An Introduction to Hierarchical Modeling in Quantitative Research

JEFF GILL
Distinguished Professor
Departments of Government and Mathematics & Statistics
American University

Advantages of Multilevel Models

- Removes the restriction that the estimated coefficients are constant across individual cases by specifying levels of additional effects.

Advantages of Multilevel Models

- Removes the restriction that the estimated coefficients are constant across individual cases by specifying levels of additional effects.
- Provides a notationally efficient way to organize groups in the model.

Advantages of Multilevel Models

- Removes the restriction that the estimated coefficients are constant across individual cases by specifying levels of additional effects.
- Provides a notationally efficient way to organize groups in the model.
- Accounting for individual versus group-level variation.

Advantages of Multilevel Models

- Removes the restriction that the estimated coefficients are constant across individual cases by specifying levels of additional effects.
- Provides a notationally efficient way to organize groups in the model.
- Accounting for individual versus group-level variation.
- Modeling variation among individual-level regression coefficients.

Advantages of Multilevel Models

- Removes the restriction that the estimated coefficients are constant across individual cases by specifying levels of additional effects.
- Provides a notationally efficient way to organize groups in the model.
- Accounting for individual versus group-level variation.
- Modeling variation among individual-level regression coefficients.
- Estimating regression coefficients for groups of interest.

Advantages of Multilevel Models

- Removes the restriction that the estimated coefficients are constant across individual cases by specifying levels of additional effects.
- Provides a notationally efficient way to organize groups in the model.
- Accounting for individual versus group-level variation.
- Modeling variation among individual-level regression coefficients.
- Estimating regression coefficients for groups of interest.
- Gets the standard errors right.

Features of Multilevel Models

- Each level of the model is its *own* regression, with its own assumptions about: functional form, linearity, independence, variance, distribution of errors, etc.

Features of Multilevel Models

- Each level of the model is its *own* regression, with its own assumptions about: functional form, linearity, independence, variance, distribution of errors, etc.
- Models are usually "mixed," meaning some coefficients are *modeled* and some are *unmodeld*.

Features of Multilevel Models

- Each level of the model is its *own* regression, with its own assumptions about: functional form, linearity, independence, variance, distribution of errors, etc.
- Models are usually "mixed," meaning some coefficients are *modeled* and some are *unmodeld*.
- Multilevel models are highly symbiotic with Bayesian specifications because the focus in both cases is on making reasonable distributional assumptions.

Features of Multilevel Models

- Each level of the model is its *own* regression, with its own assumptions about: functional form, linearity, independence, variance, distribution of errors, etc.
- Models are usually "mixed," meaning some coefficients are *modeled* and some are *unmodeld*.
- Multilevel models are highly symbiotic with Bayesian specifications because the focus in both cases is on making reasonable distributional assumptions.
- These approaches are generally more demanding of statistical estimation process (software) to produce results.

Modern Notation

- Units (smallest level of analysis): $i = 1, \ldots, n$.

Modern Notation

- Units (smallest level of analysis): $i = 1, \ldots, n$.
- Outcomes: $\mathbf{y} = y_1, \ldots, y_n$.

Modern Notation

- Units (smallest level of analysis): $i = 1, \ldots, n$.
- Outcomes: $\mathbf{y} = y_1, \ldots, y_n$.
- Explanatory variables (predictors) at the first level are collected in an $n \times k$ non-singular matrix $\mathbf{X}$, with a leading column of 1s.

Modern Notation

- Units (smallest level of analysis): $i = 1, \ldots, n$.
- Outcomes: $\mathbf{y} = y_1, \ldots, y_n$.
- Explanatory variables (predictors) at the first level are collected in an $n \times k$ non-singular matrix $\mathbf{X}$, with a leading column of 1s.
- Second-level explanatory variables are collected in an $n \times J$ non-singular matrix $\mathbf{Z}$.

Modern Notation

- Units (smallest level of analysis): $i = 1, \ldots, n$.
- Outcomes: $\mathbf{y} = y_1, \ldots, y_n$.
- Explanatory variables (predictors) at the first level are collected in an $n \times k$ non-singular matrix $\mathbf{X}$, with a leading column of 1s.
- Second-level explanatory variables are collected in an $n \times J$ non-singular matrix $\mathbf{Z}$.
- Estimation produces $k \times 1$ vector $\boldsymbol{\beta}$ and $k \times 1$ vector $\hat{\mathbf{y}}$, and higher level regression vectors $\boldsymbol{\gamma}$.

Modern Notation

- Units (smallest level of analysis): $i = 1, \ldots, n$.
- Outcomes: $\mathbf{y} = y_1, \ldots, y_n$.
- Explanatory variables (predictors) at the first level are collected in an $n \times k$ non-singular matrix $\mathbf{X}$, with a leading column of 1s.
- Second-level explanatory variables are collected in an $n \times J$ non-singular matrix $\mathbf{Z}$.
- Estimation produces $k \times 1$ vector $\boldsymbol{\beta}$ and $k \times 1$ vector $\hat{\mathbf{y}}$, and higher level regression vectors $\boldsymbol{\gamma}$.
- Individual-level result: $\mathbf{y}_i = g^{-1}(\mathbf{X}_i\boldsymbol{\beta})$.

Modern Notation

- Units (smallest level of analysis): $i = 1, \ldots, n$.
- Outcomes: $\mathbf{y} = y_1, \ldots, y_n$.
- Explanatory variables (predictors) at the first level are collected in an $n \times k$ non-singular matrix $\mathbf{X}$, with a leading column of 1s.
- Second-level explanatory variables are collected in an $n \times J$ non-singular matrix $\mathbf{Z}$.
- Estimation produces $k \times 1$ vector $\boldsymbol{\beta}$ and $k \times 1$ vector $\hat{\mathbf{y}}$, and higher level regression vectors $\boldsymbol{\gamma}$.
- Individual-level result: $\mathbf{y}_i = g^{-1}(\mathbf{X}_i\boldsymbol{\beta})$.
- For an individual predictor, the column vector of $\mathbf{X}$ is denoted $\mathbf{X}_{(\kappa)}$.

Modern Notation

- Units (smallest level of analysis): $i = 1, \ldots, n$.
- Outcomes: $\mathbf{y} = y_1, \ldots, y_n$.
- Explanatory variables (predictors) at the first level are collected in an $n \times k$ non-singular matrix $\mathbf{X}$, with a leading column of 1s.
- Second-level explanatory variables are collected in an $n \times J$ non-singular matrix $\mathbf{Z}$.
- Estimation produces $k \times 1$ vector $\boldsymbol{\beta}$ and $k \times 1$ vector $\hat{\mathbf{y}}$, and higher level regression vectors $\boldsymbol{\gamma}$.
- Individual-level result: $\mathbf{y}_i = g^{-1}(\mathbf{X}_i\boldsymbol{\beta})$.
- For an individual predictor, the column vector of $\mathbf{X}$ is denoted $\mathbf{X}_{(\kappa)}$.
- Groups are denoted $j = 1, \ldots J$ at the first level and $k = 1, \ldots K$ at the second level.

Modern Notation

- Units (smallest level of analysis): $i = 1, \ldots, n$.
- Outcomes: $\mathbf{y} = y_1, \ldots, y_n$.
- Explanatory variables (predictors) at the first level are collected in an $n \times k$ non-singular matrix $\mathbf{X}$, with a leading column of 1s.
- Second-level explanatory variables are collected in an $n \times J$ non-singular matrix $\mathbf{Z}$.
- Estimation produces $k \times 1$ vector $\boldsymbol{\beta}$ and $k \times 1$ vector $\hat{\mathbf{y}}$, and higher level regression vectors $\boldsymbol{\gamma}$.
- Individual-level result: $\mathbf{y}_i = g^{-1}(\mathbf{X}_i\boldsymbol{\beta})$.
- For an individual predictor, the column vector of $\mathbf{X}$ is denoted $\mathbf{X}_{(\kappa)}$.
- Groups are denoted $j = 1, \ldots J$ at the first level and $k = 1, \ldots K$ at the second level.
- Group membership subscripted $j[i]$, so $j[35] = 4$ means the 35th case belongs to group 4.

Modern Notation

- Units (smallest level of analysis): $i = 1, \ldots, n$.
- Outcomes: $\mathbf{y} = y_1, \ldots, y_n$.
- Explanatory variables (predictors) at the first level are collected in an $n \times k$ non-singular matrix $\mathbf{X}$, with a leading column of 1s.
- Second-level explanatory variables are collected in an $n \times J$ non-singular matrix $\mathbf{Z}$.
- Estimation produces $k \times 1$ vector $\boldsymbol{\beta}$ and $k \times 1$ vector $\hat{\mathbf{y}}$, and higher level regression vectors $\boldsymbol{\gamma}$.
- Individual-level result: $\mathbf{y}_i = g^{-1}(\mathbf{X}_i\boldsymbol{\beta})$.
- For an individual predictor, the column vector of $\mathbf{X}$ is denoted $\mathbf{X}_{(\kappa)}$.
- Groups are denoted $j = 1, \ldots J$ at the first level and $k = 1, \ldots K$ at the second level.
- Group membership subscripted $j[i]$, so $j[35] = 4$ means the 35th case belongs to group 4.
- Use `a`, `b`, `g` in code to denote $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$.

Modern Notation

- Units (smallest level of analysis): $i = 1, \ldots, n$.
- Outcomes: $\mathbf{y} = y_1, \ldots, y_n$.
- Explanatory variables (predictors) at the first level are collected in an $n \times k$ non-singular matrix $\mathbf{X}$, with a leading column of 1s.
- Second-level explanatory variables are collected in an $n \times J$ non-singular matrix $\mathbf{Z}$.
- Estimation produces $k \times 1$ vector $\boldsymbol{\beta}$ and $k \times 1$ vector $\hat{\mathbf{y}}$, and higher level regression vectors $\boldsymbol{\gamma}$.
- Individual-level result: $\mathbf{y}_i = g^{-1}(\mathbf{X}_i\boldsymbol{\beta})$.
- For an individual predictor, the column vector of $\mathbf{X}$ is denoted $\mathbf{X}_{(\kappa)}$.
- Groups are denoted $j = 1, \ldots J$ at the first level and $k = 1, \ldots K$ at the second level.
- Group membership subscripted $j[i]$, so $j[35] = 4$ means the 35th case belongs to group 4.
- Use `a`, `b`, `g` in code to denote $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$.
- Standard errors are denoted σ_y, σ_α, σ_β,....

Linear Model Illustration

- Start with a standard linear model specification indexed by subjects and a first level of grouping, the *context* level.

Linear Model Illustration

- Start with a standard linear model specification indexed by subjects and a first level of grouping, the *context* level.
- Now use a single explanatory variable that has the form:

$$y_i = \beta_{j0[i]} + \beta_{j1[i]} X_i + \epsilon_i.$$

Linear Model Illustration

- Start with a standard linear model specification indexed by subjects and a first level of grouping, the *context* level.
- Now use a single explanatory variable that has the form:

$$y_i = \beta_{j0[i]} + \beta_{j1[i]} X_i + \epsilon_i.$$

- Suppose we have group-level explanatory variables, $Z_{j.}$, in that their effect is measured at the aggregated rather than at the individual level.

Linear Model Illustration

- Start with a standard linear model specification indexed by subjects and a first level of grouping, the *context* level.
- Now use a single explanatory variable that has the form:

$$y_i = \beta_{j0[i]} + \beta_{j1[i]} X_i + \epsilon_i.$$

- Suppose we have group-level explanatory variables, $Z_{j.}$, in that their effect is measured at the aggregated rather than at the individual level.
- Now add a second level to the model that explicitly nests effects within groups and index these groups $j = 1$ to J:

$$\beta_{j0[i]} = \gamma_{00} + \gamma_{10} Z_{j0} + u_{j0}$$
$$\beta_{j1[i]} = \gamma_{01} + \gamma_{11} Z_{j1} + u_{j1},$$

Linear Model Illustration

- The two-level model is produced by inserting the context level specifications into the original linear expression for the outcome variable of interest:

$$y_i = \gamma_{00} + \gamma_{01} X_i + \gamma_{10} Z_{j0} + \gamma_{11} X_i Z_{j1} + u_{j1} X_i + u_{j0} + \epsilon_i.$$

Linear Model Illustration

- The two-level model is produced by inserting the context level specifications into the original linear expression for the outcome variable of interest:

$$y_i = \gamma_{00} + \gamma_{01} X_i + \gamma_{10} Z_{j0} + \gamma_{11} X_i Z_{j1} + u_{j1} X_i + u_{j0} + \epsilon_i.$$

- This equation shows that the composite error structure, $u_{j1} X_i + u_{j0} + \epsilon_i$, is now clearly heteroscedastic since it is conditioned on levels of the explanatory variable, causing additional estimation complexity.

Linear Model Illustration

- The two-level model is produced by inserting the context level specifications into the original linear expression for the outcome variable of interest:

$$y_i = \gamma_{00} + \gamma_{01} X_i + \gamma_{10} Z_{j0} + \gamma_{11} X_i Z_{j1} + u_{j1} X_i + u_{j0} + \epsilon_i.$$

- This equation shows that the composite error structure, $u_{j1} X_i + u_{j0} + \epsilon_i$, is now clearly heteroscedastic since it is conditioned on levels of the explanatory variable, causing additional estimation complexity.

- Notice that there is an "automatic" interaction component: $\gamma_{11} X_i Z_{j1}$.

Linear Model Illustration

- The two-level model is produced by inserting the context level specifications into the original linear expression for the outcome variable of interest:

$$y_i = \gamma_{00} + \gamma_{01}X_i + \gamma_{10}Z_{j0} + \gamma_{11}X_iZ_{j1} + u_{j1}X_i + u_{j0} + \epsilon_i.$$

- This equation shows that the composite error structure, $u_{j1}X_i + u_{j0} + \epsilon_i$, is now clearly heteroscedastic since it is conditioned on levels of the explanatory variable, causing additional estimation complexity.
- Notice that there is an "automatic" interaction component: $\gamma_{11}X_iZ_{j1}$.
- Now we are going model *distributions* for y, β_{j0}, and β_{j1}.

Linear Model Illustration

- The two-level model is produced by inserting the context level specifications into the original linear expression for the outcome variable of interest:

$$y_i = \gamma_{00} + \gamma_{01}X_i + \gamma_{10}Z_{j0} + \gamma_{11}X_iZ_{j1} + u_{j1}X_i + u_{j0} + \epsilon_i.$$

- This equation shows that the composite error structure, $u_{j1}X_i + u_{j0} + \epsilon_i$, is now clearly heteroscedastic since it is conditioned on levels of the explanatory variable, causing additional estimation complexity.
- Notice that there is an "automatic" interaction component: $\gamma_{11}X_iZ_{j1}$.
- Now we are going model *distributions* for y, β_{j0}, and β_{j1}.
- This means we will make distributional regression statements:

$$\beta_{j0} \sim f(\gamma_{00} + \gamma_{10}Z_{j0}, \sigma_{\beta_0}).$$

Vocabulary Fun

- The varying coefficients are sometimes called random effects, since they are associated with distributional statements like $\alpha_j = N(\mathbf{Z}_j\gamma, \sigma^2_\alpha)$.

Vocabulary Fun

- The varying coefficients are sometimes called random effects, since they are associated with distributional statements like $\alpha_j = N(\mathbf{Z}_j\gamma, \sigma^2_\alpha)$.
- The term fixed effects is more nebulous with different meanings from different authors:
  - ▷ coefficients that are constant across individuals (most common definition)
  - ▷ factor contrasts
  - ▷ nuisance coefficients that are uninteresting but included
  - ▷ coefficients in population models
  - ▷ realized random variables
  - ▷ MLE values assuming infinite group-level variance

(see the discussion in G&H, p.245).

Vocabulary Fun

- The varying coefficients are sometimes called random effects, since they are associated with distributional statements like $\alpha_j = N(\mathbf{Z}_j\gamma, \sigma^2_\alpha)$.
- The term fixed effects is more nebulous with different meanings from different authors:
  - ▷ coefficients that are constant across individuals (most common definition)
  - ▷ factor contrasts
  - ▷ nuisance coefficients that are uninteresting but included
  - ▷ coefficients in population models
  - ▷ realized random variables
  - ▷ MLE values assuming infinite group-level variance

  (see the discussion in G&H, p.245).
- Sometimes these models labeled as mixed effects models.

Vocabulary Fun

- The varying coefficients are sometimes called random effects, since they are associated with distributional statements like $\alpha_j = N(\mathbf{Z}_j\gamma, \sigma^2_\alpha)$.
- The term fixed effects is more nebulous with different meanings from different authors:
  - ▷ coefficients that are constant across individuals (most common definition)
  - ▷ factor contrasts
  - ▷ nuisance coefficients that are uninteresting but included
  - ▷ coefficients in population models
  - ▷ realized random variables
  - ▷ MLE values assuming infinite group-level variance

  (see the discussion in G&H, p.245).
- Sometimes these models labeled as mixed effects models.
- Prescription: use multilevel models or hierarchical models with appropriate descriptor or detailed specification.

Vocabulary Overview

- For the data matrices, X_i for individual i in cluster j, and $\mathbf{Z}_j$ for cluster j, there are five canonical models that we will look at:

"Completely Pooled"	$y_i = \beta_0 + \beta_1 \mathbf{X}_i + \gamma \mathbf{Z} + e_i$
"Fixed Effect"	$y_i = \beta_{j0[i]} + \beta_1 \mathbf{X}_i + e_i$
"Random Effect"	$y_i = \beta_{j0[i]} + \beta_1 \mathbf{X}_i + \gamma \mathbf{Z}_j + e_i$
"Random Intercept and Random Slope"	$y_i = \beta_{j0[i]} + \beta_{j1[i]} \mathbf{X}_i + e_i$
"Completely Unpooled'	$y_{j[i]} = \beta_{j0} + (\beta_{j1} \mathbf{X}_{j[i]} + \gamma \mathbf{Z}_j) + e_{j[i]}$

Vocabulary Overview

- For the data matrices, X_i for individual i in cluster j, and $\mathbf{Z}_j$ for cluster j, there are five canonical models that we will look at:

"Completely Pooled"	$y_i = \beta_0 + \beta_1 \mathbf{X}_i + \gamma \mathbf{Z} + e_i$
"Fixed Effect"	$y_i = \beta_{j0[i]} + \beta_1 \mathbf{X}_i + e_i$
"Random Effect"	$y_i = \beta_{j0[i]} + \beta_1 \mathbf{X}_i + \gamma \mathbf{Z}_j + e_i$
"Random Intercept and Random Slope"	$y_i = \beta_{j0[i]} + \beta_{j1[i]} \mathbf{X}_i + e_i$
"Completely Unpooled'	$y_{j[i]} = \beta_{j0} + (\beta_{j1} \mathbf{X}_{j[i]} + \gamma \mathbf{Z}_j) + e_{j[i]}$

- This is produced by replacing the previous γ coefficient names with common regression-style language.

Vocabulary Overview

- For the data matrices, X_i for individual i in cluster j, and $\mathbf{Z}_j$ for cluster j, there are five canonical models that we will look at:

"Completely Pooled"	$y_i = \beta_0 + \beta_1 \mathbf{X}_i + \gamma \mathbf{Z} + e_i$
"Fixed Effect"	$y_i = \beta_{j0[i]} + \beta_1 \mathbf{X}_i + e_i$
"Random Effect"	$y_i = \beta_{j0[i]} + \beta_1 \mathbf{X}_i + \gamma \mathbf{Z}_j + e_i$
"Random Intercept and Random Slope"	$y_i = \beta_{j0[i]} + \beta_{j1[i]} \mathbf{X}_i + e_i$
"Completely Unpooled'	$y_{j[i]} = \beta_{j0} + (\beta_{j1} \mathbf{X}_{j[i]} + \gamma \mathbf{Z}_j) + e_{j[i]}$

- This is produced by replacing the previous γ coefficient names with common regression-style language.

- "Fixed" and "random" can differ in definition by literature (Kreft and De Leeuw 1988, Section 1.3.3, Gelman 2005), and better notation is "random intercepts" for "fixed effect," and "varying-intercept, varying-slope" for "random intercept and random slope."

Vocabulary Overview

- For the data matrices, X_i for individual i in cluster j, and $\mathbf{Z}_j$ for cluster j, there are five canonical models that we will look at:

"Completely Pooled"	$y_i = \beta_0 + \beta_1 \mathbf{X}_i + \gamma \mathbf{Z} + e_i$
"Fixed Effect"	$y_i = \beta_{j0[i]} + \beta_1 \mathbf{X}_i + e_i$
"Random Effect"	$y_i = \beta_{j0[i]} + \beta_1 \mathbf{X}_i + \gamma \mathbf{Z}_j + e_i$
"Random Intercept and Random Slope"	$y_i = \beta_{j0[i]} + \beta_{j1[i]} \mathbf{X}_i + e_i$
"Completely Unpooled'	$y_{j[i]} = \beta_{j0} + (\beta_{j1} \mathbf{X}_{j[i]} + \gamma \mathbf{Z}_j) + e_{j[i]}$

- This is produced by replacing the previous γ coefficient names with common regression-style language.
- "Fixed" and "random" can differ in definition by literature (Kreft and De Leeuw 1988, Section 1.3.3, Gelman 2005), and better notation is "random intercepts" for "fixed effect," and "varying-intercept, varying-slope" for "random intercept and random slope."
- Best to conceptualize these specifications as members of a larger multilevel family where indices are *turned-on* or *turned-off* systematically depending on the hierarchical purpose.

Comparison with Variable Contrasts

- The most commonly used scheme for dealing with purely categorical explanatory variables is called a **treatment contrast** where one category is selected as a baseline and for k categories of the explanatory variable there are $k - 1$ estimated difference parameters.

Comparison with Variable Contrasts

- The most commonly used scheme for dealing with purely categorical explanatory variables is called a **treatment contrast** where one category is selected as a baseline and for k categories of the explanatory variable there are $k - 1$ estimated difference parameters.
- Such schema are necessary because including all categorical information creates an uninvertible $\mathbf{X}'\mathbf{X}$ matrix (perfect collinearity).

Comparison with Variable Contrasts

- The most commonly used scheme for dealing with purely categorical explanatory variables is called a **treatment contrast** where one category is selected as a baseline and for k categories of the explanatory variable there are $k-1$ estimated difference parameters.
- Such schema are necessary because including all categorical information creates an uninvertible $\mathbf{X}'\mathbf{X}$ matrix (perfect collinearity).
- This can be awkward for large k or where there is not a logical baseline category.

Comparison with Variable Contrasts

- The most commonly used scheme for dealing with purely categorical explanatory variables is called a **treatment contrast** where one category is selected as a baseline and for k categories of the explanatory variable there are $k-1$ estimated difference parameters.
- Such schema are necessary because including all categorical information creates an uninvertible $\mathbf{X}'\mathbf{X}$ matrix (perfect collinearity).
- This can be awkward for large k or where there is not a logical baseline category.
- Multilevel models allow inclusion of all categorical values through specification in a hierarchy: they become part of the model specification rather than just additional $\mathbf{X}$ columns.

Back to Pooling

- Complete-Pooling: excluding categorical predictors completely (no hierarchy).

Back to Pooling

- Complete-Pooling: excluding categorical predictors completely (no hierarchy).
- This ignores (possibly important) variation between categories.

Back to Pooling

- Complete-Pooling: excluding categorical predictors completely (no hierarchy).
- This ignores (possibly important) variation between categories.
- No-Pooling: estimating separate models for each level of the categorical predictors.

Back to Pooling

- Complete-Pooling: excluding categorical predictors completely (no hierarchy).
- This ignores (possibly important) variation between categories.
- No-Pooling: estimating separate models for each level of the categorical predictors.
- This overstates variation between categories, making them look more different than they really are (unless the categories are not meaningful).

Back to Pooling

- Complete-Pooling: excluding categorical predictors completely (no hierarchy).
- This ignores (possibly important) variation between categories.
- No-Pooling: estimating separate models for each level of the categorical predictors.
- This overstates variation between categories, making them look more different than they really are (unless the categories are not meaningful).
- Multilevel Models: a compromise between these two extremes that captures within category uniqueness and between category similarities.

Back to Pooling

- Complete-Pooling: excluding categorical predictors completely (no hierarchy).
- This ignores (possibly important) variation between categories.
- No-Pooling: estimating separate models for each level of the categorical predictors.
- This overstates variation between categories, making them look more different than they really are (unless the categories are not meaningful).
- Multilevel Models: a compromise between these two extremes that captures within category uniqueness and between category similarities.
- Running example from Gelman & Hill: Radon gas by county ($J = 85$) in Minnesota.

Presenting Results from Multilevel Models

- Often there are too many parameters to present in journal articles with realistically large models.

Word	Sequences 233			3277		
	Recovered	Missed	Imputed	Recovered	Missed	Imputed
araC	6	0	6	6	0	9
arcA	8	5	28	6	7	60
argR	15	2	24	15	2	108
cpxR	11	1	29	7	5	99
creB	8	0	9	8	0	19
crp	36	13	131	34	15	610
cspA	4	0	4	3	1	12
cytR	2	3	7	1	4	55
dnaA	7	1	41	6	2	96
fadR	7	0	8	6	1	21
fis	8	7	36	8	7	200
fliA	12	0	14	12	0	25
fnr	12	0	14	11	1	43
fruR	12	0	18	11	1	43
fur	8	1	18	8	1	69
galR	7	0	10	5	2	10
gcvA	4	0	4	4	0	6
glpR	7	6	20	6	7	71
hipB	2	2	2	0	4	2
lexA	19	0	24	19	0	46
malT	4	6	6	0	10	0
metJ	6	3	8	5	4	13
metR	5	3	10	4	4	44
nagC	6	0	9	6	0	22
narL	7	3	9	4	6	18
narP	8	0	4	8	0	7
ntrC	4	1	4	4	1	6
ompR	5	4	28	4	1	6
oxyR	4	0	4	4	0	4
phoB	10	2	12	9	3	35
purR	21	1	25	17	5	47
rpoH2	6	1	6	6	1	9
rpoH3	8	0	8	8	0	13
rpoN	6	1	11	6	1	22
rpoS17	5	10	9	1	14	4
rpoS18	4	3	8	3	4	5
soxS	11	6	22	9	8	61
torR	3	1	5	3	1	14
trpR	4	0	4	4	0	6
tus	5	0	5	5	0	5
tyrR	13	4	19	10	7	54
Total	340	90	663	296	134	2231

A binding site is identified in positions where its posterior probability is greater than 0.5.

Presenting Results from Multilevel Models

- Often there are too many parameters to present in journal articles with realistically large models.
- In some literatures, such as statistical genomics, there are thousands.

Word	Sequences 233			3277		
	Recovered	Missed	Imputed	Recovered	Missed	Imputed
araC	6	0	6	6	0	9
arcA	8	5	28	6	7	60
argR	15	2	24	15	2	108
cpxR	11	1	29	7	5	99
creB	8	0	9	8	0	19
crp	36	13	131	34	15	610
cspA	4	0	4	3	1	12
cytR	2	3	7	1	4	55
dnaA	7	1	41	6	2	96
fadR	7	0	8	6	1	21
fis	8	7	36	8	7	200
fliA	12	0	14	12	0	25
fnr	12	0	14	11	1	43
fruR	12	0	18	11	1	43
fur	8	1	18	8	1	69
galR	7	0	10	5	2	10
gcvA	4	0	4	4	0	6
glpR	7	6	20	6	7	71
hipB	2	2	2	0	4	2
lexA	19	0	24	19	0	46
malT	4	6	6	0	10	0
metJ	6	3	8	5	4	13
metR	5	3	10	4	4	44
nagC	6	0	9	6	0	22
narL	7	3	9	4	6	18
narP	8	0	4	8	0	7
ntrC	4	1	4	4	1	6
ompR	5	4	28	4	1	6
oxyR	4	0	4	4	0	4
phoB	10	2	12	9	3	35
purR	21	1	25	17	5	47
rpoH2	6	1	6	6	1	9
rpoH3	8	0	8	8	0	13
rpoN	6	1	11	6	1	22
rpoS17	5	10	9	1	14	4
rpoS18	4	3	8	3	4	5
soxS	11	6	22	9	8	61
torR	3	1	5	3	1	14
trpR	4	0	4	4	0	6
tus	5	0	5	5	0	5
tyrR	13	4	19	10	7	54
Total	340	90	663	296	134	2231

A binding site is identified in positions where its posterior probability is greater than 0.5.

Presenting Results from Multilevel Models

- Often there are too many parameters to present in journal articles with realistically large models.
- In some literatures, such as statistical genomics, there are thousands.
- Strategies:
  - ▷ give only group level summaries,
  - ▷ plot group level effects,
  - ▷ sample cases from the total,
  - ▷ identify critical cases,
  - ▷ graph summaries of individual level effects.

Word	Sequences 233			3277		
	Recovered	Missed	Imputed	Recovered	Missed	Imputed
araC	6	0	6	6	0	9
arcA	8	5	28	6	7	60
argR	15	2	24	15	2	108
cpxR	11	1	29	7	5	99
creB	8	0	9	8	0	19
crp	36	13	131	34	15	610
cspA	4	0	4	3	1	12
cytR	2	3	7	1	4	55
dnaA	7	1	41	6	2	96
fadR	7	0	8	6	1	21
fis	8	7	36	8	7	200
fliA	12	0	14	12	0	25
fnr	12	0	14	11	1	43
fruR	12	0	18	11	1	43
fur	8	1	18	8	1	69
galR	7	0	10	5	2	10
gcvA	4	0	4	4	0	6
glpR	7	6	20	6	7	71
hipB	2	2	2	0	4	2
lexA	19	0	24	19	0	46
malT	4	6	6	0	10	0
metJ	6	3	8	5	4	13
metR	5	3	10	4	4	44
nagC	6	0	9	6	0	22
narL	7	3	9	4	6	18
narP	8	0	4	8	0	7
ntrC	4	1	4	4	1	6
ompR	5	4	28	4	1	6
oxyR	4	0	4	4	0	4
phoB	10	2	12	9	3	35
purR	21	1	25	17	5	47
rpoH2	6	1	6	6	1	9
rpoH3	8	0	8	8	0	13
rpoN	6	1	11	6	1	22
rpoS17	5	10	9	1	14	4
rpoS18	4	3	8	3	4	5
soxS	11	6	22	9	8	61
torR	3	1	5	3	1	14
trpR	4	0	4	4	0	6
tus	5	0	5	5	0	5
tyrR	13	4	19	10	7	54
Total	340	90	663	296	134	2231

A binding site is identified in positions where its posterior probability is greater than 0.5.

Statistical "Significance" In This Course

- We will not be making the mistake of fixating on p-values and "stars."

Statistical "Significance" In This Course

- We will not be making the mistake of fixating on p-values and "stars."
- Melton, A.W. (1962). Editorial. *Journal of Experimental Psychology*, 64, 553-557.

Statistical "Significance" In This Course

- We will not be making the mistake of fixating on p-values and "stars."
- Melton, A.W. (1962). Editorial. *Journal of Experimental Psychology*, 64, 553-557.
- Generally statistical significance in the sciences is not very significant.

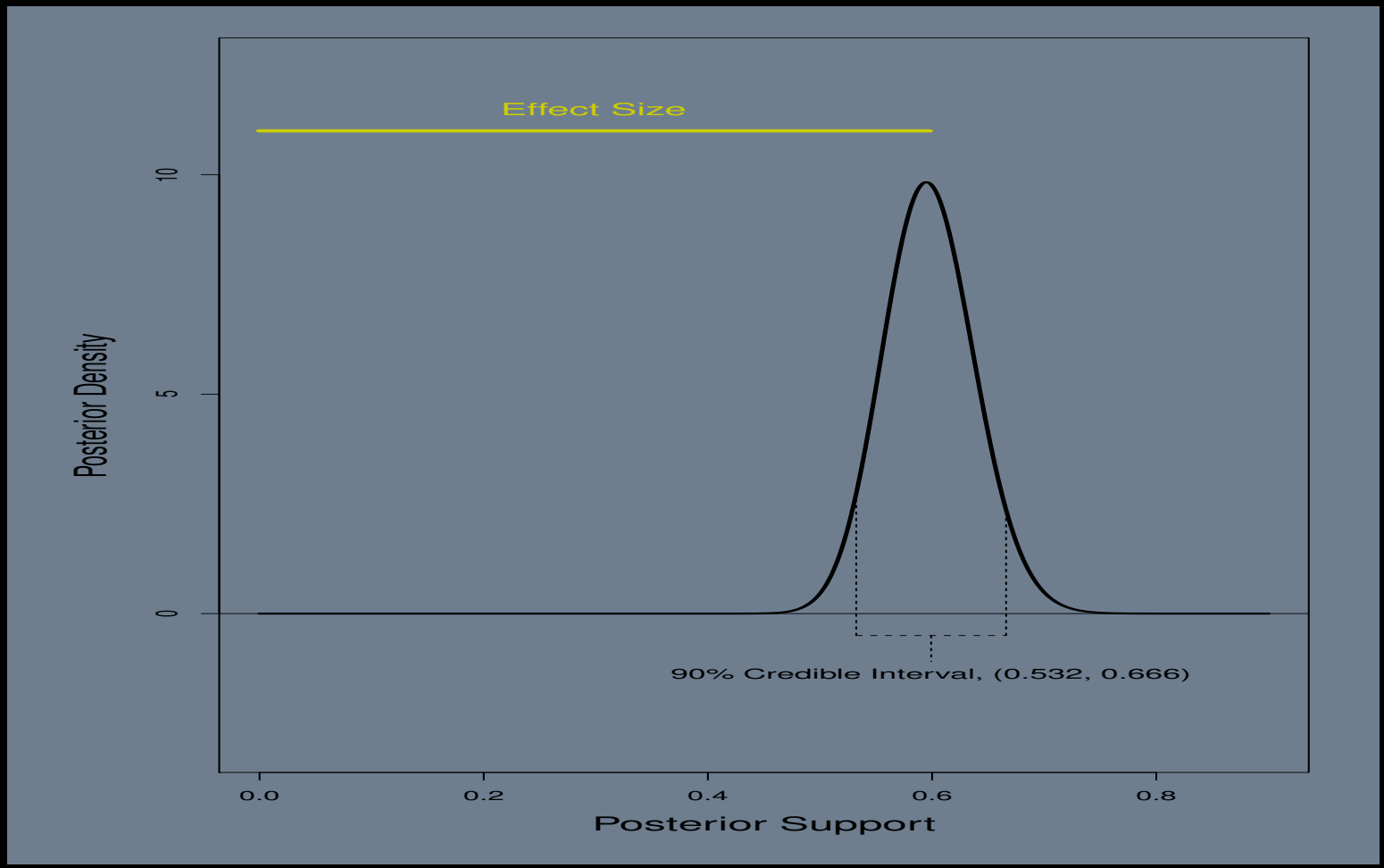
Statistical "Significance" In This Course

- We will not be making the mistake of fixating on p-values and "stars."
- Melton, A.W. (1962). Editorial. *Journal of Experimental Psychology*, 64, 553-557.
- Generally statistical significance in the sciences is not very significant.
- We will take a Bayesian approach with (mostly) vague priors and subsequently describe the resulting posterior distributions.

Statistical "Significance" In This Course

- We will not be making the mistake of fixating on p-values and "stars."
- Melton, A.W. (1962). Editorial. *Journal of Experimental Psychology*, 64, 553-557.
- Generally statistical significance in the sciences is not very significant.
- We will take a Bayesian approach with (mostly) vague priors and subsequently describe the resulting posterior distributions.
- We will therefore be thinking about more important concepts like:
  - effect size,
  - power,
  - statistical reliability,
  - posterior probability.

Simple Illustration of Bayesian Inference



Simple Illustration of Bayesian Inference

```
dur <- c(0.833, 1.070, 1.234, 1.671, 2.065, 2.080, 2.114, 2.168, 2.274, 2.629, 2.637)
N <- c(38, 28, 27, 20, 17, 15, 15, 15, 15, 14, 12)
L <- qgamma(0.05,shape=sum(N),rate=sum(N*dur))
H <- qgamma(0.95,shape=sum(N),rate=sum(N*dur))
ruler <- seq(0,0.90,length=1000)
postscript("Class.Multilevel/Images/models.figure01.ps")
par(mfrow=c(1,1),mar=c(6,6,2,2),cex.axis=1,cex.lab=1.5,bg="slategray")
plot(ruler,seq(0,12,length=length(ruler)),type="n", ylim=c(-3,12.5),
    xlab="Posterior Support",ylab="Posterior Density")
lines(ruler,dgamma(ruler,shape=sum(N),rate=sum(N*dur)),lwd=2.5); abline(h=0)
segments(L,0-.5,L,dgamma(L,shape=sum(N),rate=sum(N*dur)),lty=2)
segments(H,0-.5,H,dgamma(H,shape=sum(N),rate=sum(N*dur)),lty=2)
segments(L,0-.5,H,0-.5,lty=2)
segments((L+H)/2,0-.5,(L+H)/2,0-1.09,lty=2)
text((L+H)/2,-1.4,paste("90% Credible Interval, (",round(L,3),", ",
    round(H,3),")",sep=""),cex=1.1)
segments(0,11,sum(N)/sum(N*dur),11,col="yellow3",lwd=4)
text(sum(N)/sum(N*dur)/2,11.5,"Effect Size",col="yellow3",cex=1.3)
dev.off()
```

The pseudo-Frequentist NHST is wrong for 1-off analysis of observational data

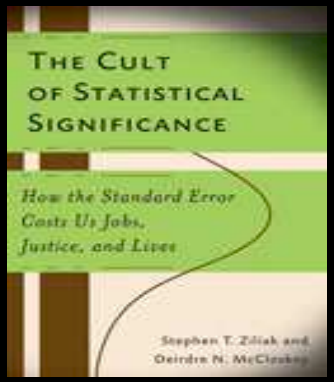
- A few authors have noted this (just a small sample): *Barnett 1973, Berger, Boukai, and Wang 1997, Berger Thomas Sellke 1987, Berkhardt and Schoenfeld 2003, Bernardo 1984, Brandstätter 1999, Carver 1978, 1993, Dar, Serlin and Omar 1994, Cohen 1988, 1994, 1992, 1977, 1962, Denis 2005, Falk and Greenbaum 1995, Gelman, Carlin, Stern, and Rubin 1995, Gigerenzer 1987, 1993, 1998, Gigerenzer and Murray 1987, Gill 1999, 2005, Gliner, Leech and Morgan 2002, Grayson 1998, Greenwald 1975, Greenwald, Gonzalez, Harris and Guthrie 1996, Hager 2000, Howson and Urbach 1993, Hunter 1997, Hunter and Schmidt 1990, Jeffreys 1961, Kirk 1996, Krueger 1999, 2001, Lindsay 1995, Loftus 1991, 1993a, 1993b, 1994, 1996, Loftus and Bamber 1990, Macdonald 1997, Meehl 1967, 1978, 1990, 1978, Nickerson 2000, Oakes 1986, Pollard 1993, Pollard and Richardson 1987, Robinson and Levin 1997, Rosnow and Rosenthal 1989, Rozeboom 1960, 1997, Schmidt 1996, Schmidt and Hunter 1977, Sedlmeier and Gigerenzer 1989, Thompson 2002, Wilkinson 1999.*

The pseudo-Frequentist NHST is wrong for 1-off analysis of observational data

- A few authors have noted this (just a small sample): *Barnett 1973, Berger, Boukai, and Wang 1997, Berger Thomas Sellke 1987, Berkhardt and Schoenfeld 2003, Bernardo 1984, Brandstätter 1999, Carver 1978, 1993, Dar, Serlin and Omar 1994, Cohen 1988, 1994, 1992, 1977, 1962, Denis 2005, Falk and Greenbaum 1995, Gelman, Carlin, Stern, and Rubin 1995, Gigerenzer 1987, 1993, 1998, Gigerenzer and Murray 1987, Gill 1999, 2005, Gliner, Leech and Morgan 2002, Grayson 1998, Greenwald 1975, Greenwald, Gonzalez, Harris and Guthrie 1996, Hager 2000, Howson and Urbach 1993, Hunter 1997, Hunter and Schmidt 1990, Jeffreys 1961, Kirk 1996, Krueger 1999, 2001, Lindsay 1995, Loftus 1991, 1993a, 1993b, 1994, 1996, Loftus and Bamber 1990, Macdonald 1997, Meehl 1967, 1978, 1990, 1978, Nickerson 2000, Oakes 1986, Pollard 1993, Pollard and Richardson 1987, Robinson and Levin 1997, Rosnow and Rosenthal 1989, Rozeboom 1960, 1997, Schmidt 1996, Schmidt and Hunter 1977, Sedlmeier and Gigerenzer 1989, Thompson 2002, Wilkinson 1999.*

- Why?

1. Artificial Model Selection Criteria
2. The Arbitrariness of Alpha
3. Replication Fallacy
4. Asymmetry and Accepting the Null Hypothesis
5. Probabilistic Modus Tollens
6. Inverse Probability Problem



Sample Analysis, Data

- 29 Incarcerated Women with Substance Use Disorder and Post-traumatic Stress Disorder in Providence, Rhode Island, 1999-2001.

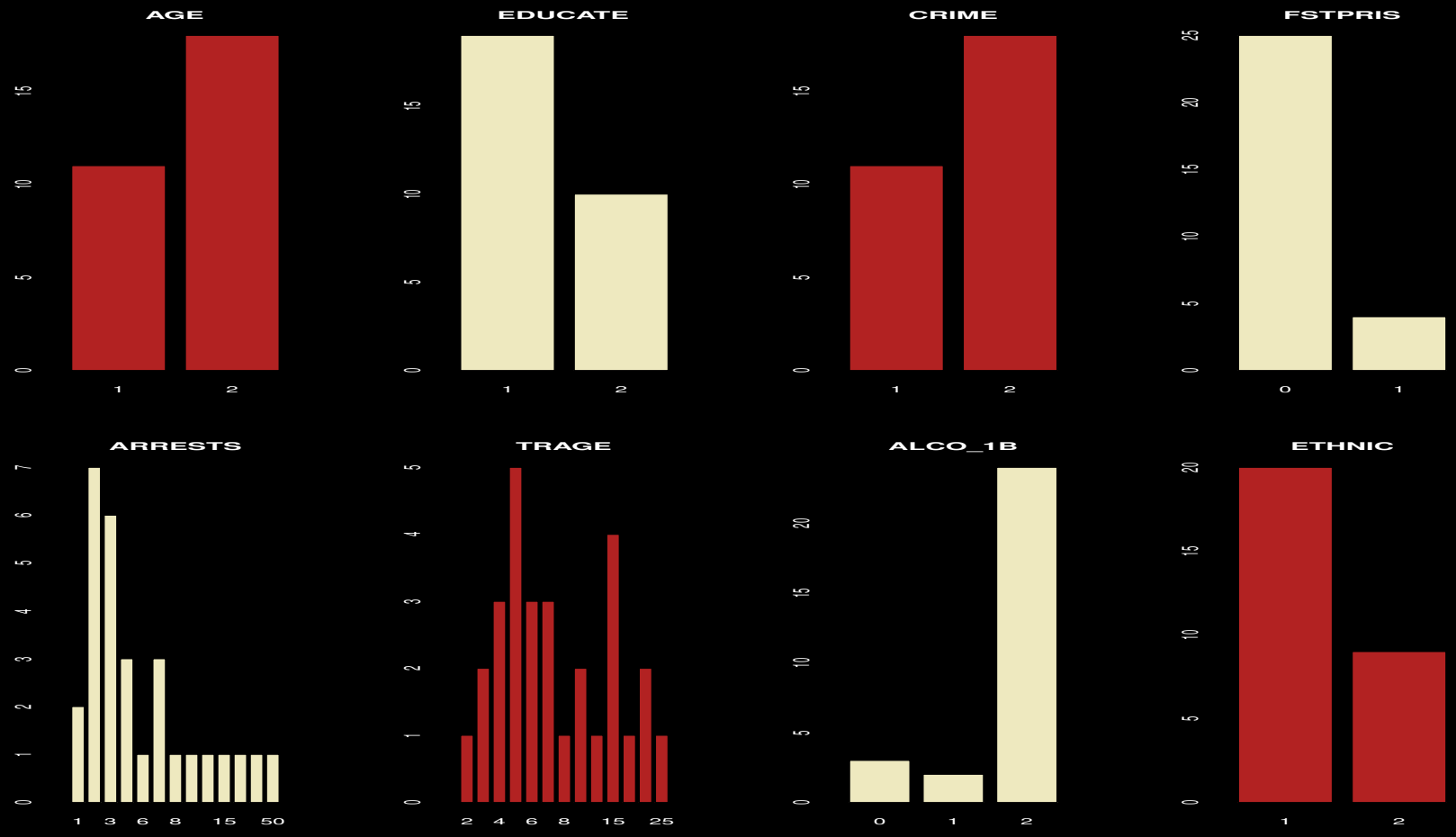
Sample Analysis, Data

- 29 Incarcerated Women with Substance Use Disorder and Post-traumatic Stress Disorder in Providence, Rhode Island, 1999-2001.
- Outcome variable: `PTSD diagnosis` (13 negative, 16 positive).

Sample Analysis, Data

- 29 Incarcerated Women with Substance Use Disorder and Post-traumatic Stress Disorder in Providence, Rhode Island, 1999-2001.
- Outcome variable: `PTSD diagnosis` (13 negative, 16 positive).
- Explanatory variables
  - `AGE`, 1 for 20-29, 2 for 30+
  - `ETHNIC`, 1 for white (nonhispanic), 2 for nonwhite
  - `EDUCATE`, 1 for no HS diploma, 2 for HS diploma
  - `FSTPRIS`, 0 for in prison before, 1 for first time
  - `CRIME`, 0 for misdomeanor, 1 for felony
  - `ARRESTS`, the number of arrests with convictions
  - `TRAGE`, age of first trauma (robbery/mugging, sexual abuse, physical abuse)
  - `ALCO1B`, alchohol issue: 0=None, 1=Abuse, 2=Dependent

Sample Analysis, Data



Sample Analysis, Model

$$
\begin{aligned}
p(y_i = 1) = \text{logit}^{-1} \big(& \beta^0 + \beta^{\text{AGE}} \cdot \texttt{AGE}_i + \beta^{\text{EDUCATE}} \cdot \texttt{EDUCATE}_i \\
& + \beta^{\text{AGE·EDUCATE}} \cdot (\texttt{AGE}_i \cdot \texttt{EDUCATE}_i) + \beta^{\text{CRIME}} \cdot \texttt{CRIME}_i + \beta^{\text{FSTPRIS}} \cdot \texttt{FSTPRIS}_i \\
& + \beta^{\text{ARRESTS}} \cdot \log(\texttt{ARRESTS}_i) + \beta^{\text{TRAGE}} \cdot \exp(\texttt{TRAGE}_i) + \beta^{\text{ALCO1B}} \cdot \texttt{ALCO1B}_i + \alpha_{j[i]}^{\text{ETHNIC}} \big)
\end{aligned}
$$

Sample Analysis, Model

$$p(y_i = 1) = \text{logit}^{-1}\Big(\beta^0 + \beta^{\text{AGE}} \cdot \texttt{AGE}_i + \beta^{\text{EDUCATE}} \cdot \texttt{EDUCATE}_i$$
$$+\beta^{\text{AGE·EDUCATE}} \cdot (\texttt{AGE}_i \cdot \texttt{EDUCATE}_i) + \beta^{\text{CRIME}} \cdot \texttt{CRIME}_i + \beta^{\text{FSTPRIS}} \cdot \texttt{FSTPRIS}_i$$
$$+\beta^{\text{ARRESTS}} \cdot \log(\texttt{ARRESTS}_i) + \beta^{\text{TRAGE}} \cdot \exp(\texttt{TRAGE}_i) + \beta^{\text{ALCO1B}} \cdot \texttt{ALCO1B}_i + \alpha^{\text{ETHNIC}}_{j[i]}\Big)$$

$$\alpha^{\text{ETHNIC}}_j \sim N\left(\alpha_0 + \alpha^{\text{AGE}}_{m[j]} \cdot \texttt{AGE} + \alpha^{\text{EDUCATE}}_{m[j]} \cdot \texttt{EDUCATE},\ \sigma^2_{\text{ETHNIC}}\right)$$

Sample Analysis, Model

$$p(y_i = 1) = \text{logit}^{-1}\Big(\beta^0 + \beta^{\text{AGE}} \cdot \texttt{AGE}_i + \beta^{\text{EDUCATE}} \cdot \texttt{EDUCATE}_i + \beta^{\text{AGE·EDUCATE}} \cdot (\texttt{AGE}_i \cdot \texttt{EDUCATE}_i) + \beta^{\text{CRIME}} \cdot \texttt{CRIME}_i + \beta^{\text{FSTPRIS}} \cdot \texttt{FSTPRIS}_i + \beta^{\text{ARRESTS}} \cdot \log(\texttt{ARRESTS}_i) + \beta^{\text{TRAGE}} \cdot \exp(\texttt{TRAGE}_i) + \beta^{\text{ALCO1B}} \cdot \texttt{ALCO1B}_i + \alpha^{\text{ETHNIC}}_{j[i]}\Big)$$

$$\alpha^{\text{ETHNIC}}_j \sim N\left(\alpha_0 + \alpha^{\text{AGE}}_{m[j]} \cdot \texttt{AGE} + \alpha^{\text{EDUCATE}}_{m[j]} \cdot \texttt{EDUCATE},\ \sigma^2_{\text{ETHNIC}}\right)$$

Fixed effects:

	Estimate	Std. Error	z value
(Intercept)	-5.02e+00	4.14e+00	-1.214
AGE	5.12e+00	2.78e+00	1.841
EDUCATE	7.37e+00	3.76e+00	1.959
CRIME	-2.00e+00	9.01e-01	-2.221
FSTPRIS	2.48e-01	1.11e+00	0.222
log(ARRESTS)	-1.19e+00	6.17e-01	-1.927
exp(TRAGE)	5.63e-08	3.48e-08	1.617
ALCO_1B	1.10e-01	4.18e-01	0.264
AGE:EDUCATE	-3.98e+00	1.99e+00	-2.003

Random effects:

Groups	Name	Variance	Std.Dev.
ETHNIC	(Intercept)	0.1839	0.429
	AGE	0.0643	0.253
	EDUCATE	0.0904	0.301
	Residual	0.7299	0.854

AIC	BIC	logLik	deviance
57.2	77.7	-13.6	27.2

Sample Analysis, Code

```
trauma.short.complete <- read.table("http://jgill.wustl.edu/data/trauma.short.dat",
                         header=TRUE)
library(nlme); library(arm)
trauma.out <- lmer(PTSD2 ~ CRIME + FSTPRIS + log(ARRESTS) + exp(TRAGE) + ALCO_1B
                + (1 + AGE + EDUCATE | ETHNIC),
                family=binomial(link="probit"),
                data=trauma.short.complete)
summary(trauma.out)
```

Specifications with the `lmer()` Function

- Start with an outcome variable `Y`, a continuous variable `X1`. and a categorical grouping variable `X2` Then model M1 is:

```
Y ~ X1 + (1|X2)
```

  This gives estimates for β_0 a global (constant) intercept, β_1 a slope estimate corresponding to `X1`, and a set of group-level intercepts that for the deviation from the global intercept.

- We can add another hierarchy definition for M2:

```
Y ~ X1 + (1|X2) + (0+X1|X2)
```

  This provides everything in M1 and also give the effect of `X1` within each level of `X2`, which is a set of group-level deviations from the slope.

- M2 assumed that there are no correlations between the two sets of deviations. To relax this specify M3:

```
Y ~ X1 + (1+X1|X2)
```

  which gives the correlation between the two sets of deviations.

Specifications with the `lmer()` Function

- Adding another continuous paramater `X3` according to M4:

```
Y ~ X1*X3 + (1+X1+X3|X2)
```

  gives:

  - a global intercept, β_0
  - a single global estimate for the effect of `X1`, β_1
  - a single global estimate for the effect of `X3`, β_2
  - a single global estimate for the interaction between `X1` and `X3`, β_3
  - deviations from the global intercept in each level of `X2`, γ_1
  - deviations of the slope effect from β_1 in each level of `X2`, γ_2
  - deviations of the slope effect from β_3 in each level of `X2`, γ_3

Specifications with the `lmer()` Function

- Continuing M4...
  - correlation between γ_1 and γ_2 across levels of `X2`
  - correlation between γ_1 and γ_3 across levels of `X2`
  - correlation between γ_1 and γ_4 across levels of `X2`
  - correlation between γ_2 and γ_3 across levels of `X2`
  - correlation between γ_2 and γ_4 across levels of `X2`
  - correlation between γ_3 and γ_4 across levels of `X2`

  plus standard errors for these coefficients.

- A model that did not give deviations from the global intercept but did give deviations from the slopes is specified by M5:

```
Y ~ X1 + (0+X1|X2)
```

The Bayesian Take On Hierarchical Models

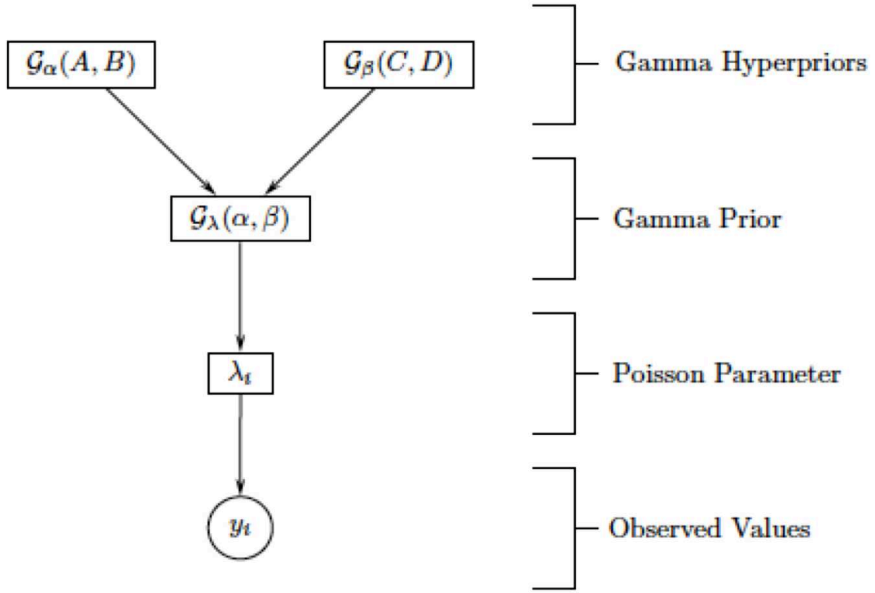


FIGURE 12.1: Poisson-Gamma Hierarchical Model

the example model, this representation is:

$$\begin{aligned} y_i &\sim \mathcal{P}(\lambda_i) \\ \lambda_i &\sim \mathcal{G}(\alpha,\beta) \\ \alpha &\sim \mathcal{G}(A,B) \\ \beta &\sim \mathcal{G}(C,D), \end{aligned} \tag{12.24}$$