What are Multilevel Models, and Why Do We Fit Them?

Brady T. West

Overview

Fitting statistical models to **dependent data**
observations ***correlated*** due to feature of study design

Overview

Fitting statistical models to **dependent data**
observations ***correlated*** due to feature of study design

- Several observations collected at one time point from **sampled clusters** of analytic units (neighborhoods, schools, clinics, etc.)
- Several observations collected over time from **same individuals** in **longitudinal** studies

Overview

Fitting statistical models to **dependent data**
observations ***correlated*** due to feature of study design

- Several observations collected at one time point from **sampled clusters** of analytic units (neighborhoods, schools, clinics, etc.)
- Several observations collected over time from **same individuals** in **longitudinal** studies

Models need to reflect the **correlations**!

What are Multilevel Models?

General class of statistical models to **model dependent data**, where **observations within a randomly sampled cluster may be correlated** with each other.

What are Multilevel Models?

General class of statistical models to **model dependent data**, where **observations within a randomly sampled cluster may be correlated** with each other.

Coefficients allowed to **randomly vary** across **randomly sampled** higher-level clusters

What are Multilevel Models?

Coefficients allowed to **randomly vary** across **randomly sampled** higher-level clusters

Longitudinal Study Example

Predictor Time → **Outcome of Interest**

intercept and slope allowed to randomly vary across randomly sampled subjects

→ each subject have own unique intercept and slope!

What are Multilevel Models?

Estimate **regression parameters** describing overall relationships of predictors with outcomes

+

What are Multilevel Models?

Estimate **regression parameters** describing overall relationships of predictors with outcomes

+

Estimate **variance** of coefficients allowed to vary across clusters at higher levels

What are Multilevel Models?

Estimate **regression parameters** describing overall relationships of predictors with outcomes

+

Estimate **variance** of coefficients allowed to vary across clusters at higher levels

→ Expand types of inferences:

- What are the relationships between predictors and outcomes?
- How variable are coefficients in larger population from which clusters (schools, clinics, etc.) were randomly sampled? **NEW!**
- Can we explain that variance with cluster-level variables? **NEW!**

What are Multilevel Models?

Q: What changes allow coefficients to randomly vary?

A: **Random effects** of higher-level, randomly sampled clusters!

Level 1: $$y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + e_{ij}$$

Random coefficients (**not parameters!**)

Level 2: $$\beta_{0j} = \beta_0 + u_{0j}$$
$$\beta_{1j} = \beta_1 + u_{1j}$$

What are Multilevel Models?

Level 1: $y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + e_{ij}$

Level 2:
$$\beta_{0j} = \beta_0 + u_{0j}$$
$$\beta_{1j} = \beta_1 + u_{1j}$$

Random effects (*u*) allow each cluster denoted by *j* to have unique coefficients!

Random effects are **random variables:** values for different clusters assumed to be random (depending on which clusters randomly sampled!) from normal distribution with mean 0 and some variance.

Interested in **estimating that variance**!

What are Multilevel Models?

- Explicit inclusion of random effects → observations coming from same cluster **correlated** with each other statistically.

What are Multilevel Models?

- Explicit inclusion of random effects → observations coming from same cluster **correlated** with each other statistically.
- Without random effects, assuming observations from same cluster are **independent**

What are Multilevel Models?

- Explicit inclusion of random effects → observations coming from same cluster **correlated** with each other statistically.
- Without random effects, assuming observations from same cluster are **independent**
- Accounting for correlations often **substantially improves model fit** when working with dependent data!

What are Multilevel Models?

- Decomposing unexplained variance in given outcome into **between-** and **within-cluster** variance that isn't accounted for by predictors.

What are Multilevel Models?

- Decomposing unexplained variance in given outcome into **between-** and **within-cluster** variance that isn't accounted for by predictors.
- **How much** of unexplained variance due to **between-cluster variance** in intercepts or slopes for given model?

Key research question!

What are Multilevel Models?

- Decomposing unexplained variance in given outcome into **between-** and **within-cluster** variance that isn't accounted for by predictors.
- **How much** of unexplained variance due to **between-cluster variance** in intercepts or slopes for given model?

Key research question!

- **Need explicit research interest in estimating variances of these random coefficients**; otherwise, should consider other models for dependent data.

Why do we Fit Multilevel Models?

All need to be true to warrant multilevel modeling…

- **Data organized into clusters** (clinics, subjects, etc.) where several **correlated** observations are collected from each cluster.

Why do we Fit Multilevel Models?

All need to be true to warrant multilevel modeling…

- **Data organized into clusters** (clinics, subjects, etc.) where several **correlated** observations are collected from each cluster.
- **Clusters randomly sampled** from larger population of clusters *(variables like gender, race/ethnicity, etc. are **NOT** cluster variables),* and wish to make inference about that larger population.

Why do we Fit Multilevel Models?

All need to be true to warrant multilevel modeling…

- **Data organized into clusters** (clinics, subjects, etc.) where several **correlated** observations are collected from each cluster.
- **Clusters randomly sampled** from larger population of clusters *(variables like gender, race/ethnicity, etc. are* ***NOT*** *cluster variables),* and wish to make inference about that larger population.
- Wish to **explicitly model correlation** of observations within same cluster.

Why do we Fit Multilevel Models?

All need to be true to warrant multilevel modeling…

- **Data organized into clusters** (clinics, subjects, etc.) where several **correlated** observations are collected from each cluster.
- **Clusters randomly sampled** from larger population of clusters *(variables like gender, race/ethnicity, etc. are* ***NOT*** *cluster variables),* and wish to make inference about that larger population.
- Wish to **explicitly model correlation** of observations within same cluster.
- Have **explicit research interest** in estimating between-cluster variance in selected regression coefficients in our model.

Why do we Fit Multilevel Models?

- Have **explicit research interest** in estimating between-cluster variance in selected regression coefficients in our model…

Examples:

How much of unexplained variance among hospitals in mean patient satisfaction is due to the size of the hospital?

How much variance is there in long-term trends of substance use for a sample of drug users?

Why do we Fit Multilevel Models?

Advantages over other approaches for dependent data:

- Estimate **one parameter** representing variance of given random coefficient across clusters, rather than **unique regression coefficient for every possible cluster.** **More Efficient!**

- Clusters with smaller sample sizes do not have as pronounced of an effect on variance estimate as larger clusters; their effects shrink toward overall mean of outcome when using random effects.

What are Multilevel Models?

Estimating variance in given random coefficient across higher-level clusters → can add cluster-level predictors to those Level-2 equations for random coefficients, and explain variance in random effects!

Example: y = outcome, x = age, t = time point, i = subject

Level 1: $$y_{ti} = \beta_{0i} + \beta_{1i} x_{1ti} + e_{ti}$$

Level 2: $$\beta_{0i} = \beta_{00} + \beta_{01} T_i + u_{0i}$$
$$\beta_{1i} = \beta_{10} + \beta_{11} T_i + u_{1i}$$

What are Multilevel Models?

Example: y = outcome, x = age, t = time point, i = subject

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i} x_{1ti} + e_{ti}$

Level 2: $\beta_{0i} = \beta_{00} + \beta_{01} T_i + u_{0i}$

$\beta_{1i} = \beta_{10} + \beta_{11} T_i + u_{1i}$

Add regression parameters for subject-level covariate T (treatment), to **explain variance** in random intercepts and random slopes!

View Level 2 equations for random coefficients like mini-regression models; adding cluster-level predictors like T → explain variance in random effects denoted by *u*

Test hypotheses about regression parameters for T → **"45% of between-subject variance in the age – Y relationship is due to T!"**

What's Next?

- **Visualize** ideas multilevel modeling **online**!
- **Details** about fitting multilevel models to *different kinds of dependent variables*
- **Examples**

Remember: *need explicit research interest in estimating between-cluster variance in regression coefficients.*
Other modeling approaches for dependent (correlated) data don't need random effects!