Fitting and understanding multilevel (hierarchical) models

Andrew Gelman
Department of Statistics and Department of Political Science
Columbia University

8 December 2004

Making more use of existing information

- The problem: not enough data to estimate effects with confidence
- The solution: make your studies *broader* and *deeper*
  - Broader: extend to other countries, other years, other outcomes, . . .
  - Deeper: inferences for individual states, demographic subgroups, components of outcomes, . . .
- The solution: multilevel modeling

Making more use of existing information

- The problem: not enough data to estimate effects with confidence
- The solution: make your studies *broader* and *deeper*
  - Broader: extend to other countries, other years, other outcomes, . . .
  - Deeper: inferences for individual states, demographic subgroups, components of outcomes, . . .
- The solution: multilevel modeling

Making more use of existing information

- The problem: not enough data to estimate effects with confidence
- The solution: make your studies *broader* and *deeper*
  - Broader: extend to other countries, other years, other outcomes, ...
  - Deeper: inferences for individual states, demographic subgroups, components of outcomes, ...
- The solution: multilevel modeling

Making more use of existing information

- The problem: not enough data to estimate effects with confidence
- The solution: make your studies *broader* and *deeper*
  - Broader: extend to other countries, other years, other outcomes, . . .
  - Deeper: inferences for individual states, demographic subgroups, components of outcomes, . . .
- The solution: multilevel modeling
  - Regression with coefficients grouped into batches
    No such thing as "too many predictors"

Making more use of existing information

- The problem: not enough data to estimate effects with confidence
- The solution: make your studies *broader* and *deeper*
  - Broader: extend to other countries, other years, other outcomes, . . .
  - Deeper: inferences for individual states, demographic subgroups, components of outcomes, . . .
- The solution: multilevel modeling
  - Regression with coefficients grouped into batches
  - No such thing as "too many predictors"

Making more use of existing information

- The problem: not enough data to estimate effects with confidence
- The solution: make your studies *broader* and *deeper*
  - Broader: extend to other countries, other years, other outcomes, . . .
  - Deeper: inferences for individual states, demographic subgroups, components of outcomes, . . .
- The solution: multilevel modeling
  - Regression with coefficients grouped into batches
  - No such thing as "too many predictors"

Making more use of existing information

- The problem: not enough data to estimate effects with confidence
- The solution: make your studies *broader* and *deeper*
  - Broader: extend to other countries, other years, other outcomes, ...
  - Deeper: inferences for individual states, demographic subgroups, components of outcomes, ...
- The solution: multilevel modeling
  - Regression with coefficients grouped into batches
  - No such thing as "too many predictors"

Making more use of existing information

- The problem: not enough data to estimate effects with confidence
- The solution: make your studies *broader* and *deeper*
  - Broader: extend to other countries, other years, other outcomes, . . .
  - Deeper: inferences for individual states, demographic subgroups, components of outcomes, . . .
- The solution: multilevel modeling
  - Regression with coefficients grouped into batches
  - No such thing as "too many predictors"

Fitting and understanding multilevel models

- The effectiveness of multilevel models
- Multilevel models in unexpected places
- Multilevel models as a way of life
- collaborators:
  - Iain Pardoe, Dept of Decision Sciences, University of Oregon
  - David Park, Dept of Political Science, Washington University
  - Joe Bafumi, Dept of Political Science, Columbia University
  - Boris Shor, Dept of Political Science, Columbia University
  - Noah Kaplan, Dept of Political Science, University of Houston
  - Shouhao Zhao, Dept of Statistics, Columbia University
  - Zaiying Huang, Circulation, New York Times

Fitting and understanding multilevel models

- The effectiveness of multilevel models
- Multilevel models in unexpected places
- Multilevel models as a way of life
- collaborators:
  - Iain Pardoe, Dept of Decision Sciences, University of Oregon
  - David Park, Dept of Political Science, Washington University
  - Joe Bafumi, Dept of Political Science, Columbia University
  - Boris Shor, Dept of Political Science, Columbia University
  - Noah Kaplan, Dept of Political Science, University of Houston
  - Shouhao Zhao, Dept of Statistics, Columbia University
  - Zaiying Huang, Circulation, New York Times

Fitting and understanding multilevel models

- The effectiveness of multilevel models
- Multilevel models in unexpected places
- Multilevel models as a way of life
- collaborators:
  - Iain Pardoe, Dept of Decision Sciences, University of Oregon
  - David Park, Dept of Political Science, Washington University
  - Joe Bafumi, Dept of Political Science, Columbia University
  - Boris Shor, Dept of Political Science, Columbia University
  - Noah Kaplan, Dept of Political Science, University of Houston
  - Shouhao Zhao, Dept of Statistics, Columbia University
  - Zaiying Huang, Circulation, New York Times

Fitting and understanding multilevel models

- The effectiveness of multilevel models
- Multilevel models in unexpected places
- Multilevel models as a way of life
- collaborators:
  - Iain Pardoe, Dept of Decision Sciences, University of Oregon
  - David Park, Dept of Political Science, Washington University
  - Joe Bafumi, Dept of Political Science, Columbia University
  - Boris Shor, Dept of Political Science, Columbia University
  - Noah Kaplan, Dept of Political Science, University of Houston
  - Shouhao Zhao, Dept of Statistics, Columbia University
  - Zaiying Huang, Circulation, New York Times

Fitting and understanding multilevel models

- The effectiveness of multilevel models
- Multilevel models in unexpected places
- Multilevel models as a way of life
- collaborators:
  - Iain Pardoe, Dept of Decision Sciences, University of Oregon
  - David Park, Dept of Political Science, Washington University
  - Joe Bafumi, Dept of Political Science, Columbia University
  - Boris Shor, Dept of Political Science, Columbia University
  - Noah Kaplan, Dept of Political Science, University of Houston
  - Shouhao Zhao, Dept of Statistics, Columbia University
  - Zaiying Huang, Circulation, New York Times

Outline of talk

- The effectiveness of multilevel models
  - State-level opinions from national polls
    (crossed multilevel modeling and poststratification)
- Multilevel models in unexpected places
- Multilevel models as a way of life

Outline of talk

- The effectiveness of multilevel models
  - State-level opinions from national polls
    (crossed multilevel modeling and poststratification)
- Multilevel models in unexpected places
  - Estimating incumbency advantage and its variation
    Before-after studies
- Multilevel models as a way of life

Outline of talk

- The effectiveness of multilevel models
  - State-level opinions from national polls (crossed multilevel modeling and poststratification)
- Multilevel models in unexpected places
  - Estimating incumbency advantage and its variation
  - Before-after studies
- Multilevel models as a way of life

Outline of talk

- The effectiveness of multilevel models
  - State-level opinions from national polls
    (crossed multilevel modeling and poststratification)
- Multilevel models in unexpected places
  - Estimating incumbency advantage and its variation
  - Before-after studies
- Multilevel models as a way of life

Outline of talk

- The effectiveness of multilevel models
  - State-level opinions from national polls
    (crossed multilevel modeling and poststratification)
- Multilevel models in unexpected places
  - Estimating incumbency advantage and its variation
  - Before-after studies
- Multilevel models as a way of life

Outline of talk

- The effectiveness of multilevel models
  - State-level opinions from national polls
    (crossed multilevel modeling and poststratification)
- Multilevel models in unexpected places
  - Estimating incumbency advantage and its variation
  - Before-after studies
- Multilevel models as a way of life
  - Building and fitting models
  - Displaying and summarizing inferences

Outline of talk

- The effectiveness of multilevel models
  - State-level opinions from national polls
    (crossed multilevel modeling and poststratification)
- Multilevel models in unexpected places
  - Estimating incumbency advantage and its variation
  - Before-after studies
- Multilevel models as a way of life
  - Building and fitting models
  - Displaying and summarizing inferences

Outline of talk

- The effectiveness of multilevel models
  - State-level opinions from national polls
    (crossed multilevel modeling and poststratification)
- Multilevel models in unexpected places
  - Estimating incumbency advantage and its variation
  - Before-after studies
- Multilevel models as a way of life
  - Building and fitting models
  - Displaying and summarizing inferences

Outline of talk

- The effectiveness of multilevel models
  - State-level opinions from national polls
    (crossed multilevel modeling and poststratification)
- Multilevel models in unexpected places
  - Estimating incumbency advantage and its variation
  - Before-after studies
- Multilevel models as a way of life
  - Building and fitting models
  - Displaying and summarizing inferences

National opinion trends

State-level opinion trends

- Goal: estimating time series within each state
- One poll at a time: small-area estimation
- It works! Validated for pre-election polls
- Combining surveys: model for parallel time series
- Multilevel modeling + poststratification
- Poststratification cells: sex × ethnicity × age × education × state

State-level opinion trends

- Goal: estimating time series within each state
- One poll at a time: small-area estimation
- It works! Validated for pre-election polls
- Combining surveys: model for parallel time series
- Multilevel modeling + poststratification
- Poststratification cells: sex × ethnicity × age × education × state

State-level opinion trends

- Goal: estimating time series within each state
- One poll at a time: small-area estimation
- It works! Validated for pre-election polls
- Combining surveys: model for parallel time series
- Multilevel modeling + poststratification
- Poststratification cells: sex × ethnicity × age × education × state

State-level opinion trends

- Goal: estimating time series within each state
- One poll at a time: small-area estimation
- It works! Validated for pre-election polls
- Combining surveys: model for parallel time series
- Multilevel modeling + poststratification
- Poststratification cells: sex × ethnicity × age × education × state

State-level opinion trends

- Goal: estimating time series within each state
- One poll at a time: small-area estimation
- It works! Validated for pre-election polls
- Combining surveys: model for parallel time series
- Multilevel modeling + poststratification
- Poststratification cells: sex × ethnicity × age × education × state

State-level opinion trends

- Goal: estimating time series within each state
- One poll at a time: small-area estimation
- It works! Validated for pre-election polls
- Combining surveys: model for parallel time series
- Multilevel modeling + poststratification
- Poststratification cells: sex $\times$ ethnicity $\times$ age $\times$ education $\times$ state

Multilevel modeling of opinions

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- X includes demographic and geographic predictors
- Group-level model for the 16 age $\times$ education predictors
- Group-level model for the 50 state predictors
- Bayesian inference, summarize by posterior simulations of β:

Simulation	θ_1	$\cdots$	θ_{75}
1	**	$\cdots$	**
$\vdots$	$\vdots$	$\ddots$	$\vdots$
1000	**	$\cdots$	**

Multilevel modeling of opinions

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- X includes demographic and geographic predictors
- Group-level model for the 16 age $\times$ education predictors
- Group-level model for the 50 state predictors
- Bayesian inference, summarize by posterior simulations of β:

Simulation	θ_1	$\cdots$	θ_{75}
1	**	$\cdots$	**
$\vdots$	$\vdots$	$\ddots$	$\vdots$
1000	**	$\cdots$	**

Multilevel modeling of opinions

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- X includes demographic and geographic predictors
- Group-level model for the 16 age $\times$ education predictors
- Group-level model for the 50 state predictors
- Bayesian inference, summarize by posterior simulations of β:

Simulation	θ_1	$\cdots$	θ_{75}
1	**	$\cdots$	**
$\vdots$	$\vdots$	$\ddots$	$\vdots$
1000	**	$\cdots$	**

Multilevel modeling of opinions

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- X includes demographic and geographic predictors
- Group-level model for the 16 age $\times$ education predictors
- Group-level model for the 50 state predictors
- Bayesian inference, summarize by posterior simulations of β:

Simulation	θ_1	$\cdots$	θ_{75}
1	**	$\cdots$	**
$\vdots$	$\vdots$	$\ddots$	$\vdots$
1000	**	$\cdots$	**

Multilevel modeling of opinions

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- X includes demographic and geographic predictors
- Group-level model for the 16 age $\times$ education predictors
- Group-level model for the 50 state predictors
- Bayesian inference, summarize by posterior simulations of β:

Simulation	θ_1	$\cdots$	θ_{75}
1	**	$\cdots$	**
$\vdots$	$\vdots$	$\ddots$	$\vdots$
1000	**	$\cdots$	**

Multilevel modeling of opinions

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- X includes demographic and geographic predictors
- Group-level model for the 16 age $\times$ education predictors
- Group-level model for the 50 state predictors
- Bayesian inference, summarize by posterior simulations of β:

Simulation	θ_1	$\cdots$	θ_{75}
1	**	...	**
$\vdots$	$\vdots$	$\ddots$	$\vdots$
1000	**	...	**

Interlude: why "multilevel" $\neq$ "hierarchical"

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- X includes demographic and geographic predictors
- Group-level model for the 16 age $\times$ education predictors
- Group-level model for the 50 state predictors
- Crossed (nonnested) structure of age, education, state
- Several overlapping "hierarchies"

Interlude: why "multilevel" $\neq$ "hierarchical"

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- X includes demographic and geographic predictors
- Group-level model for the 16 age $\times$ education predictors
- Group-level model for the 50 state predictors
- Crossed (nonnested) structure of age, education, state
- Several overlapping "hierarchies"

Interlude: why "multilevel" $\neq$ "hierarchical"

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- X includes demographic and geographic predictors
- Group-level model for the 16 age $\times$ education predictors
- Group-level model for the 50 state predictors
- Crossed (nonnested) structure of age, education, state
- Several overlapping "hierarchies"

Interlude: why "multilevel" $\neq$ "hierarchical"

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- X includes demographic and geographic predictors
- Group-level model for the 16 age $\times$ education predictors
- Group-level model for the 50 state predictors
- Crossed (nonnested) structure of age, education, state
- Several overlapping "hierarchies"

Interlude: why "multilevel" $\neq$ "hierarchical"

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- X includes demographic and geographic predictors
- Group-level model for the 16 age $\times$ education predictors
- Group-level model for the 50 state predictors
- Crossed (nonnested) structure of age, education, state
- Several overlapping "hierarchies"

Interlude: why "multilevel" $\neq$ "hierarchical"

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- X includes demographic and geographic predictors
- Group-level model for the 16 age $\times$ education predictors
- Group-level model for the 50 state predictors
- Crossed (nonnested) structure of age, education, state
- Several overlapping "hierarchies"

Interlude: why "multilevel" $\neq$ "hierarchical"

- Logistic regression: $\Pr(y_i = 1) = \text{logit}^{-1}((X\beta)_i)$
- X includes demographic and geographic predictors
- Group-level model for the 16 age $\times$ education predictors
- Group-level model for the 50 state predictors
- Crossed (nonnested) structure of age, education, state
- Several overlapping "hierarchies"

Poststratification to estimate state opinions

- Implied inference for $\theta_j = \text{logit}^{-1}(X\beta)$ in each of 3264 cells j (e.g., black female, age 18–29, college graduate, Massachusetts)
- Poststratification
  - Within each state s, average over 64 cells:
    $\sum_{j \in s} N_j \theta_j \Big/ \sum_{j \in s} N_j$
    N_j = population in cell j (from Census)
    1000 simulation draws propagate to uncertainty for each θ_s

Poststratification to estimate state opinions

- Implied inference for $\theta_j = \text{logit}^{-1}(X\beta)$ in each of 3264 cells j (e.g., black female, age 18–29, college graduate, Massachusetts)
- Poststratification
  - Within each state s, average over 64 cells: $\sum_{j \in s} N_j \theta_j \Big/ \sum_{j \in s} N_j$
  - N_j = population in cell j (from Census)
  - 1000 simulation draws propagate to uncertainty for each θ_s

Poststratification to estimate state opinions

- Implied inference for $\theta_j = \text{logit}^{-1}(X\beta)$ in each of 3264 cells j (e.g., black female, age 18–29, college graduate, Massachusetts)
- Poststratification
  - Within each state s, average over 64 cells: $\sum_{j \in s} N_j \theta_j \Big/ \sum_{j \in s} N_j$
  - $N_j =$ population in cell j (from Census)
  - 1000 simulation draws propagate to uncertainty for each θ_j

Poststratification to estimate state opinions

- Implied inference for $\theta_j = \text{logit}^{-1}(X\beta)$ in each of 3264 cells j (e.g., black female, age 18–29, college graduate, Massachusetts)
- Poststratification
  - Within each state s, average over 64 cells: $\sum_{j \in s} N_j \theta_j \Big/ \sum_{j \in s} N_j$
  - N_j = population in cell j (from Census)
  - 1000 simulation draws propagate to uncertainty for each θ_j

Poststratification to estimate state opinions

- Implied inference for $\theta_j = \text{logit}^{-1}(X\beta)$ in each of 3264 cells j (e.g., black female, age 18–29, college graduate, Massachusetts)
- Poststratification
  - Within each state s, average over 64 cells: $\sum_{j\in s} N_j\theta_j \Big/ \sum_{j \in s} N_j$
  - N_j = population in cell j (from Census)
  - 1000 simulation draws propagate to uncertainty for each θ_j

Poststratification to estimate state opinions

- Implied inference for $\theta_j = \text{logit}^{-1}(X\beta)$ in each of 3264 cells j (e.g., black female, age 18–29, college graduate, Massachusetts)
- Poststratification
  - Within each state s, average over 64 cells: $\sum_{j \in s} N_j \theta_j \Big/ \sum_{j \in s} N_j$
  - N_j = population in cell j (from Census)
  - 1000 simulation draws propagate to uncertainty for each θ_j

CBS/New York Times pre-election polls from 1988

- Validation study: fit model on poll data and compare to election results
- Competing estimates:
  - No pooling: separate estimate within each state
  - Complete pooling: no state predictors
  - Hierarchical model and poststratify
- Mean absolute state errors:

CBS/New York Times pre-election polls from 1988

- Validation study: fit model on poll data and compare to election results
- Competing estimates:
  - No pooling: separate estimate within each state
  - Complete pooling: no state predictors
  - Hierarchical model and poststratify
- Mean absolute state errors:

CBS/New York Times pre-election polls from 1988

- Validation study: fit model on poll data and compare to election results
- Competing estimates:
  - No pooling: separate estimate within each state
  - Complete pooling: no state predictors
  - Hierarchical model and poststratify
- Mean absolute state errors:

CBS/New York Times pre-election polls from 1988

- Validation study: fit model on poll data and compare to election results
- Competing estimates:
  - No pooling: separate estimate within each state
  - Complete pooling: no state predictors
  - Hierarchical model and poststratify
- Mean absolute state errors:

CBS/New York Times pre-election polls from 1988

- Validation study: fit model on poll data and compare to election results
- Competing estimates:
  - No pooling: separate estimate within each state
  - Complete pooling: no state predictors
  - Hierarchical model and poststratify
- Mean absolute state errors:
  - No pooling: 10.4%
  - Complete pooling: 5.4%
  - Hierarchical model with poststratification: 4.5%

CBS/New York Times pre-election polls from 1988

- Validation study: fit model on poll data and compare to election results
- Competing estimates:
  - No pooling: separate estimate within each state
  - Complete pooling: no state predictors
  - Hierarchical model and poststratify
- Mean absolute state errors:
  - No pooling: 10.4%
  - Complete pooling: 5.4%
  - Hierarchical model with poststratification: 4.5%

CBS/New York Times pre-election polls from 1988

- Validation study: fit model on poll data and compare to election results
- Competing estimates:
  - No pooling: separate estimate within each state
  - Complete pooling: no state predictors
  - Hierarchical model and poststratify
- Mean absolute state errors:
  - No pooling: 10.4%
  - Complete pooling: 5.4%
  - Hierarchical model with poststratification: 4.5%

CBS/New York Times pre-election polls from 1988

- Validation study: fit model on poll data and compare to election results
- Competing estimates:
  - No pooling: separate estimate within each state
  - Complete pooling: no state predictors
  - Hierarchical model and poststratify
- Mean absolute state errors:
  - No pooling: 10.4%
  - Complete pooling: 5.4%
  - Hierarchical model with poststratification: 4.5%

CBS/New York Times pre-election polls from 1988

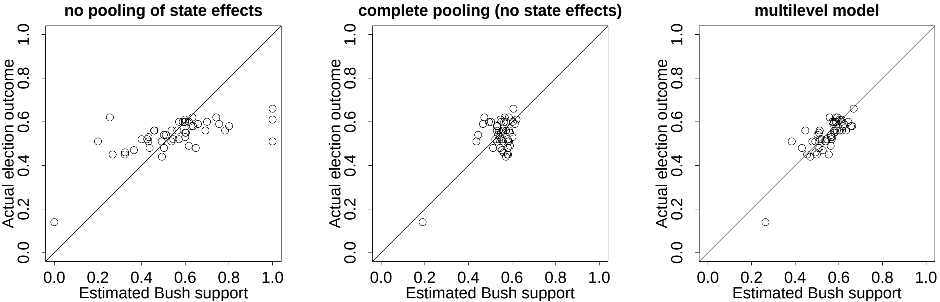
- Validation study: fit model on poll data and compare to election results
- Competing estimates:
  - No pooling: separate estimate within each state
  - Complete pooling: no state predictors
  - Hierarchical model and poststratify
- Mean absolute state errors:
  - No pooling: 10.4%
  - Complete pooling: 5.4%
  - Hierarchical model with poststratification: 4.5%

CBS/New York Times pre-election polls from 1988

- Validation study: fit model on poll data and compare to election results
- Competing estimates:
  - No pooling: separate estimate within each state
  - Complete pooling: no state predictors
  - Hierarchical model and poststratify
- Mean absolute state errors:
  - No pooling: 10.4%
  - Complete pooling: 5.4%
  - Hierarchical model with poststratification: 4.5%

Validation study: comparison of state errors

1988 election outcome vs. poll estimate



Multilevel models always

- Anything worth doing is worth doing repeatedly
- A "method" is any procedure applied more than once
- City planning
  - Outward expansion: fitting a model to other countries, other years, other outcomes, . . .
  - Infilling: inferences for individual states, demographic subgroups, components of data, . . .
- "Frequentist" statistical theory of repeated inferences

Multilevel models always

- Anything worth doing is worth doing repeatedly
- A "method" is any procedure applied more than once
- City planning
  - Outward expansion: fitting a model to other countries, other years, other outcomes, ...
  - Infilling: inferences for individual states, demographic subgroups, components of data, ...
- "Frequentist" statistical theory of repeated inferences

Multilevel models always

- Anything worth doing is worth doing repeatedly
- A “method” is any procedure applied more than once
- City planning
  - Outward expansion: fitting a model to other countries, other years, other outcomes, ...
  - Infilling: inferences for individual states, demographic subgroups, components of data, ...
- “Frequentist” statistical theory of repeated inferences

Multilevel models always

- Anything worth doing is worth doing repeatedly
- A "method" is any procedure applied more than once
- City planning
  - Outward expansion: fitting a model to other countries, other years, other outcomes, ...
  - Infilling: inferences for individual states, demographic subgroups, components of data, ...
- "Frequentist" statistical theory of repeated inferences

Multilevel models always

- Anything worth doing is worth doing repeatedly
- A "method" is any procedure applied more than once
- City planning
  - Outward expansion: fitting a model to other countries, other years, other outcomes, . . .
  - Infilling: inferences for individual states, demographic subgroups, components of data, . . .
- "Frequentist" statistical theory of repeated inferences

Multilevel models always

- Anything worth doing is worth doing repeatedly
- A "method" is any procedure applied more than once
- City planning
  - Outward expansion: fitting a model to other countries, other years, other outcomes, ...
  - Infilling: inferences for individual states, demographic subgroups, components of data, ...
- "Frequentist" statistical theory of repeated inferences

Incumbency advantage in U.S. House elections

- Regression approach (Gelman and King, 1990):
  - For any year, compare districts with and without incs running
  - Control for vote in previous election
  - Control for incumbent *party*
  - $v_{it} = \beta_0 + \beta_1 v_{i,t-1} + \beta_2 P_{it} + \psi I_{it} + \epsilon_{it}$
- Other estimates (sophomore surge, etc.) have selection bias

Incumbency advantage in U.S. House elections

- Regression approach (Gelman and King, 1990):
  - For any year, compare districts with and without incs running
  - Control for vote in previous election
  - Control for incumbent *party*
  - $v_{it} = \beta_0 + \beta_1 v_{i,t-1} + \beta_2 P_{it} + \psi I_{it} + \epsilon_{it}$
- Other estimates (sophomore surge, etc.) have selection bias

Incumbency advantage in U.S. House elections

- Regression approach (Gelman and King, 1990):
  - For any year, compare districts with and without incs running
  - Control for vote in previous election
  - Control for incumbent *party*
  - $v_{it} = \beta_0 + \beta_1 v_{i,t-1} + \beta_2 P_{it} + \psi I_{it} + \epsilon_{it}$
- Other estimates (sophomore surge, etc.) have selection bias

Incumbency advantage in U.S. House elections

- Regression approach (Gelman and King, 1990):
  - For any year, compare districts with and without incs running
  - Control for vote in previous election
  - Control for incumbent *party*
  - $v_{it} = \beta_0 + \beta_1 v_{i,t-1} + \beta_2 P_{it} + \psi I_{it} + \epsilon_{it}$
- Other estimates (sophomore surge, etc.) have selection bias

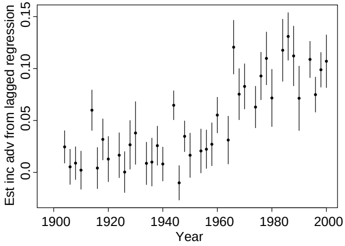
Incumbency advantage in U.S. House elections

- Regression approach (Gelman and King, 1990):
  - For any year, compare districts with and without incs running
  - Control for vote in previous election
  - Control for incumbent *party*
  - $v_{it} = \beta_0 + \beta_1 v_{i,t-1} + \beta_2 P_{it} + \psi I_{it} + \epsilon_{it}$
- Other estimates (sophomore surge, etc.) have selection bias

Incumbency advantage in U.S. House elections

- Regression approach (Gelman and King, 1990):
  - For any year, compare districts with and without incs running
  - Control for vote in previous election
  - Control for incumbent *party*
  - $v_{it} = \beta_0 + \beta_1 v_{i,t-1} + \beta_2 P_{it} + \psi I_{it} + \epsilon_{it}$
- Other estimates (sophomore surge, etc.) have selection bias

Estimated incumbency advantage from lagged regressions



Can we do better?

- Regression estimate: $v_{it} = \beta_0 + \beta_1 v_{i,t-1} + \beta_2 P_{it} + \psi I_{it} + \epsilon_{it}$
- "Political science" problem: ψ is assumed to be same in all districts
- "Statistics" problem: the model doesn't fit the data
- We'll show pictures of the model not fitting
- We'll set up a model allowing inc advantage to vary

Can we do better?

- Regression estimate: $v_{it} = \beta_0 + \beta_1 v_{i,t-1} + \beta_2 P_{it} + \psi I_{it} + \epsilon_{it}$
- "Political science" problem: ψ is assumed to be same in all districts
- "Statistics" problem: the model doesn't fit the data
- We'll show pictures of the model not fitting
- We'll set up a model allowing inc advantage to vary

Can we do better?

- Regression estimate: $v_{it} = \beta_0 + \beta_1 v_{i,t-1} + \beta_2 P_{it} + \psi I_{it} + \epsilon_{it}$
- "Political science" problem: ψ is assumed to be same in all districts
- "Statistics" problem: the model doesn't fit the data
- We'll show pictures of the model not fitting
- We'll set up a model allowing inc advantage to vary

Can we do better?

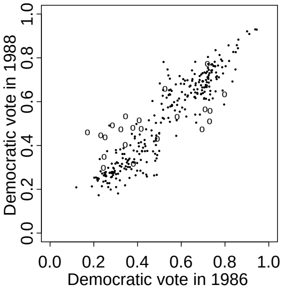
- Regression estimate: $v_{it} = \beta_0 + \beta_1 v_{i,t-1} + \beta_2 P_{it} + \psi I_{it} + \epsilon_{it}$
- "Political science" problem: ψ is assumed to be same in all districts
- "Statistics" problem: the model doesn't fit the data
- We'll show pictures of the model not fitting
- We'll set up a model allowing inc advantage to vary

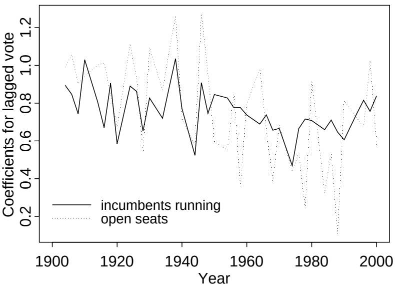
Can we do better?

- Regression estimate: $v_{it} = \beta_0 + \beta_1 v_{i,t-1} + \beta_2 P_{it} + \psi I_{it} + \epsilon_{it}$
- "Political science" problem: ψ is assumed to be same in all districts
- "Statistics" problem: the model doesn't fit the data
- We'll show pictures of the model not fitting
- We'll set up a model allowing inc advantage to vary

Model misfit

Under the model, parallel lines are fitted to the circles (open seats) and dots (incs running for reelection)





Multilevel model

- for $t = 1, 2$: $v_{it} = 0.5 + \delta_t + \alpha_i + \phi_{it} I_{it} + \epsilon_{it}$
  - $\delta_2 - \delta_1$ is the national vote swing
  - α_i is the "normal vote" for district i: mean 0, sd σ_α.
  - ϕ_{it} is the inc advantage in district i at time t: mean ψ, sd σ_ϕ
  - ϵ_{it}'s are independent errors: mean 0 and sd σ_ϵ.
- Candidate-level incumbency effects:

Multilevel model

- for $t = 1, 2$: $v_{it} = 0.5 + \delta_t + \alpha_i + \phi_{it} I_{it} + \epsilon_{it}$
  - $\delta_2 - \delta_1$ is the national vote swing
  - α_i is the "normal vote" for district i: mean 0, sd σ_α.
  - ϕ_{it} is the inc advantage in district i at time t: mean ψ, sd σ_ϕ
  - ϵ_{it}'s are independent errors: mean 0 and sd σ_ϵ.
- Candidate-level incumbency effects:

Multilevel model

- for $t = 1, 2$: $v_{it} = 0.5 + \delta_t + \alpha_i + \phi_{it} I_{it} + \epsilon_{it}$
  - $\delta_2 - \delta_1$ is the national vote swing
  - α_i is the "normal vote" for district i: mean 0, sd σ_α.
  - ϕ_{it} is the inc advantage in district i at time t: mean ψ, sd σ_ϕ
  - ϵ_{it}'s are independent errors: mean 0 and sd σ_ϵ.
- Candidate-level incumbency effects:

Multilevel model

- for $t = 1, 2$: $v_{it} = 0.5 + \delta_t + \alpha_i + \phi_{it} I_{it} + \epsilon_{it}$
  - $\delta_2 - \delta_1$ is the national vote swing
  - α_i is the "normal vote" for district i: mean 0, sd σ_α.
  - ϕ_{it} is the inc advantage in district i at time t: mean ψ, sd σ_ϕ
  - ϵ_{it}'s are independent errors: mean 0 and sd σ_ϵ.
- Candidate-level incumbency effects:
  - If the same incumbent is running in years 1 and 2, then $\phi_{i2} = \phi_{i1}$.
    Otherwise, ϕ_{i1} and ϕ_{i2} are independent.

Multilevel model

- for $t = 1, 2$: $v_{it} = 0.5 + \delta_t + \alpha_i + \phi_{it} I_{it} + \epsilon_{it}$
  - $\delta_2 - \delta_1$ is the national vote swing
  - α_i is the "normal vote" for district i: mean 0, sd σ_α.
  - ϕ_{it} is the inc advantage in district i at time t: mean ψ, sd σ_ϕ
  - ϵ_{it}'s are independent errors: mean 0 and sd σ_ϵ.
- Candidate-level incumbency effects:
  - If the same incumbent is running in years 1 and 2, then $\phi_{i2} = \phi_{i1}$
  - Otherwise, ϕ_{i1} and ϕ_{i2} are independent

Multilevel model

- for $t = 1, 2$: $v_{it} = 0.5 + \delta_t + \alpha_i + \phi_{it} I_{it} + \epsilon_{it}$
  - $\delta_2 - \delta_1$ is the national vote swing
  - α_i is the "normal vote" for district i: mean 0, sd σ_α.
  - ϕ_{it} is the inc advantage in district i at time t: mean ψ, sd σ_ϕ
  - ϵ_{it}'s are independent errors: mean 0 and sd σ_ϵ.
- Candidate-level incumbency effects:
  - If the same incumbent is running in years 1 and 2, then $\phi_{i2} \equiv \phi_{i1}$
  - Otherwise, ϕ_{i1} and ϕ_{i2} are independent

Multilevel model

- for $t = 1, 2$: $v_{it} = 0.5 + \delta_t + \alpha_i + \phi_{it} I_{it} + \epsilon_{it}$
  - $\delta_2 - \delta_1$ is the national vote swing
  - α_i is the "normal vote" for district i: mean 0, sd σ_α.
  - ϕ_{it} is the inc advantage in district i at time t: mean ψ, sd σ_ϕ
  - ϵ_{it}'s are independent errors: mean 0 and sd σ_ϵ.
- Candidate-level incumbency effects:
  - If the same incumbent is running in years 1 and 2, then $\phi_{i2} \equiv \phi_{i1}$
  - Otherwise, ϕ_{i1} and ϕ_{i2} are independent

Multilevel model

- for $t = 1, 2$: $v_{it} = 0.5 + \delta_t + \alpha_i + \phi_{it} I_{it} + \epsilon_{it}$
  - $\delta_2 - \delta_1$ is the national vote swing
  - α_i is the "normal vote" for district i: mean 0, sd σ_α.
  - ϕ_{it} is the inc advantage in district i at time t: mean ψ, sd σ_ϕ
  - ϵ_{it}'s are independent errors: mean 0 and sd σ_ϵ.
- Candidate-level incumbency effects:
  - If the same incumbent is running in years 1 and 2, then $\phi_{i2} \equiv \phi_{i1}$
  - Otherwise, ϕ_{i1} and ϕ_{i2} are independent

Fitting the multilevel model

- Bayesian inference
- Linear parameters: national vote swings, district effects, incumbency effects
- 3 variance parameters: district effects, incumbency effects, residual errors
- Need to model a selection effect: information provided by the incumbent party at time 1
- Solve analytically for Pr(inclusion), include factor in the likelihood
- Gibbs-Metropolis sampling, program in Splus

Fitting the multilevel model

- Bayesian inference
- Linear parameters: national vote swings, district effects, incumbency effects
- 3 variance parameters: district effects, incumbency effects, residual errors
- Need to model a selection effect: information provided by the incumbent party at time 1
- Solve analytically for Pr(inclusion), include factor in the likelihood
- Gibbs-Metropolis sampling, program in Splus

Fitting the multilevel model

- Bayesian inference
- Linear parameters: national vote swings, district effects, incumbency effects
- 3 variance parameters: district effects, incumbency effects, residual errors
- Need to model a selection effect: information provided by the incumbent party at time 1
- Solve analytically for Pr(inclusion), include factor in the likelihood
- Gibbs-Metropolis sampling, program in Splus

Fitting the multilevel model

- Bayesian inference
- Linear parameters: national vote swings, district effects, incumbency effects
- 3 variance parameters: district effects, incumbency effects, residual errors
- Need to model a selection effect: information provided by the incumbent party at time 1
- Solve analytically for Pr(inclusion), include factor in the likelihood
- Gibbs-Metropolis sampling, program in Splus

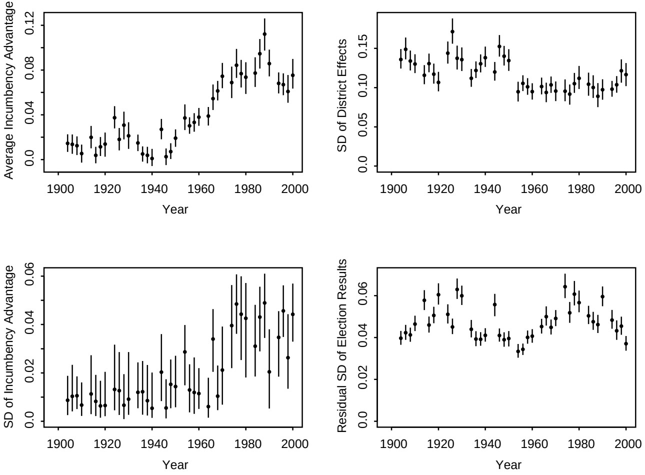
Fitting the multilevel model

- Bayesian inference
- Linear parameters: national vote swings, district effects, incumbency effects
- 3 variance parameters: district effects, incumbency effects, residual errors
- Need to model a selection effect: information provided by the incumbent party at time 1
- Solve analytically for Pr(inclusion), include factor in the likelihood
- Gibbs-Metropolis sampling, program in Splus

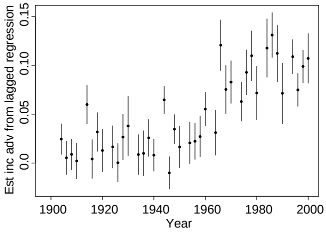
Fitting the multilevel model

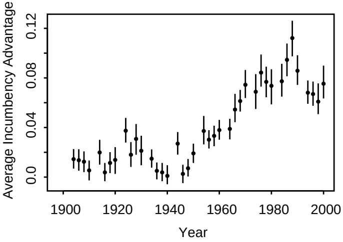
- Bayesian inference
- Linear parameters: national vote swings, district effects, incumbency effects
- 3 variance parameters: district effects, incumbency effects, residual errors
- Need to model a selection effect: information provided by the incumbent party at time 1
- Solve analytically for Pr(inclusion), include factor in the likelihood
- Gibbs-Metropolis sampling, program in Splus

Estimated incumbency advantage and its variation



Compare old and new estimates



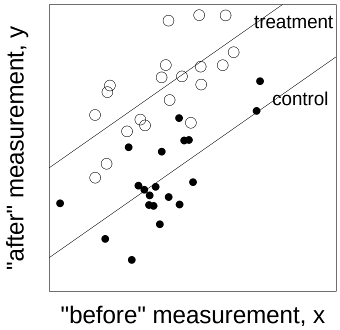


No-interaction model

- Before-after data with treatment and control groups
- Default model: constant treatment effects
  - Fisher's classical null hyp: effect is zero for all cases
  - Regression model: $y_i = T_i\theta + X_i\beta + \epsilon_i$

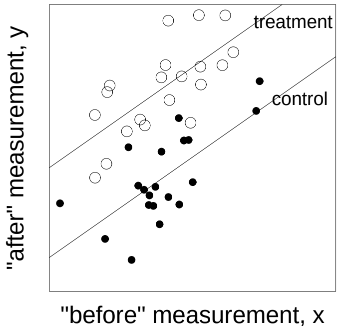
No-interaction model

- Before-after data with treatment and control groups
- Default model: constant treatment effects
  - Fisher's classical null hyp: effect is zero for all cases
  - Regression model: $y_i = T_i\theta + X_i\beta + \epsilon_i$



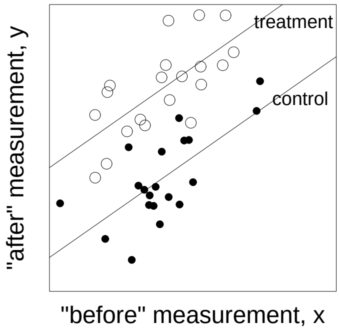
No-interaction model

- Before-after data with treatment and control groups
- Default model: constant treatment effects
  - Fisher's classical null hyp: effect is zero for all cases
  - Regression model: $y_i = T_i\theta + X_i\beta + \epsilon_i$



No-interaction model

- Before-after data with treatment and control groups
- Default model: constant treatment effects
  - Fisher's classical null hyp: effect is zero for all cases
  - Regression model: $y_i = T_i\theta + X_i\beta + \epsilon_i$



Actual data show interactions

- Treatment interacts with "before" measurement
- Before-after correlation is higher for *controls* than for *treated* units
- Examples
  - An observational study of legislative redistricting
    An experiment with pre-test, post-test data
    Congressional elections with incumbents and open seats

Actual data show interactions

- Treatment interacts with "before" measurement
- Before-after correlation is higher for *controls* than for *treated* units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data
  - Congressional elections with incumbents and open seats

Actual data show interactions

- Treatment interacts with "before" measurement
- Before-after correlation is higher for *controls* than for *treated* units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data
  - Congressional elections with incumbents and open seats

Actual data show interactions

- Treatment interacts with "before" measurement
- Before-after correlation is higher for *controls* than for *treated* units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data
  - Congressional elections with incumbents and open seats

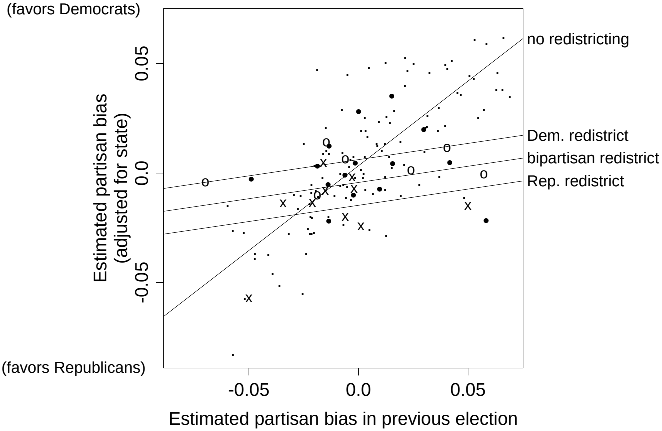
Actual data show interactions

- Treatment interacts with "before" measurement
- Before-after correlation is higher for *controls* than for *treated* units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data
  - Congressional elections with incumbents and open seats

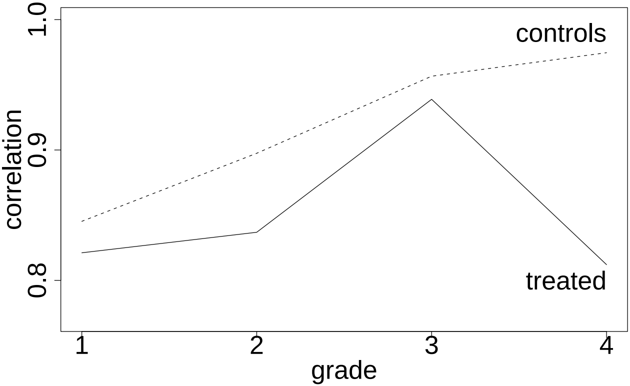
Actual data show interactions

- Treatment interacts with "before" measurement
- Before-after correlation is higher for *controls* than for *treated* units
- Examples
  - An observational study of legislative redistricting
  - An experiment with pre-test, post-test data
  - Congressional elections with incumbents and open seats

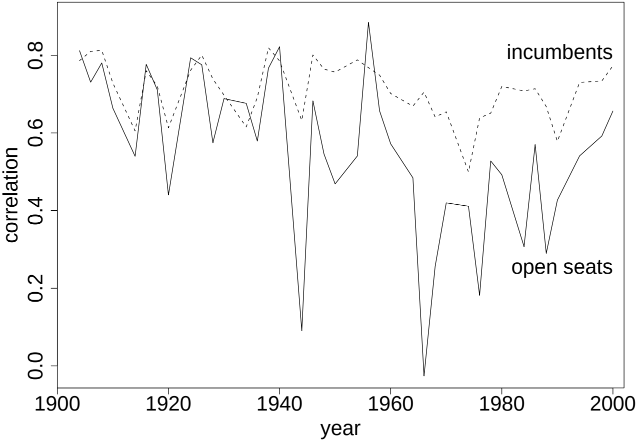
Observational study of legislative redistricting before-after data



Experiment: correlation between pre-test and post-test data for controls and for treated units



Correlation between two successive Congressional elections for incumbents running (controls) and open seats (treated)



Interactions as variance components

Unit-level "error term" η_i

- For control units, η_i persists from time 1 to time 2
- For treatment units, η_i changes:
  - Subtractive treatment error (η_i only at time 1)
  - Additive treatment error (η_i only at time 2)
  - Replacement treatment error
- Under all these models, the before-after correlation is higher for controls than treated units

Interactions as variance components

Unit-level "error term" η_i

- For control units, η_i persists from time 1 to time 2
- For treatment units, η_i changes:
  - Subtractive treatment error (η_i only at time 1)
  - Additive treatment error (η_i only at time 2)
  - Replacement treatment error
- Under all these models, the before-after correlation is higher for controls than treated units

Interactions as variance components

Unit-level "error term" η_i

- For control units, η_i persists from time 1 to time 2
- For treatment units, η_i changes:
  - Subtractive treatment error (η_i only at time 1)
  - Additive treatment error (η_i only at time 2)
  - Replacement treatment error
- Under all these models, the before-after correlation is higher for controls than treated units

Interactions as variance components

Unit-level "error term" η_i

- For control units, η_i persists from time 1 to time 2
- For treatment units, η_i changes:
  - Subtractive treatment error (η_i only at time 1)
  - Additive treatment error (η_i only at time 2)
  - Replacement treatment error
- Under all these models, the before-after correlation is higher for controls than treated units

Interactions as variance components

Unit-level “error term” η_i

- For control units, η_i persists from time 1 to time 2
- For treatment units, η_i changes:
  - Subtractive treatment error (η_i only at time 1)
  - Additive treatment error (η_i only at time 2)
  - Replacement treatment error
- Under all these models, the before-after correlation is higher for controls than treated units

Interactions as variance components

Unit-level "error term" η_i

- For control units, η_i persists from time 1 to time 2
- For treatment units, η_i changes:
  - Subtractive treatment error (η_i only at time 1)
  - Additive treatment error (η_i only at time 2)
  - Replacement treatment error
- Under all these models, the before-after correlation is higher for controls than treated units

Interactions as variance components

Unit-level "error term" η_i

- For control units, η_i persists from time 1 to time 2
- For treatment units, η_i changes:
  - Subtractive treatment error (η_i only at time 1)
  - Additive treatment error (η_i only at time 2)
  - Replacement treatment error
- Under all these models, the before-after correlation is higher for controls than treated units

Some new tools

- Building and fitting multilevel models
- Displaying and summarizing inferences

Some new tools

- Building and fitting multilevel models
- Displaying and summarizing inferences

Some new tools

- Building and fitting multilevel models
- Displaying and summarizing inferences

Building and fitting multilevel models

- A reparameterization can change a model (even if it leaves the likelihood unchanged)
- Redundant additive parameterization
- Redundant multiplicative parameterization
- Weakly-informative prior distribution for group-level variance parameters

Building and fitting multilevel models

- A reparameterization can change a model (even if it leaves the likelihood unchanged)
- Redundant additive parameterization
- Redundant multiplicative parameterization
- Weakly-informative prior distribution for group-level variance parameters

Building and fitting multilevel models

- A reparameterization can change a model (even if it leaves the likelihood unchanged)
- Redundant additive parameterization
- Redundant multiplicative parameterization
- Weakly-informative prior distribution for group-level variance parameters

Building and fitting multilevel models

- A reparameterization can change a model (even if it leaves the likelihood unchanged)
- Redundant additive parameterization
- Redundant multiplicative parameterization
- Weakly-informative prior distribution for group-level variance parameters

Building and fitting multilevel models

- A reparameterization can change a model
  (even if it leaves the likelihood unchanged)
- Redundant additive parameterization
- Redundant multiplicative parameterization
- Weakly-informative prior distribution for group-level variance parameters

Redundant parameterization

- Data model: $\Pr(y_i = 1) = \text{logit}^{-1}\left(\beta^0 + \beta^{\text{age}}_{\text{age(i)}} + \beta^{\text{state}}_{\text{state(i)}}\right)$
- Usual model for the coefficients:

$$\begin{aligned} \beta^{\text{age}}_j &\sim \text{N}(0, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\ \beta^{\text{state}}_j &\sim \text{N}(0, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50 \end{aligned}$$

- Additively redundant model:

$$\begin{aligned} \beta^{\text{age}}_j &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\ \beta^{\text{state}}_j &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50 \end{aligned}$$

- Why add the redundant $\mu_{\text{age}}, \mu_{\text{state}}$?

Redundant parameterization

- Data model: $\Pr(y_i = 1) = \text{logit}^{-1}\left(\beta^0 + \beta^{\text{age}}_{\text{age(i)}} + \beta^{\text{state}}_{\text{state(i)}}\right)$
- Usual model for the coefficients:

$$\begin{aligned}\beta_j^{\text{age}} &\sim \text{N}(0, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4\\ \beta_j^{\text{state}} &\sim \text{N}(0, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50\end{aligned}$$

- Additively redundant model:

$$\begin{aligned}\beta_j^{\text{age}} &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4\\ \beta_j^{\text{state}} &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50\end{aligned}$$

- Why add the redundant $\mu_{\text{age}}, \mu_{\text{state}}$?

Redundant parameterization

- Data model: $\Pr(y_i = 1) = \text{logit}^{-1}\left(\beta^0 + \beta^{\text{age}}_{\text{age(i)}} + \beta^{\text{state}}_{\text{state(i)}}\right)$
- Usual model for the coefficients:

$$\begin{aligned} \beta^{\text{age}}_j &\sim \text{N}(0, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \dots, 4 \\ \beta^{\text{state}}_j &\sim \text{N}(0, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \dots, 50 \end{aligned}$$

- Additively redundant model:

$$\begin{aligned} \beta^{\text{age}}_j &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \dots, 4 \\ \beta^{\text{state}}_j &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \dots, 50 \end{aligned}$$

- Why add the redundant $\mu_{\text{age}}, \mu_{\text{state}}$?
  - Iterative algorithm moves more smoothly

Redundant parameterization

- Data model: $\Pr(y_i = 1) = \text{logit}^{-1}\left(\beta^0 + \beta^{\text{age}}_{\text{age(i)}} + \beta^{\text{state}}_{\text{state(i)}}\right)$
- Usual model for the coefficients:

$$\begin{aligned} \beta^{\text{age}}_j &\sim \text{N}(0, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\ \beta^{\text{state}}_j &\sim \text{N}(0, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50 \end{aligned}$$

- Additively redundant model:

$$\begin{aligned} \beta^{\text{age}}_j &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\ \beta^{\text{state}}_j &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50 \end{aligned}$$

- Why add the redundant $\mu_{\text{age}}, \mu_{\text{state}}$?
  - Iterative algorithm moves more smoothly

Redundant parameterization

- Data model: $\Pr(y_i = 1) = \text{logit}^{-1}\left(\beta^0 + \beta^{\text{age}}_{\text{age(i)}} + \beta^{\text{state}}_{\text{state(i)}}\right)$
- Usual model for the coefficients:

$$\begin{aligned}\beta_j^{\text{age}} &\sim \text{N}(0, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4\\ \beta_j^{\text{state}} &\sim \text{N}(0, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50\end{aligned}$$

- Additively redundant model:

$$\begin{aligned}\beta_j^{\text{age}} &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4\\ \beta_j^{\text{state}} &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50\end{aligned}$$

- Why add the redundant $\mu_{\text{age}}, \mu_{\text{state}}$?
  - Iterative algorithm moves more smoothly

Redundant parameterization

- Data model: $\Pr(y_i = 1) = \text{logit}^{-1}\left(\beta^0 + \beta^{\text{age}}_{\text{age(i)}} + \beta^{\text{state}}_{\text{state(i)}}\right)$
- Usual model for the coefficients:

$$
\begin{aligned}
\beta_j^{\text{age}} &\sim \text{N}(0, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\
\beta_j^{\text{state}} &\sim \text{N}(0, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50
\end{aligned}
$$

- Additively redundant model:

$$
\begin{aligned}
\beta_j^{\text{age}} &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\
\beta_j^{\text{state}} &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50
\end{aligned}
$$

- Why add the redundant $\mu_{\text{age}}, \mu_{\text{state}}$?
  - Iterative algorithm moves more smoothly

Redundant additive parameterization

- Model

$$
\begin{aligned}
\Pr(y_i = 1) &= \operatorname{logit}^{-1}\left(\beta^0 + \beta^{\text{age}}_{\text{age(i)}} + \beta^{\text{state}}_{\text{state(i)}}\right) \\
\beta^{\text{age}}_j &\sim \mathrm{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\
\beta^{\text{state}}_j &\sim \mathrm{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50
\end{aligned}
$$

- Identify using centered parameters:

$$
\begin{aligned}
\tilde{\beta}^{\text{age}}_j &= \beta^{\text{age}}_j - \bar{\beta}^{\text{age}}, \quad \text{for } j = 1, \ldots, 4 \\
\tilde{\beta}^{\text{state}}_j &= \beta^{\text{state}}_j - \bar{\beta}^{\text{state}}, \quad \text{for } j = 1, \ldots, 50
\end{aligned}
$$

- Redefine the constant term:

$$\tilde{\beta}^0 = \beta^0 + \bar{\beta}^{\text{age}} + \bar{\beta}^{\text{age}}$$

Redundant additive parameterization

- Model

$$\begin{aligned}
\Pr(y_i = 1) &= \text{logit}^{-1}\left(\beta^0 + \beta^{\text{age}}_{\text{age(i)}} + \beta^{\text{state}}_{\text{state(i)}}\right) \\
\beta^{\text{age}}_j &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \dots, 4 \\
\beta^{\text{state}}_j &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \dots, 50
\end{aligned}$$

- Identify using centered parameters:

$$\begin{aligned}
\tilde{\beta}^{\text{age}}_j &= \beta^{\text{age}}_j - \bar{\beta}^{\text{age}}, \quad \text{for } j = 1, \dots, 4 \\
\tilde{\beta}^{\text{state}}_j &= \beta^{\text{state}}_j - \bar{\beta}^{\text{state}}, \quad \text{for } j = 1, \dots, 50
\end{aligned}$$

- Redefine the constant term:

$$\tilde{\beta}^0 = \beta^0 + \bar{\beta}^{\text{age}} + \bar{\beta}^{\text{age}}$$

Redundant additive parameterization

- Model

$$\begin{aligned}\Pr(y_i = 1) &= \text{logit}^{-1}\left(\beta^0 + \beta^{\text{age}}_{\text{age(i)}} + \beta^{\text{state}}_{\text{state(i)}}\right) \\ \beta^{\text{age}}_j &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\ \beta^{\text{state}}_j &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50\end{aligned}$$

- Identify using centered parameters:

$$\begin{aligned}\tilde{\beta}^{\text{age}}_j &= \beta^{\text{age}}_j - \bar{\beta}^{\text{age}}, \quad \text{for } j = 1, \ldots, 4 \\ \tilde{\beta}^{\text{state}}_j &= \beta^{\text{state}}_j - \bar{\beta}^{\text{state}}, \quad \text{for } j = 1, \ldots, 50\end{aligned}$$

- Redefine the constant term:

$$\tilde{\beta}^0 = \beta^0 + \bar{\beta}^{\text{age}} + \bar{\beta}^{\text{age}}$$

Redundant multiplicative parameterization

- New model

$$\begin{aligned} \Pr(y_i = 1) &= \text{logit}^{-1}\left(\beta^0 + \xi^{\text{age}}\beta^{\text{age}}_{\text{age(i)}} + \xi^{\text{state}}\beta^{\text{state}}_{\text{state(i)}}\right) \\ \beta^{\text{age}}_j &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\ \beta^{\text{state}}_j &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50 \end{aligned}$$

- Identify using centered and scaled parameters:

$$\begin{aligned} \tilde{\beta}^{\text{age}}_j &= \xi^{\text{age}}(\beta^{\text{age}}_j - \bar{\beta}^{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\ \tilde{\beta}^{\text{state}}_j &= \xi^{\text{state}}\left(\beta^{\text{state}}_j - \bar{\beta}^{\text{state}}\right), \quad \text{for } j = 1, \ldots, 50 \end{aligned}$$

- Faster convergence
- More general model, connections to factor analysis

Redundant multiplicative parameterization

- New model

$$
\begin{aligned}
\Pr(y_i = 1) &= \text{logit}^{-1}\left(\beta^0 + \xi^{\text{age}}\beta^{\text{age}}_{\text{age(i)}} + \xi^{\text{state}}\beta^{\text{state}}_{\text{state(i)}}\right) \\
\beta^{\text{age}}_j &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\
\beta^{\text{state}}_j &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50
\end{aligned}
$$

- Identify using centered and scaled parameters:

$$
\begin{aligned}
\tilde{\beta}^{\text{age}}_j &= \xi^{\text{age}}(\beta^{\text{age}}_j - \bar{\beta}^{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\
\tilde{\beta}^{\text{state}}_j &= \xi^{\text{state}}\left(\beta^{\text{state}}_j - \bar{\beta}^{\text{state}}\right), \quad \text{for } j = 1, \ldots, 50
\end{aligned}
$$

- Faster convergence
- More general model, connections to factor analysis

Redundant multiplicative parameterization

- New model

$$
\begin{aligned}
\Pr(y_i = 1) &= \text{logit}^{-1}\left(\beta^0 + \xi^{\text{age}}\beta^{\text{age}}_{\text{age(i)}} + \xi^{\text{state}}\beta^{\text{state}}_{\text{state(i)}}\right) \\
\beta^{\text{age}}_j &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\
\beta^{\text{state}}_j &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50
\end{aligned}
$$

- Identify using centered and scaled parameters:

$$
\begin{aligned}
\tilde{\beta}^{\text{age}}_j &= \xi^{\text{age}}(\beta^{\text{age}}_j - \bar{\beta}^{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\
\tilde{\beta}^{\text{state}}_j &= \xi^{\text{state}}\left(\beta^{\text{state}}_j - \bar{\beta}^{\text{state}}\right), \quad \text{for } j = 1, \ldots, 50
\end{aligned}
$$

- Faster convergence
- More general model, connections to factor analysis

Redundant multiplicative parameterization

- New model

$$
\begin{aligned}
\Pr(y_i = 1) &= \text{logit}^{-1}\left(\beta^0 + \xi^{\text{age}}\beta^{\text{age}}_{\text{age(i)}} + \xi^{\text{state}}\beta^{\text{state}}_{\text{state(i)}}\right) \\
\beta^{\text{age}}_j &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\
\beta^{\text{state}}_j &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50
\end{aligned}
$$

- Identify using centered and scaled parameters:

$$
\begin{aligned}
\tilde{\beta}^{\text{age}}_j &= \xi^{\text{age}}(\beta^{\text{age}}_j - \bar{\beta}^{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\
\tilde{\beta}^{\text{state}}_j &= \xi^{\text{state}}\left(\beta^{\text{state}}_j - \bar{\beta}^{\text{state}}\right), \quad \text{for } j = 1, \ldots, 50
\end{aligned}
$$

- Faster convergence
- More general model, connections to factor analysis

Weakly informative prior distribution for the multilevel variance parameter

- Redundant multiplicative parameterization:

$$\begin{aligned}
\Pr(y_i = 1) &= \text{logit}^{-1}\left(\beta^0 + \xi^{\text{age}}\beta^{\text{age}}_{\text{age(i)}} + \xi^{\text{state}}\beta^{\text{state}}_{\text{state(i)}}\right) \\
\beta^{\text{age}}_j &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \dots, 4 \\
\beta^{\text{state}}_j &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \dots, 50
\end{aligned}$$

- Separate prior distributions on the ξ and σ parameters:
  - Normal on ξ
  - Inverse-gamma on σ^2
- Generalizes and fixes problems with the standard choices of prior distributions

Weakly informative prior distribution for the multilevel variance parameter

- Redundant multiplicative parameterization:

$$\begin{aligned}
\Pr(y_i = 1) &= \text{logit}^{-1}\left(\beta^0 + \xi^{\text{age}}\beta^{\text{age}}_{\text{age(i)}} + \xi^{\text{state}}\beta^{\text{state}}_{\text{state(i)}}\right) \\
\beta^{\text{age}}_j &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\
\beta^{\text{state}}_j &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50
\end{aligned}$$

- Separate prior distributions on the ξ and σ parameters:
  - Normal on ξ
  - Inverse-gamma on σ^2
- Generalizes and fixes problems with the standard choices of prior distributions

Weakly informative prior distribution for the multilevel variance parameter

- Redundant multiplicative parameterization:

$$\begin{aligned}
\Pr(y_i = 1) &= \text{logit}^{-1}\left(\beta^0 + \xi^{\text{age}}\beta^{\text{age}}_{\text{age(i)}} + \xi^{\text{state}}\beta^{\text{state}}_{\text{state(i)}}\right) \\
\beta^{\text{age}}_j &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\
\beta^{\text{state}}_j &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50
\end{aligned}$$

- Separate prior distributions on the ξ and σ parameters:
  - Normal on ξ
  - Inverse-gamma on σ^2
- Generalizes and fixes problems with the standard choices of prior distributions

Weakly informative prior distribution for the multilevel variance parameter

- Redundant multiplicative parameterization:

$$\begin{aligned}
\Pr(y_i = 1) &= \text{logit}^{-1}\left(\beta^0 + \xi^{\text{age}}\beta^{\text{age}}_{\text{age(i)}} + \xi^{\text{state}}\beta^{\text{state}}_{\text{state(i)}}\right) \\
\beta^{\text{age}}_j &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\
\beta^{\text{state}}_j &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50
\end{aligned}$$

- Separate prior distributions on the ξ and σ parameters:
  - Normal on ξ
  - Inverse-gamma on σ^2
- Generalizes and fixes problems with the standard choices of prior distributions

Weakly informative prior distribution for the multilevel variance parameter

- Redundant multiplicative parameterization:

$$\begin{aligned}
\Pr(y_i = 1) &= \text{logit}^{-1}\left(\beta^0 + \xi^{\text{age}}\beta^{\text{age}}_{\text{age(i)}} + \xi^{\text{state}}\beta^{\text{state}}_{\text{state(i)}}\right) \\
\beta^{\text{age}}_j &\sim \text{N}(\mu_{\text{age}}, \sigma^2_{\text{age}}), \quad \text{for } j = 1, \ldots, 4 \\
\beta^{\text{state}}_j &\sim \text{N}(\mu_{\text{state}}, \sigma^2_{\text{state}}), \quad \text{for } j = 1, \ldots, 50
\end{aligned}$$

- Separate prior distributions on the ξ and σ parameters:
  - Normal on ξ
  - Inverse-gamma on σ^2
- Generalizes and fixes problems with the standard choices of prior distributions

Displaying and summarizing inferences

- Displaying parameters in groups rather than as a long list
- Average predictive effects
- R^2 and partial pooling factors
- Analysis of variance

Displaying and summarizing inferences

- Displaying parameters in groups rather than as a long list
- Average predictive effects
- R^2 and partial pooling factors
- Analysis of variance

Displaying and summarizing inferences

- Displaying parameters in groups rather than as a long list
- Average predictive effects
- R^2 and partial pooling factors
- Analysis of variance

Displaying and summarizing inferences

- Displaying parameters in groups rather than as a long list
- Average predictive effects
- R^2 and partial pooling factors
- Analysis of variance

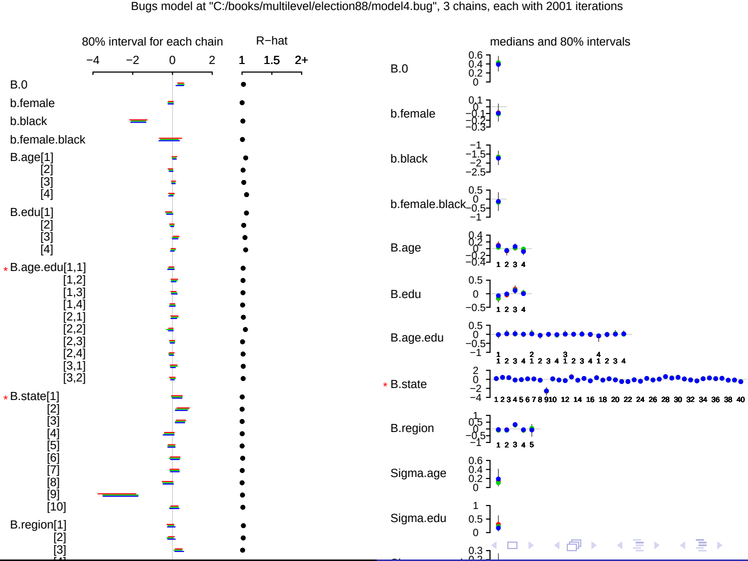
Displaying and summarizing inferences

- Displaying parameters in groups rather than as a long list
- Average predictive effects
- R^2 and partial pooling factors
- Analysis of variance

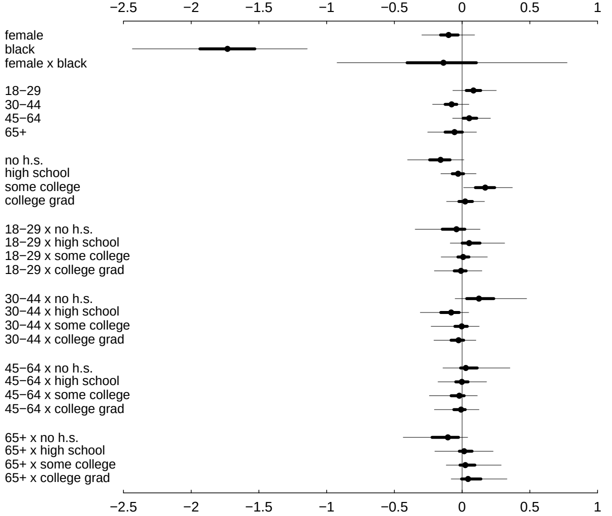
Raw display of inference

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
B.0	0.402	0.147	0.044	0.326	0.413	0.499	0.652	1.024	110
b.female	-0.094	0.102	-0.283	-0.162	-0.095	-0.034	0.107	1.001	1000
b.black	-1.701	0.305	-2.323	-1.910	-1.691	-1.486	-1.152	1.014	500
b.female.black	-0.143	0.393	-0.834	-0.383	-0.155	0.104	0.620	1.007	1000
B.age[1]	0.084	0.088	-0.053	0.012	0.075	0.140	0.277	1.062	45
B.age[2]	-0.072	0.087	-0.260	-0.121	-0.054	-0.006	0.052	1.017	190
B.age[3]	0.044	0.077	-0.105	-0.007	0.038	0.095	0.203	1.029	130
B.age[4]	-0.057	0.096	-0.265	-0.115	-0.052	0.001	0.133	1.076	32
B.edu[1]	-0.148	0.131	-0.436	-0.241	-0.137	-0.044	0.053	1.074	31
B.edu[2]	-0.022	0.082	-0.182	-0.069	-0.021	0.025	0.152	1.028	160
B.edu[3]	0.148	0.112	-0.032	0.065	0.142	0.228	0.370	1.049	45
B.edu[4]	0.023	0.090	-0.170	-0.030	0.015	0.074	0.224	1.061	37
B.age.edu[1,1]	-0.044	0.133	-0.363	-0.104	-0.019	0.025	0.170	1.018	1000
B.age.edu[1,2]	0.059	0.123	-0.153	-0.011	0.032	0.118	0.353	1.016	580
B.age.edu[1,3]	0.049	0.124	-0.146	-0.023	0.022	0.104	0.349	1.015	280
B.age.edu[1,4]	0.001	0.116	-0.237	-0.061	0.000	0.052	0.280	1.010	1000
B.age.edu[2,1]	0.066	0.152	-0.208	-0.008	0.032	0.124	0.449	1.022	140
B.age.edu[2,2]	-0.081	0.127	-0.407	-0.137	-0.055	0.001	0.094	1.057	120
B.age.edu[2,3]	-0.004	0.102	-0.226	-0.048	0.000	0.041	0.215	1.008	940
B.age.edu[2,4]	-0.042	0.108	-0.282	-0.100	-0.026	0.014	0.157	1.017	170
B.age.edu[3,1]	0.034	0.135	-0.215	-0.030	0.009	0.091	0.361	1.021	230
B.age.edu[3,2]	0.007	0.102	-0.213	-0.039	0.003	0.052	0.220	1.019	610
B.age.edu[3,3]	0.033	0.130	-0.215	-0.029	0.009	0.076	0.410	1.080	61
B.age.edu[3,4]	-0.009	0.109	-0.236	-0.064	-0.005	0.043	0.214	1.024	150
B.age.edu[4,1]	-0.141	0.190	-0.672	-0.224	-0.086	-0.003	0.100	1.036	270
B.age.edu[4,2]	-0.014	0.119	-0.280	-0.059	-0.008	0.033	0.239	1.017	240
B.age.edu[4,3]	0.046	0.132	-0.192	-0.024	0.019	0.108	0.332	1.010	210
B.age.edu[4,4]	0.042	0.142	-0.193	-0.022	0.016	0.095	0.377	1.015	160
B.state[1]	0.201	0.211	-0.131	0.047	0.172	0.326	0.646	1.003	960

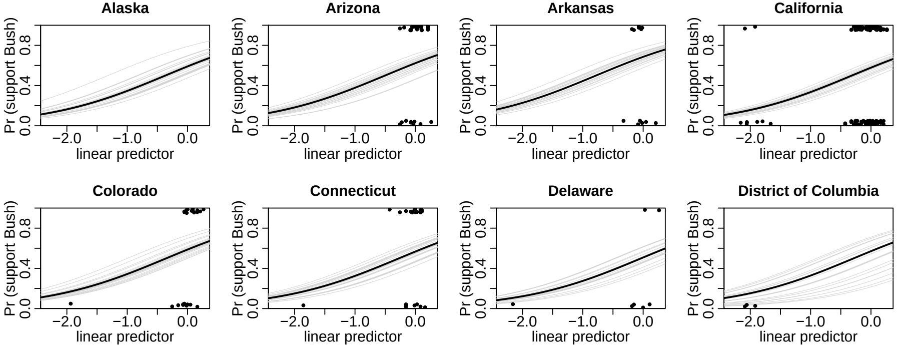
Raw graphical display



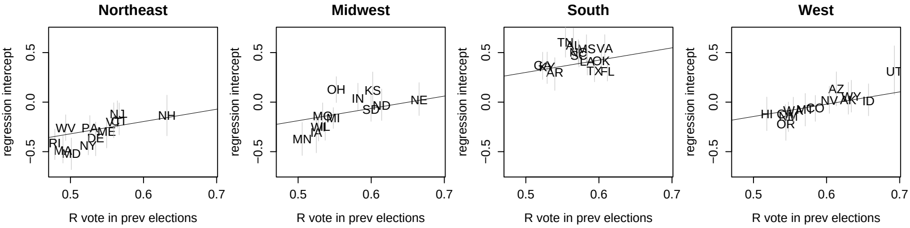
Better graphical display 1: demographics



Better graphical display 2: within states

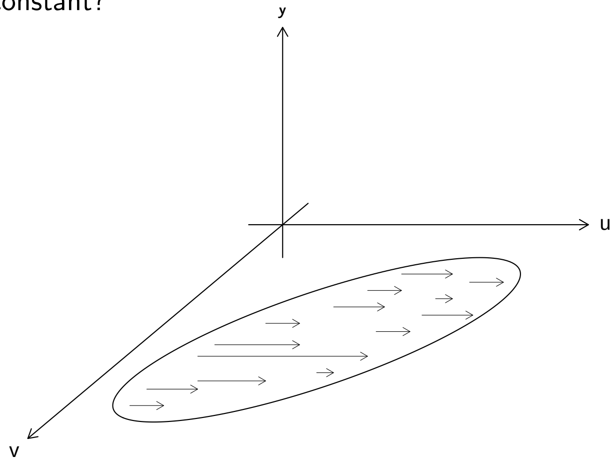


Better graphical display 3: between states



Average predictive effects

- What is $E(y \mid x_1 = \text{high}) - \text{E}(\text{y} \mid \text{x}_1 = \text{low})$, with all other x's held constant?



Average predictive effects

- What is $E(y \,|\, x_1 = \mathrm{high}) - \mathrm{E(y \,|\, x_1 = low)}$, with all other x's held constant?
- In general, difference can depend on x
- Average over distribution of x in the data
  - You can't just use a central value of x
- Compute APE for each input variable x
- Multilevel factors are categorical input variables

Average predictive effects

- What is $E(y \,|\, x_1 = \mathrm{high}) - \mathrm{E}(\mathrm{y} \,|\, \mathrm{x}_1 = \mathrm{low})$, with all other x's held constant?
- In general, difference can depend on x
- Average over distribution of x in the data
  - You can't just use a central value of x
- Compute APE for each input variable x
- Multilevel factors are categorical input variables

Average predictive effects

- What is $E(y \mid x_1 = \mathrm{high}) - \mathrm{E(y \mid x_1 = low)}$, with all other x's held constant?
- In general, difference can depend on x
- Average over distribution of x in the data
  - You can't just use a central value of x
- Compute APE for each input variable x
- Multilevel factors are categorical input variables

Average predictive effects

- What is $E(y \,|\, x_1 = \text{high}) - \text{E}(\text{y} \,|\, \text{x}_1 = \text{low})$, with all other x's held constant?
- In general, difference can depend on x
- Average over distribution of x in the data
  - You can't just use a central value of x
- Compute APE for each input variable x
- Multilevel factors are categorical input variables

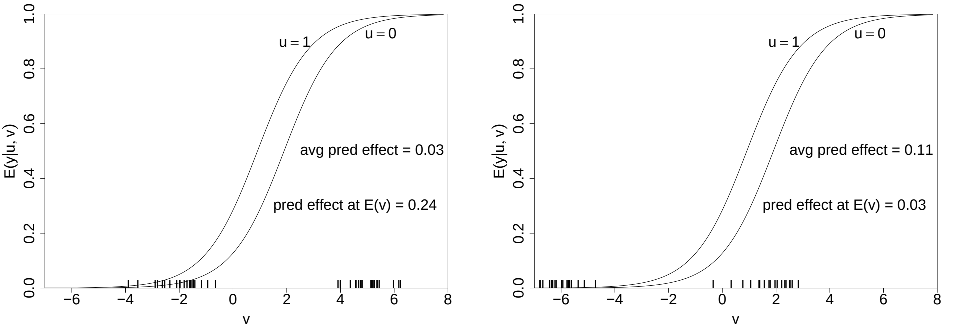
Average predictive effects

- What is $E(y \mid x_1 = \mathrm{high}) - \mathrm{E}(\mathrm{y} \mid \mathrm{x}_1 = \mathrm{low})$, with all other x's held constant?
- In general, difference can depend on x
- Average over distribution of x in the data
  - You can't just use a central value of x
- Compute APE for each input variable x
- Multilevel factors are categorical input variables

Average predictive effects

- What is $E(y \mid x_1 = \mathrm{high}) - \mathrm{E}(\mathrm{y} \mid \mathrm{x}_1 = \mathrm{low})$, with all other x's held constant?
- In general, difference can depend on x
- Average over distribution of x in the data
  - You can't just use a central value of x
- Compute APE for each input variable x
- Multilevel factors are categorical input variables

APE: why you can't just use a central value of x



Understanding sources of variation

- Generalization of R^2 (explained variance), defined at each level of the model
- Partial pooling factor, defined at each level
- Analysis of variance
- Open question: how to construct models with deep interaction structures?

Understanding sources of variation

- Generalization of R^2 (explained variance), defined at each level of the model
- Partial pooling factor, defined at each level
- Analysis of variance
  - Summarize the scale of each batch of predictors
  - Go beyond classical null hypothesis testing framework
- Open question: how to construct models with deep interaction structures?

Understanding sources of variation

- Generalization of R^2 (explained variance), defined at each level of the model
- Partial pooling factor, defined at each level
- Analysis of variance
  - Summarize the scale of each batch of predictors
  - Go beyond classical null-hypothesis-testing framework
- Open question: how to construct models with deep interaction structures?

Understanding sources of variation

- Generalization of R^2 (explained variance), defined at each level of the model
- Partial pooling factor, defined at each level
- Analysis of variance
  - Summarize the scale of each batch of predictors
  - Go beyond classical null-hypothesis-testing framework
- Open question: how to construct models with deep interaction structures?

Understanding sources of variation

- Generalization of R^2 (explained variance), defined at each level of the model
- Partial pooling factor, defined at each level
- Analysis of variance
  - Summarize the scale of each batch of predictors
  - Go beyond classical null-hypothesis-testing framework
- Open question: how to construct models with deep interaction structures?

Understanding sources of variation

- Generalization of R^2 (explained variance), defined at each level of the model
- Partial pooling factor, defined at each level
- Analysis of variance
  - Summarize the scale of each batch of predictors
  - Go beyond classical null-hypothesis-testing framework
- Open question: how to construct models with deep interaction structures?

Understanding sources of variation

- Generalization of R^2 (explained variance), defined at each level of the model
- Partial pooling factor, defined at each level
- Analysis of variance
  - Summarize the scale of each batch of predictors
  - Go beyond classical null-hypothesis-testing framework
- Open question: how to construct models with deep interaction structures?

Conclusions

- Multilevel modeling is not just for grouped data
- New ideas needed to fit, understand, display, and summarize each level of the model

- General framework for modeling treatment effects that vary
- It's not just about "data fitting" or "getting the right standard errors"

Conclusions

- Multilevel modeling is not just for grouped data
- New ideas needed to fit, understand, display, and summarize each level of the model

- General framework for modeling treatment effects that vary
- It's not just about "data fitting" or "getting the right standard errors"

Conclusions

- Multilevel modeling is not just for grouped data
- New ideas needed to fit, understand, display, and summarize each level of the model

- General framework for modeling treatment effects that vary
- It's not just about "data fitting" or "getting the right standard errors"

Conclusions

- Multilevel modeling is not just for grouped data
- New ideas needed to fit, understand, display, and summarize each level of the model

- General framework for modeling treatment effects that vary
- It's not just about "data fitting" or "getting the right standard errors"

Conclusions

- Multilevel modeling is not just for grouped data
- New ideas needed to fit, understand, display, and summarize each level of the model

- General framework for modeling treatment effects that vary
- It's not just about "data fitting" or "getting the right standard errors"

Multilevel modeling

- Why?
  - Make use of lots of information out there that's already collected
  - Use MLM to adjust for time effects, state effects, survey-organization effects
  - Don't freak out because you have "too many predictors"
- How?

Multilevel modeling

- Why?
  - Make use of lots of information out there that's already collected
  - Use MLM to adjust for time effects, state effects, survey-organization effects
  - Don't freak out because you have "too many predictors"
- How?

Multilevel modeling

- Why?
  - Make use of lots of information out there that's already collected
  - Use MLM to adjust for time effects, state effects, survey-organization effects
  - Don't freak out because you have "too many predictors"
- How?

Multilevel modeling

- Why?
  - Make use of lots of information out there that's already collected
  - Use MLM to adjust for time effects, state effects, survey-organization effects
  - Don't freak out because you have "too many predictors"
- How?
  - Make your studies broader and deeper
  - Make plots and compute summaries to understand the estimated model

Multilevel modeling

- Why?
  - Make use of lots of information out there that's already collected
  - Use MLM to adjust for time effects, state effects, survey-organization effects
  - Don't freak out because you have "too many predictors"
- How?
  - Make your studies broader and deeper
  - Make plots and compute summaries to understand the estimated model

Multilevel modeling

- Why?
  - Make use of lots of information out there that's already collected
  - Use MLM to adjust for time effects, state effects, survey-organization effects
  - Don't freak out because you have "too many predictors"
- How?
  - Make your studies broader and deeper
  - Make plots and compute summaries to understand the estimated model

Multilevel modeling

- Why?
  - Make use of lots of information out there that's already collected
  - Use MLM to adjust for time effects, state effects, survey-organization effects
  - Don't freak out because you have "too many predictors"
- How?
  - Make your studies broader and deeper
  - Make plots and compute summaries to understand the estimated model

Multilevel modeling

- Why?
  - Make use of lots of information out there that's already collected
  - Use MLM to adjust for time effects, state effects, survey-organization effects
  - Don't freak out because you have "too many predictors"
- How?
  - Make your studies broader and deeper
  - Make plots and compute summaries to understand the estimated model