PoliSci 4782 Political Analysis II

Hierarchical Model

Seamus Wagner

The Ohio State University

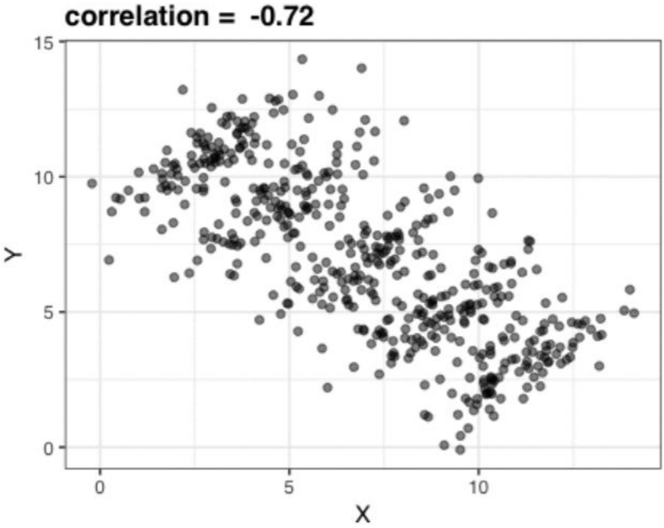
Clustering and Multilevel Data

- Nested multilevel data: individual observations (i) in groups (j), so we may have both individual attributes and group-level attributes (individuals are the subsets of groups):
  - students in different classes in different schools
  - voters in different districts in different states

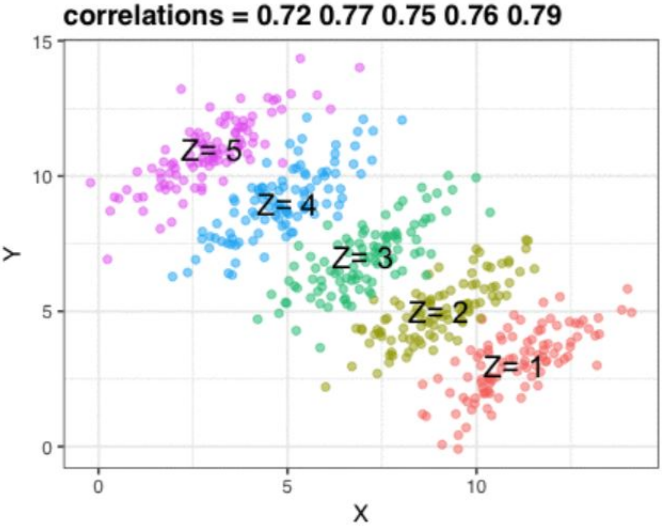
- Non-nested multilevel data: individual observations (i) are characterized by overlapping categories of attributes:
  - time-series cross-sectional data: each observation is indexed by unit i and time t

Question: How do we account for the variation between groups/clusters and ideally put it as part of our model?

Why Identifying Groups Matters



Why Identifying Groups Matters



Different Approaches

- Pooling model: completely ignore groups/clusters and focus on individual-level attributes
- No-pooling model: completely ignore individual-level attributes
- Fixed-effects model: a pooling model with group-level indicators as fixed effects on the outcome
- Hierarchical (Multilevel) model: a model with both individual and group-level variables in which group-level variations can be functions (or random effects) of individual variables

Pooling Model

$$y_i = \alpha + \beta \boldsymbol{X}_i + \epsilon_i$$

- outcome variable Y and a vector of explanatory variables $\boldsymbol{X}$
- individual index i
- group index j

No-pooling Model

$$y_i = \alpha + \gamma_1 J[1] + \gamma_2 J[2] + \cdots + \gamma_{j-1} J(j-1) + \epsilon$$

or more succinctly

$$y_i = \alpha_{j-1} + \epsilon$$

- outcome variable Y and a vector of explanatory variables $\boldsymbol{X}$
- individual index i (J refers to binary indicators for specific groups)
- group index j

Fixed-effects Model

$$y_i = \beta \boldsymbol{X}_i + \gamma_1 J[1] + \gamma_2 J[2] + \cdots + \gamma_{j-1} J(j-1) + \epsilon_i$$

or more succinctly

$$y_i = \beta \boldsymbol{X}_i + \gamma_j + \epsilon_i$$

- outcome variable Y and a vector of explanatory variables $\boldsymbol{X}$
- individual index i (J refers to binary indicators for specific groups)
- group index j
- γ is the group-level fixed-effects (serve as a group-varying intercept in the model)

Building from Fixed-effects Model

Fixed-effects model:

$$y_i = \beta \boldsymbol{X}_i + \gamma_j + \epsilon_i$$

- γ is more like a group separator in this model
- Its value is given by the average of Y in each group
- We are neither modeling γ nor consider its endogenous relationships with $\boldsymbol{X}$ in each group
  - Whether it is an important limitation is a judgment call.
- Hierarchical models instead try to estimate the relationships across levels, enabling us to have changing intercepts and coefficients across groups (more detail later)

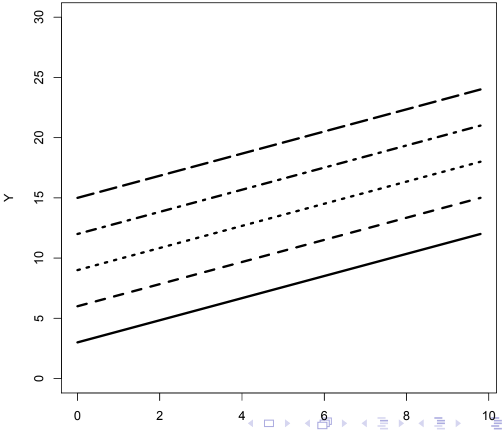
Comparison

- Pooling model: all the weight is on the individual level (groups/clusters do not matter)
- No-pooling model: all the weight is on the group level (individual attributes do not matter)
- Fixed-effects and hierarchical models lie between the two extremes
  - fixed-effects models treat all the groups *equally* and focus on estimating the within-group variation (every group has a "hard constraint")
  - hierarchical models model the group-level variation with both systematic and stochastic components (every group has a "soft constraint" that follows a probability distribution)

Hierarchical Model: Varying Intercept

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i$$

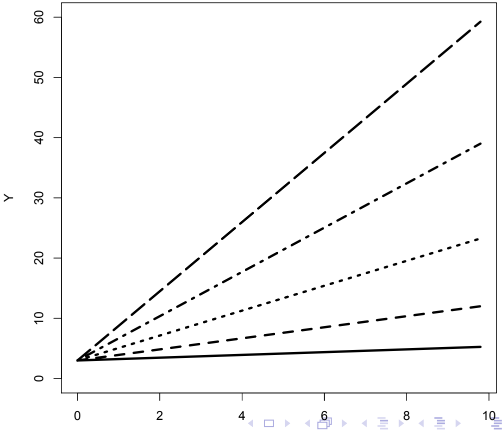
- All groups have same slope
- Each group has a different intercept



Hierarchical Model: Varying Slope

$$y_i = \alpha + \beta_{j[i]} x_i + \epsilon_i$$

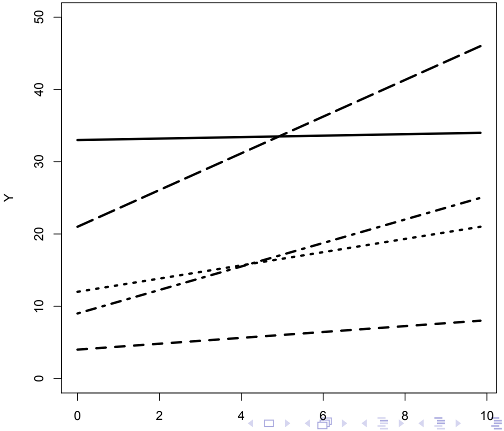
- All groups have the same intercept
- Each group has a different slope



Hierarchical Model: Varying Intercept, Varying Slope

$$y_i = \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i$$

- All groups have different slopes
- Each group has a different intercept
- Provides interactions between X and the group designations
- Also, easy to extend to other group-level variables



Basic Hierarchical Analysis with Varying Intercepts

- Hierarchical models have a "soft constraint" applied to α_j: they are assigned a probability distribution similar to other parts of the systematic component

$$\alpha_j \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2), \quad \text{for } j = 1, \ldots, J$$

  with their mean μ_α and the standard deviation σ_α estimated from the data
- This has the effect of pulling estimates of α_j towards the mean μ_α, but not all the way (*partial pooling* so to speak)
- As $\sigma_\alpha \to \infty$, the soft constraint does nothing and the model becomes a no-pooling one (all explained by group indicators)
- As $\sigma_\alpha \to 0$, they pull the estimates all the way to zero (complete pooling, no variation at group level)

Building More Complex Hierarchical Models

- Group level predictors are included in the second level of the model

$$\alpha_j \sim \mathcal{N}(\gamma_0 + \gamma_1 z_j, \sigma_\alpha^2)$$

which we can re-write as

$$\alpha_j = \gamma_0 + \gamma_1 z_j + \eta_j, \quad \text{with } \eta_j \sim \mathcal{N}(0, \sigma_\alpha^2)$$

- The errors in this model represent variation *among groups* that is not explained by the local/individual and group level predictors

Putting Together A Hierarchical Model

- Let's add the individual level predictor x_i to the model
- We use the formulation

$$y_i \sim \mathcal{N}(\alpha_{j[i]} + \beta x_i, \sigma_y^2), \quad \text{for } i = 1, \dots, n$$
$$\alpha_j \sim \mathcal{N}(\gamma_0 + \gamma_1 z_{j[i]}, \sigma_\alpha^2), \quad \text{for } j = 1, \dots, J$$

- Essentially, it is a two-step model, except that both steps are fitted at once.

Varying-intercepts and Varying-slope Hierarchical Model

Model can be written as the following:

$$
\begin{aligned}
y_i &\sim \mathcal{N}(\alpha_{j[i]} + \beta_{j[i]} x_i, \sigma_y^2), \quad \text{for } i = 1, \ldots, n \\
\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right), \quad \text{for } j = 1, \ldots, J
\end{aligned}
$$

It allows variation in both the α_j and the β_j, as well as a between-group correlation parameter ρ

Hierarchical Model based on GLM

Our example is based on OLS liner regression, but the same approach applies to any type of generalized linear model (despite that estimation of hierarchical GLM could be much more computational challenging):

$$
\begin{aligned}
y_i &\sim f(\alpha_{j[i]} + \beta_{j[i]} x_i, \sigma_y^2), \quad \text{for } i = 1, \ldots, n \\
\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} &\sim f\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right), \quad \text{for } j = 1, \ldots, J
\end{aligned}
$$

When to Use Hierarchical Models

- With only one or two groups, multilevel reduces to classical regression (e.g. pooling or fixed-effects)
  - Typically, in this situation you would would run a classical regression with an indicator (i.e. an indicator for gender rather than a multilevel model with male and female groups)
- When your interest mostly lies in identifying the causal relationship at the individual level within each group, fixed-effects might be more helpful
  - e.g. you want to examine whether economic crises increase the chance of democratic backslide, with a county-year dataset
- Use hierarchical models when you care about the variation between groups
  - e.g. you want to examine whether economic crises make different political impacts in different regions in the world, with a county-year dataset

Coming Up

- Lab tutorial on hierarchical model this week
- Review session next week