ \gamma \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{X}\beta \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{U}\boldsymbol{\Sigma}_\Gamma\mathbf{U}^T + \boldsymbol{\Sigma}_\epsilon & \mathbf{U}\boldsymbol{\Sigma}_\Gamma \\ \boldsymbol{\Sigma}_\Gamma\mathbf{U}^T & \boldsymbol{\Sigma}_\Gamma \end{pmatrix}\right)$$

- BLUP of the random effect is

$$\tilde{\gamma} = \boldsymbol{\Sigma}_\Gamma\mathbf{U}^T\mathbf{V}^{-1}\left[\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\right]\mathbf{y}$$

- Which is equivalent to the posterior mean of $\boldsymbol{\gamma}$ given the data.

$$Var\begin{pmatrix} \tilde{\beta} - \beta \\ \tilde{\gamma} - \gamma \end{pmatrix} \sim \begin{pmatrix} \mathbf{X}^T\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{X} & \mathbf{X}^T\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{U} \\ \mathbf{U}^T\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{X} & \boldsymbol{\Sigma}_\Gamma^{-1} + \mathbf{U}^T\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{U} \end{pmatrix}^{-1}$$

For random intercept model

- For

$$\mathbf{y} \sim N\left(\mathbf{X}\beta, \mathbf{V} = \mathbf{U}\boldsymbol{\Sigma}_{\Gamma}\mathbf{U}^{T} + \sigma_y^2\mathbf{I}\right)$$

- Where

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_{n_1} \\ \vdots \\ \mathbf{1}_{n_J} \end{pmatrix}, \mathbf{U} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_{n_J} \end{pmatrix}, \gamma = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_J \end{pmatrix}, \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_J \end{pmatrix}$$

- Calculate $\boldsymbol{\Sigma}_{\Gamma} = \sigma_\gamma^2\mathbf{I}$

$$\tilde{\gamma} = \boldsymbol{\Sigma}_{\Gamma}\mathbf{U}^{T}\mathbf{V}^{-1}\left[\mathbf{I} - \mathbf{X}(\mathbf{X}^{T}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{T}\mathbf{V}^{-1}\right]\mathbf{y}$$

REML (Residual ML)

- Instead of maximizing

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{U}\gamma + \epsilon$$

- Maximizes $\mathbf{L}\mathbf{y}$ such that $\mathbf{LX}\boldsymbol{\beta}=\mathbf{0}$ then $\mathbf{L}\mathbf{y} = \mathbf{L}(\mathbf{U}\boldsymbol{\gamma}+\boldsymbol{\varepsilon})$ has no $\boldsymbol{\beta}$.
- If we let $\mathbf{L}=\mathbf{I}-\mathbf{P}_{\mathrm{x}}$ where $\mathbf{P}_{\mathrm{x}}$ is the projection onto the $\mathrm{span}(\mathbf{X})$.
- $\mathbf{L}\mathbf{y} = (\mathbf{I}-\mathbf{P}_{\mathrm{x}})\mathbf{y} = \mathbf{y}-\hat{\mathbf{y}}$ hence Residual ML