Hierarchical Models

Donald E. Brown

School of Data Science
University of Virginia
Charlottesville, VA 22904

Agenda

Bayesian Multilevel Models

- Multilevel models are generalizations of regression models where the parameters vary by group

Bayesian Multilevel Models

- Multilevel models are generalizations of regression models where the parameters vary by group
- In Bayesian multilevel models, the parameters are random with distributions

Bayesian Multilevel Models

- Multilevel models are generalizations of regression models where the parameters vary by group
- In Bayesian multilevel models, the parameters are random with distributions
- Hierarchical models are special cases of multilevel models where the groups are nested

Bayesian Multilevel Models

- Multilevel models are generalizations of regression models where the parameters vary by group
- In Bayesian multilevel models, the parameters are random with distributions
- Hierarchical models are special cases of multilevel models where the groups are nested
- Multilevel models allow us to answer questions about grouped data where there are varying sizes to the groups, e.g.,

Bayesian Multilevel Models

- Multilevel models are generalizations of regression models where the parameters vary by group
- In Bayesian multilevel models, the parameters are random with distributions
- Hierarchical models are special cases of multilevel models where the groups are nested
- Multilevel models allow us to answer questions about grouped data where there are varying sizes to the groups, e.g.,
  - Schools and student or teacher scores

Bayesian Multilevel Models

- Multilevel models are generalizations of regression models where the parameters vary by group
- In Bayesian multilevel models, the parameters are random with distributions
- Hierarchical models are special cases of multilevel models where the groups are nested
- Multilevel models allow us to answer questions about grouped data where there are varying sizes to the groups, e.g.,
  - Schools, and student or teacher scores
  - Hospitals, and patient or caregiver results

Bayesian Multilevel Models

- Multilevel models are generalizations of regression models where the parameters vary by group
- In Bayesian multilevel models, the parameters are random with distributions
- Hierarchical models are special cases of multilevel models where the groups are nested
- Multilevel models allow us to answer questions about grouped data where there are varying sizes to the groups, e.g.,
  - Schools, and student or teacher scores
  - Hospitals, and patient or caregiver results
  - Political and geographic units, and economic or social outcomes

Problem Structure for Hierarchical Models

- Suppose we have N observations of a target variable $y_i, i = 1, \ldots, N$ with predictors in an $N \times k$ matrix $\mathbf{X}$.

Problem Structure for Hierarchical Models

- Suppose we have N observations of a target variable $y_i, i = 1, \ldots, N$ with predictors in an $N \times k$ matrix $\mathbf{X}$.
- Let X_i be the i^{th} row of $\mathbf{X}$.

Problem Structure for Hierarchical Models

- Suppose we have N observations of a target variable $y_i, i = 1, \ldots, N$ with predictors in an $N \times k$ matrix $\mathbf{X}$.
- Let X_i be the i^{th} row of $\mathbf{X}$.
- Suppose we have J groups indexed by $j = 1, \ldots, J$

Problem Structure for Hierarchical Models

- Suppose we have N observations of a target variable $y_i, i = 1, \ldots, N$ with predictors in an $N \times k$ matrix $\mathbf{X}$.
- Let X_i be the i^{th} row of $\mathbf{X}$.
- Suppose we have J groups indexed by $j = 1, \ldots, J$
- Traditional regression model with dummy variables, $z_1, \ldots, z_{J-1}$

$$y_i = \alpha + \boldsymbol{\theta}\mathbf{X}_i + z_1 + \cdots + z_{J-1} + \epsilon_i$$

Pooling Models

- A complete pooling model - ignore the groups

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

Pooling Models

- A complete pooling model - ignore the groups

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

- A no pooling model - one equation for each group, $j = 1, \ldots, J$

$$y_{j[i]} = \alpha_{j[i]} + \boldsymbol{\theta}_{j[i]}\mathbf{X}_{j[i]} + \epsilon_{j[i]}$$

Partial Pooling

- Partial pooling - no predictors

$$\hat{\alpha}_j \approx \frac{\frac{N_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{\text{all}}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

where σ_y^2 is the within -group variance computed using either an average or a value for each group (i.e., $\sigma_{y,j}^2$) and σ_α^2 is the variance of the average values, $\bar{y}_j$, in the groups. $\bar{y}_{\text{all}}$ is the overall average.

Partial Pooling

- Partial pooling - no predictors

$$\hat{\alpha}_j \approx \frac{\frac{N_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{\text{all}}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

  where σ_y^2 is the within– group variance computed using either an average or a value for each group (i.e., $\sigma_{y,j}^2$) and σ_α^2 is the variance of the average values, $\bar{y}_j$, in the groups. $\bar{y}_{\text{all}}$ is the overall average.
- Averages from groups with less observations carry less weight and will "shrink" toward the overall average

Partial Pooling

- Partial pooling - no predictors

$$\hat{\alpha}_j \approx \frac{\frac{N_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{\text{all}}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

where σ_y^2 is the within– group variance computed using either an average or a value for each group (i.e., $\sigma_{y,j}^2$) and σ_α^2 is the variance of the average values, $\bar{y}_j$, in the groups. $\bar{y}_{\text{all}}$ is the overall average.

- Averages from groups with less observations carry less weight and will “shrink” toward the overall average
- Averages from bigger groups carry more weight and are closer to their unpooled estimates

Basic Hierarchical Models

- A basic hierarchical model with varying intercepts and varying slopes is given by

$$y_i = \alpha_{j[i]} + \boldsymbol{\theta}_{j[i]}\mathbf{X}_i + \epsilon_i$$

where $\alpha_{j[i]}, \boldsymbol{\theta}_{j[i]}$ are the parameters for the group, say j, associated with observation i.

Summary

- Complete pooling cannot provide estimates for groups in data

Summary

- Complete pooling cannot provide estimates for groups in data
- No pooling gives poor and possible extreme estimates for groups with a small size

Summary

- Complete pooling cannot provide estimates for groups in data
- No pooling gives poor and possible extreme estimates for groups with a small size
- Partial pooling captures differences in groups and better estimates smaller-sized groups but does not use predictor variables

Summary

- Complete pooling cannot provide estimates for groups in data
- No pooling gives poor and possible extreme estimates for groups with a small size
- Partial pooling captures differences in groups and better estimates smaller–sized groups but does not use predictor variables
- Hierarchical modeling uses the available information in the data, i.e., predictor variables, to better predict group target or response values, even in small groups

Agenda

Example - Radon Testing (Gellman & Hill, 2006)

- Radon is a radioactive gas that enters homes through contact points with the ground.

Example - Radon Testing (Gellman & Hill, 2006)

- Radon is a radioactive gas that enters homes through contact points with the ground.
- It is a carcinogen that is the primary cause of lung cancer in non-smokers.

Example - Radon Testing (Gellman & Hill, 2006)

- Radon is a radioactive gas that enters homes through contact points with the ground.
- It is a carcinogen that is the primary cause of lung cancer in non-smokers.
- Varying levels between houses.

Example - Radon Testing (Gellman & Hill, 2006)

- Radon is a radioactive gas that enters homes through contact points with the ground.
- It is a carcinogen that is the primary cause of lung cancer in non-smokers.
- Varying levels between houses.
- EPA measured levels in 80k houses - example data just from Minnesota

Example - Radon Testing (Gellman & Hill, 2006)

- Radon is a radioactive gas that enters homes through contact points with the ground.
- It is a carcinogen that is the primary cause of lung cancer in non-smokers.
- Varying levels between houses.
- EPA measured levels in 80k houses - example data just from Minnesota
- Two important predictors:

Example - Radon Testing (Gellman & Hill, 2006)

- Radon is a radioactive gas that enters homes through contact points with the ground.
- It is a carcinogen that is the primary cause of lung cancer in non-smokers.
- Varying levels between houses.
- EPA measured levels in 80k houses - example data just from Minnesota
- Two important predictors:
  - Floor - basement or first

Example - Radon Testing (Gellman & Hill, 2006)

- Radon is a radioactive gas that enters homes through contact points with the ground.
- It is a carcinogen that is the primary cause of lung cancer in non-smokers.
- Varying levels in between houses.
- EPA measured levels in 80k houses - example data just from Minnesota
- Two important predictors:
  - Floor - basement or first
  - County uranium levels