

IDS 702: MODULE 4.2

MULTILEVEL/HIERARCHICAL LINEAR MODELS

DR. OLANREWAJU MICHAEL AKANDE

MOTIVATING EXAMPLE: THE RADON ANALYSIS

- As a motivating example, we will look at data on radon levels of houses within each of 85 counties in Minnesota.
- The data is in the file `Radon.txt` on Sakai.
- The full data actually includes data for more states but we will focus on just Minnesota like the textbook.
- The U.S. Environmental Protection Agency and the Surgeon General's Office have estimated that as many as 20,000 lung cancer deaths are caused each year by exposure to radon (reference [here](#)).
- Radon is a cancer-causing radioactive gas and is the second leading cause of lung cancer. Unfortunately, you cannot see, smell or taste it. The most commonly used device for making short-term radon measurements in homes is the charcoal canister
- Radon occurs naturally as an indirect decay product of uranium.
- Given that counties are nested within states, thinking about a hierarchical model here makes sense.

MOTIVATING EXAMPLE: THE RADON ANALYSIS

| Variable | Description |
|-------------|---|
| radon | radon levels for each house |
| log_radon | log(radon) |
| state | state |
| floor | lowest living area of each house: 0 for basement, 1 for first floor |
| countyname | county names |
| countyID | ID for the county names (1-85) |
| fips | state + county fips code |
| uranium | county-level soil uranium |
| log_uranium | log(uranium) |

The response variable, radon (or log_radon) is continuous, so we need a (hierarchical/multilevel) regression framework.

To ascertain that we need a multilevel model here, we should check for differences across counties during EDA.

HIERARCHICAL LINEAR MODELS

- Hierarchical models (like the model for the school data) can be applied to regression contexts where observations are grouped
- First we will only focus on models for linear regression.
- However, the same ideas apply to logistic regression (as we will see soon), Poisson regression, etc.
- Recall that a standard linear model with one predictor can be written as

$$y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i; \quad \epsilon_i \sim N(0, \sigma^2); \quad i = 1, \dots, n.$$

- Now suppose that the observations fall into J groups, indexed by j .
- Then there are several ways to take advantage of the group within the context of hierarchical models.

RANDOM INTERCEPTS MODEL

- First, we can let the intercept alone vary by group, if we think that the predictor has the same effect on each group, but the overall intercept (grand mean of the response) is different for each group.
- This is known as the **random intercepts model** or the **varying-intercept model**, and is often written as

$$\begin{aligned}y_{ij} &= \beta_{0j} + \beta_1 x_{1ij} + \epsilon_{ij}; \quad i = 1, \dots, n_j; \quad j = 1, \dots, J \\ \epsilon_{ij} &\sim N(0, \sigma^2) \\ \beta_{0j} &\sim N(\beta_0, \tau_0^2).\end{aligned}$$

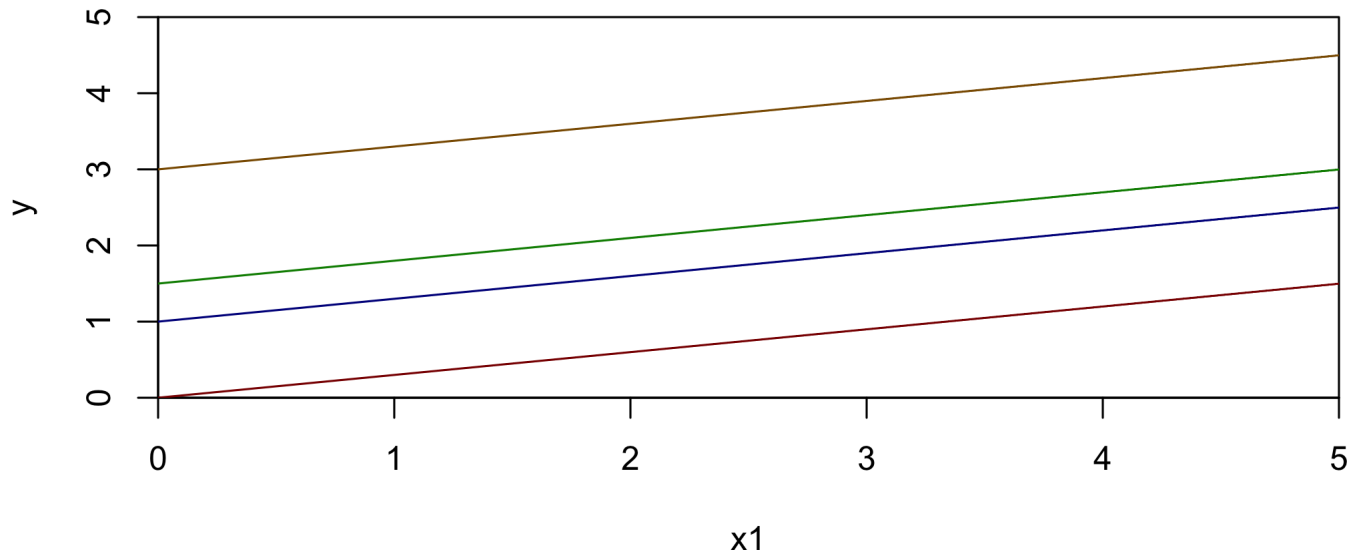
where i indexes observations and j indexes groups.

- The model can also be written as

$$\begin{aligned}y_{ij} &= (\beta_0 + \gamma_{0j}) + \beta_1 x_{1ij} + \epsilon_{ij}; \quad i = 1, \dots, n_j; \quad j = 1, \dots, J \\ \epsilon_{ij} &\sim N(0, \sigma^2) \\ \gamma_{0j} &\sim N(0, \tau_0^2).\end{aligned}$$

RANDOM INTERCEPTS MODEL

- Allows separate intercepts for each group, but shrinks estimates towards common value.



- Useful for repeated measurements, when the "groups" are individuals, e.g., we take a subject's blood pressure three times and include all three measurements in the model).
- Also useful when some groups have small sample sizes, so that estimation of individual group means is highly variable.

RANDOM SLOPES MODEL

- We may want to let only the slopes vary by group, if we think that the predictor has a different effect on each group, but the overall intercept is the same for each group.
- This is known as the **random slopes model** or the **varying-slope model**, and is often written as

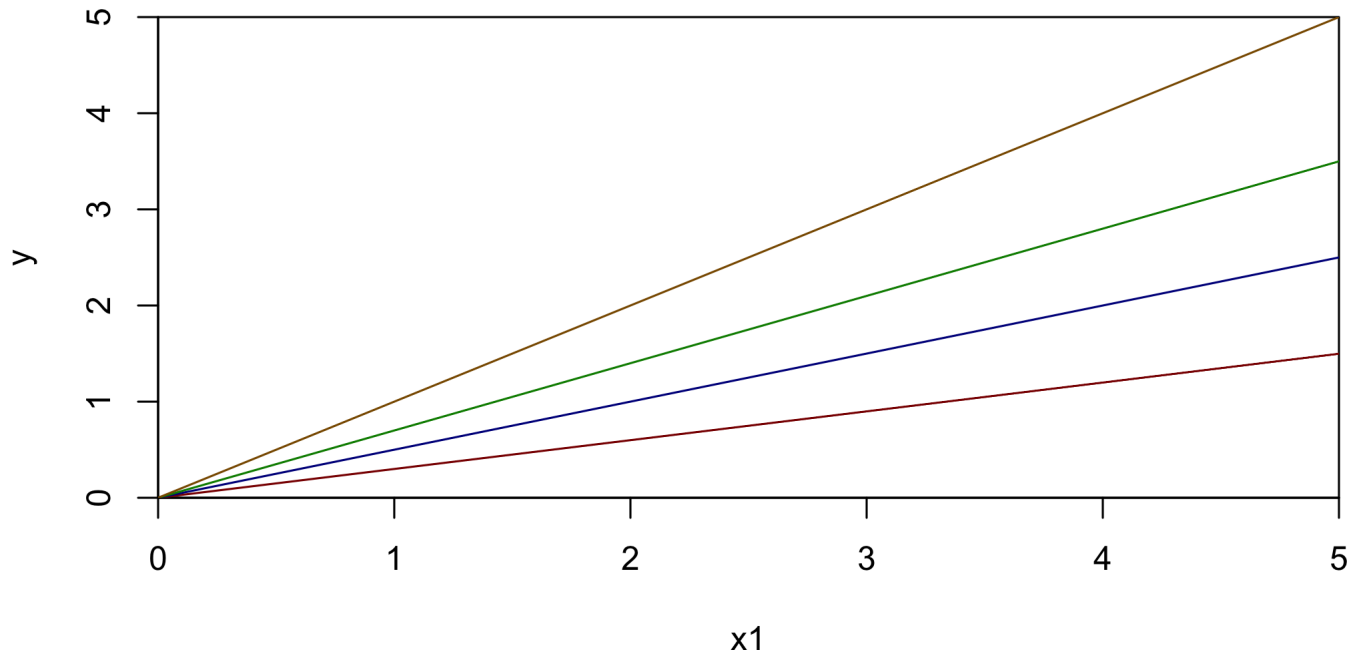
$$\begin{aligned}y_{ij} &= \beta_0 + \beta_{1j}x_{1ij} + \epsilon_{ij}; \quad i = 1, \dots, n_j; \quad j = 1, \dots, J \\ \epsilon_{ij} &\sim N(0, \sigma^2) \\ \beta_{1j} &\sim N(\beta_1, \tau_1^2).\end{aligned}$$

- The model can also be written as

$$\begin{aligned}y_{ij} &= \beta_0 + (\beta_1 + \gamma_{1j})x_{1ij} + \epsilon_{ij}; \quad i = 1, \dots, n_j; \quad j = 1, \dots, J \\ \epsilon_{ij} &\sim N(0, \sigma^2) \\ \gamma_{1j} &\sim N(0, \tau_1^2).\end{aligned}$$

RANDOM SLOPES MODEL

- Allows separate slopes for each group, but shrinks estimates towards common value.
- Also useful when some groups have small sample sizes, so that estimation of slopes is highly variable.
- The model implies the same intercept for each group.



RANDOM SLOPES AND INTERCEPTS MODEL

- We can also combine both ideas, that is, allow for the slopes and intercepts to both vary by group, if we think that the predictor has a different effect on each group, and the overall intercept is also different for each group.
- This is known as the **random slopes and intercepts model** or the **varying-slope, varying-intercept model**, and is often written as

$$\begin{aligned}y_{ij} &= \beta_{0j} + \beta_{1j}x_{1ij} + \epsilon_{ij}; & i = 1, \dots, n_j; & \quad j = 1, \dots, J \\ \epsilon_{ij} &\sim N(0, \sigma^2) \\ (\beta_{0j}, \beta_{1j}) &\sim N_2((\beta_0, \beta_1), \Sigma).\end{aligned}$$

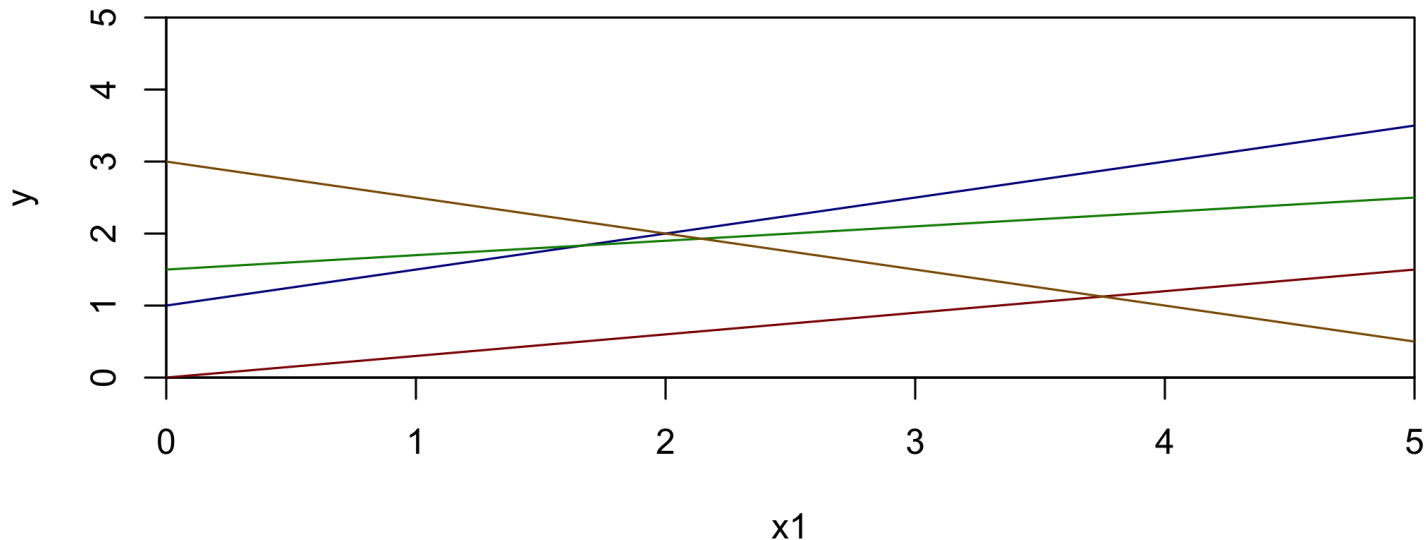
where $N_2(\boldsymbol{\mu}, \Sigma)$ is the bivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ .

- The model can also be written as

$$\begin{aligned}y_{ij} &= (\beta_0 + \gamma_{0j}) + (\beta_1 + \gamma_{1j})x_{1ij} + \epsilon_{ij}; & i = 1, \dots, n_j; & \quad j = 1, \dots, J \\ \epsilon_{ij} &\sim N(0, \sigma^2) \\ (\gamma_{0j}, \gamma_{1j}) &\sim N_2(\mathbf{0}, \Sigma).\end{aligned}$$

RANDOM SLOPES AND INTERCEPTS MODEL

- Allows for separate slopes and intercepts for each group, but shrinks estimates towards common value
- Useful when some groups have small sample sizes, so that estimation of slopes and intercepts is highly variable
- $(\gamma_{0j}, \gamma_{1j})$ are called **random effects** while (β_0, β_1) are called **fixed effects**. Models with fixed and random effects are often called **mixed effects models**.



MIXED EFFECTS MODEL

- Use the `lmer` command in the `lme4` package in R to estimate the parameters using maximum likelihood (ML) or restricted maximum likelihood (REML) estimation.
- Take STA 601/602 and/or STA 610 for information on fitting these models using Bayesian methods.
- Also, note that the terms fixed effects, random effects, and mixed effects can have (very) different meanings in different fields.
- So, we will not get too carried away with the terminology.
- For us, the important thing will be to be able to distinguish between parameters that vary by group and those that do not.

MODEL ASSESSMENT AND VALIDATION

- Model assessment and validation from linear regression carries over.
- You should still have linearity (by each group for varying slopes), independence of the errors (and also of the varying effects for each predictor), equal variance, and normality.
- You should still look out for outliers and check for multicollinearity.
- Model comparison between two multi-level models does not quite work the same way.
- We will not dive deeply into estimation but basically, ML produces unbiased estimates for the fixed effects but not the random effects whereas REML produces unbiased estimates for the random effects.
- When using the `anova` function in R, keep the random effects part the same when comparing two models (so you'll be comparing fixed effects). Use AIC or BIC to decide the form of the random effects.

WHAT'S NEXT?

MOVE ON TO THE READINGS FOR THE NEXT MODULE!