Multilevel/Hierarchical Modeling

Advanced Topics in Quant Social Research

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?

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**.

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)	$Y = \alpha_i + \beta_i X + \epsilon$

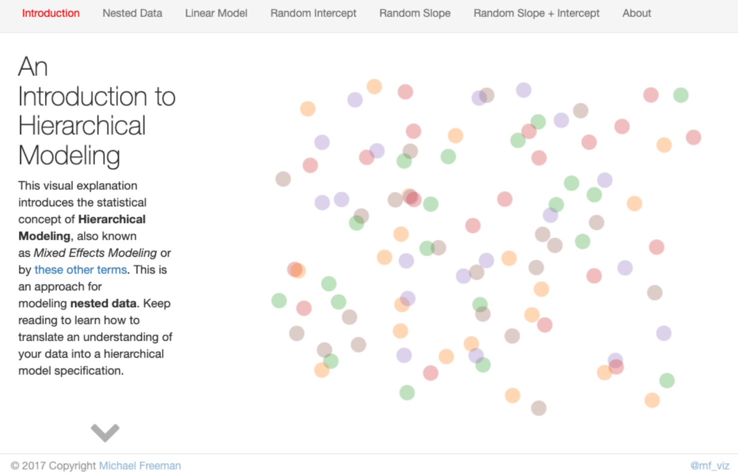
	`R` code
Linear (fixed α and β)	`lm(y~x)`
Linear Multilevel (varying α)	`lmer(y~x+(1|state))`
Linear Multilevel (varying β)	`lmer(y~x+(0+x|state))`
Linear Multilevel (varying α and β)	`lmer(y~x+(1+x|state))`

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/

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

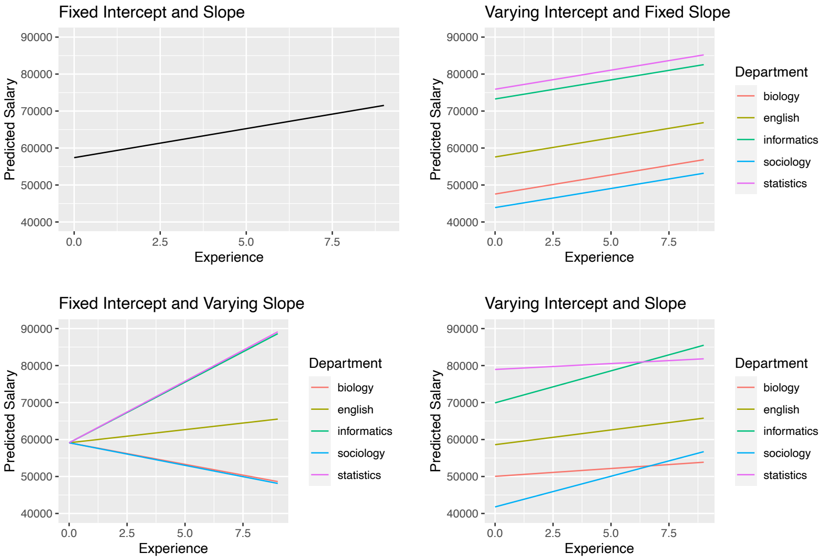
- As in the case of all other modeling techniques, we should keep the following caveats in mind:
  - Hypothesis testing is a headache, if not impossible – rigorous thinking is needed to derive the p-values and their corresponding levels of statistical significance. Potential solution: **Bayesian** multilevel 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.

Recap: Multilevel/hierarchical modeling

- Linear multilevel regression is an extension of (classical) linear regression (OLS).
- Linear multilevel regression can be used to analyze **nested data** (see below) by allowing observations in each cluster or group to have a unique fitted line (i.e., varying intercepts and/or slopes).
- Pros and cons
  - Pros: Versatile, flexible and nuanced.
  - Cons: Formal and computational tractability (complicated and time-consuming); hypothesis testing can be challenging (so go **Bayesian** please).

Level 1		Level 2
students		classrooms
corporations		nations
spouses	nested within	couples
voters		districts
repeated measures		individuals

Example: Years of experiences and salaries across 5 depts



Multivariate v multilvel linear regression

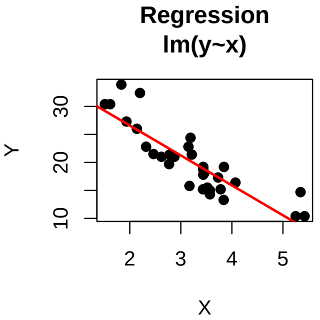
We want to use regression to explain var(Y); the model's explanatory power can be improved by

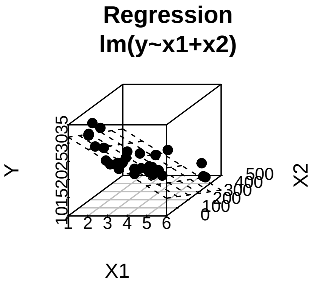
- Bivariate v multivariate: increasing the number of predictors (dimensions).
- Classical (single-) v multi-level: increasing the number of levels (hierarchies).

	Classical	Multi-level
Bivariate (one predictor)	`lm()`	`lmer()`
Multivariate (more than predictor)	`lm()`	`lmer()`

Classicial regression (fixed intercept and slopes)

- Bivariate (one predictor): $Y = \alpha + \beta X + \epsilon$.
- Multivariate (two predictors): $Y = \alpha + \beta_1 X_1 + +\beta_2 X_2 + \epsilon$.





Multilevel regression (varying intercept and/or slopes)

- Bivariate (one predictor; $Y = \alpha + \beta X + \epsilon$): there will be 3 multilevel models if one would like to varying α and/or β across groups.
  - Varying α: $Y = \alpha_j + \beta X + \epsilon$.
  - Varying β: $Y = \alpha + \beta_j X + \epsilon$.
  - Varying α and β: $Y = \alpha_j + \beta_j X + \epsilon$.
- Multivariate (two predictors; $Y = \alpha + \beta_1 X_1 + + \beta_2 X_2 + \epsilon$): there will be 7 multilevel models if one would like to varying α, β_1, and/or β_2 across groups.

Multilevel regression (varying intercept and/or slopes)

- Bivariate (one predictor; $Y = \alpha + \beta X + \epsilon$): there will be 3 multilevel models if one would like to varying α and/or β across groups.
  - Varying α: $Y = \alpha_j + \beta X + \epsilon$.
  - Varying β: $Y = \alpha + \beta_j X + \epsilon$.
  - Varying α and β: $Y = \alpha_j + \beta_j X + \epsilon$.
- Multivariate (two predictors; $Y = \alpha + \beta_1 X_1 + +\beta_2 X_2 + \epsilon$): there will be 7 multilevel models if one would like to varying α, β_1, and/or β_2 across groups.
- For any baseline linear regression (fixed intercepts and slopes), there will be $2^k - 1$ multilevel models one can fit (where k refers to the number of predictors).

Why linear "multilevel" regression

- Practical motivation.
  - When observations can be nested into different groups (or clusters).
  - The statistical relationship between Y and predictors (Xs) can be different across different groups.
- Theoretical motivation.
  - When observations can be nested into different groups (or clusters).
  - Using linear regression to fit nested data can generate errors that violate the assumptions for OLS to be BLUE.

Assumptions for linear regression to be BLUE (skip)

- Linearity and addivity: Y is a linear function of the predictor(s).
- Normality, homoscedasticity, and independence of ϵ.
  - The residuals should be normally distributed; that is, probabilistically the model not produce extreme errors.
  - The residuals should have equal variance; that is, errors should not be predictable; if violated, $\widehat{\beta}$ can be biased.
  - The residuals should have equal variance; that is, the errors should not depend on each other.

	Diagnostics
PG Certificate	V
Normality ϵ	Quantile-quantile plot
Homoscedasticity of ϵ	Residual plots ($\widehat{Y}$ or X vs ϵ)
Independence of ϵ	Ad hoc statistics (e.g., Durbin–Watson)

Assumptions for linear regression to be BLUE

- In real life, nested data is quite common.
- Using linear regression to analysis nested data can be problematic, because the error of an observation may be determined by the group to which it belongs to (hence no **homoscedasticity** and/or **independence**).
- Linear multilevel regression allows us to use linear functions to model nested data.

Example: Gelman et al (2008)

- "How is income related to individual's vote choice?"
- Results from multilevel regression show that **income matters more in red (i.e., Republican) America**.

Quarterly Journal of Political Science, 2007, 2: 345–367

Rich State, Poor State, Red State, Blue State: What's the Matter with Connecticut?*

Andrew Gelman[1], Boris Shor[2], Joseph Bafumi[3] and David Park[4]

[1]*Department of Statistics and Department of Political Science, Columbia University, New York, USA; gelman@stat.columbia.edu; www.stat.columbia.edu/~gelman*
[2]*Harris School of Public Policy Studies, University of Chicago, USA*
[3]*Department of Government, Dartmouth College, USA*
[4]*Department of Political Science, George Washington University, USA*

Example: Blaydes and Linzer (2012)

- "How is religiosity related to anti-Americanism among Muslims?"
- Religiosity is positively correlated with sentiment against the US when **a country is polarized over religious–secular issue dimension**.

American Political Science Review Vol. 106, No. 2 May 2012
doi:10.1017/S0003055412000135

Elite Competition, Religiosity, and Anti-Americanism in the Islamic World

LISA BLAYDES *Stanford University*
DREW A. LINZER *Emory University*

The battle for public opinion in the Islamic world is an ongoing priority for U.S. diplomacy. The current debate over why many Muslims hold anti-American views revolves around whether they dislike fundamental aspects of American culture and government, or what Americans do in international affairs. We argue, instead, that Muslim anti-Americanism is predominantly a domestic, elite-led phenomenon that intensifies when there is greater competition between Islamist and secular-nationalist political factions within a country. Although more observant Muslims tend to be more anti-American, paradoxically the most anti-American countries are those in which Muslim populations are less religious overall, and thus more divided on the religious–secular issue dimension. We provide case study evidence consistent with this explanation, as well as a multilevel statistical analysis of public opinion data from nearly 13,000 Muslim respondents in 21 countries.

Extra: Advanced multilevel modeling

- What if we want to include group-level predictors (Gelman and Hill 2007)?
- What if there are clusters across levels – should we include all of them (Finch et al 2014)? For instance,
  - In administrative data, **towns** can be nested into **municipalities**, which can then be nested into **provinces**.
  - In a cross-national survey, **respondents** can be nested into **countries**, which can then be bested into **continents**.
- What if the observations can be nested into groups and time (Fairbrother 2013)?
- What if we want to include group-level predictors? – group-level estimates can be problematic when we do not have too many groups (Bryan et al 2016).

Extra: Advanced multilevel modeling

- What if we want to include group-level predictors (Gelman and Hill 2007)?
- What if there are clusters across levels – should we include all of them (Finch et al 2014)?
- What if the observations can be nested into groups and time (Fairbrother 2013)?
  - In longitudinal cross-national survey, you may have **repeated measures of each respondent** (e.g., education and income) across different **countries** over the **years**.
  - In longitudinal school data, you may have **repeated measures of each student** (e.g., race and exam mark) across different **schools** over the **semesters**.
- What if we want to include group-level predictors? – the estimates at the group level can be problematic when we do not have too many groups (Bryan et al 2016).