ICPSR Summer Program in Quantitative Methods of Social Research

Sentencing and Other Federal Case Data Analysis

Brian D. Johnson
University of Maryland

© Brian Johnson 2009
Conceptual Rationale for Multilevel Modeling
Multilevel Models - Overview

- Introduction to Multilevel Models
  - Notes on Terminology
  - The Many Uses of Multilevel Models
  - Multilevel Data Structures
  - Conceptual and Theoretical Justification

- Statistical Overview
  - Single vs. Multilevel Regression
  - Building the Multilevel Model
    - Null Model, Random Intercepts, & Random Coefficients
  - Extensions of the Multilevel Model
    - Generalized Linear Models, 3 Level Models, Data Over Time

- HLM Data Analysis with USSC Data
  - Reading Data into the Program
  - The Unconditional Model
  - Basic Two Level Analysis
  - Generalized Linear Models
  - The Three Level Model
Introduction to Multilevel Models

A Model By Any Other Name?
- Multilevel Models are Statistical Models for Nested Data
  - Regression Analyses for Multiple Levels of Analysis (That’s It!)

Multiple Monikers
- Multilevel Models (MLM)
- Hierarchical Linear Models (HLM)
- Nested Models
- Mixed Models

Multiple Programs
- General Multilevel Programs (HLM, MLwinN)
- Specialized Multilevel Programs (aML, WINBUGS)
- Common Stat Packages (STATA, SAS, SPSS)
- Other Options (MIXed up suite, R, LIMDEP, M+, S+, SYSTAT)
## Multilevel Analyses in *Criminology* 2005-08

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Year</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson, Ulmer and Kramer</td>
<td>2008</td>
<td>Federal Guidelines Departures</td>
</tr>
<tr>
<td>Xie and McDowall</td>
<td>2008</td>
<td>Residential Turnover and Victimization</td>
</tr>
<tr>
<td>Mears, Wang, Hay and Bales</td>
<td>2008</td>
<td>Social Context and Recidivism</td>
</tr>
<tr>
<td>Zhang, Messner and Liu</td>
<td>2007</td>
<td>Crime Reporting in China</td>
</tr>
<tr>
<td>Kreager</td>
<td>2007</td>
<td>School Violence and Peer Acceptance</td>
</tr>
<tr>
<td>Wilcox, Madensen and Tillyer</td>
<td>2007</td>
<td>Guardianship and Burglary Victimization</td>
</tr>
<tr>
<td>Chiricos, Barrick, Bales, and Bontranger</td>
<td>2007</td>
<td>Labeling and Felony Recidivism</td>
</tr>
<tr>
<td>Osgood and Shreck</td>
<td>2007</td>
<td>Stability and Specialization in Violence</td>
</tr>
<tr>
<td>Rosenfeld, Fornango, and Rengifo</td>
<td>2007</td>
<td>Order-Maintenance Policing and Crime</td>
</tr>
<tr>
<td>Bernburg and Thorlindsson</td>
<td>2007</td>
<td>Community Structure and Delinquency</td>
</tr>
<tr>
<td>Warner</td>
<td>2007</td>
<td>Social Context and Calls to Police</td>
</tr>
<tr>
<td>Hay and Forrest</td>
<td>2006</td>
<td>The Stability of Self-Control</td>
</tr>
<tr>
<td>Doherty</td>
<td>2006</td>
<td>Self-Control, Social Bonds, and Desistence</td>
</tr>
<tr>
<td>Griffin and Wooldredge</td>
<td>2006</td>
<td>Sex Disparities in Imprisonment</td>
</tr>
<tr>
<td>Sampson, Laub, and Wimer</td>
<td>2006</td>
<td>Marriage and Crime Reduction</td>
</tr>
<tr>
<td>Ulmer and Bradley</td>
<td>2006</td>
<td>Trial Penalties</td>
</tr>
<tr>
<td>Johnon</td>
<td>2006</td>
<td>Judge and Court Context in Sentencing</td>
</tr>
<tr>
<td>Kubrin and Stewart</td>
<td>2006</td>
<td>Neighborhood Context and Recidivism</td>
</tr>
<tr>
<td>Simons, Simons, Burt, Brody, and Cutrona</td>
<td>2005</td>
<td>Collective Efficacy, Parenting and Delinquency</td>
</tr>
<tr>
<td>Slocum, Simpson, and Smith</td>
<td>2005</td>
<td>Strain and Offending</td>
</tr>
<tr>
<td>Wright and Beaver</td>
<td>2005</td>
<td>Parental Influence and Self-Control</td>
</tr>
<tr>
<td>Bontranger, Bales, and Chiricos</td>
<td>2005</td>
<td>Race and Adjudicated Guilt</td>
</tr>
<tr>
<td>Kleck, Sever, Li, and Gertz</td>
<td>2005</td>
<td>Perceptions of Punishment</td>
</tr>
<tr>
<td>Johnson</td>
<td>2005</td>
<td>Sentencing Guidelines Departures</td>
</tr>
</tbody>
</table>
Multilevel Analyses of Federal Sentencing

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Year</th>
<th>USSC Data</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson et al.</td>
<td>2008</td>
<td>FY1997-00</td>
<td>Variation in Use of Federal Guidelines Departures</td>
</tr>
<tr>
<td>USSC</td>
<td>2004</td>
<td>FY2001</td>
<td>Inter-Judge and Inter-District Disparity in Sentence Length</td>
</tr>
<tr>
<td>Kautt</td>
<td>2002</td>
<td>FY1999</td>
<td>Federal Sentence Lengths for Drug Trafficking</td>
</tr>
</tbody>
</table>

“Widespread disparities in the sentences imposed by Federal courts…in different parts of the country and between adjoining districts” (Frankel, 1972: 62)

“Studies … indicate that overt and more subtle forms of bias against minority defendants occur,” but only “in some social contexts” (Zatz 1987: 70)
Hierarchical Nature of Social Data

“The most pervasive fallacy of philosophic thinking goes back to neglect of context” (John Dewey, 1931)

- Multiple Levels of Analysis
  - Individual Offenders
  - Judges
  - District Courts
  - Circuit Courts
Conceptual Model of Nested Data

Judge 1: J. Judy
- Offender 1
- Offender 2
- Offender 3

Judge 2: J. Ito
- Offender 4
- Offender 5
- Offender 6

Judge 3: J. Cowell
- Offender 7
- Offender 8
- Offender 9
Data File for Nested Data Structure

- **Individual Criminal Cases**
  - (n=20)

- **Nested In Districts**
  - (n=10)

- **Nested in Circuits**
  - (n=5)
Conceptual Rationales for Multilevel Modeling

- More Realistic Portrayal of the Social World
  - Theories Suggest Micro-Effects Vary across Macro Units
- Cross Level Inferences are Problematic
  - Ecological and Individualistic Fallacies
    - Robinson (1950) Immigrant Literacy

This article assesses the link between country music and metropolitan suicide rates. Country music is hypothesized to nurture a suicidal mood through its concerns with problems common in the suicidal population, such as marital discord, alcohol abuse, and alienation from work. The results of a multiple regression analysis of 49 metropolitan areas show that the greater the airtime devoted to country music, the greater the white suicide rate. The effect is independent of divorce, southernness, poverty, and gun availability. The existence of a country music subculture is thought to reinforce the link between country music and suicide. Our model explains 51% of the variance in urban white suicide rates.
Statistical Rationales for Multilevel Modeling

- Nested Data Often Violate Traditional Regression Assumptions
  - Statistical Dependence Among Errors
    - Contexts Share Unobserved Similarities
    - Correlated Error Terms Violate OLS Assumptions
    - Misestimated Standard Errors Inflate Type I Error
  - Statistical Significance Tests
    - Individual vs. Contextual Degrees of Freedom
  - Homogeneity of Regression Coefficients
    - Regression Effects May Vary Across Contexts
  - Interaction Effects
    - Biased Estimates of Cross-Level Interactions
  - Inefficiency in Group Estimates
    - Small Within Group Sample Sizes
Statistical Advantages of Multilevel Models

- Multilevel Models Address These Limitations
  - Incorporate Multiple Levels of Analysis
    - Accommodates Both Within and Between Group Information
    - Individual Regression Equations for Each Group
    - Overall Regression for All Groups
    - Joint Estimates Weighted by Sample Size/Reliability
  - Corrected Standard Errors and Significance Tests
    - Adjusts for Correlated Error Terms
    - Adjusts for Degrees of Freedom
  - Model Heterogeneity in Regression Effects
    - Estimates of “Variance Components”
    - Random Intercepts and Coefficients
  - Cross-Level Interactions
    - Estimates Moderating Effects Across Levels of Analysis
  - Improved Estimation of Effects
    - Conditional Shrinkage and “Borrowed” Statistical Power
    - Empirical Bayes Estimates
Statistical Rationale for Multilevel Modeling
Statistical Overview

- OLS Regression vs. Multilevel Models
  - Ordinary Regression Approaches
    - 1) Pooled Analysis Ignoring Group Structure
      - Ignores Important Group-Level Variability
      - Violates Core Assumptions of OLS
    - 2) Separate Un-pooled Analysis for Each Group
      - Difficult to Manage with Large Number of Groups
      - Problematic with Small Group Sample Sizes
    - 3) Aggregate Analysis of Group Means
      - Ignores Within-Group Variation
      - Requires Large Number of Groups

- All Rely on Limited Amounts of Available Information
- Can Produce Erroneous Statistical Inferences
- Limited Ability to Answer Multilevel Research Questions
Statistical Overview

- Example: Homework and Math Scores
- Pooled Analysis Across Schools

Regression Ignoring Hierarchical Structure

\[ \text{math} = 44.074 + 3.5719 \times \text{homework} \]
Statistical Overview

- Example 1: Homework and Math Scores
- Individual School Level Regressions

Reproduced from: *Models for Clustered Data* by Carolyn J. Anderson
Statistical Overview

Example 2: Variation in Federal Sentencing

Boxplots for Months of Incarceration in 20 Federal Courts
Statistical Overview

- Variation in “Trial Penalties” Across Courts
- Heterogeneity in Regression Effects
- Need Statistical Models that Account for Group Level Differences
From Regression to Multilevel Analysis

- Statistical Models are Mathematical Models
  - Based on Assumptions about Data Generation Processes
- Mathematical Models Require Estimators
  - Classical Linear Regression uses OLS
  - Multilevel Estimators Typically Utilize Maximum Likelihood (MLE)

- MLM are Simple Extensions of Traditional Regression

- **The Traditional Regression Model**
  \[ Y_i = \beta_0 + \beta_1 X_i + r_i \]
  - Represents Expected Change in Y for a Unit Change in X
  - Two Key Model Assumptions
    - Linear Relationship
    - Independent Errors – Nothing Systematic About Residuals
  - *Both Assumptions May Be Wrong with Multilevel Data!*
Testing the Assumption of Independent Errors

- Nested Data Often Results in Group Level Dependence

- The Assumption of Independence Can be Tested
  - Estimate the OLS Regression Equation for your Outcome of Interest
  - Save the Residuals as a New Variable
  - Run Analysis of Variance (ANOVA) on Residuals by Group
  - The Null Hypothesis is that Residuals are Independent of Groups

- Analysis of Variance for Independent Errors
  - Example: Sentence Lengths Across District Courts, USSC Data 2007
    - USSC 2007 ICPSRS HLM DATA ANALYSIS.SPS

**ANCOVA Comparing Regression Residuals Across Federal District Courts**

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>690.950</td>
<td>93</td>
<td>7.430</td>
<td>24.071</td>
</tr>
<tr>
<td>Within Groups</td>
<td>19941.807</td>
<td>64610</td>
<td>.309</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>20632.757</td>
<td>64703</td>
<td>.309</td>
<td></td>
</tr>
</tbody>
</table>
From Regression to Multilevel Analysis

Consequences of Violating OLS Model Assumptions
- False Power at Both Levels of Analysis
- Less Information in Clustered Sample
- Number of Level 2 Observations Exaggerated
- Misestimated Standard Errors
  - Clustered Observations Share Unobserved Similarities
  - Results in Mean Differences in Residuals Between Groups
  - Violates Assumption of Independent Errors

The Multilevel Adaptation
- MLM Conceptually Equivalent to Regression with Extra Error Terms
  - $u_j$ = the Group Level (i.e. Level 2) Error Term, Variance, or Residual
  - The $u_j$ Parameter Accounts for Group Level Dependence
  - $r_{ij}$ Now Independent Because Group Difference Captured by $u_j$
- The Two-Level Model Format is a Simple Notational Convenience

\[ Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \]
\[ \beta_{0j} = \gamma_{00} + u_{0j} \]

OR

\[ Y_{ij} = \gamma_{00} + \beta_{1j} X_{ij} + r_{ij} + u_j \]
From Regression to Multilevel Analysis

Some Advantages of the Multilevel Model

- Results Provide $\beta$’s and Two Error Variances
- Optimal Weighting of Level 1 and Level 2 Data
- When Assumptions Met, Provides Most Precise, Most Efficient Estimator
- Standard Errors are Corrected for Clustering of Observations
- Significance Tests are Adjusted across Levels of Analysis
- Explicitly Models Variation in Level 1 Parameters Across Level 2 Units
- Allows for Inclusion of Level 1 and Level 2 Predictors
Building the Multilevel Model

Several Different Types of Multilevel Models
- Varying Complexity
  - Unconditional or Null Model – No Predictors
  - Random Intercept Model – Mean Differences Across Groups
  - Random Coefficient Model – Varying Effects Across Groups
  - Cross-Level Interaction Model – Group Moderation of Effects

A Note on Notation
- \( i \) Indexes Level 1 Units (e.g. Individuals)
- \( j \) Indexes Level 2 Units (e.g. Schools)
- \( j \) Applies to All Observations in the Level 2 Unit
- \( \beta \) = Level 1 Regression Coefficient  \( X \) = Level 1 Explanatory Variable
- \( \gamma \) = Level 2 Regression Coefficient  \( W \) = Level 2 Explanatory Variable

\[
\begin{align*}
\text{Level 1} & \quad Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \\
\text{Level 2} & \quad \beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}
\end{align*}
\]
The Unconditional Model

- **Fully Unconditional (or Null) Model**
  - No Predictors at Either Level of Analysis
  - “Unconditional” = Estimates Not Conditional on Any Predictors
  - Also Called One-Way Anova with Random Effects
  - Estimates Overall Mean of $Y_{ij}$
  - Estimates Level 1 ($\sigma^2$) and Level 2 ($\tau_{00}$) Variance Components
  - Parcels Variance Between Levels of Analysis
    - $\text{Var}(Y_{ij}) = \text{Var}(u_{0j} + r_{ij}) = \tau_{00} + \sigma^2$
  - Investigates If There Are Mean Differences to be Explained
  - Provides Baseline Comparison for Conditional Models

Level 1: $Y_{ij} = \beta_{0j} + r_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

By Substituting Level 2 into Level 1

OR

$Y_{ij} = (\gamma_{00} + u_{j}) + r_{ij}$
Random Intercept Models

- Random Intercept Models
  - Estimates Between-Group Variance in Outcome
    - E.g. Race on Sentence Length Across Districts
    - Focus is On Individual Relationships, Corrected for Clustering
  - Why Not Just Do Pooled Regression?
    - 1) Statistical Dependence
    - 2) Theoretical Interest in Group Variation
- The Level 1 Model
  - \( X_{ij} \) = Level 1 Explanatory Variable for Individual \( i \) in Group \( j \)
  - Both \( r_{ij} \) and \( u_j \) are Residuals; Can Be Reduced by \( X_{ij} \)’s
  - “Compositional” Effects at Level 2

\[
Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}
\]

\[
\beta_{0j} = \gamma_{00} + u_{0j} \quad \text{OR} \quad Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + u_j + r_{ij}
\]

\[
\beta_{1j} = \gamma_{10}
\]
Random Intercept Models

- Random Intercept Models with Level 2 Variable
  - Explains Individual and Mean Differences in Outcome
  - Individual Offender Characteristics Predict Individual Outcome
  - Aggregate Court Characteristics Predict Mean Differences

- Jointly Estimates Individual and Contextual Effects
  - Explains Variance at Level 1 and Level 2
  - Easily Extended to Case of Multiple $X_{ij}$s and Multiple $W_{j}$s

\[ Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij} \]
\[ \beta_{0j} = \gamma_{00} + \gamma_{01}W_{j} + u_{0j} \]
\[ \beta_{1j} = \gamma_{10} \]

\[ Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_{j} + u_{j} + r_{ij} \]
Building the Multilevel Model

- Random Intercept vs. Random Coefficient Models

**VARYING INTERCEPTS ONLY**

<table>
<thead>
<tr>
<th>Independent Variable (X)</th>
<th>Dependent Variable (Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

**VARYING INTERCEPTS & SLOPES**

<table>
<thead>
<tr>
<th>Independent Variable (X)</th>
<th>Dependent Variable (Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
Random Coefficient Models

- Model Variation in Effects of Level 1 Variables Across Level 2 Units
  - E.g. The Effect of Race is Different in Different District Courts

- The Random Intercept Model
  - Is There 1 Uniform Effect or Separate Effects in Each District or Circuit?
  - If Individual Effects Vary by District = Additional Source of Dependence!
  - Produces Unaccounted-For Similarity in Residuals

- Why Does it Matter?
  - Leads to Generalizability Problems
    - 75,000 vs. 89 Independent Observations About Race Effects
    - More Variability in Race Effect = Closer to 89 than 75,000 Cases
  - Standard Errors Underestimated for both OLS and Random Intercept Model
Random Coefficient Models

- Random Coefficient Specified Just Like Random Intercept
  \[ Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \]
  \[ \beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \]
  \[ \beta_{1j} = \gamma_{10} + u_{1j} \leftarrow \text{New Random Effect in the Model} \]

- \( u_{1j} \) Captures Variation in the Effect of \( X_{ij} \) on \( Y_{ij} \)
  - E.g. Trial Penalty Might Vary Across District Courts

- But this Complicates the Error Structure of the Model
  - Adds Both Variance and Covariance to the Model
  - Model Allows for Correlation Among Level 2 Error Terms

\[
\text{Var} \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix}
\]

- \( \tau_{11} \) is the new variance component for the random coefficient (\( \beta_{1j} \))
Random Coefficient Models

- To Be (Random) or Not to Be (Random)
  - Theory, Theory, Theory
    - Theory Should Guide Model Specification
  - Chi-Square Significance Tests
    - Crude Approximations but Useful Starting Point
  - Deviance Test/Likelihood Ratio Tests
    - Compare Log Likelihood Model Fit Statistics
    - Nested Models use Chi-Square Distribution
    - Degrees of Freedom = Difference in # of Parameters
  - Robust Standard Errors
    - Useful Indicator of Model Misspecification
    - Large Differences Suggests Omitted Random Effects
  - Model Estimation and Parsimony
    - Be Judicious in Specification of Random Effects
    - Dramatically Complicates Model
    - Can Make Estimates Unstable
Building the Multilevel Model (Almost Done!)

- The Fully Specified Multilevel Model
  - Random Coefficient Means Level 1 Effect Varies By Level 2 Unit
    - But Why? How Can We Explain This Variation?
  - Treat the Slope at Level 1 as an *Outcome* at Level 2
  - “Cross-Level Interaction”
    - Adds a Level 2 Predictor to Explain Variation in the Level 1 Effect
    - E.g. Maybe Trial Penalties Vary With Caseload Pressure of the District

Full Random Coefficient Model

\[
Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \\
\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \\
\beta_{1j} = \gamma_{10} + u_{1j}
\]

Cross-Level Interaction Model

\[
Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \\
\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \\
\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j}
\]

\( W_j \) is Added to the Model as a Predictor of \( B_{1j} \)
Cross Level Interaction Model

- **Different Types of Interactions**
  - Interaction of Level 1 Variables (Race*Gender)
  - Interaction of Level 1 & Level 2 Variable (South*Black)
  - Interaction of Level 2 Variables (South*%Black)

- **Why Examine Cross-Level Interactions?**
  - Exploratory Analysis
    - Try to Explain Variation in Random Coefficients
  - Confirmatory Analysis
    - Theory Predicts a Cross-Level Relationship
    - Can Test Interaction Without Significant Variation in Level 1 Effect ($\tau_{11}$)
      - Cross-Level Interaction Test is More Powerful than Variance Test
      - So Significant Interactions Can Occur w/out Random Coefficients
Centering in Multilevel Models

- Hierarchical Models and “Centering”
  - What is Centering?
    - Subtracting the Mean (Or Another Constant) from $X$ or $W$
    - The Subtracted Value (e.g. Mean) Becomes the New “Zero” Point
    - Grand Mean Centering = $(X_{ij} - \bar{X}_{..})$
    - Group Mean Centering = $(X_{ij} - \bar{X}_{.j})$
  - Why Should You Center?
    - Make Intercept Substantively Meaningful
    - Make Main Effects More Meaningful (e.g. With Interactions)
    - Reduce Collinearity (e.g. Polynomials and Interactions)
    - Facilitates Model Convergence (e.g. Nonlinear Models)
    - Simplifies Graphical Displays
  - When Should You Center?
    - Grand Mean Centering is Often Useful and Rarely Detrimental
    - Group Mean Centering Should Be Used Cautiously
      - Changes Meaning of Coefficients and Variance Components
    - Often Useful to Center Dummy Variables
    - Often Useful to Center Level 2 Variables As Well
Multilevel Model Questions

- A Lot to Conceptually Swallow!
  - Questions???
  - Clarifications???
  - Comments???
Basic Multilevel Analysis of USSC Data
Section 3: Applying HLM to USSC Data

- **HLM Software**
  - A Free 15 Day or Student Version of HLM is Available at: http://www.ssicentral.com/hlm/student.html

- *One distinguishing feature of recent research on sentencing is a deeper appreciation for the salience of macro-social contexts* (Sampson & Lauritsen, 1997: 349)

- **USSC Data**
  - **Level 1 Data**
    - 25,000 Offenders
  - **Level 2 Data**
    - 89 District Courts
  - **Level 3 Data**
    - 11 Circuit Courts
Using the HLM Program

- The Full Complexity of the Multilevel Model
The Importance of District Context

- Johnson (2008)
  - USSC FY1997-FY2000
  - Demonstrate with These Data

"I sentence you to ten years — thank you for choosing the Fourth District Court."

<table>
<thead>
<tr>
<th>Sentence Length</th>
<th>California South 25.5 months</th>
<th>Florida North 120.2 months</th>
<th>Total 57.9 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downward Departures</td>
<td>West Virginia North 2.2%</td>
<td>Arizona 64.3%</td>
<td>Total 16.2%</td>
</tr>
<tr>
<td>5K1 Departures</td>
<td>Utah 7.1%</td>
<td>New York North 52.5%</td>
<td>Total 20.8%</td>
</tr>
<tr>
<td>Upward Departures</td>
<td>Oklahoma East 0%</td>
<td>Wisconsin West 4.7%</td>
<td>Total 0.7%</td>
</tr>
</tbody>
</table>
Using the HLM Program

- Reading Data into HLM
  - Prepare the Data Beforehand
  - No Data Transformation in HLM
  - Sensitive to Missing Cases
  - Data Must Be Properly Sorted
    - Sort Both Files by Level 2 ID
  - SPSS Bug in HLM 6.02
    - Read Data in as “Anything Else”

![Table and Diagram]

The other commandments didn't have unique IDs.
Using the HLM Program

The Level 1 Dataset
- Data Sorting Procedure
- SORT CASES BY DISTRICT (A).

The Level 2 Dataset
- Create “Aggregate Measures” Using Level 1 Data
  - E.g. AGGREGATE
    /OUTFILE='C:\...\USSC 2007 Level 2\USSC LEVEL 2.0.sav'
    /BREAK=DISTRICT
    /perdrug = MEAN(Drugs) /trialrat = MEAN(Trial2)
    /subasrat = MEAN(SubAsist) /cases=N.
- Add “Global Measures” from Other Data Sources
  - E.g. Population Demographic Measures from Census Etc.
  - E.g. Federal Prison Capacity?
- Open USSC Level 2 25K

*The prisons are overcrowded, so I'm sentencing you to twenty years in Detroit.*
Using the HLM Program

Reading Data into HLM

Go to File; Make New MDM File; Stat Package Input

- HLM2 is used for two-level linear and non-linear (HGLM) models.
- HLM3 is used for three-level linear and non-linear (HGLM) models.
- HMLM is used for multivariate normal models (MNM) with incomplete data.
- HMLM2 is used for MNM with persons nested in higher-level units.
- HCM2 is used for cross-classified models with two higher level classifications.
Reading Data into HLM

- Click Browse to Select Level 1 and Level 2 Data Files
  - Each Level of Analysis Is a Separate Data File
  - Each Data File Must Be Sorted By Unique Identifier
  - HLM Reads In Two Data Sets and Merges By ID Variable

SPSS BUG!

READ IN DATA AS “ANYTHING ELSE”
Reading Data into HLM

- Data Cannot Be Manipulated in HLM
  - All Coding Must Be Done Beforehand
  - Address Missing Data Beforehand
  - Limit Variable Names to 8 (12) Characters
- Data Files Must Be Sorted by Level 2 ID
  - Both Files Sorted by “District”

### Level 1 Data File

<table>
<thead>
<tr>
<th>circuit</th>
<th>district</th>
<th>district</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>maine</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>maine</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>maine</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>massachusetts</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>massachusetts</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>massachusetts</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>new hampshire</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>new hampshire</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>rhode island</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>rhode island</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>connecticut</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>new york north</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>new york north</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>new york north</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>new york north</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>new york north</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>new york east</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>new york east</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>new york east</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>new york east</td>
</tr>
<tr>
<td>21</td>
<td>2</td>
<td>new york east</td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>new york south</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>new york south</td>
</tr>
<tr>
<td>24</td>
<td>2</td>
<td>new york south</td>
</tr>
</tbody>
</table>

### Level 2 Data File

<table>
<thead>
<tr>
<th>moncirc</th>
<th>district</th>
<th>district</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>maine</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>massachusetts</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>new hampshire</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>rhode island</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>connecticut</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>new york north</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>new york east</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>new york south</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>new york west</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>vermont</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>delaware</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>new jersey</td>
</tr>
</tbody>
</table>
### Reading Data into HLM

- **Choose Level 1 Variables**
  - District = ID
  - Others = in MDM

- **Choose Level 2 Variables**
  - District = ID
  - Others = in MDM

- **Save HLM Data Files**
  - Name MDM File
  - Save Template (mdmt file)
  - Make MDM
  - Check Descriptive Statistics
Using the HLM Program

- The Unconditional (Null) Model
  - Choose Dependent Variable (Sentence Length)
  - No Predictors in the Null Model
  - Default Estimator is Restricted Maximum Likelihood (RML)
  - Run Analysis; Save and Run
Using the HLM Program

- The Unconditional (Null) Model
  - Interpreting the Unconditional Model
    - Model Intercept = Mean Sentence Length Across Courts
    - Level 1 Variance ($r_{ij}$) = Between Individual Variation in Sentence Length
    - Level 2 Variance ($u_{oj}$) = Between District Variation in Sentence Length
    - Intraclass Correlation rho ($\rho$) = $\tau_{00}/(\sigma^2 + \tau_{00}) = 267 / (267 + 4630) = .055$
    - 5.5% of Total Variation in Sentence Length is Between Federal Districts

Unconditional Model Output from HLM

Unconditional Model of Federal Sentence Lengths

Sentence Length in Months

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>$b$</th>
<th>S.E.</th>
<th>df</th>
<th>T-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>52.54</td>
<td>1.83</td>
<td>88</td>
<td>28.76</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>$s^2$</th>
<th>S.D.</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p-value</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 ($r_{ij}$)</td>
<td>4630.04</td>
<td>68.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2 ($u_{oj}$)</td>
<td>267.08</td>
<td>16.34</td>
<td>88</td>
<td>1420.32</td>
<td>0.00</td>
<td>0.055</td>
</tr>
</tbody>
</table>

Deviance = 282173.7
Parameters = 2
N=25,000
Using the HLM Program

- **Random Intercept Models**
  - Add Level 1 Predictor (Severity) Grand Mean Centered to the Model
  - $(u_{ij})$ Allows Intercept (Group Means) to Vary Across Districts
  - Effects Can be Added and Removed By Clicking Them
  - An Increase in Severity Increases Length by 5.6 Months
  - Variance Components are Now Residuals

### Random Intercept Model Output from HLM

#### Random Coefficient Model of Federal Sentence Lengths

**Sentence Length in Months**

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>b</th>
<th>S.E.</th>
<th>df</th>
<th>T-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept $(\gamma_{00})$</td>
<td>51.00</td>
<td>1.08</td>
<td>88</td>
<td>48.20</td>
<td>0.00</td>
</tr>
<tr>
<td>Severity $(\beta_1)$</td>
<td>5.56</td>
<td>0.19</td>
<td>24998</td>
<td>29.27</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>$s^2$</th>
<th>S.D.</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 $(r_{ij})$</td>
<td>2228.65</td>
<td>47.21</td>
<td>88</td>
<td>1420.32</td>
<td>0.00</td>
</tr>
<tr>
<td>Level 2 $(u_{ij})$</td>
<td>93.15</td>
<td>9.65</td>
<td>88</td>
<td>1420.32</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Deviance = 263875.9  
Parameters = 2  
N=25,000

**Mixed Model**

\[
\text{SENTLENG} = \gamma_{00} + \gamma_{10} * \text{SEVERITY} + u_{0} + r
\]
Using the HLM Program

- Random Effects in the Null and Random Intercept Models
  - Severity Reduces Both Level 1 and Level 2 Variance

<table>
<thead>
<tr>
<th>Unconditional Model</th>
<th>of Federal Sentence Lengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Effects</td>
<td>$s^2$</td>
</tr>
<tr>
<td>Level 1 ($r_{ij}$)</td>
<td>4630.04</td>
</tr>
<tr>
<td>Level 2 ($u_{0j}$)</td>
<td>267.08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Coefficient Model</th>
<th>of Federal Sentence Lengths (with Severity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Effects</td>
<td>$s^2$</td>
</tr>
<tr>
<td>Level 1 ($r_{ij}$)</td>
<td>2228.65</td>
</tr>
<tr>
<td>Level 2 ($u_{0ij}$)</td>
<td>93.15</td>
</tr>
</tbody>
</table>

- Level 1 & Level 2 $R^2$

$$R^2_{Lev1} = \frac{\sigma^2_{unc} - \sigma^2_{cond}}{\sigma^2_{unc}} = \frac{4630 - 2229}{4630} = .52$$

$$R^2_{Lev2} = \frac{\tau_{unc} - \tau_{cond}}{\tau_{unc}} = \frac{267 - 93}{267} = .65$$

- Severity Explains 52% of Individual Variation in Sentencing
- It Explains 65% of District Level Variation in Sentencing
- The Latter is a “Compositional” Effect
- Inter-District Variation Still Significant
Using the HLM Program

- Random Intercept Models
  - Level 2 Aggregate Analysis
    - Add Level 2 Predictor (South) with No Level 1 Controls
    - Only Explains Mean Differences Across Level 2 Units
    - Mean Sentences in Southern Districts are 12.6 Months Longer

\[
R^2_{\text{lev}1} = \frac{4630 - 4630}{4630} = .00
\]

\[
R^2_{\text{lev}2} = \frac{267 - 233}{267} = .13
\]

Random Intercept Model Output from HLM

Random Coefficient Model of Federal Sentence Lengths

<table>
<thead>
<tr>
<th>Sentence Length in Months</th>
<th>Fixed Effects</th>
<th>b</th>
<th>S.E.</th>
<th>df</th>
<th>T-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ((\gamma_{00}))</td>
<td>52.45</td>
<td>1.69</td>
<td>87</td>
<td></td>
<td>30.96</td>
<td>0.00</td>
</tr>
<tr>
<td>South ((\gamma_{01}))</td>
<td>12.63</td>
<td>3.75</td>
<td>87</td>
<td></td>
<td>3.37</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>(s^2)</th>
<th>S.D.</th>
<th>df</th>
<th>(\chi^2)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 ((r_0))</td>
<td>4630.15</td>
<td>68.04</td>
<td></td>
<td>1420.32</td>
<td>0.00</td>
</tr>
<tr>
<td>Level 2 ((u_0))</td>
<td>232.54</td>
<td>15.25</td>
<td>88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Deviance = 263875.9
Parameters = 2
N=25,000
Using the HLM Program

- The Full Random Intercept Model
  - Including Both Level 1 (Severity) and Level 2 (South) Predictors
    - Severity Can Explain Both Level 1 and Level 2 Variance
    - South Can Explain Only Level 2 Residual Variance
    - Severity Still Increases Sentence Length by 5.56 Months
    - Controlling for Mean Differences in Severity, South Increases Sentence Length by 7 Additional Months

Random Coefficient Model of Federal Sentence Lengths

Sentence Length in Months

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>b</th>
<th>S.E.</th>
<th>df</th>
<th>T-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>50.93</td>
<td>1.00</td>
<td>87</td>
<td>30.96</td>
<td>0.00</td>
</tr>
<tr>
<td>South ($\gamma_{01}$)</td>
<td>7.07</td>
<td>2.25</td>
<td>87</td>
<td>3.37</td>
<td>0.00</td>
</tr>
<tr>
<td>Severity ($\beta_1$)</td>
<td>5.56</td>
<td>0.19</td>
<td>24997</td>
<td>29.30</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>$s^2$</th>
<th>S.D.</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 ($\epsilon_i$)</td>
<td>2228.69</td>
<td>47.21</td>
<td>87</td>
<td>1145.49</td>
<td>0.00</td>
</tr>
<tr>
<td>Level 2 ($u_j$)</td>
<td>82.45</td>
<td>9.08</td>
<td>87</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Deviance = 263860.9
Parameters = 2
N=25,000
Using the HLM Program

- The Random Coefficient Model
  - Deviance Tests and $\chi^2$ Value Indicate Significant Variation in Severity
  - So Effect of Severity on Sentence Length Differs Across District Courts
  - Model Variation in This Effect with New Random Parameter ($u_{1j}$)

- How Much Does the Effect Vary?
  - One Standard Deviation in Effect
  - $(5.66-1.2) = 4.46$
  - $(5.66+1.2) = 6.86$
  - Across 68% of Courts

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>$s^2$</th>
<th>S.D.</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 ($r_{ij}$)</td>
<td>2098.66</td>
<td>45.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2 ($u_0$)</td>
<td>78.75</td>
<td>8.87</td>
<td>88</td>
<td>1228.16</td>
<td>0.00</td>
</tr>
<tr>
<td>Severity ($u_1$)</td>
<td>1.44</td>
<td>1.20</td>
<td>88</td>
<td>1643.69</td>
<td>0.00</td>
</tr>
</tbody>
</table>
The Random Coefficient Model

- The Deviance and $\chi^2$ Test Statistics
  - Used to Test for Significance of Random Coefficient
  - Deviance Limited to Nested Models with Extra Variance Component
  - Differences in Deviance = $263876 - 262530 = 1346$ with 2 df
  - $\chi^2$ Approximation = 1644
  - $\chi^2_{crit} = 5.99$ Both Reject Null – Random Coefficient Improves Model Fit

**Random Intercept**

Deviance = 263875.9  
Parameters = 2

**Random Coefficient**

Deviance = 262530.1 
Parameters = 4

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>$s^2$</th>
<th>S.D.</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 ($r_{ij}$)</td>
<td>2098.66</td>
<td>45.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2 ($u_0$)</td>
<td>78.75</td>
<td>8.87</td>
<td>88</td>
<td>1228.16</td>
<td>0.00</td>
</tr>
<tr>
<td>Severity ($u_1$)</td>
<td>1.44</td>
<td>1.20</td>
<td>88</td>
<td>1643.69</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Using the HLM Program

- The Full Random Coefficient Model
  - Includes Random Intercept and Random Coefficient
  - Includes Both Level 1 and Level 2 Predictors
  - Easily Extended to Case of Multiple $X_{ij}s$ and Multiple $W_{ij}s$
- Comparing the Full Random Intercept and Random Coefficient Models
  - RIM Deviance $263,861$ – RCM Deviance $262,521 = \chi^2 = 1340; 2 \text{ df}$

Random Coefficient Model of Federal Sentence Lengths

Sentence Length in Months

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>b</th>
<th>S.E.</th>
<th>df</th>
<th>T-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>49.58</td>
<td>0.97</td>
<td>87</td>
<td>49.14</td>
<td>0.00</td>
</tr>
<tr>
<td>South ($\gamma_{01}$)</td>
<td>3.42</td>
<td>1.43</td>
<td>87</td>
<td>3.37</td>
<td>0.02</td>
</tr>
<tr>
<td>Severity ($\beta_1$)</td>
<td>5.66</td>
<td>0.13</td>
<td>88</td>
<td>42.45</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>$\sigma^2$</th>
<th>S.D.</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 ($r_0$)</td>
<td>2098.56</td>
<td>45.81</td>
<td>87</td>
<td>1130.14</td>
<td>0.00</td>
</tr>
<tr>
<td>Level 2 ($u_0$)</td>
<td>71.65</td>
<td>8.46</td>
<td>87</td>
<td>1644.40</td>
<td>0.00</td>
</tr>
<tr>
<td>Severity ($u_1$)</td>
<td>1.45</td>
<td>1.21</td>
<td>88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Deviance = 262521
Parameters = 4
Chi-square statistic = 1339.89851
$N_1=25,000$
Number of degrees of freedom = 2
$N_2=89$
P-value = 0.00
Using the HLM Program

- Estimating Cross-Level Interactions
  - Specify the Full Random Coefficient Model
  - Add Select Interaction Terms to Test Cross-Level Moderation
    - E.g. Does Severity Have a Stronger Effect in Southern Districts?
    - Does Being Black Have a Stronger Effect in the South?

Random Coefficient Model of Federal Sentence Lengths

Sentence Length in Months

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>b</th>
<th>S.E.</th>
<th>df</th>
<th>T-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>49.57</td>
<td>0.96</td>
<td>87</td>
<td>49.14</td>
<td>0.00</td>
</tr>
<tr>
<td>South ($\gamma_{01}$)</td>
<td>6.26</td>
<td>1.96</td>
<td>87</td>
<td>3.37</td>
<td>0.02</td>
</tr>
<tr>
<td>Severity ($\beta_1$)</td>
<td>5.66</td>
<td>0.13</td>
<td>88</td>
<td>42.45</td>
<td>0.00</td>
</tr>
<tr>
<td>South*Severity ($\gamma_{11}$)</td>
<td>0.58</td>
<td>0.28</td>
<td>87</td>
<td>2.07</td>
<td>0.04</td>
</tr>
</tbody>
</table>

There is a Small but Significant Interaction

Offense Severity Has a Stronger Effect in the South

Cross-Level Interaction (Severity*South)
Advanced Multilevel Analysis of USSC Data
Advanced HLM Modeling

- Hierarchical Generalized Linear Models
  - Models for Non-Normal Dependent Variables
    - E.g. Binary Dependent Variable (Prison vs. No Prison)
  - “Basic Settings” – Select “Bernoulli” for Binary Outcome
    - Applies the Logistic Link Function to the Dependent Variable
    - Can Also Specify Binomial, Poisson, Ordinal/Multinomial Distribution
  - Outcome is Now the Log Odds of Receiving a Prison Sentence

![Image of HLM modeling interface]
Advanced HLM Modeling

- Hierarchical Logistic Regression Models
  - Analytical Steps are Same Beginning with Null Model
  - Some Important Differences from Linear Model
    - No Level 1 Variance Component for Binary Outcome
    - Default Estimator is “Penalized” Quasi Likelihood
      - Can Be Useful to Relax Convergence Criteria for Exploratory Analysis

Unconditional Logistic Model

\[
\text{LEVEL 1 MODEL (bold: group-mean centered)}
\]
\[
\text{Prob(PRISON=1|\beta)} = \varphi
\]
\[
\log[\varphi/(1 - \varphi)] = \eta
\]
\[
\eta = \beta_0
\]

\[
\text{LEVEL 2 MODEL (bold italic: grand-mean)}
\]
\[
\beta_0 = \gamma_{00} + u_0
\]
**Advanced HLM Modeling**

- Interpreting Hierarchical Logistic Regression Models
  - **The Unconditional Model**
    - No Simple Intraclass Correlation
    - Unit Specific vs. Population Average Results
    - Intercept \((\gamma_{00})\) is the Average Log Odds of Imprisonment Across Districts
    - Likelihood of Imprisonment Varies Significantly Across Districts \((u_{0j})\)
  - **The Random Intercept Model**
    - Coefficients are Equivalent to Logistic Regression
    - Female Offenders are .28 Times as Likely as Males to Be Incarcerated

---

### Random Intercept Logistic Model

Unconditional Logistic Model of Federal Prison Sentences

Prison vs. No Prison (Unit-Specific Model with Robust Standard Errors)

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>b</th>
<th>S.E.</th>
<th>df</th>
<th>T-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ((\gamma_{00}))</td>
<td>1.55</td>
<td>0.06</td>
<td>88</td>
<td>28.11</td>
<td>0.00</td>
</tr>
<tr>
<td>(\hat{\gamma}_{ij}^{10})</td>
<td>.23</td>
<td>0.48</td>
<td>88</td>
<td>826.76</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Random Intercept Logistic Model of Federal Prison Sentences

Prison vs. No Prison (Unit-Specific Model with Robust Standard Errors)

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>b</th>
<th>S.E.</th>
<th>df</th>
<th>T-ratio</th>
<th>p-value</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ((\gamma_{00}))</td>
<td>1.64</td>
<td>0.06</td>
<td>88</td>
<td>28.98</td>
<td>0.00</td>
<td>0.28</td>
</tr>
<tr>
<td>Female ((\gamma_{10}))</td>
<td>-1.27</td>
<td>0.06</td>
<td>24998</td>
<td>-22.44</td>
<td>0.00</td>
<td>0.28</td>
</tr>
<tr>
<td>(\hat{\gamma}_{ij}^{10})</td>
<td>.23</td>
<td>0.48</td>
<td>88</td>
<td>776.32</td>
<td>0.00</td>
<td>58</td>
</tr>
</tbody>
</table>
Advanced HLM Modeling

- Interpreting Hierarchical Logistic Regression Models
  - The Full Random Coefficient Model with Level 1 & Level 2 Predictors
  - Test Random Coefficients at Level 1
    - The Effects of Both Offense Severity and Gender Vary Across Districts
    - Controlling for Severity, Females are .44 Times as Likely to be Imprisoned
  - Add Level 2 Predictors
    - Small but Significant Effect for Court Size
  - Level 2 Variance Increased from Unconditional Model (.23 vs. .42)
    - Introduction of Random Coefficients Can Increase Level 2 Variance!
    - R² Calculations Can Sometimes Result in Negative Explained Variance!
  - Could Also Add Cross-Level Interactions to the Model

---

**Random Intercept Logistic Model of Federal Prison Sentences**

<table>
<thead>
<tr>
<th>Prison vs. No Prison (Unit-Specific Model with Robust Standard Errors)</th>
<th>Fixed Effects</th>
<th>b</th>
<th>S.E.</th>
<th>df</th>
<th>T-ratio</th>
<th>p-value</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ( (\gamma_{00}) )</td>
<td>2.85</td>
<td>0.08</td>
<td>87</td>
<td>33.84</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Court Size ( (\gamma_{10}) )</td>
<td>-0.02</td>
<td>0.01</td>
<td>87</td>
<td>-2.08</td>
<td>0.04</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>Severity ( (\gamma_{11}) )</td>
<td>0.25</td>
<td>0.01</td>
<td>88</td>
<td>30.99</td>
<td>0.00</td>
<td>1.29</td>
<td></td>
</tr>
<tr>
<td>Female ( (\gamma_{20}) )</td>
<td>-0.81</td>
<td>0.06</td>
<td>88</td>
<td>-12.67</td>
<td>0.00</td>
<td>0.44</td>
<td></td>
</tr>
</tbody>
</table>

**Random Effects**

<table>
<thead>
<tr>
<th>( s^2 )</th>
<th>S.D.</th>
<th>df</th>
<th>( \chi^2 )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 2 Intercept ( (\gamma_{ij}) )</td>
<td>0.42</td>
<td>0.64</td>
<td>87</td>
<td>474.15</td>
</tr>
<tr>
<td>Severity Slope ( (\gamma_j) )</td>
<td>0.003</td>
<td>0.06</td>
<td>88</td>
<td>301.56</td>
</tr>
<tr>
<td>Female Slope ( (\gamma_{2j}) )</td>
<td>0.11</td>
<td>0.32</td>
<td>88</td>
<td>135.22</td>
</tr>
</tbody>
</table>

---

### Model Specifications

- **File:** hln2.MDM
- **File:** USSC.MDM
- **Command File:** whnt

### Data Structure

- Outcome
- Level-1
- Level-2

---

**Statistics Table**

- **Outcome:** Federal Prison Sentences
- **Level-1:** Partially Robust Standard Error Estimation
- **Level-2:** Partially Robust Standard Error Estimation

---

**Model Equations**

- **LEVEL 1 MODEL** (bold group-mean centering, bold)
  \[ \text{Prob}(\text{PRISON}=1|\beta) = \varphi \]
  \[ \log(\varphi/(1-\varphi)) = \eta \]
  \[ \eta = \beta_0 + \beta_1 (\text{SEVERITY}) + \beta_2 (\text{FEMALE}) \]

- **LEVEL 2 MODEL** (italic group-mean centering)
  \[ \beta_0 = \gamma_{00} + \gamma_{01} (\text{COURTSIZ}) + \gamma_{02} (\text{COURT IMM}) \]
  \[ \beta_1 = \gamma_{10} (\text{COURTSIZ}) + \gamma_{11} (\text{COURT IMM}) \]
  \[ \beta_2 = \gamma_{20} + \gamma_{21} (\text{COURT IMM}) \]
Advanced HLM Modeling

- **Other Generalized Linear Models**
  - **Multinomial Models**
    - Uses Multinomial Logit Link: \( \eta_{mij} = \log \left( \frac{\text{prob}(R_{ij} = m)}{\text{prob}(R_{ij} = M)} \right) \)
    - HLM Uses Last Category as Reference
    - Requires Identical Error Structure Across Contrasts
  - **Poisson Models**
    - Uses Poisson Log Link: \( \eta_{ij} = \log(\lambda_{ij}) \)
    - Can Include Extra Parameter for Over-Dispersion
    - Must Specify an Exposure Variable (Usually Measure of Time)

Choosing the Multinomial Model

Choosing the Poisson Model
Advanced HLM Modeling

- 3 Level Hierarchical Models
  - Add Circuit Courts As Level 3 to the Equation
  - Have to Remake the MDM File from SPSS (or Other) Data
    - Data Must Be Sorted By Level 3, Then Level 2
  - Now Circuit = Level 3 ID; District = Level 2 ID
Advanced HLM Modeling

- 3 Level Hierarchical Models
  - The 3 Level Unconditional Model
    - Decomposes Variance Across All 3 Levels
    - Mean Sentence Length Varies Across Both District and Circuit Courts
  - Calculate Intraclass Correlation for Both Level 2 and Level 3
    - Level 2: \( \rho_{Lvl 2} = \frac{\tau_{\pi}}{\sigma^2 + \tau_{\pi} + \tau_{\beta}} = \frac{173}{4,888} = .035 \)
    - Level 3: \( \rho_{Lvl 3} = \frac{\tau_{\beta}}{\sigma^2 + \tau_{\pi} + \tau_{\beta}} = \frac{85}{4,888} = .017 \)
  - 3.5% of Variance Between District; 1.7% Between Circuits

### 3 Level Unconditional Model of Sentence Length

Sentence Length in Months

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>( b )</th>
<th>S.E.</th>
<th>df</th>
<th>T-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (( \gamma_{000} ))</td>
<td>52.57</td>
<td>3.20</td>
<td>10</td>
<td>16.43</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>( s^2 )</th>
<th>S.D.</th>
<th>df</th>
<th>( \chi^2 )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 (( \sigma^2 ))</td>
<td>4630.08</td>
<td>68.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2 (( \tau_{\pi} ))</td>
<td>172.55</td>
<td>15.25</td>
<td>78</td>
<td>229.93</td>
<td>0.00</td>
</tr>
<tr>
<td>Level 3 (( \tau_{\beta} ))</td>
<td>85.24</td>
<td>9.23</td>
<td>10</td>
<td>50.80</td>
<td>0.00</td>
</tr>
</tbody>
</table>

\[ \sum = 4,888 \]
Advanced HLM Modeling

- 3 Level Hierarchical Models
- Random Intercept Model
  - Add Level 1 Predictors (e.g. Severity, Black)
- Random Coefficient Model
  - Test for Random Coefficients
  - Coefficients Can Vary Across Both Level 2 and Level 3
  - Be Careful of Compounding Complexity!
  - Effect of Black Does Not Vary Across Circuits – “Fix” This Effect

Fully Random Coefficient Model

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>$s^2$</th>
<th>S.D.</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercepts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1 ($\sigma^2$)</td>
<td>2071.67</td>
<td>45.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2 ($\tau_1$)</td>
<td>49.66</td>
<td>7.05</td>
<td>77</td>
<td>248.87</td>
<td>0.00</td>
</tr>
<tr>
<td>Level 3 ($\tau_2$)</td>
<td>14.74</td>
<td>3.84</td>
<td>10</td>
<td>30.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severity ($\tau_{\pi_1}$)</td>
<td>1.10</td>
<td>1.05</td>
<td>77</td>
<td>363.10</td>
<td>0.00</td>
</tr>
<tr>
<td>Black ($\tau_{\pi_2}$)</td>
<td>33.08</td>
<td>5.75</td>
<td>10</td>
<td>164.59</td>
<td>0.00</td>
</tr>
<tr>
<td>Level 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severity ($\tau_{\beta_1}$)</td>
<td>0.24</td>
<td>0.49</td>
<td>77</td>
<td>28.33</td>
<td>0.00</td>
</tr>
<tr>
<td>Black ($\tau_{\beta_2}$)</td>
<td>2.20</td>
<td>1.48</td>
<td>10</td>
<td>7.59</td>
<td>&gt;.500</td>
</tr>
</tbody>
</table>

\begin{align*}
\text{LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)}
\quad & \text{SENLENG} = \pi_0 + \pi_1(\text{SEVERITY}) + \pi_2(\text{BLACK}) + e \\
\quad & \pi_0 = \beta_{00} + r_0 \\
\quad & \pi_1 = \beta_{10} + r_1 \\
\quad & \pi_2 = \beta_{20} + r_2 \\
\text{LEVEL 2 MODEL (bold: group-mean centering; bold italic: grand-mean centering)}
\quad & \beta_{00} = r_{000} + \nu_{00} \\
\quad & \beta_{10} = r_{100} + \nu_{10} \\
\quad & \beta_{20} = r_{200} + \nu_{20}
\end{align*}
Advanced HLM Modeling

- 3 Level Hierarchical Models
  - Add Level 2 and Level 3 Predictors
  - Individual Severity Increases Sentence But Mean Severity Decreases It
  - Robust Standard Errors Unreliable with Few Groups
    - Model Limited by Only 11 Circuits at Level 3
    - Too Few Level 3 Units for Level 3 Predictors
  - Could Add Cross Level Interactions at Level 2 and Level 3
  - Could Calculate $R^2$s at Level 1, Level 2 and Level 3 Now
  - Could Specify Generalized Linear Models with 3 Levels Too

### 3 Level Random Coefficient Model with Level 2 Predictor

Sentence Length in Months

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>b</th>
<th>S.E.</th>
<th>df</th>
<th>T-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>48.70</td>
<td>1.78</td>
<td>10</td>
<td>27.34</td>
<td>0.00</td>
</tr>
<tr>
<td>Mean Severity ($\gamma_{01}$)</td>
<td>-2.24</td>
<td>0.26</td>
<td>87</td>
<td>-8.60</td>
<td>0.00</td>
</tr>
<tr>
<td>Severity ($\gamma_{02}$)</td>
<td>5.58</td>
<td>0.19</td>
<td>10</td>
<td>29.08</td>
<td>0.00</td>
</tr>
<tr>
<td>Black ($\gamma_{30}$)</td>
<td>11.89</td>
<td>1.03</td>
<td>88</td>
<td>11.59</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Concluding Thoughts & Cautionary Tales

- Multilevel Models Require Care and Understanding
- Small Changes in Model Specification May Alter Findings
  - Especially for Level 2 Predictors
  - Essential to Address Collinearity Among Level 2 Variables
  - Essential to Perform Model Diagnostics and Robustness Checks
- Three Level and HGLM Can Quickly Become Unwieldy
  - Complicated Models May Have Convergence Problems
  - Especially True for HGLM with Multiple Random Effects
  - Be Judicious in Use of Random Coefficients
- Research Questions Should Dictate Methodology, Not the Reverse
  - Sometimes Simpler Methods Can Be Equally Effective
  - E.g. Fixed Effects Models and Clustered Standard Errors
- Statistical Sophistication Can’t Substitute for Theorizing & Conceptualization
Concluding Thoughts & Cautionary Tales

“Regional variation in sentencing has been, and will likely continue to be, a lively area of research and debate” (USSC 2004: 133)

- Very Few Multilevel Analyses of USSC Data
  - No Post-
  - No Post-
  - No Post-
  - No Post-
  - Only One Offense-Specific HLM Analysis

- Relatively Poor Measures of Social Context in Prior Work

- Multilevel Models Can Offer Numerous Advantages Over Traditional Regression Techniques when Properly Applied
Independent Data Exercises

- Read 2007 USSC Data into HLM
  - 2 or 3 Level Model

- Run Unconditional Models

- Run Random Intercept Models

- Run Random Coefficient Models

- Add Level 2 Variable(s)

- Try Including Cross-Level Interactions