IQSR Multilevel/Hierarchical Modeling (Part I)

Chao-Yo Cheng

Birkbeck
UNIVERSITY OF LONDON

Plan ahead

- Week 3: Thinking about your project
- Weeks 4 and 5: Analyzing survey data
- Weeks 8 and 9: Multilevel and hierarchical modeling
  - **Week 8: The fundamentals**
  - Week 9: Assumptions and diagnostics
- Week 10: Concluding remarks, Q&A, and post-term activities (TBD)

IQSR: Linear regression and its extension

- What is linear regression ($Y = \alpha + \beta X + \epsilon$)?
- What can linear regression do?
  - For **prediction**: The main concern is $\widehat{Y}$;
  - To show the marginal effect of X on Y: The main concern is to identify the conditions under which $\widehat{\beta}$ is causally valid;
  - Searching for the "best" **data generation process** (DGP) of Y: The main concern is the functional form; go for **non-linear** and multilevel modeling.
- What are the key assumptions? Why do we need them? Which assumptions are the most (or least) important ones? How should we test these assumptions – can we and should we? What will happen if we relax (any of) the assumptions?

IQSR: Linear regression and its extension

- For **prediction**: Go for **maching learning** (forecasting, pattern recognition and classification).
- To show the marginal effect of X on Y: Go for **causal inference**.
- Searching for the "best" **data generation process** (DGP) of Y: Go for **non-linear** and multilevel modeling.
  - Linear regression with polynomial terms, say

$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \epsilon.$$

  - Generalized linear models for discrete and categorical variables, say

$$\log\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X + \epsilon.$$

  - Multilevel (aka hierarchical aka mixed effects) models for when we have **nested data**.

Birkbeck
UNIVERSITY OF LONDON

Why multilevel modeling

- Very often, we can detect some nuanced (not necessarily complicated) structure in our data. Say:
  - Vote choices across different respondents in a national survey can be grouped by state.
  - Exam marks across different students in a secondary school can be grouped by teacher.
  - Local public goods provision records across different villages can be grouped by constituency.
  - ...What else?

Why multilevel modeling

- Very often, we can detect some nuanced (not necessarily complicated) structure in our data. Say:
  - Vote choices across different respondents in a national survey can be grouped by state.
  - Exam marks across different students in a secondary school can be grouped by teacher.
  - Local public goods provision records across different villages can be grouped by constituency.
  - ...What else?
- How does multilevel modeling matter or come to rescue?
  - **Practical motivation**: The underlying intercept and slope we try to uncover via OLS may depend on the group of focus.
  - **Theoretical motivation**: Nested data may and can violate the basic assumptions of linear regression (esp. equal variance and independence of errors).

Prototypes of multilevel models

- We want to know whether race influences individual's approval rating (0-100 pts) for the President of the United States.
  - Predictor (X): 'race' (e.g., Asian American, etc).
  - Outcome (Y): 'approval'.
  - Grouping: 'state' (e.g., California).

- Let's consider several modeling choices.

Prototypes of multilevel models

	Formal expression
Linear (fixed α and β)	$Y = \alpha + \beta X + \epsilon$
Linear Multilevel (varying α only)	$Y = \alpha_i + \beta X + \epsilon$
Linear Multilevel (varying β only)	$Y = \alpha + \beta_i X + \epsilon$
Linear Multilevel (varying α and β only)	$Y = \alpha_i + \beta_i X + \epsilon$

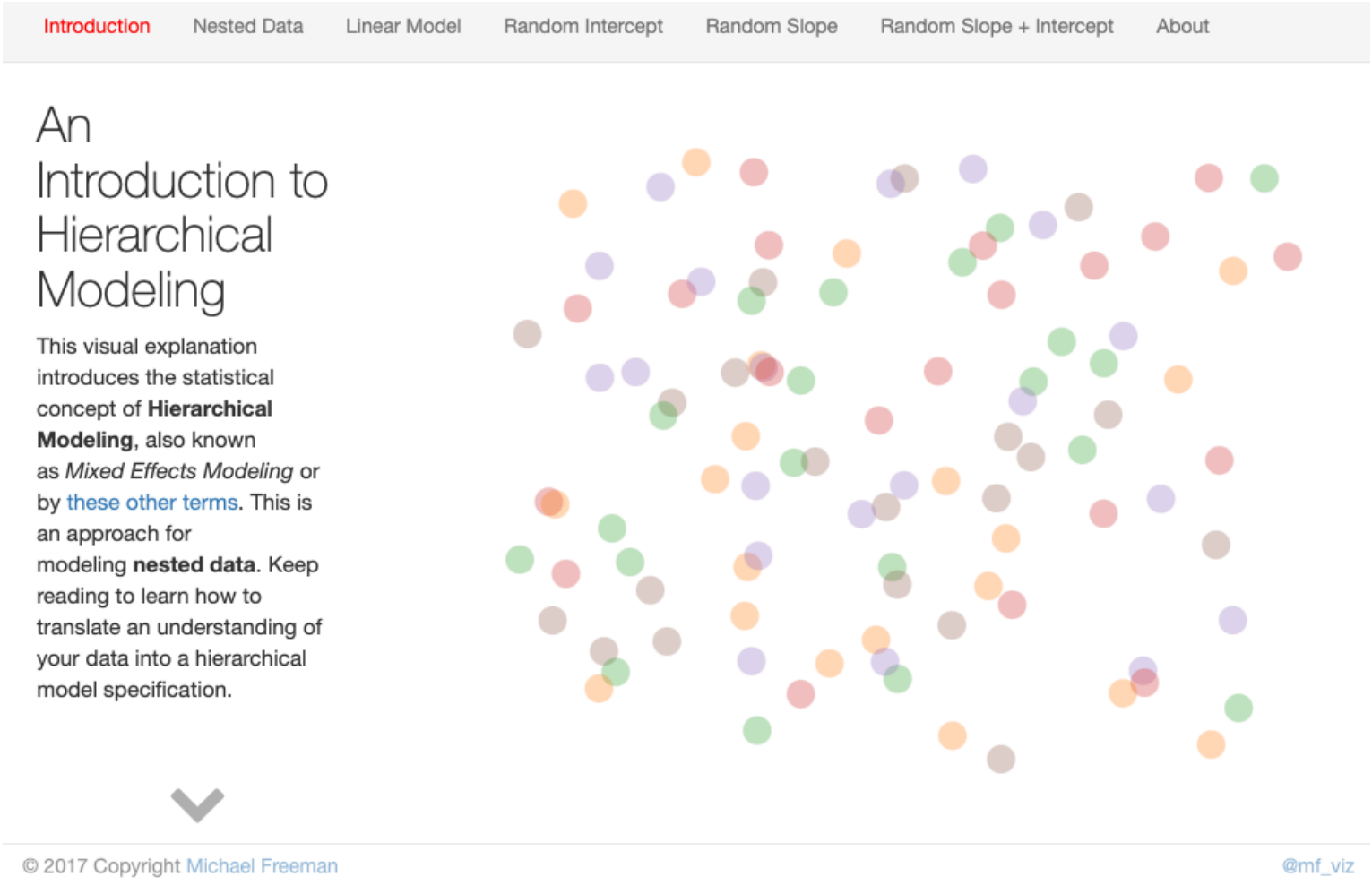
	R code
Linear (fixed α and β)	`lm(y~x)`
Linear Multilevel (varying α)	`lmer(y~x+(1|state))`
Linear Multilevel (varying β)	`lmer(y~x+(0+x|state))`
Linear Multilevel (varying α and β)	`lmer(y~x+(1+x|state))`

Why different names?

- Multilevel or hierarchical models: Used to highlight the fact that that we are trying to analyze **nested** data.

- Mixed-effect models: Used to highlight the fact that α and β can "vary" by group, depending on how we set up the model.
  - The varying α consists of a fixed element (same for all groups) and a random element (each group has a different value).
  - The varying β consists of a fixed element (same for all groups) and a random element (each group has a different value).

Visualization of (linear) multilevel models



http://mfviz.com/hierarchical-models/

Birkbeck
UNIVERSITY OF LONDON

Some additiional modeling considerations

- Should we include **group**-level predictors? Short answer: Yes (many researchers do so).
- Should we use **linear** multilevel models? Short answer: Of course (there is logit multilevel regression).
- What if there are **multiple** levels or many alternative ways of grouping observations, such as **time** and **location**? Short answer: Be my guest (but you want to think about the trade-off).
- Can we use multilevel modeling for **longitudinal data analysis**? Short answer: Yes (but be extremely cautious – you may be entering a territory where few people are in now).

Caveats

- As in the case of all other modeling techniques, we should keep the following caveats in mind:
  - Hypothesis testing is a headache, if not impossible – rigorous thinking is needed to derive the p-values and their corresponding levels of statistical significance. Potential solution: **Bayesian** multilevel modeling.
  - The results may be driven by influential observations or outliers when the number of observations varies too much across different groups.
  - The model may become unnecessarily intractable (formally and computationally) when it gets too complicated.
  - It may not be easy to have a set of clear priors to make a decision which model performs or works better – again, what does that mean to say a model is "better" than other alternatives?
- Extra: For more discussion, see **Gelman and Hill** (2007) – focus on Chapters 11-13.

Looking ahead

- Today in tutorial:
  - We will use **World Values Survey** (Wave 7) to study how **education** is related to **happiness**.
  - We will group the respondents by the **type of political regimes** (e.g., autocracy, anocracy, and democracy).
- Next week, we will use the assumptions of linear regression to motivate why multilevel or hierarchical modeling *theoretically*.
- We may be able to discuss a real example of multilevel modeling.