Basics of Multilevel Models

- Generalize linear regression, allowing intercepts, or slopes, or … to vary by group ID.
- In a regression, include a categorical predictor which may interact with other predictors.

Complete pooling ignores groups,
No pooling fits a model for each group separately.
We want to explore midlle ground between these 2 extremes.

Notation

- Units numbered $i = 1, \ldots, n$. (Not necessarily experimental or observational units, as those may be measured repeatedly.)
- Response is y_i
- Data matrix $\mathbf{X}$ of dimension n by k. The ith row is called $\mathbf{X}_i$, kth column is $\mathbf{X}_{(k-1)}$, as the first column is $\mathbf{X}_{(0)} = \mathbf{1}$.
- Coefficient vector $\boldsymbol{\beta}^{\mathsf{T}} = (\beta_0 \ \cdots \ \beta_k)$ such that $E(\mathbf{y}) = \mathbf{X}\beta$.

Multilevel Extensions

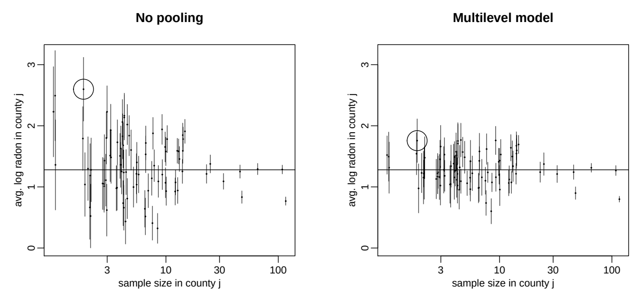
- Groups $j = 1, \ldots, J$, e.g. centers at which our clinical trials are run.
- Possible second level of grouping $k = 1, \ldots, K$ (could be confused with predictor label).
- Indexing $j[i]$ tells us which group the ith obs. belongs to.

More Notation

- Coefficient vector can be separated as in $\boldsymbol{\beta}^{\mathsf{T}} = (\alpha \ \beta)$
  Group level coefficients may be called $\boldsymbol{\gamma}$.
  In R and BUGs code, $\alpha \to$ a, $\beta \to$ b, $\gamma \to$ g
- With multiple predictors, $y_i = \mathbf{X}_i\mathbf{B} + \epsilon_i$ where $\mathbf{B}$ is a matrix of coefficients.
- In addition to $\sigma^2 = \text{Var}(\epsilon_i)$, we have other variances $\sigma^2_\alpha, \ \sigma^2_\beta$.
- Some predictors are measured only at the group level.
  $\alpha_j \sim N(U_j\gamma, \sigma^2_\alpha)$ where $\mathbf{U}$ has J rows.

Partial Pooling

Radon seeps from soil into people's homes.
A study of radon level was done in 85 Minnesota counties.
Consider the 85 county means (sample sizes vary from 1 to 116).



No pooling means we estimate each mean independently.
Partial pooling uses a weighted average of the overall mean (1.3) and the county mean. As n increases, the overall mean gets less weight.

Multilevel Pooling

For counties with few measurements, it really helps to "borrow" information from other counties. Pooling pulls in extremes and reduces variance.

$$\widehat{a}_j \approx \frac{\frac{n_j}{\sigma_y^2}\overline{y}_{\cdot j} + \frac{1}{\sigma_\alpha^2}\overline{y}_{\cdot\cdot}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} \text{ with variance: } \frac{1}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

The weights applied to the two means are inverse variances (called precisions). $\text{Var}(\overline{y}_{\cdot j}) = \frac{\sigma_y^2}{n_j}$ and $\text{Var}(\overline{y}_{\cdot\cdot}) = \sigma_\alpha^2$

Complete Pooling With Predictors

A predictor of interest is "floor" where 0 means basement, 1 = ground floor. We expect higher radon readings in basements. First use complete pooling:

```
> lm.pooled ← lm (log.radon ~ floor)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.327	0.030	44.640	0.000
floor	-0.613	0.073	-8.421	0.000

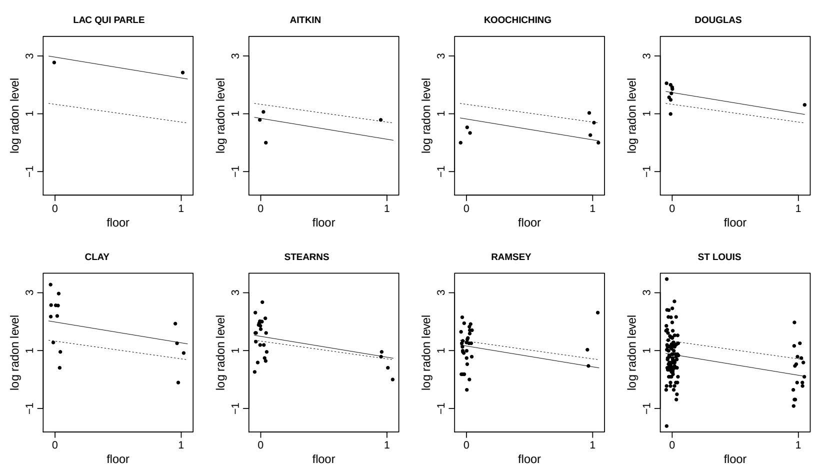
```
> lm.unpooled ← lm (log.radon ~ floor + factor(county) −1)
```

	Estimate	Std. Error	t value	Pr(>|t|)
floor	-0.721	0.074	-9.800	0.000
factor(county)1	0.841	0.379	2.220	0.027
factor(county)2	0.875	0.105	8.333	0.000
factor(county)85	1.187	0.535	2.218	0.027

Note that slope changes when we add county-level intercepts.

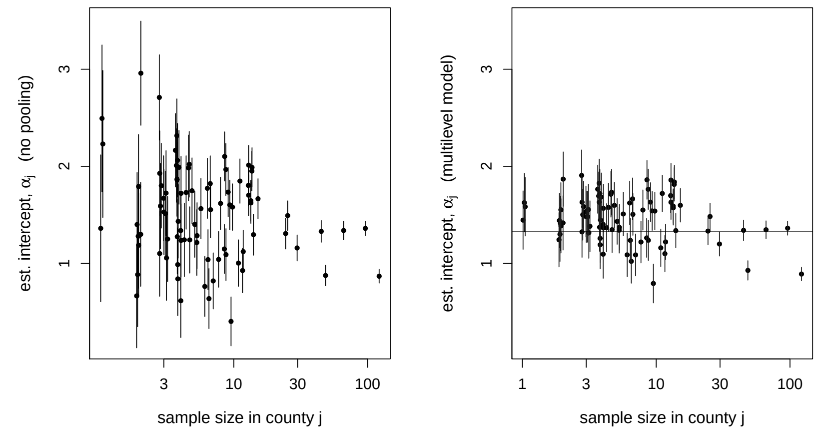
Complete versus No Pooling

Solid line: No pooling, dashed: complete pooling



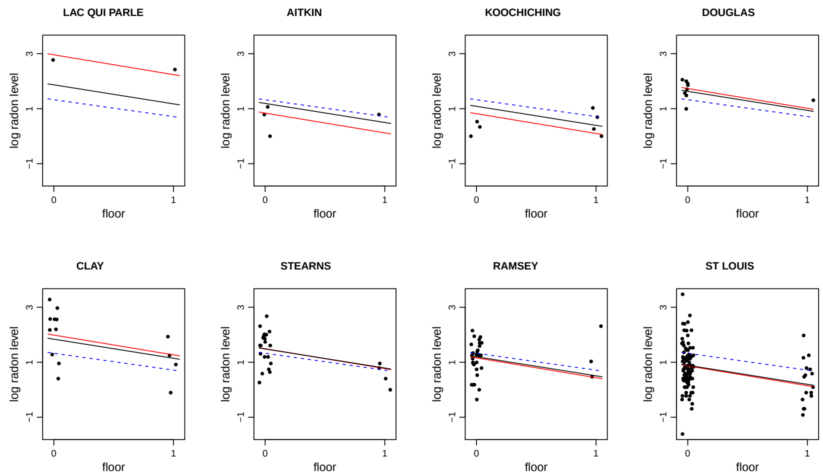
Intercept comparison

Slope is all the same, but intercepts are allowed to vary by county.



Similar to figure 12.1, but we did add one slope.

Pooled Effects



Red = unpooled, blue = completely pooled, black = partially pooled. Slopes are still all the same.

Parameter Estimates

Individual county intercepts are not of interest. Variance is.

Parameter	Estimate
σ_α	0.33
μ_α	1.46
β	-0.69
σ_y	0.76

Total variance, $\text{Var}(y_i) = \text{Var}(\alpha_{j[i]}) + \text{Var}(\epsilon_i) = \sigma_\alpha^2 + \sigma_y^2$ estimated as $0.33^2 + 0.76^2 = 0.687$. Of that, what proportion is due to county-to-county variation? $0.33^2/0.687 = 0.16$
Two houses from different counties have covariance: _____
Two houses from the same county share $\alpha_{j[i]}$, and have covariance _____.
Intraclass correlation between two houses in the same county is: _____

Variance Interpretation

$0.33^2/0.76^2 = 0.19$ so counties are about one-fifth as variable as individuals.
When we average individuals together into a smaple mean, we also reduce variation. County-to-county variation is comparable to the variance of a mean of 5 observations.
If a county has less than 5 obs, then it's sample mean is more variable than that of α_j, so it's pooled intercept should be closer to the complete-pooled intercept than to its individual unpooled estimate. At n=5, they should get averaged together. With larger sample sizes, partial pooled estimate should be closer to the unpooled values.
"Shrinkage" describes the movement of estimators toward a common value.