# Applied Bayesian Modeling module 8:
# **Bayesian multilevel regression models**

Leontine Alkema, lalkema@umass.edu
Fall 2022

# Outline

- Module 7:
  - Introduction to Bayesian multilevel models:
    a 2-level hierarchical model for estimating group means using normal distributions
- This module:
  - Predictions (for yet-to-be-sampled units or group-level parameters)
  - Bayesian multilevel regression models

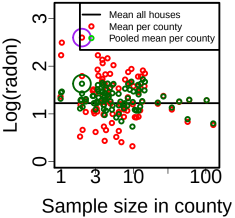
# Module 7: Bayesian multilevel model

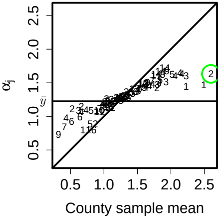
- $y_i|\mu_i, \sigma^2 \sim N(\mu_i, \sigma^2)$; now data are organized into groups
  - index $j[i]$ denotes the group for unit $i$ (counties for radon data)
- A multilevel model:

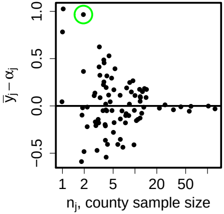
$$y_i|\alpha_{j[i]}, \sigma_y \overset{i.i.d}{\sim} N(\alpha_{j[i]}, \sigma_y^2), \tag{1}$$

$$\alpha_j|\mu_\alpha, \sigma_\alpha \overset{i.i.d}{\sim} N(\mu_\alpha, \sigma_\alpha^2), \tag{2}$$

  with priors for the model parameters $\sigma_y, \mu_\alpha, \sigma_\alpha$.
- Estimates for group means $\alpha_j$ from a multilevel model are *shrunk* from the group sample mean $\bar{y}_j$ towards the overall mean; smaller sample size $\Rightarrow$ more shrinkage



# Predictions for unsampled groups and units

- Suppose we fitted a Bayesian multilevel model with group mean $\alpha_j$:

$$\begin{aligned} y_i|\mu, \sigma^2 &\sim N(\alpha_{j[i]}, \sigma_y^2), \\ \alpha_j|\mu_\alpha, \sigma_\alpha^2 &\sim N(\mu_\alpha, \sigma_\alpha^2). \end{aligned}$$

- We can obtain posterior samples $\alpha_j^{(s)} \sim p(\alpha_j|\boldsymbol{y})$ for $s = 1, 2, \ldots, S$ for all groups $j$ with data
- What about predicting outcomes for
  - a group without data?
  - a house in a group with or without data?

# Predicting log-radon in a non-sampled house

- Suppose we want to predict radon in a new house $k$ in county $j[k]$.
- Assume that for the new house, the same sampling distribution holds true

$$\tilde{y}_k | \alpha_{j[k]}, \sigma_y^2 \quad \sim \quad N(\alpha_{j[k]}, \sigma_y^2),$$

  where the $\sim$ is added to indicate this is not-yet-observed observation.
- Suppose that we have other data in county $j[k]$, so we already have obtained samples $(\alpha_{j[k]}^{(s)}, \sigma_y^{(s)}) \sim p(\alpha_{j[k]}, \sigma_y |, \boldsymbol{y})$.
- Then what? What do we want?
- We want to sample $\tilde{y}_k \sim p(\tilde{y}_k | \boldsymbol{y})$... can we do that?
- Yes!

# Predicting mean log-radon in a non-sampled county

- We want to sample $\tilde{y}_k \sim p(\tilde{y}_k|\boldsymbol{y})$ when

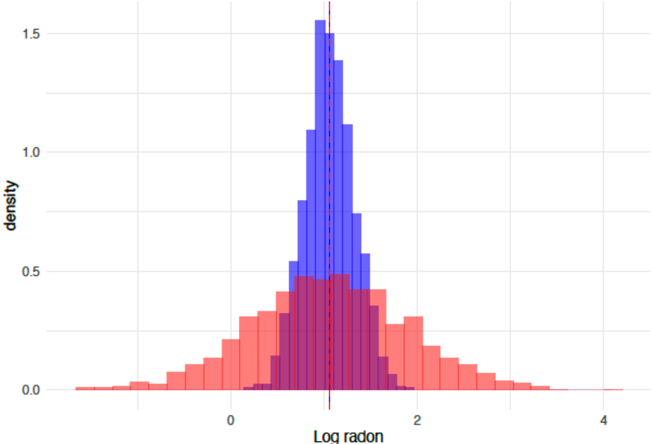
$$\tilde{y}_k|\alpha_{j[k]}, \sigma_y^2 \quad \sim \quad N(\alpha_{j[k]}, \sigma_y^2),$$

- We can sample $\tilde{y}_k \sim p(\tilde{y}_k|\boldsymbol{y})$ in two steps:
  (1) Sample $(\alpha_{j[k]}^{(s)}, \sigma_y^{(s)}) \sim p(\alpha_{j[k]}, \sigma_y|, \boldsymbol{y})$,
  (2) Sample $\tilde{y}_k^{(s)} \sim p(\tilde{y}_k|\alpha_{j[k]}^{(s)}, \sigma_y^{(s)})$.

  We already have samples $(\alpha_{j[k]}^{(s)}, \sigma_y^{(s)}) \sim p(\alpha_{j[k]}, \sigma_y|\boldsymbol{y})$ from fitting the model, so just need to do the 2nd step.
- Details: this produces a sample $\tilde{y}_k \sim p(\tilde{y}_k|\boldsymbol{y})$ because $p(\tilde{y}_k|\boldsymbol{y}) = \int \int p(\tilde{y}_k|\alpha_{j[k]}, \sigma_y)p(\alpha_{j[k]}, \sigma_y|, \boldsymbol{y})d\alpha_{j[k]}d\sigma_y$

# Predicting log-radon in a non-sampled house: results

- Posterior density of $\alpha_{j[k]}$ (blue) and predictive density for $\tilde{y}_k$ (red) with estimates $E(\alpha_{j[k]}|\boldsymbol{y}) \approx 1/S \sum \alpha_{j[k]}^{(s)}$ and $E(\tilde{y}_k|\boldsymbol{y}) \approx 1/S \sum \tilde{y}_k^{(s)}$
- How do the two densities compare wrt their mean and variance? Is that what you expected?



# Predicting mean log-radon in a non-sampled county

- Assume that for the new county, the same hierarchical distribution holds true:

$$\tilde{\alpha}_h | \mu_\alpha, \sigma^2_\alpha \sim N(\mu_\alpha, \sigma^2_\alpha),$$

  where the $\sim$ is added to indicate this is not-yet-observed group mean, and using index $h$ for that group.
- Then what?
- Can we sample $\tilde{\alpha}_h^{(s)} \sim p(\tilde{\alpha}_h | \boldsymbol{y})$?
- Yes!

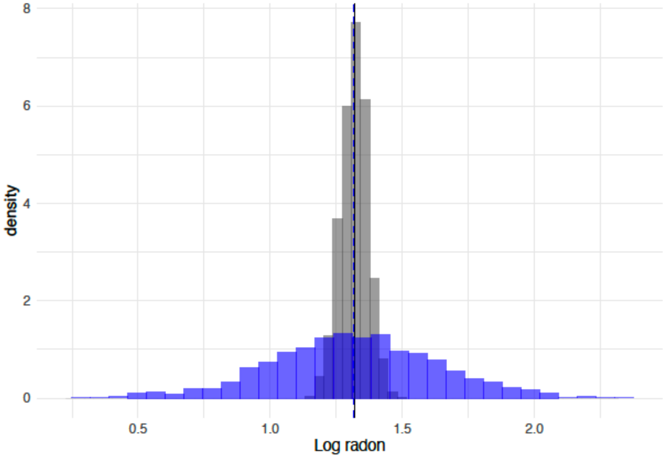
# Predicting mean log-radon in a non-sampled county

- Assume that for the new county, the same hierarchical distribution holds true: $\tilde{\alpha}_h|\mu_\alpha, \sigma_\alpha^2 \sim N(\mu_\alpha, \sigma_\alpha^2)$.
- We can sample $\tilde{\alpha}_h^{(s)} \sim p(\tilde{\alpha}_h|\boldsymbol{y})$ in two steps:
  (1) Sample $(\mu_\alpha^{(s)}, \sigma_\alpha^{2(s)}) \sim p(\mu_\alpha, \sigma_\alpha|\boldsymbol{y})$,
  (2) Sample $\tilde{\alpha}^{(s)} \sim p(\tilde{\alpha}|\mu_\alpha^{(s)}, \sigma_\alpha^{2(s)})$.

  We already have samples $(\mu_\alpha^{(s)}, \sigma_\alpha^{(s)})$ from fitting the model, so just need to draw the $\tilde{\alpha}_h^{(s)}$
- Details: this produces a sample $\tilde{\alpha}_h^{(s)} \sim p(\tilde{\alpha}_h|\boldsymbol{y})$ because $p(\tilde{\alpha}_h|\boldsymbol{y}) = \int\int p(\tilde{\alpha}_h|\mu_\alpha, \sigma_\alpha)p(\mu_\alpha, \sigma_\alpha|\boldsymbol{y})d\mu_\alpha d\sigma_\alpha$

# Predicting mean log-radon in a non-sampled county

- Posterior density of $\mu_\alpha$ (black) and predictive density for $\tilde{\alpha}_h$ (blue)
- How do the two densities compare wrt their mean and variance? Is that what you expected?



# Predicting log-radon in a non-sampled house in a non-sampled county

- Suppose that we are interested in predicting radon in a non-sampled house $k$ in a non-sampled county $h = j[k]$, with

$$\begin{aligned} \tilde{y}_k | \tilde{\alpha}_{j[k]}, \sigma_y^2 &\sim N(\tilde{\alpha}_{j[k]}, \sigma_y^2), \\ \tilde{\alpha}_h | \mu_\alpha, \sigma_\alpha^2 &\sim N(\mu_\alpha, \sigma_\alpha^2). \end{aligned}$$

- Can you sample $\tilde{y}_k^{(s)} \sim p(\tilde{y}_k | \boldsymbol{y})$?
- Yes! and you get to do it in the HW :)

# Hierarchical models with predictors

- For the radon data
  - The measurements are not exactly comparable across houses because in some houses, measurements are taken in the basement, while in other houses, 1st floor measurement are taken.
  - Additionally, county-level uranium measurements are probably informative for across-county differences in mean levels.
- To do: include predictors into our Bayesian model!
- For data with hierarchical structures, we can consider group-specific regression coefficients as well.

# Including unit-level predictors

- Model w/o predictors (extending notation to make it easier to introduce predictors):

$$
\begin{aligned}
y_i|\mu_i, \sigma_y &\overset{i.i.d}{\sim} N(\mu_i, \sigma_y^2), && (3)\\
\mu_i &= \alpha + \eta_{0,j[i]}, && (4)\\
\eta_{0,j}|\sigma_{\eta,0} &\overset{i.i.d}{\sim} N(0, \sigma_{\eta,0}^2). && (5)
\end{aligned}
$$

- Let unit-level predictor $x_i =$ house-level first-floor indicator (with $x_i = 0$ for basements, 1 otherwise).
- We can include house-level predictors in the house-level mean as follows:

$$
\mu_i = \alpha + \eta_{0,j[i]} + \beta_1 x_i,
$$

  this assumes that the difference in $\mu_i$ based on $x_i$ is the same across all counties.
- What if the relation between $x_i$ and $\mu_i$ varies by county?

## Including unit-level predictors - ctd

- If the relation between $x_i$ and $\mu_i$ varies by county, then we can consider

$$\begin{aligned} \mu_i &= \alpha + \eta_{0,j[i]} + (\beta_1 + \eta_{1,j[i]})x_i, \\ \eta_{1,j}|\sigma_{\eta,1} &\overset{i.i.d}{\sim} N(0, \sigma^2_{\eta,1}). \end{aligned}$$

  where $\eta_{1,j[i]}$ capture county-specific deviations in the relationship between $x_i$ and $\mu_i$.

- There may be correlation between $\eta_{0,j}$ and $\eta_{1,j}$, we can estimate and account for this correlation if we use a bivariate hierarchical distribution:

$$\boldsymbol{\eta}_{0:1,j}|\boldsymbol{\Sigma} \sim N_2(\mathbf{0}, \boldsymbol{\Sigma}),$$

  with a prior on the variance-covariance matrix $\boldsymbol{\Sigma}$.

# Including unit-level predictors: Model fitting

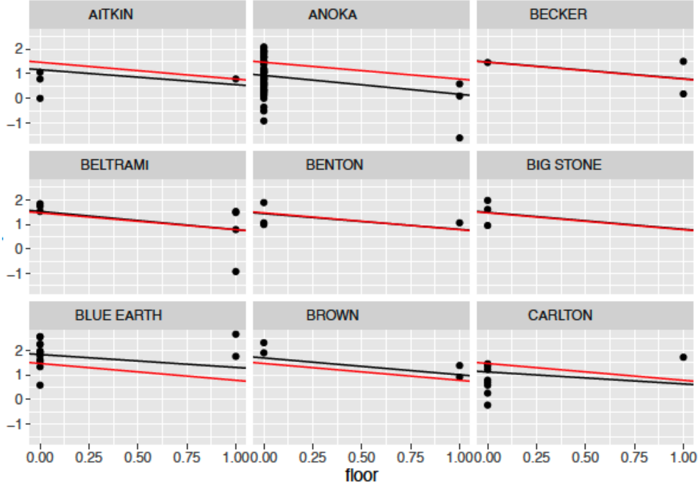
- Model: $\mu_i = \alpha + \eta_{0,j[i]} + (\beta_1 + \eta_{1,j[i]})x_i$
- brm call: `brm(y ~ (1+floor|county) + floor, ...)`

```
Group level effects ~ county
```

```
##                      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## sd(Intercept)            0.36      0.05     0.27     0.47 1.00     1330
## sd(floor)                0.26      0.15     0.02     0.55 1.01      530
## cor(Intercept,floor)    -0.18      0.38    -0.86     0.69 1.00     1984
##                      Tail_ESS
## sd(Intercept)            2118
## sd(floor)                1417
## cor(Intercept,floor)     1779
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept     1.47      0.06     1.36     1.58 1.00     1757     2221
## floor        -0.68      0.08    -0.85    -0.52 1.00     3675     2704
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.76      0.02     0.72     0.80 1.00     3710     2723
```

# Some results

- Relation between $x$ and $\mu$ for selected counties: data (dots), county-specific regression line (black, $\alpha + \eta_{0,j[i]} + (\beta_1 + \eta_{1,j[i]})x$), and mean regression line (red, $\alpha + \beta_1 x$)



# Including group-level predictors

- County-level log-uranium measurements $u_j$ are probably informative for across-county differences in mean levels.
- We can include group-level predictors in the group-level mean as follows:

$$\mu_i \quad = \quad \alpha + \eta_{0,j[i]} + (\beta_1 + \eta_{1,j[i]})x_i + \beta_2 u_{j[i]}$$

- Would it make sense to consider group-level coefficients for $u_{j[i]}$? No: there is only one value for $u_j$ per group $j$!
- What if the association between $x_i$ and $\mu_i$ depends on $u_j$? This introduces an interaction term:

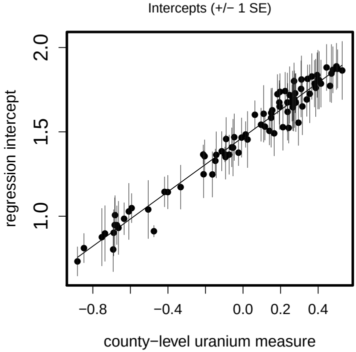
$$\mu_i \quad = \quad \alpha + \eta_{0,j[i]} + (\beta_1 + \eta_{1,j[i]})x_i + \beta_2 u_{j[i]} + \beta_3 u_{j[i]} x_i,$$

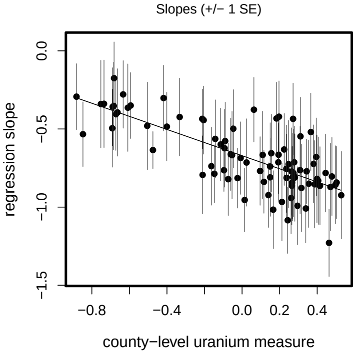
  where, for radon,
  - $\beta_2$ refers to the association between $u_{j[i]}$ and $\mu_i$ when $x_i = 0$ (basement measurements)
  - $\beta_2 + \beta_3$ captures the association for $x_i = 1$ (1st floor measurements).

# Interpretation and visualization of how $\mu$ changes with $u_j$

intercept $= \alpha + \beta_2 u_j (+\eta_{0,j})$

here slope $=$ coefficient for $x_i = \beta_1 + \beta_3 u_j (+\eta_{1,j})$




- As county-uranium increases, county-level means increase.
- County-level slopes (the log-radon difference between 1st floor and basement measurements) decrease away from 0, hence relative differences in radon levels increase with county-level uranium.

# Look how far we gotten already!

- We discussed Bayesian multilevel regression models with varying intercepts and/or slopes, for data that are normally distributed.
- For a research question (in words) and data set, you are able to specify such a Bayesian multilevel regression model in Greek and fit it to data, using `brm`
- Given such a multilevel model in Greek and model output, you are able to interpret the parameter estimates and create predictions.
- Modules 9 onwards:
  - Model checking.
  - Further model extensions, e.g., what if $\sigma_y$ varies across counties, what if data are not normally distributed?