University of Zurich[UZH]

Institut für Politikwissenschaft

The Hierarchical Linear Model

Derivation, Estimation, Interpretation, and Testing

Marco R. Steenbergen

A Note on Notation

Parameter	Hox	Steenbergen
Fixed Effects	γ_{pq}	γ_{pq}
Random Effects	β_{pj}	β_{pj}
Level-1 Error Term	e_{ij}	ϵ_{ij}
Level-2 Error Term	u_{pj}	δ_{pj}
Level-1 Variance Component	σ_e^2	σ^2
Level-2 Variance Component	$\sigma_{u_p}^2$	τ_{pp}

Part I

Model Derivation

I.A

Random Effects ANOVA

Motivating Example

- Question: How much did evaluations of Barack Obama vary across states in 2008?
- Data: ANES feeling thermometer ratings (Y_{ij}) and a state indicator.
- Fixed effects approach: Create group dummies, run a regression, and perform an F-test on the dummies.
- That would be an OK approach if (1) the group dummies exhaust the population or (2) the available groups are the only ones of theoretical interest.
- However, the ANES give us a sample of 33 states, which do not necessarily meet the aforementioned criteria.

Inspiration from Experimental Design

- In experimental design, the groups are made up of different treatments and controls.
- Sometimes, these groups cover everything that is of theoretical interest.
- Often, however, treatment levels and controls are chosen randomly and other values could have been selected.
- This random sampling is incorporated into the analysis by estimating a random effects analysis of variance (ANOVA).
- This approach has a long history in psychology (e.g., Searle et al. 2006).
- It can also be used for our problem.

Random Effects ANOVA

Level-1 Model

Let Y_{ij} be an outcome of interest, which varies across randomly selected individuals $i = 1, \cdots, n_j$ and groups $j = 1, \cdots, J$. Then

$$y_{ij} = \beta_{0j} + \epsilon_{ij}$$

where β_{0j} is a group mean and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ captures individual variation around the group mean.

Level-2 Model

Further,

$$\beta_{0j} = \gamma_{00} + \delta_{0j}$$

where γ_{00} is the grand mean and $\delta_{0j} \sim \mathcal{N}(0, \tau_{00})$ captures group variation around the grand mean.

Some More Jargon

- γ_{00} is a fixed effect.
- β_{0j} is a random effect.
- σ^2 and τ_{00} are variance components.

General Observations about the Model

1. There are two sources of variation—groups and individuals.
2. In general, level-1 models are always stated in terms of Y_{ij} and level-2 models in terms of parameters.
3. We could combine the equations into a mixed model:

$$y_{ij} = \gamma_{00} + \underbrace{\delta_{0j} + \epsilon_{ij}}_{\nu_{ij}}$$

Assumptions: Exchangeability

Exchangeability means that δ_{0j} is driven by factors that

- Are similar across contexts—level-2 units should be comparable
- Operate independently between contexts so that

$$E[\delta_{0j}, \delta_{0k}] \quad = \quad 0$$

for $j \neq k$.

Assumptions: Level-1 and Level-2 Errors

- Level-1 errors:
  - The level-1 errors are normally distributed
  - They have a mean of 0
  - They have a constant variance
  - There is no autocorrelation
- Level-2 errors:
  - The level-2 errors are normally distributed
  - They have a mean of 0
  - They have a constant variance
  - There is no autocorrelation
- The level-1 and level-2 errors are uncorrelated.

Implied Mean Structure

- $E[\beta_{0j}] = \gamma_{00}$ so that we can say that the random effects are drawn from a distribution with common parameter γ_{00}.
- $E[y_{ij}] = \gamma_{00}$, so that the expected value is the grand mean.

Implied Variance and Covariance Structures

- The implied variance/covariance structure is one of compound symmetry.
- The variance is:

$$Var(y_{ij}) = \sigma^2 + \tau_{00}$$

- The covariance is:

$$Cov(y_{ij}, y_{kl}) = \begin{cases} \tau_{00} & \text{if } j = l \\ 0 & \text{if } j \neq l \end{cases}$$

The Intra-Class Correlation

- By definition, the correlation is the ratio of the covariance over the product of the standard deviations.
- Applying this definition, we obtain the following ICC:

$$\rho = \frac{\tau_{00}}{\sigma^2 + \tau_{00}}$$

- We can interpret this as the portion of the total variance that is due to between-group variation.

Back to the Example

– In Stata,
  the random effects ANOVA model can be estimated using `xtmixed`, e.g.,

  xtmixed obamafeel, || state:, covariance(indep) var

– Doing so yields:

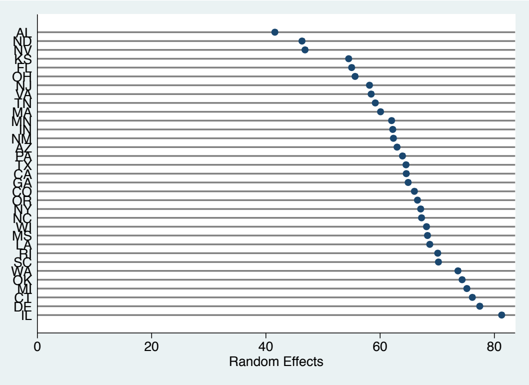
$$\begin{array}{lr} \hat{\gamma}_{00} & 64.05 \\ \hat{\tau}_{00} & 93.57 \\ \hat{\sigma}^2 & 748.11 \end{array}$$

– The estimated ICC is

$$\hat{\rho} = .111,$$

  i.e., about 11 percent of the variance is due to cross-state differences.

Random Effects in the Example



I.B

Random Coefficient Models

Motivating Example

- In analyzing the evaluations of Obama, we would like to consider the impact of a person's race (X_{ij}).
- We also would like to allow the effect to vary across states.
- The random effects ANOVA does not accommodate covariates.
- However, random coefficient models do.
- These have a long history in econometrics (e.g., Swamy & Tavlas 1995).

The Random Coefficient Model: Level-1

Ordinary Regression Specification

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij}$$

with $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$.

Modified Regression Specification

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \epsilon_{ij}$$

Intercept and slope are now allowed to vary across level-2 units.

The Random Coefficient Model: Level-2

Level-2 Model

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \delta_{0j}\\ \beta_{1j} &= \gamma_{10} + \delta_{1j}\end{aligned}$$

Error Structure

$$\begin{pmatrix} \delta_{0j} \\ \delta_{1j} \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{01} & \tau_{11} \end{bmatrix} \right)$$

The Random Coefficient Model: Mixed Model and Assumptions

Mixed Model

$$
\begin{aligned}
y_{ij} &= (\gamma_{00} + \delta_{0j}) + (\gamma_{10} + \delta_{1j})\, x_{ij} + \epsilon_{ij} \\
&= \gamma_{00} + \gamma_{10} x_{ij} + \underbrace{\delta_{0j} + \delta_{1j} x_{ij} + \epsilon_{ij}}_{\nu_{ij}}
\end{aligned}
$$

Assumptions

The assumptions are identical to that of the random effects ANOVA.

Implied Mean Structure

- Since $E[\delta_{0j}] = 0$, the random intercepts are drawn from a distribution with common intercept γ_{00}.
- Since $E[\delta_{1j}] = 0$, the random slopes are drawn from a distribution with common slope γ_{10}.
- Since $E[\epsilon_{ij}] = 0$, the expected outcome is $E[y_{ij}] = \gamma_{00} + \gamma_{10} x_{ij} = \mu_{ij}$.

Implied Variance and Covariance Structures

– The variance is:

$$Var(y_{ij}) = \tau_{00} + 2x_{ij}\tau_{01} + x_{ij}^2\tau_{11} + \sigma^2$$

– The covariance is:

$$Cov(y_{ij}, y_{kl}) = \begin{cases} \tau_{00} + (x_{ij} + x_{kl})\tau_{01} + x_{ij}x_{kl}\tau_{11} & \text{if } j = l \\ 0 & \text{if } j \neq l \end{cases}$$

Intra-Class Correlation

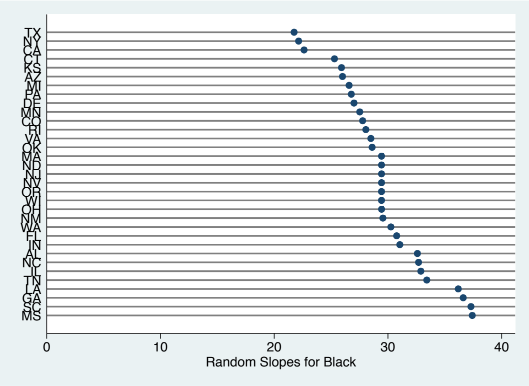
$$\rho = \begin{cases} \frac{\tau_{00}+(x_{ij}+x_{kl})\tau_{01}+x_{ij}x_{kl}\tau_{11}}{\sqrt{\tau_{00}+2x_{ij}\tau_{01}+x_{ij}^2\tau_{11}+\sigma^2}\sqrt{\tau_{00}+2x_{kl}\tau_{01}+x_{kl}^2\tau_{11}+\sigma^2}} & \text{if } j = l \\ 0 & \text{if } j \neq l \end{cases}$$

Back to the Example

Parameter	Notation	Estimate
Constant	γ_{00}	55.24
Black	γ_{10}	29.44
Level-1 Variance	σ^2	623.87
Variance in Intercepts	τ_{00}	59.56
Variance in Slopes	τ_{11}	45.84

Notes: τ_{01} constrained to 0. The number of level-2 units is 33. The number of level-1 units is 2267. The cluster sizes vary between 20 and 362.

Random Slopes in the Example



I.C

Adding Level-2 Covariates—the Hierarchical Linear Model

Motivating Example

- So far we have discovered that:
  - Evaluations of Obama vary across states
  - The effect of race on those evaluations varies across states
- However, we lack an explanation for either phenomenon.
- Imagine we believe that the state characteristic "percentage blacks" (Z_j) can account for both phenomena.
- How do we integrate this into the model?

The Hierarchical Linear Model

Level-1 Model

The level-1 model is identical to the random coefficients model:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \epsilon_{ij}$$

Level-2 Model

The level-2 model receives a modification by adding a level-2 covariate:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} z_j + \delta_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} z_j + \delta_{1j} \end{aligned}$$

The Hierarchical Linear Model Cont'd

Mixed Model

$$
\begin{aligned}
y_{ij} &= (\gamma_{00} + \gamma_{01} z_j + \delta_{0j}) + (\gamma_{10} + \gamma_{11} z_j + \delta_{1j}) x_{ij} + \epsilon_{ij} \\
&= \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \gamma_{11} z_j x_{ij} + \underbrace{\delta_{0j} + \delta_{1j} x_{ij} + \epsilon_{ij}}_{\nu_{ij}}
\end{aligned}
$$

Assumptions

All of the assumptions of the random coefficients model apply.

The Cross-Level Interaction

- $z_j x_{ij}$ is the cross-level interaction and γ_{11} is the cross-level interaction effect.
- This captures "causal" heterogeneity—variation in the effects of level-1 covariates across level-2 units.
- The ability to incorporate such interactions is a particularly powerful feature of multilevel models.

The Status of the Level-2 Errors

- In the HLM, the level-2 errors δ_{0j} and δ_{1j} have the status of residuals.
- That is, they pick up variation in intercepts and slopes that is left unexplained by Z_j.
- If Z_j has any explanatory power at all, then τ_{00} and τ_{11} should be reduced relative to the random coefficients model.
- We shall exploit this idea later to develop a R^2 measure.

Back to the Example

Parameter	Notation	Estimate
Constant	γ_{00}	64.71
Percent Black	γ_{01}	-0.73
Black	γ_{10}	14.10
Percent Black $\times$ Black	γ_{11}	0.96
Level-1 Variance	σ^2	621.91
Variance in Intercepts	τ_{00}	29.09
Variance in Slopes	τ_{11}	0.00

Notes: τ_{01} constrained to 0. The number of level-2 units is 33. The number of level-1 units is 2267. The cluster sizes vary between 20 and 362.

Extension to Multiple Covariates

Level-1 Model

With P level-1 covariates, the model is

$$
\begin{aligned}
y_{ij} &= \beta_{0j} + \beta_{1j} x_{1ij} + \beta_{2j} x_{2ij} + \cdots + \beta_{Pj} x_{Pij} + \epsilon_{ij} \\
&= \beta_{0j} + \sum_{p=1}^{P} \beta_{pj} x_{pij} + \epsilon_{ij}
\end{aligned}
$$

Extension to Multiple Covariates Cont'd

Level-2 Model

With Q level-2 covariates, the most general formulation of the level-2 model is

$$
\begin{aligned}
\beta_{0j} &= \gamma_{00} + \gamma_{01} z_{1j} + \gamma_{02} z_{2j} + \cdots + \gamma_{0Q} z_{Qj} + \delta_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11} z_{1j} + \gamma_{12} z_{2j} + \cdots + \gamma_{1Q} z_{Qj} + \delta_{1j} \\
\beta_{2j} &= \gamma_{20} + \gamma_{21} z_{1j} + \gamma_{22} z_{2j} + \cdots + \gamma_{2Q} z_{Qj} + \delta_{2j} \\
\vdots &\;\vdots\; \vdots \\
\beta_{Pj} &= \gamma_{P0} + \gamma_{P1} z_{1j} + \gamma_{P2} z_{2j} + \vdots + \gamma_{PQ} z_{Qj} + \delta_{Pj}
\end{aligned}
$$

Extension to Multiple Covariates Cont'd

Mixed Model

$$y_{ij} = \gamma_{00} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{p=1}^{P} \sum_{q=1}^{Q} \gamma_{pq} z_{qj} x_{pij} + \underbrace{\delta_{0j} + \sum_{p=1}^{P} \delta_{pj} x_{pij} + \epsilon_{ij}}_{\nu_{ij}}$$

Implied Variance-Covariance Structure

Variance

$$
\begin{aligned}
Var(y_{ij}) \quad = \quad & \tau_{00} + \sum_p x_{pij}^2 \tau_{pp} + \sigma^2 + \\
& 2 \sum_p x_{pij} \tau_{0p} + 2 \sum_p \sum_{q>p} x_{pij} x_{qij} \tau_{pq}
\end{aligned}
$$

Implied Variance-Covariance Structure Cont'd

Covariance

For units i and j from the same cluster k, the covariance is

$$\begin{aligned} Cov(y_{ik}, y_{jk}) &= \tau_{00} + \sum_p (x_{pik} + x_{pjk})\tau_{0p} + \sum_p x_{pik} x_{pjk} \tau_{pp} + \\ & \sum_p \sum_{q>p} (x_{pik} x_{qjk} + x_{pjk} x_{qik}) \tau_{pq} \end{aligned}$$

Everywhere else the covariance is zero.

Restrictions

- We can constrain some parameters to be fixed by retaining only γ_{p0}.
- Even when we include random slopes, they do not have to be a function of all of the Z variables—we constrain certain γ_{pq}s.
- Indeed, for identification purposes, it is usually unavoidable to engage in these strategies.

Part II

Identification

Number of Estimated Parameters

With P level-1 covariates and Q level-2 covariates, a fully specified HLM contains the following numbers of parameters:

Fixed Effects	$(P+1)(Q+1)$
Level-1 Variance Component	1
Level-2 Variance and Covariance Components	$.5(P+1)(P+2)$

This quickly becomes a very large number.

Identification

- Identification of the effects of level-1 covariates, including cross-level interactions, depends on $n = \sum_j n_j$—this is usually large in electoral studies.
- Identification of the effects of level-2 covariates depends on J—this is usually small in electoral studies.
- Identification of the variance and covariance components also depends on J.

An Example

- The 2009 EES contains $J = 27$ countries and up to $n = 27,069$ individuals.
- With n being this large, there is practically no limitation on the number of level-1 covariates that can be specified.
- Up to 26 level-2 covariates can be included in the model (one fewer than J because we need to reserve one degree of freedom for the intercept).
- Up to 27 level-2 variance and covariance components can be specified. This means that the subset of *random* level-1 covariates is limited to $P = 5$.

An Important Implication

- It is usually not possible to decide based on empirics which effects are random.
- Instead, this determination has ti be made based on theory and past evidence.

Some Identification Strategies

- Only focus on random intercepts
  - This is what much of the literature does
  - However, it can easily produce specification errors
- Do not estimate covariance components
  - This reduces the number of paramaters drastically, from $.5(P+1)(P+2)$ to $P+1$
  - Covariance components often display problems (corner solutions)
  - But be careful with categorical predictors

Part III

Estimation

III.A

Estimation of Fixed Effects

Fixed Effects

Fixed effects can be estimated in a variety of ways:

- Least squares (e.g., De Leeuw & Kreft 1986; Lewis & Linzer 2005; Jusko & Shively 2005)
- Maximum likelihood (e.g., Goldstein 1986; Longford 1987; Raudenbush & Bryk 1986)
- Bayesian inference (e.g., Seltzer et al. 1996)

Estimation in Large Clusters

– The least squares estimation procedures should not be ignored, especially not when cluster sizes are large and primary interest is in the fixed effects.

– Particularly useful here is the paper by Lewis & Linzer (2005), which outlines a procedure to correct least squares standard errors.

III.B

Estimation of Variance Components

Variance and Covariance Components

Variance and covariance components are best estimated using

- Maximum likelihood:
  - Full information maximum likelihood (FIML)
  - Restricted maximum likelihood (REML)
- Bayesian inference
- In special cases, other estimators may be available.

A Note on FIML

- In the HLM, the joint distribution of the data is given by a multivariate normal.
- The likelihood function, L, is derived from this distribution and gives the likelihood of observing the data given the fixed effects and (co)variance components.
- The maximum likelihood estimates are those values of the fixed effects and (co)variance components, which maximize the likelihood.
- These can be derived through an optimizer (usually Newton-Raphson).

A Note on REML

- The FIML estimator of the variance components is downwardly biased when J is small.
- REML corrects for this bias.
- One way to think of this is that it adds a penalty term to the likelihood for estimating fixed effects.

III.C

Estimation of Random Effects

Empirical Bayes Estimation

- The random effects are typically estimated using empirical Bayes (EB) estimation.
- Here the mean of the posterior serves as the estimator but, unlike Bayesian estimation as such, the prior is informed by the ML estimates.
- One can also think of this as a "compromise" between two different estimates:
  - Within-group estimates
  - Fixed effects estimates
- A weighted average of these estimates is taken, whereby the weight is a function of the reliability of each.

Example: Random Effects Anova

– Consider again the model

$$\begin{aligned} y_{ij} &= \beta_{0j} + \epsilon_{ij} \\ \beta_{0j} &= \gamma_{00} + \delta_{0j} \end{aligned}$$

– We could estimate β_{0j} in each group, as per equation 1; the resulting ML estimator would be $\hat{\beta}_{0j} = \bar{y}_{.j}$, i.e., the group mean.

– We could also estimate β_{0j} using the fixed effect from the second equation; this yields $\hat{\gamma}_{00} = \bar{y}_{..}$, i.e., the grand mean.

Example Cont'd

- The precision of the grand mean is given by τ_{00}, which indicates how much the group means vary from the grand mean.
- The precision of the group mean is given by σ^2/n_j, the usual formula for the variance of a mean.
- This suggests the following weighted average

$$\tilde{\beta}_{0j} = \lambda_j \bar{y}_{.j} + (1 - \lambda_j)\bar{y}_{..}$$

with

$$\lambda_j = \frac{\tau_{00}}{\tau_{00} + \frac{\sigma^2}{n_j}}$$

- This is the reliability of the between estimator.

Example Cont'd

- A Limiting case:
  - Imagine $\sigma^2 = 0$
  - Then $\lambda_j = 1$
  - Consequently, $\tilde{\beta}_{0j} = \bar{y}_{.j}$
- Another limiting case:
  - Imagine $\tau_{00} = 0$
  - Then $\lambda_j = 0$
  - Consequently, $\tilde{\beta}_{0j} = \bar{y}_{..}$

Some General Observations

- The EB estimator is equal to the mean of the posterior distribution.
- This is also known as the BLUP = best linear unbiased predictor.
- Another label is shrinkage estimator, as it shrinks $\tilde{\beta}_{0j}$ in the direction of the most reliable estimator.
- The shrinkage principle is sometimes touted as a major benefit of multilevel analysis as it allows one to borrow strength.

Borrowing Strength

- Sometimes within-group estimation is impossible because n_j is too small.
- Because we can borrow strength, however, from other groups through the fixed effects estimator, we may still be able to compute a random effect.
- This is useful in cases where n_j is small.
- This is usually not a problem in electoral research.

Empirical Bayes Residuals

- Many programs compute the EB residuals, $\tilde{\delta}_{pj}$.
- This is true of Stata, for example.
- To turn these into EB estimates of the random coefficients, just add the fixed portion of the model.
- In the random effects ANOVA, for example, $\tilde{\beta}_{0j} = \hat{\gamma}_{00} + \tilde{\delta}_{0j}$.
- This can be shown to be equal to

$$\tilde{\delta}_{0j} = \lambda_j(\hat{\beta}_{0j} - \hat{\gamma}_{00})$$

EB Residuals in Stata

To be run after the `xtmixed` command:

predict b, reffects level(state)*

Part IV

Interpretation

IV.A

Interpreting Fixed Effects

Nature of the Interpretation

– Consider the simple HLM

$$y_{ij} = \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \gamma_{11} z_j x_{ij} + \delta_{0j} + x_{ij}\delta_{1j} + \epsilon_{ij}$$

– When we average over level-2 and level-1 units, only the fixed effects remain:

$$E[y_{ij}] = \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \gamma_{11} z_j x_{ij} = \mu_{ij}$$

– Most papers rely on this fixed effects interpretation.

Interpretation Without Cross-Level Interactions

- Imagine $\gamma_{11} = 0$, so that

$$\mu_{ij} = \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij}$$

- The interpretation is straightforward:
  - γ_{01} gives the expected change in Y for a unit change in Z
  - γ_{10} gives the expected change in Y for a unit change in X

Interpretation With Cross-Level Interactions

- Let $\gamma_{11} \neq 0$, then the cross-level interaction can be probed through the simple slope equation (e.g., Bauer & Curran 2005).
- This can take on two forms:

$$\omega_{X|Z} = \frac{\partial \mu_{ij}}{\partial x_{ij}} = \gamma_{10} + \gamma_{11} z_j \qquad Z \text{ serves as the moderator}$$
$$\omega_{Z|X} = \frac{\partial \mu_{ij}}{\partial z_j} = \gamma_{01} + \gamma_{11} x_{ij} \qquad X \text{ serves as the moderator}$$

- Note that the statistical main effects of x_{ij} and z_j are, in fact, conditional effects that come about when the moderator is 0.

Standard Error and Test Statistic of the Simple Slope

– Since the simple slope is a function only of the fixed effects, its standard error can be obtained quite simply:

$$
\begin{aligned}
V\left[\hat{\omega}_{X|Z}\right] &= Var(\hat{\gamma}_{10}) + 2z_j Cov(\hat{\gamma}_{10}, \hat{\gamma}_{11}) + z_j^2 Var(\hat{\gamma}_{11}) \\
V\left[\hat{\omega}_{Z|X}\right] &= Var(\hat{\gamma}_{01}) + 2x_{ij} Cov(\hat{\gamma}_{01}, \hat{\gamma}_{11}) + x_{ij}^2 Var(\hat{\gamma}_{11})
\end{aligned}
$$

– Moreover, it can be demonstrated that

$$
\begin{aligned}
\frac{\hat{\omega}_{X|Z}}{\sqrt{V\left[\hat{\omega}_{X|Z}\right]}} &\overset{asy}{\sim} \mathcal{N}(0,1) \\
\frac{\hat{\omega}_{Z|X}}{\sqrt{V\left[\hat{\omega}_{Z|X}\right]}} &\overset{asy}{\sim} \mathcal{N}(0,1)
\end{aligned}
$$

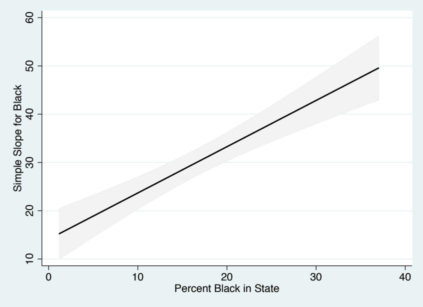
so that a confidence region can be constructed.

An Example: Ratings of Barack Obama

- Earlier we estimated a model of Obama ratings with a cross-level interaction between race of the respondent (X) and percent black in the state (Z).
- The simple slope for race is $\hat{\omega}_{X|Z} = 14.10 + .96z_j$.
- Its variance can be computed using the `vce` command in Stata, which yields

$$\begin{array}{rr} Var(\hat{\gamma}_{10}) & 7.96 \\ Var(\hat{\gamma}_{11}) & 0.02 \\ Cov(\hat{\gamma}_{10}, \hat{\gamma}_{11}) & -0.37 \end{array}$$

Example Cont'd



Another Example: Race as a Moderator

- What if we condition the effect of percent black in the state on the race of the respondent?
- Using the same notation as before, we get $\hat{\omega}_{Z|X} = -.73 + .96x_{ij}$.
- This works out to

White	-0.73
Black	0.23

- The variances and covariances of the estimators are

$Var(\hat{\gamma}_{01})$	0.02
$Var(\hat{\gamma}_{11})$	0.02
$Cov(\hat{\gamma}_{10}, \hat{\gamma}_{11})$	-0.01

Example Cont'd

Race	Slope	95% CI	
White	-0.73	-0.99	-0.47
Black	0.23	-0.08	0.54

IV.B

Interpreting Random Effects

Motivating Example

- The fixed effects interpretation does not show us how race plays out in different states.
- A random effects interpretation can provide this picture.

A Random Slope and Intercept Model

– Consider the model
  $y_{ij} = \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \gamma_{11} z_j x_{ij} + \delta_{0j} + \delta_{1j} x_{ij} + \epsilon_{ij}$.

– The regression line for a particular level-2 unit is estimated as

$$\hat{y}_{ij} = \hat{\gamma}_{00} + \hat{\gamma}_{01} z_j + \hat{\gamma}_{10} x_{ij} + \hat{\gamma}_{11} z_j x_{ij} + \tilde{\delta}_{0j} + \tilde{\delta}_{1j} x_{ij}$$

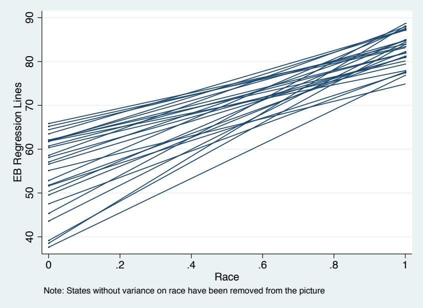
– This can be computed based on the fixed effects estimates and the EB residuals.

Example: The Effect of Race on Feelings Toward Obama Across States

Stata syntax:

```
xtmixed obamafeel black pctblack interact, || state:
black, covariance(indep) var
predict feelhat, fitted
sort state black
twoway (line feelhat black, connect(ascending)
xtitle(Race) ytitle(EB Regression Lines))
```

Example Cont'd



IV.C

Centering and Interpretation

Why Center?

- Centering refers to transforming a variable into mean-deviation form.
- The goal of centering is to aid in the interpretation of fixed effects.
- This is especially important for intercept terms.
- Centering is more critical for level-1 covariates—less so for level-2 covariates.

Types of Centering

Type	Transformation
Uncentered	$x^*_{ij} = x_{ij}$
Centered about the Grand Mean	$x^*_{ij} = x_{ij} - \bar{x}_{..}$
Centered about the Group Mean	$x^*_{ij} = x_{ij} - \bar{x}_{.j}$

Example: Random Effects ANCOVA

- In uncentered form the random effects ANCOVA model is

$$y_{ij} \quad = \quad \beta_{0j} + \beta_{1j} x_{ij} + \epsilon_{ij}$$

- Aggregating over i, we now get

$$\mu_j \quad = \quad \beta_{0j} + \beta_{1j} \bar{x}_{.j}$$

  where μ_j is the mean of Y in group j.

- Here, β_{0j} is the value of μ_j if $\bar{x}_{.j} = 0$.
- But this is interpretable only if zero is a feasible value of the covariate.

Example Cont'd

- With centering around the grand mean we have:

$$y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij} - \bar{x}_{..}) + \epsilon_{ij},$$

which is the usual formulation of the ANCOVA model.

- Aggregating over i we get

$$\mu_j = \beta_{0j} + \beta_{1j}(\bar{x}_{.j} - \bar{x}_{..})$$

- Consequently,

$$\beta_{0j} = \mu_j - \beta_{1j}(\bar{x}_{.j} - \bar{x}_{..})$$

which is the adjusted mean of Y.

- This can always be interpreted.

Example: ANCOVA

- With centering around the group mean we have:

$$y_{ij} \quad = \quad \beta_{0j} + \beta_{1j}(x_{ij} - \bar{x}_{.j}) + \epsilon_{ij}$$

- Aggregating over i we get

$$\mu_j \quad = \quad \beta_{0j} + \beta_{1j}(\bar{x}_{.j} - \bar{x}_{.j}) = \beta_{0j}$$

- Thus, β_{0j} is equal to the group mean of Y.
- This can always be interpreted.

Part V

Hypothesis Testing

Testing Fixed Effects

Typical Null Hypothesis

$$H_0 : \gamma_{pq} \quad = \quad 0$$

Asymptotic Test Statistic

$$z \quad = \quad \frac{\hat{\gamma}_{pq}}{\sqrt{\widehat{V}[\hat{\gamma}_{pq}]}} \overset{asy}{\sim} N(0,1)$$

Alternative Tests

- The test statistic is based on the asymptotic normality of the ML fixed effects estimators.
- This is what Stata implements.
- Bryk & Raudenbush (1992) suggest a more conservative approach, which is asymptotically equivalent:

$$\frac{\hat{\gamma}_{pq}}{\sqrt{\widehat{V}[\hat{\gamma}_{pq}]}} \sim t_{J-P-1}$$

- One can also obtain more precise approximations of the degrees of freedom based on the error variances (Satterthwaite 1946).

Testing Variance Components

- The typical null hypothesis is: $H_0 : \tau_{pp} = 0$.
- In most cases, this is best evaluated using a deviance/likelihood ratio test.
- Here, we estimate models with and without variance components and compare their fit.

The Deviance

Definition

The deviance is

$$D = -2\ell$$

where ℓ is the log-likelihood. This may be viewed as a measure of lack of fit.

The Deviance Test

– Consider two models, M_U and M_R, where M_R is a subset of M_U.

– This subset comes about by imposing restrictions on some of the parameters in M_U.

– We now estimate both models and obtain their deviances.

– If the restrictions are valid, i.e., consistent with the data, then $D_R = D_U$.

– If the restrictions are invalid, then $D_R > D_U$.

– We now obtain the test statistic

$$LR \quad = \quad D_R - D_U \overset{asy}{\sim} \chi^2_r$$

where r denotes the number of restrictions.

The Deviance Test for Variance Components

Null Hypothesis

$$H_0 : \tau_{jk} \quad = \quad 0$$

for $j, k = 0, \cdots, P$.

Analytic Strategy

- Estimate a model, M_U, containing the variance and covariance components
- Estimate a second model, M_R, imposing the restrictions implied by H_0
- Compute the deviance test statistic and its p-value

Halving the p-Value

- When we test for the significance of a variance component, we should remember that this is always a one-tailed test.
- The reason is that variance components can only deviate from 0 in a positive direction.
- Thus, we should halve the p-value, as we would with any one-tailed test.
- Note that Stata does not do this; hence its test is "conservative."

Example: Back to Obama

- In our model, we included race of the respondent and, at one point, decided to let it have a random slope.
- Do we need this added complexity?
- When we add the variance component, we obtain $\ell_U = -10548.992$ and $D_U = 21097.984$.
- When we exclude the variance component, we obtain $\ell_R = -10553.962$ and $D_R = 21107.924$.
- Hence, $LR = 21107.924 - 21097.984 = 9.94$.
- When referred to χ^2_1, we obtain $p = .002$. Halving this, we get $p = .001$.
- We reject H_0, i.e., there is evidence that the effect of race varies across states.

Some Cautionary Remarks

- Sample:
  - Always make sure the estimation sample is the same for M_U and M_R
- REML:
  - With REML, the deviance test is valid only if the fixed portion of the model remains identical
- Special Situations:
  - In some cases only quasi-likelihood estimation is possible
  - Under these circumstances, one should refrain from conducting a deviance test
  - A Wald test may be more appropriate

Part VI

Model Evaluation

A Plethora of R^2s

- There are many R^2 measures in HLMs.
- There is one level-1 R^2.
- There are as many level-2 R^2s as there are random coefficients.

General Logic

- Estimate a model without covariates that can account for the variance of the outcome or parameter. Call the variance V_0.
- Estimate a model with covariates that can account for the variance of the outcome or parameter. Call the variance V_1.
- Compute the proportional reduction in error

$$\frac{V_0 - V_1}{V_0} = 1 - \frac{V_1}{V_0} = R^2$$

- For a level-1 R^2 look at the proportional reduction of the variance in Y.
- For a level-2 R^2 look at the proportional reduction of the variance in β_{pj}.

Level-1 R^2 in a Random Intercept Model

Estimate	M_0	M_1
σ^2	748.110	633.880
τ_{00}	93.570	40.350
$Var(y) = \sigma^2 + \tau_{00}$	841.680	674.230

Notes: M_0 = random effects ANOVA; M_1 = random intercept model. The explained variance is

$$R_1^2 = 1 - \frac{674.230}{841.680} = 0.199$$

Level-2 R^2 in a Random Intercept Model

Estimate	M_0	M_1
$\tau_{00} = Var(\beta_{0j})$	40.350	25.998

Notes: M_0 = random intercept model; M_1 = random intercept model with level-2 covariate. The explained variance is

$$R_2^2 = 1 - \frac{25.998}{40.350} = 0.356$$

Part VII

Caveats

Overview

Three important considerations:

1. Theoretical requirements
2. Sample size requirements
3. Measurement requirements

Theoretical Demands

- The theoretical demands for proper multilevel model specification are steep.
- They include:
  - The selection of levels (how many and which?)
  - The specification of a level-1 model (which level-1 covariates?)
  - The specification of a level-2 model (which effects are random; which level-2 covariates are relevant?)
- Inappropriate decisions at each of these decision points can result in invalid inferences.
- Therefore: Take model specification very seriously.

Theoretical Demands Cont'd

- Incorrect specification of the level-1 model can lead to false indications of random effects.
- See Bauer (2005).

Sample Size Requirements

- In multilevel analysis, the cluster size is less relevant than the number of clusters.
- Ideally, the number of clusters should be large (some say $J \geq 50$).
- A small J can result in: (1) biased estimates of variance components and (2) low statistical power.
- A common dilemma in electoral research: N is large but J is small.
- To some extent, the problem may be overcome using a Bayesian approach.

Sample Size Requirements Cont'd

- We have been assuming that the level-2 units are sampled randomly.
- But this is often not the case. How critical is the assumption?
- It depends on one's perspective:
  - Frequentists—this is a big deal
  - Bayesians—nothing to worry about

Measurement Equivalence

- Measurement equivalence means that a survey measure behaves similarly in different contexts (e.g., Davidov et al. 2011).
- Without measurement equivalence, multilevel models may give a false indication of random intercepts and slopes.
- What we need is scalar invariance:
  - Survey items measure the same thing in different contexts
  - They have identical metrics across contexts
  - There are no differential item biases

A Measurement Model

The Model

For a measure Y of construct η,

$$y_{ij} = \tau_j + \lambda_j \eta_{ij} + \epsilon_{ij}$$

Scalar Invariance

$$\begin{aligned} \lambda_1 &= \lambda_2 = \cdots = \lambda_J \\ \tau_1 &= \tau_2 = \cdots = \tau_J \end{aligned}$$

Multilevel Analysis Without Invariance

- Consider the following HLM for the latent variable:

$$\eta_{ij} \quad = \quad \gamma_{00} + \gamma_{10} x_{ij} + \; zeta_{ij}$$

  Note that intercept ans slope are fixed.

- Now imagine the following measurement model:

$$y_{ij} \quad = \quad \tau_j + \lambda_j \eta_{ij} + \epsilon_{ij}$$

  Note that the measures lack invariance.

- The HLM for Y is now

$$y_{ij} \quad = \quad \underbrace{\tau_j + \lambda_j \gamma_{00}}_{Intercept} + \underbrace{\lambda_j \gamma_{10}}_{Slope} x_{ij} + \underbrace{\lambda_j \zeta_{ij} + \epsilon_{ij}}_{Error}$$

  so that both the intercept and slope vary.