Determining the Nesting Structure for Hierarchical Models Fitted to Education Data

Ulrike Genschel, Jillian Downey and Mark Kaiser

Iowa State University

August 3, 2016

Data Applications: Education Data

Significant effort on development of different modes of instruction

- examples include
  - Flipped classrooms
  - Team based learning
  - Randomization based introductory statistics
- Implementation can be costly (time or monetarily)
- **Need to evaluate effectiveness**

Challenges with Educational Data

Limitations of traditional group comparison methodology:

- Hard to randomize students to treatment
  - confounding variables likely introduce hidden bias
- Experimental units versus observational units
- Uncontrollable factors recognized as potentially introducing variation/structure
  - school characteristics (type, size, location)
  - instructor characteristics (experience)
  - classroom characteristics (semester, time of day)
- **Inferential scope is restrictive!**

Alternative Methodologies

- Random Effects Models
  - requires reasonable sample size
  - requires balanced data
- Hierarchical models account for nesting structures
  - requires identification of nesting structure

Description of Data

- Two factors: instructor and semester
  - assume multiple semesters and multiple instructors
  - one class per semester and instructor combination

- **response** of interest Y_{ijkl}
  - teaching approach i, $i = 1, 2$
  - in semester j, $j = 1, \ldots, J_i$
  - with instructor k, $k = 1, \ldots, K_{ij}$
  - student l, $l = 1, 2, \ldots, L_{ijk}$

Question of Interest:

What is the most appropriate nesting structure?

- Model 1: instructors nested within semester
- Model 2: semesters nested within instructor

Model Formulations:

Model 1: Instructors nested within semester

$$\begin{aligned} Y_{ijkl}|\theta_{ijk} &\sim \mathcal{N}(\theta_{ijk}, \delta^2), \\ \theta_{ijk}|\mu_{ij} &\sim \mathcal{N}(\mu_{ij}, \tau^2), \\ \mu_{ij}|\lambda_i &\sim \mathcal{N}(\lambda_i, \psi^2). \end{aligned}$$

Model 2: semesters nested within instructor

$$\begin{aligned} Y_{ijkl}|\theta_{ijk} &\sim \mathcal{N}(\theta_{ijk}, \delta^2), \\ \theta_{ijk}|\alpha_{ik} &\sim \mathcal{N}(\alpha_{ik}, \epsilon^2), \\ \alpha_{ik}|\lambda_i &\sim \mathcal{N}(\lambda_i, \sigma^2). \end{aligned}$$

Covariance Structure Comparison

Model 1: Instructors nested within semester

- $\mathrm{Var}(Y_{ijkl}) = \psi^2 + \tau^2 + \delta^2$
- $\mathrm{Cov}(Y_{ijkl}, Y_{ijkl'}) = \psi^2 + \tau^2$ for $l \neq l'$ (same class)
- $\mathrm{Cov}(Y_{ijkl}, Y_{ijk'l}) = \psi^2$ for $k \neq k'$ (same semester, different instructors)
- $\mathrm{Cov}(Y_{ijkl}, Y_{ij'kl}) = 0$ for $j \neq j'$ (same instructor, different semesters)

Model 2: Semesters nested within instructor

- $\mathrm{Var}(Y_{ijkl}) = \sigma^2 + \epsilon^2 + \delta^2$
- $\mathrm{Cov}(Y_{ijkl}, Y_{ijkl'}) = \sigma^2 + \epsilon^2$ for $l \neq l'$ (same class)
- $\mathrm{Cov}(Y_{ijkl}, Y_{ijk'l}) = 0$ for $k \neq k'$ (same semester, different instructors)
- $\mathrm{Cov}(Y_{ijkl}, Y_{ij'kl}) = \sigma^2$ for $j \neq j'$ (same instructor, different semesters)

Covariance Matrices - Students within Class

Model 1: Instructors nested within semester

$$\boldsymbol{A} = \begin{bmatrix} \psi^2+\tau^2+\delta^2 & \psi^2+\tau^2 & \dots & \psi^2+\tau^2 \\ \psi^2+\tau^2 & \psi^2+\tau^2+\delta^2 & \dots & \psi^2+\tau^2 \\ \vdots & \vdots & \ddots & \vdots \\ \psi^2+\tau^2 & \psi^2+\tau^2 & \dots & \psi^2+\tau^2+\delta^2 \end{bmatrix}$$

Model 2: Semesters nested within instructor

$$\boldsymbol{B} = \begin{bmatrix} \sigma^2+\epsilon^2+\delta^2 & \sigma^2+\epsilon^2 & \dots & \sigma^2+\epsilon^2 \\ \sigma^2+\epsilon^2 & \sigma^2+\epsilon^2+\delta^2 & \dots & \sigma^2+\epsilon^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2+\epsilon^2 & \sigma^2+\epsilon^2 & \dots & \sigma^2+\epsilon^2+\delta^2 \end{bmatrix}$$

Full Covariance Matrix - Model 1 (Instructors nested within semester)

For illustration, assume $J_i = 3$ (semesters) and $K_{ij} = 2$ (instructor)

$$\begin{bmatrix}
A & \psi^2 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\psi^2 1 & A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & A & \psi^2 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \psi^2 1 & A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & A & \psi^2 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \psi^2 1 & A & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & A & \psi^2 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \psi^2 1 & A & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & A & \psi^2 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \psi^2 1 & A & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & A & \psi^2 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \psi^2 1 & A
\end{bmatrix}$$

Full Covariance Matrix - Model 2 (Semesters nested within instructor)

For illustration, assume $J_i = 3$ (semesters) and $K_{ij} = 2$ (instructors)

$$\begin{bmatrix}
\boldsymbol{B} & 0 & \sigma^2 1 & 0 & \sigma^2 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \boldsymbol{B} & 0 & \sigma^2 1 & 0 & \sigma^2 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\sigma^2 1 & 0 & \boldsymbol{B} & 0 & \sigma^2 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \sigma^2 1 & 0 & \boldsymbol{B} & 0 & \sigma^2 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\sigma^2 1 & 0 & \sigma^2 1 & 0 & \boldsymbol{B} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \sigma^2 1 & 0 & \sigma^2 1 & 0 & \boldsymbol{B} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{B} & 0 & \sigma^2 1 & 0 & \sigma^2 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{B} & 0 & \sigma^2 1 & 0 & \sigma^2 1 \\
0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 1 & 0 & \boldsymbol{B} & 0 & \sigma^2 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 1 & 0 & \boldsymbol{B} & 0 & \sigma^2 1 \\
0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 1 & 0 & \sigma^2 1 & 0 & \boldsymbol{B} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 1 & 0 & \sigma^2 1 & 0 & \boldsymbol{B}
\end{bmatrix}$$

Estimating Elements of Covariance Matrices

- Non-trivial when only have single vector of observations
- Moment based estimates using sample variances of students within different groupings (unbiased)
- Notation:
  - Let $L(a_1, a_2, a_3)$ be a linear combination of elements a_1, a_2, and a_3
  - V_C = average variance of class
  - V_S = average variance of semester
  - V_I = average variance of instructor
  - V_A = variance of all observations
  - M_1 = mean of all observations in method 1
  - M_2 = mean of all observations in method 2

Illustration of Variance Groupings - V_S

grouping observations by semester

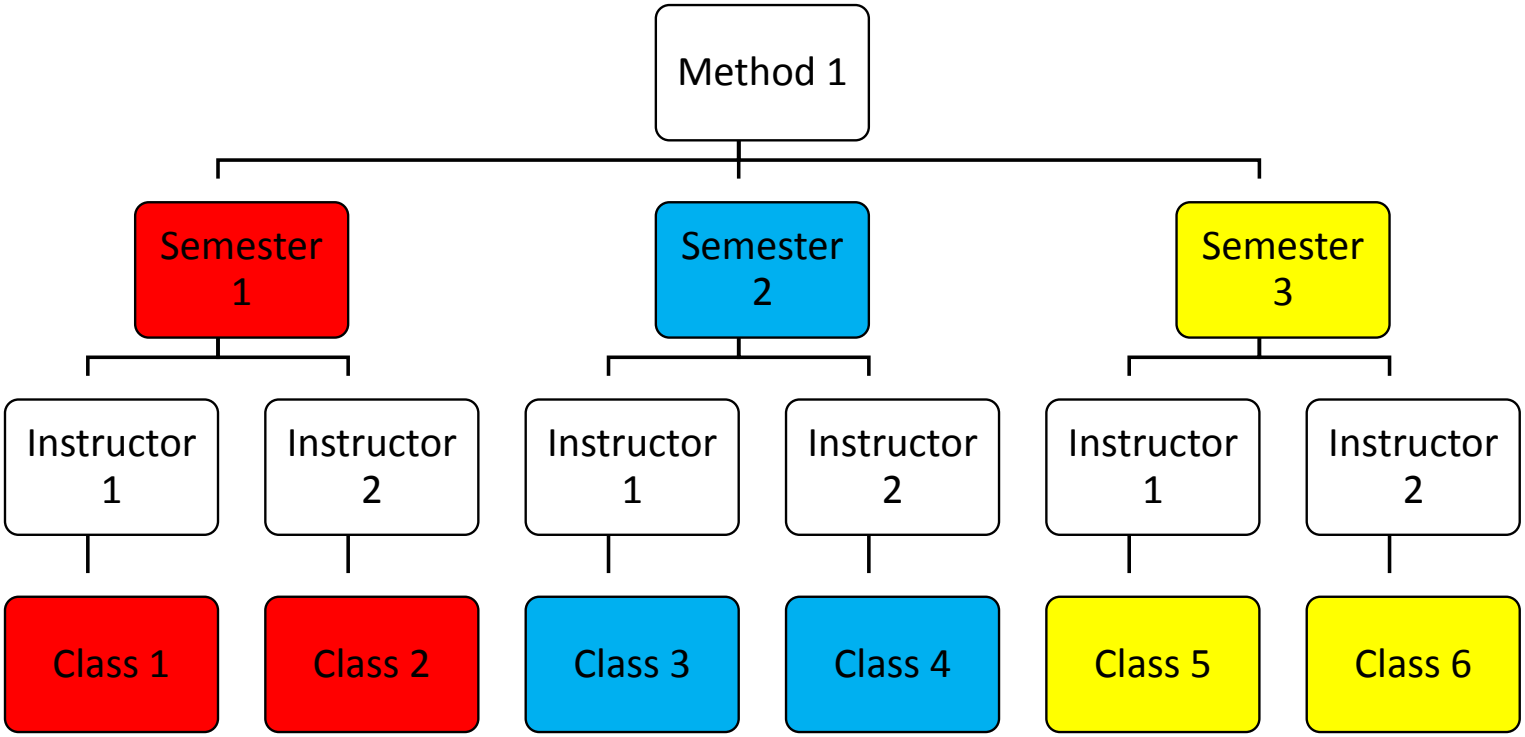
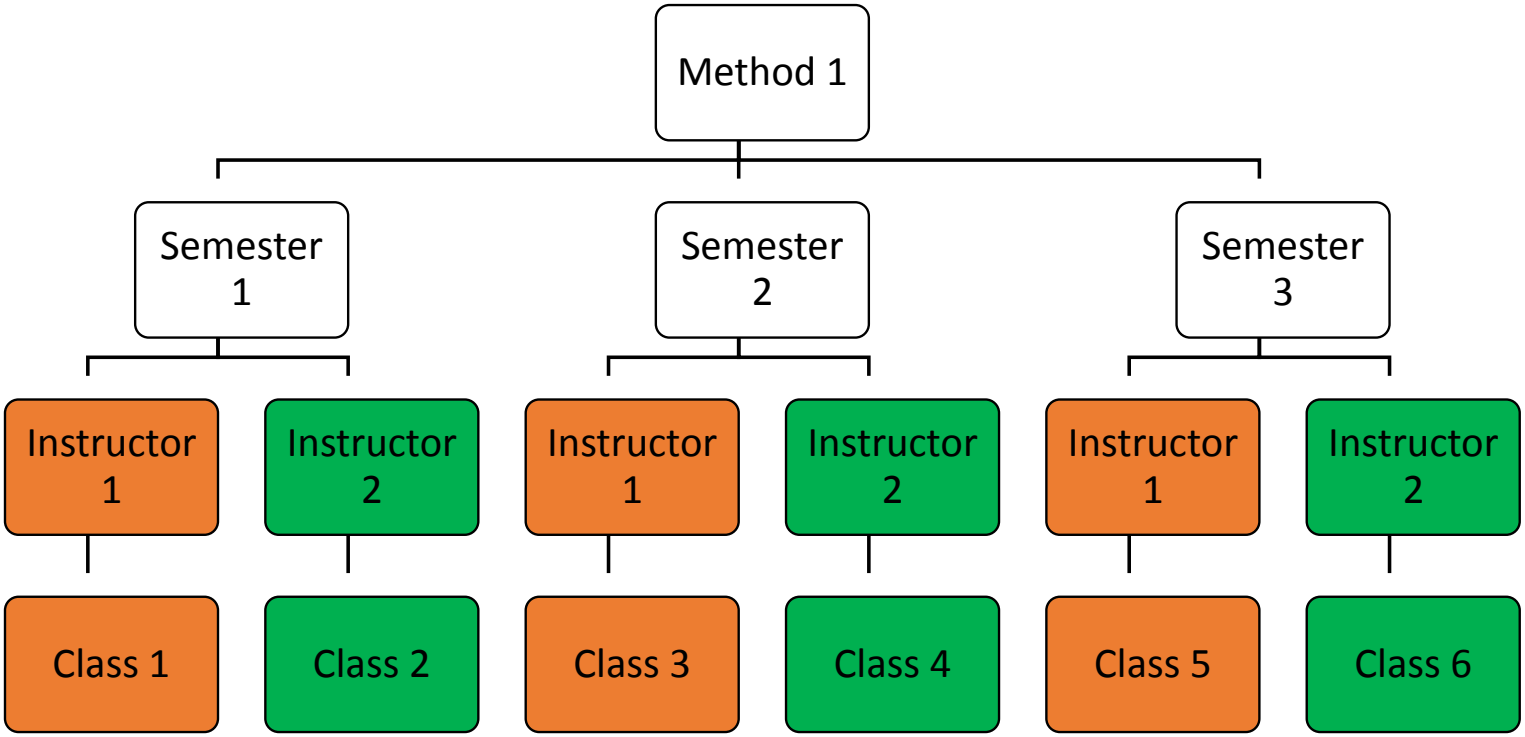


Illustration of Variance Groupings - V_I

grouping observations by instructor



Moment Based Estimates of Variances

Both Models: $\widehat{\delta}^2 = L(V_C)$ — V_C – average variance within class

Model 1: $\widehat{\tau}^2 = L(\widehat{\delta}^2, V_S)$ — V_S – average variance of semester

Model 2: $\widehat{\epsilon}^2 = L(\widehat{\delta}^2, V_I)$ — V_I – average variance of instructors

Model 1: $\widehat{\psi}^2 = L(\hat{\delta}^2, \widehat{\tau}^2, M_1, M_2, V_A)$ — V_A – average variance among all obs

Model 2: $\widehat{\sigma}^2 = L(\widehat{\delta}^2, \widehat{\epsilon}^2, M_1, M_2, V_A)$ — M_1, M_2 – mean of all obs of method 1/2, respectively.

Estimating technique allows for negative variance estimates

- Hypothesized that negative variance estimates more common under incorrect model

Analysis of Moment Based Variance Estimators

Procedure:

1. Simulate data from Model 1 (instructors nested within semesters)
2. Calculate $\widehat{\delta}^2$, $\widehat{\tau}^2$, $\widehat{\epsilon}^2$, $\widehat{\psi}^2$, and $\widehat{\sigma}^2$
3. Repeat above steps 5,000 times
4. Calculate the proportion of times each moment based estimator was less than 0

Varying Sample Size and Variance Magnitudes in Simulated Data

Simulated data from 8 different situations

- Sample Sizes used:
  1. 3 Semesters, 3 Instructors - 18 total classes
  2. 3 Semesters, 10 Instructors - 60 total classes
  3. 10 Semesters, 3 Instructors - 60 total classes
  4. 10 Semesters, 10 Instructors - 200 total classes

- Variance magnitudes used:

Parameter	Small Difference in Magnitudes	Large Difference in Magnitudes
δ^2	$12^2 = 144$	$16^2 = 256$
τ^2	$10^2 = 100$	$7^2 = 49$
ψ^2	$9^2 = 81$	$4^2 = 16$

Results- Small Difference in Variance Magnitudes

	Both Models	Model 1 (Correct)		Model 2 (Incorrect)	
	δ^2	τ^2	ψ^2	ϵ^2	σ^2
3 Sem, 3 Ins	0	0	0.1758	0	0.8558
3 Sem, 10 Ins	0	0	0.0332	0	0.9074
10 Sem, 3 Ins	0	0	0.0058	0	0.9188
10 Sem, 10 Ins	0	0	0	0	0.9682

Table: Proportion of times (out of 5,000) each variance component was estimated to be less than zero.

Results- Large Difference in Variance Magnitudes

	Both Models	Model 1 (Correct)		Model 2 (Incorrect)	
	δ^2	τ^2	ψ^2	ϵ^2	σ^2
3 Sem, 3 Ins	0	0.0004	0.3764	0	0.7582
3 Sem, 10 Ins	0	0	0.1486	0	0.7632
10 Sem, 3 Ins	0	0	0.0960	0	0.8048
10 Sem, 10 Ins	0	0	0.0010	0	0.8300

Table: Proportion of times (out of 5,000) each variance component was estimated to be less than zero.

Conclusions

- Negative variance estimates arise under the incorrect model due to incorrect groupings of observations
- As sample sizes increase, phenomena becomes more pronounced
- As among student variance at smallest grouping increases, phenomena becomes less pronounced
- Moving forward we want to explore a statistic based on variance estimates to determine appropriate nesting structure.

Thank You

Questions?