In this set of notes:

- Example Data Sets
- Quick Introduction to logistic regression.
- Marginal Model: Population-Average Model
- Random Effects Model: Subject-specific Model
- 3-level multilevel logistic regression

Reading/References:

- Snijders & Bosker, Chapter 14
More References


Data

- Clustered, nested, hierarchical, longitudinal.
- The response/outcome variable is **dichotomous**.
- **Examples:**
  - Longitudinal study of patients in treatment for depression: normal or abnormal
  - Responses to items on an exam (correct/incorrect)
  - Admission decisions for graduate programs in different departments.
  - Longitudinal study of respiratory infection in children
  - Whether basketball players make free-throw shots.
  - Whether “cool” kids are tough kids.
  - others
Respiratory Infection Data

- Preschool children from Indonesia who were examined up to 6 consecutive quarters for respiratory infection.
- Predictors/explanatory/covariates:
  - Age in months
  - Xerophthalmia as indicator of chronic vitamin A deficiency (dummy variable)—night blindness & dryness of membranes → dryness of cornea → softening of cornea
  - Cosine of annual cycle (ie., season of year)
  - Sine of annual cycle (ie., season of year).
  - Gender
  - Height (as a percent)
  - Stunted
### Longitudinal Depression Example

- Comparison of new drug with a standard drug for treating depression.
- Classified as N= Normal and A= Abnormal at 1, 2 and 4 weeks.

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>$R_x$</th>
<th>Response at Each of 3 Time Points</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NNN</td>
</tr>
<tr>
<td>Mild</td>
<td>std</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>new</td>
<td>31</td>
</tr>
<tr>
<td>Severe</td>
<td>std</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>new</td>
<td>7</td>
</tr>
</tbody>
</table>
“Cool” Kids


- **Clustering**: Kids within peer groups within classrooms.
- **Response variable**: Whether a kid nominated by peers is classified as a model (ideal) student.
- **Predictors**: Nominator’s
  - Popularity
  - Gender
  - Race
  - Classroom aggression level
Law School Admissions data: 5 items, $N = 1000$

<table>
<thead>
<tr>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
<th>$Y_4$</th>
<th>$Y_5$</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>61</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>298</td>
</tr>
</tbody>
</table>
General Social Survey

- Data are responses to 10 vocabulary items from the 2004 General Social Survey from $n = 1155$ respondents.

```r
data vocab;
input age educ degree gender wordA wordB wordC wordD wordE wordF wordG wordH wordI wordJ none elementary hsplus;
datalines;
 52 14 1 1 1 1 1 1 1 0 1 1 0 1 0
 34 17 3 2 1 1 1 1 1 0 0 1 1 0 0 1
 26 14 2 1 1 1 0 1 1 1 0 1 0 0 0 1
 32 10 0 1 1 1 0 1 1 1 0 0 1 0 1 0
 29 11 1 1 1 0 1 1 1 0 0 1 0 0 1 0
```

- Possible predictors of vocabulary knowledge:
  - Age
  - Education
Logistic Regression

The logistic regression model is a generalized linear model with

- **Random component**: The response variable is **binary**. \( Y_i = 1 \) or 0 (an event occurs or it doesn’t). We are interested in probability that \( Y_i = 1 \); that is, \( P(Y_i = 1|x_i) = \pi(x_i) \).
- The distribution of \( Y_i \) is **Binomial**.

- **Systematic component**: A linear predictor such as

  \[
  \alpha + \beta_1 x_{1i} + \ldots + \beta_j x_{ki}
  \]

  The explanatory or predictor variables may be quantitative (continuous), qualitative (discrete), or both (mixed).

- **Link Function**: The log of the odds that an event occurs, otherwise known as the **logit**:

  \[
  \text{logit}(\pi_i(x_i)) = \log \left( \frac{\pi_i(x_i)}{1 - \pi_i(x_i)} \right)
  \]

  The logistic regression model is

  \[
  \text{logit}(\pi(x_i)) = \log \left( \frac{\pi(x_i)}{1 - \pi(x_i)} \right) = \alpha + \beta_1 x_{1i} + \ldots + \beta_j x_{ki}
  \]
The Binomial Distribution

Assume that the number of “trials” is fixed and we count the number of “successes” or events that occur.

Preliminaries: Bernoulli random variables

- $X$ is a random variable where $X = 1$ or $0$
- The probability that $X = 1$ is $\pi$
- The probability that $X = 0$ is $(1 - \pi)$

Such variables are called Bernoulli random variables.
Bernoulli Random Variable

The mean of a Bernoulli random variable is

$$\mu_X = E(X) = 1\pi + 0(1 - \pi) = \pi$$

The variance of $X$ is

$$\text{var}(X) = \sigma_X^2 = E[(X - \mu_X)^2]$$
$$= (1 - \pi)^2\pi + (0 - \pi)^2(1 - \pi)$$
$$= \pi(1 - \pi)$$
Bernoulli Variance vs Mean

Bernoulli Random Variable

Variance

Mean (possible values of P_i)
Example of Bernoulli Random Variable

Suppose that a coin is

- “not fair” or is “loaded”
- The probability that it lands on heads equals .40 and the probability that it lands on tails equals .60.
- If this coin is flipped many, many, many times, then we would expect that it would land on heads 40% of the time and tails 60% of the time.

We define our Bernoulli random variable as

\[ X = \begin{cases} 1 & \text{if Heads} \\ 0 & \text{if Tails} \end{cases} \]

where \( \pi = P(X = 1) = .40 \) and \( (1 - \pi) = P(X = 0) = .60 \).

**Note:** Once you know \( \pi \), you know the mean and variance of the distribution of \( X \).
Binomial Distribution

A binomial random variable is the sum of \( n \) independent Bernoulli random variables. We will let \( Y \) represent a binomial random variable and by definition

\[
Y = \sum_{i=1}^{n} X_i
\]

The mean of a Binomial random variable is

\[
\mu_y = E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = E(X_1) + E(X_2) + \ldots + E(X_n)
\]

\[
= \mu_x + \mu_x + \ldots + \mu_x
\]

\[
= n \pi
\]
Variance of Binomial Random Variable

...and the variance of a Binomial random variable is

\[ \text{var}(Y) = \sigma_y^2 = \text{var}(X_1 + X_2 + \ldots + X_n) \]

\[ = \left( \sum_{i=1}^{n} \text{var}(X_i) \right) \]

\[ = n \pi (1-\pi) + \pi (1-\pi) + \ldots + \pi (1-\pi) \]

\[ = n \pi (1-\pi) \]

**Note:** Once you know \( \pi \) and \( n \), you know the mean and variance of the Binomial distribution.
Variance vs Mean

Binomial Random Variable (n=20)
### Binomial Distribution Function by Example

- Toss the unfair coin with $\pi = .40$ coin $n = 3$ times.
- $Y =$ number of heads.
- The tosses are independent of each other.

<table>
<thead>
<tr>
<th>Possible Outcomes $X_1 + X_2 + X_3 = Y$</th>
<th>Probability of a Sequence $P(X_1, X_2, X_3)$</th>
<th>$P(Y)$ $P(Y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 + 1 + 1 = 3</td>
<td>$(.4)(.4)(.4) = (.4)^3(.6)^0 = .064$</td>
<td>.064</td>
</tr>
<tr>
<td>1 + 1 + 0 = 2</td>
<td>$(.4)(.4)(.6) = (.4)^2(.6)^1 = .096$</td>
<td>3(.096) = .288</td>
</tr>
<tr>
<td>1 + 0 + 1 = 2</td>
<td>$(.4)(.6)(.4) = (.4)^2(.6)^1 = .096$</td>
<td>3(.096) = .288</td>
</tr>
<tr>
<td>0 + 1 + 1 = 2</td>
<td>$(.6)(.4)(.4) = (.4)^2(.6)^1 = .096$</td>
<td>3(.096) = .288</td>
</tr>
<tr>
<td>1 + 0 + 0 = 1</td>
<td>$(.4)(.6)(.6) = (.4)^1(.6)^2 = .144$</td>
<td>3(.144) = .432</td>
</tr>
<tr>
<td>0 + 1 + 0 = 1</td>
<td>$(.6)(.4)(.6) = (.4)^1(.6)^2 = .144$</td>
<td>3(.144) = .432</td>
</tr>
<tr>
<td>0 + 0 + 1 = 1</td>
<td>$(.6)(.6)(.4) = (.4)^1(.6)^2 = .144$</td>
<td>3(.144) = .432</td>
</tr>
<tr>
<td>0 + 0 + 0 = 0</td>
<td>$(.6)(.6)(.6) = (.4)^0(.6)^3 = .216$</td>
<td>.216</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Binomial Distribution Function

The formula for the probability of a Binomial random variable is

\[ P(Y = a) = \left( \begin{array}{c} \text{the number of ways that} \\ Y = a \text{ out of } n \text{ trials} \end{array} \right) P(X = 1)^a P(X = 0)^{n-a} \]

\[ = \left( \begin{array}{c} n \\ a \end{array} \right) \pi^a (1 - \pi)^{n-a} \]

where

\[ \left( \begin{array}{c} n \\ a \end{array} \right) = \frac{n!}{a!(n-a)!} = \frac{n(n-1)(n-2) \ldots 1}{a(a-1) \ldots 1((n-a)(n-a-1) \ldots 1) \ldots 1} \]

which is called the “binomial coefficient.”

For example, the number of ways that you can get \( Y = 2 \) out of 3 tosses is

\[ \left( \begin{array}{c} 3 \\ 2 \end{array} \right) = \frac{3(2)(1)}{2(1)(1)} = 3 \]
The Systematic Component

The “Linear Predictor”.

- A linear function of the explanatory variables:
  \[ \eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_K x_{Ki} \]

- The \( x \)'s could be
  - Metric (numerical, “continuous”)
  - Discrete (dummy or effect codes)
  - Products (Interactions): e.g., \( x_{3i} = x_{1i} x_{2i} \)
  - Quadratic, cubic terms, etc: e.g., \( x_{3i} = x_{2i}^2 \)
  - Transformations: e.g., \( x_{3i} = \log(x_{3i}^*) \), \( x_{3i} = \exp(x_{3i}^*) \)

- Foreshadowing random effects models:
  \[ \eta_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + \beta_{2j} x_{2ij} + \ldots + \beta_{Kj} x_{Ki} \]

where \( i \) is index of level 1 and \( j \) is index of level 2.
The Link Function:

Problem:
- Probabilities must be between 0 and 1.
- $\eta_i$ could be between $-\infty$ to $\infty$.

Solution:
- Use (inverse of) cumulative distribution function (cdf’s) of a continuous variable to “link” the linear predictor and the mean of the response variable.
- cdf’s are $P(\text{random variable} \leq \text{specific value})$, which are between 0 and 1
  - Normal $\rightarrow$ “probit” link
  - Logistic $\rightarrow$ “logit” link
  - Gumbel (extreme value) $\rightarrow$ Complementary log-log link $\log[-\log(1 - \pi)]$
Some Example cdf’s

![Graph showing cumulative distribution functions (CDF) for Normal, Extreme Value/Gumbel, and Logistic distributions.](image)
Putting All the Components Together

\[
\log \left( \frac{P(Y_i = 1|x_i)}{P(Y_i = 0|x_i)} \right) = \logit(P(Y_i = 1|x_i))
= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K
\]

where \( x_i = (x_{0i}, x_{1i}, \ldots, x_{Ki}) \).

or in-terms of probabilities

\[
E(Y_i|x_i) = P(Y_i = 1|x_i) = \frac{\exp[\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K]}{1 + \exp[\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K]}
\]

Implicit assumption (for identification):

For \( P(Y_i = 0|x_i) \): \( \beta_0 = \beta_1 = \ldots = \beta_K = 0 \).
Interpretation of the Parameters

Simple example:

\[ P(Y_i = 1|x_i) = \frac{\exp[\beta_0 + \beta_1 x_i]}{1 + \exp[\beta_0 + \beta_1 x_i]} \]

The ratio of the probabilities is the _odds_

\[ \text{(odds of } Y_i = 1 \text{ vs } Y = 0) = \frac{P(Y_i = 1|x_i)}{P(Y_i = 0|x_i)} = \exp[\beta_0 + \beta_1 x_i] \]

For a 1 unit increase in \( x_i \) the odds equal

\[ \frac{P(Y_i = 1|(x_i + 1))}{P(Y_i = 0|(x_i + 1))} = \exp[\beta_0 + \beta_1 (x_i + 1)] \]

The "_odds ratio_" for a 1 unit increase in \( x_i \) equal

\[ \frac{P(Y_i = 1|(x_i + 1))}{P(Y_i = 0|(x_i + 1))} / \frac{P(Y_i = 1|x_i)}{P(Y_i = 0|x_i)} = \frac{\exp[\beta_0 + \beta_1 (x_i + 1)]}{\exp[\beta_0 + \beta_1 x_i]} = \exp(\beta_1) \]
Example 1: Respiratory Data

One with a continuous explanatory variable (for now)

- Response variable
  - $Y = \text{whether person has had a respiratory infection } P(Y = 1)$
  - Binomial with $n = 1$
  - Note: models can be fit to data at the level of the individual (i.e., $Y_i = 1$ where $n = 1$) or to collapsed data (i.e., $i$ index for everyone with same value on explanatory variable, and $Y_i = y$ where $n = n_i$).

- Systematic component
  \[ \beta_0 + \beta_1 (\text{age})_i \]

  where age was been centered around 36 (I don’t know why).

- Link $\longrightarrow$ logit
Example 1: The model for respiratory data

Our logit model

\[ P(Y_i = 1 | \text{age}_i) = \frac{\exp(\beta_0 + \beta_1 (\text{age}_i))}{1 + \exp(\beta_0 + \beta_1 (\text{age}_i))} \]

We’ll ignore the clustering and use MLE to estimate this model, which yields

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Conf. Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.3436</td>
<td>0.1053</td>
<td>-2.55</td>
<td>-2.14</td>
<td>495.34</td>
</tr>
<tr>
<td>age</td>
<td>-0.0248</td>
<td>0.0056</td>
<td>-0.04</td>
<td>-0.01</td>
<td>19.90</td>
</tr>
</tbody>
</table>

Interpretation: The odds of an infection equals \( \exp(-0.0248) = 0.98 \) times that for a person one year younger.

OR The odds of no infection equals \( \exp(0.0248) = 1/0.98 = 1.03 \) times the odds for a person one year older.
Probability of Infection

![Graph showing the probability of infection over age in months. The probability decreases over time.](image)
Probability of NO infection

Probability of NO Infection

Fitted Probability of Infection

Age in Months

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

-40 -30 -20 -10 0 10 20 30 40 50
Example 2: Longitudinal Depression Data

- Model Normal versus Abnormal at 1, 2 and 4 weeks.
- Also, whether mild/servere \((s = 1\) for severe) and standard/new drug \((d = 1\) for new).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>(\exp(\hat{\beta}))</th>
<th>Std. Error</th>
<th>(X^2)</th>
<th>(\text{Pr} &gt; \chi^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnose</td>
<td>1</td>
<td>-1.3139</td>
<td>0.27</td>
<td>0.1464</td>
<td>80.53</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Drug</td>
<td>1</td>
<td>-0.0596</td>
<td>0.94</td>
<td>0.2222</td>
<td>0.07</td>
<td>0.7885</td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>0.4824</td>
<td>1.62</td>
<td>0.1148</td>
<td>17.67</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Drug*Time</td>
<td>1</td>
<td>1.0174</td>
<td>2.77</td>
<td>0.1888</td>
<td>29.04</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

- The odds of normal when diagnosis is severe is 0.27 times the odds when diagnosis is mild (or \(1/0.27 = 3.72\)).
- For **new drug**, the odds ratio of normal for 1 week later:
  \[
  \exp[-0.0596 + 0.4824 + 1.0174] = \exp[1.4002] = 4.22
  
- For the **standard drug**, the odds ratio of normal for 1 week later:
  \[
  \exp[0.4824] = 1.62
  
What does \(\exp(-0.0596) \exp(0.4824) \exp(1.0174)\) equal?
SAS and fitting Logit models

title 'MLE ignoring repeated aspect of the data';
proc genmod descending;
    model outcome = diagnose treat time treat*time
        / dist=bin link=logit type3 obstats;
    output out=fitted pred=fitvalues StdResChi=haberman;

Or
proc genmod descending;
    class diagnose(ref=First) treat(ref=First);
    model outcome = diagnose treat time treat*time
        / dist=bin link=logit type3 obstats;
    output out=fitted pred=fitvalues StdResChi=haberman;

Or
proc logistic descending;
    model outcome = diagnose treat time treat*time
        / lackfit influence;

Can also use the class statement in proc logistic
R and fitting Logit models

- Simplest method is to use `glm`.
- Suppose the data looks like:

```
  id  time severe Rx y
  1   0    0   0 1
  1   1    0   0 1
  1   2    0   0 1

  27  0    0   0 1
  27  1    0   0 1
  27  2    0   0 0

  220 0    1   0 0
  220 1    1   0 0
  220 2    1   0 1
```

```
simple <- glm(y ~ severe + Rx + time + Rx*time, data=depress, family=binomial)
```
Two Major Approaches to deal with Clustering

- **“Population-averaged”**

  \[
  P(Y_{ij} = 1 | x_{ij}) = \frac{\exp(\beta_0 + \beta_1 x_{1ij} + \ldots + \beta_K x_{Kij})}{1 + \exp(\beta_0 + \beta_1 x_{1ij} + \ldots + \beta_K x_{Kij})}
  \]

  Clustering a nuisance.

  - Use generalized estimating equations (GEEs). Only estimate the first 2 moments.

- **Random Effects: “subject-specific”**

  \[
  P(Y_{ij} = 1 | x_{ij}, U_j) = \frac{\exp(\beta_{0j} + \beta_{1j} x_{1ij} + \ldots + \beta_{Kj} x_{Kij})}{1 + \exp(\beta_{0j} + \beta_{1j} x_{1ij} + \ldots + \beta_{Kj} x_{Kij})}
  \]

  - The level 2 model, we specify models for the $\beta_{kj}$'s.
  - The implied marginal of this random effects model when there is only a random intercept yields

  \[
  P(Y_{ij} = 1 | x_{ij}) = \int_{U_0} \frac{\exp(\gamma_0 + \gamma_{10} x_{1ij} + \ldots + \gamma_{K0} x_{Kij} + U_0)}{1 + \exp(\gamma_0 + \gamma_{10} x_{1ij} + \ldots + \gamma_{K0} x_{Kij} + U_0)} f(U_0) dU_0
  \]
Demonstration via Simulation

The following random model was simulated:

\[
P(Y_{ij} = 1|x_{ij}) = \frac{\exp(1.0 + 2.0x_{ij} + U_{0j})}{1 + \exp(1.0 + 2.0x_{ij} + U_{0j})}
\]

- \(x_{ij} = x_i^* + \epsilon_{ij}\) where \(x_i^* \sim \mathcal{N}(0, 4)\) and \(\epsilon_{ij} \sim \mathcal{N}(0, .01)\).
- \(U_0 \sim \mathcal{N}(0, 4)\) i.i.d.
- \(x_i^*, \epsilon_{ij}\) and \(U_{0j}\) all independent.
- Number of macro units \(j = 1, \ldots, 50\).
- Number of replications (micro units) \(i = 1, \ldots, 4\).
- The logit models were fit by
  - MLE ignoring clustering (PROC GENMOD).
  - GEE using “exchangable” correlation matrix (PROC GENMOD)
  - MLE of random effects model (PROC NLMIXED)
## Simulation: Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>Estimate</th>
<th>Error</th>
<th>Estimate</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.545</td>
<td>0.205</td>
<td>0.535</td>
<td>0.314</td>
<td>0.801</td>
<td>0.483</td>
</tr>
<tr>
<td>x</td>
<td>1.396</td>
<td>0.206</td>
<td>1.370</td>
<td>0.293</td>
<td>2.278</td>
<td>0.543</td>
</tr>
</tbody>
</table>

From GEE: correlation = 0.42

From Random effects: $\sqrt{\hat{\tau}_o^2} = 2.1636$ (s.e. = 0.6018) and $\hat{\tau}_o^2 = 4.6811$

What do you notice?
Simulation: Fitted Values

- Conditional
- Marginal
- Ignore clustering

![Graph showing fitted values for different models](image)
Conditional vs Marginal Models

Fitted Probability of Y=1

x

Conditional vs Marginal Modeling

Marginal (GEE)
Explanation of Difference

or Why the “population averaged” model (GEE) has weaker effects than the random effects model:

- The subject- (or cluster-) specific or conditional curves $P(Y_{ij} = 1 | x_{ij}, U_{0j})$ exhibit quite a bit of variability (and dependency within cluster).
- For a fixed $x$, there is considerable variability in the probability, $P(Y_{ij} = 1 | U_{0j})$.

For example, consider $x = 0$, the fitted probabilities range from about .3 to almost 1.0.

- The average of the $P(Y_{ij} = 1)$ averaged over $j$ has a less steep “slope”, weaker effect.
- The greater the variability between the cluster specific curves (i.e. the larger $\tau_0^2$ and larger correlation within cluster), the greater the difference.
Population Averaged Model

- Have repeated measures data or nested data $\rightarrow$ correlated observations.
- Use Generalized Estimating Equations (GEE) method (some cases MLE possible)
- In GLM, we assumed binomial distribution for binary data, which determines the relationship between the mean $E(Y)$ and the variance $\text{var}(Y)$ of the response variable.
- For the GEE part, we need to specify (guess) what the correlational structure is for the observations. “working correlation” matrix.
  - **Independent**: no correlation between observations.
  - **Exchangeable**: correlation between pairs of observations are same within clusters (and is the same within all clusters)
  - **Autoregressive**: for time $t$ and $t'$, correlation between $Y_t$ and $Y_{t'}$ equals $\rho^{t-t'}$
  - **Unstructured**: correlations between all pairs within clusters can differ
The Working Correlation Matrix

- GEE assumes a distribution for each marginal (e.g., $P(Y_{ij} = 1)$ for all $j$) but does not assume distribution for joint (i.e., $P(Y_{i1}, Y_{i2}, \ldots, Y_{iN})$). There’s no multivariate generalizations of discrete data distributions like there is for the normal distribution.
- Data is used to estimate the dependency between observations within a cluster. (the dependency assumed to be the same within all clusters)
- Choosing a Working Correlation Matrix
  - If available, use information you know.
  - If lack information and $n$ is small, then try unstructured to give you an idea of what might be appropriate.
  - If lack information and $n$ is large, then unstructured might require too many parameters.
- If you choose wrong, then
  - still get valid standard errors because these are based on data (empirical).
  - If the correlation/dependency is small, all choices will yield very similar results.
### GEE Example: Longitudinal Depression

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Exchangeable</th>
<th>Unstructured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$-0.0280 \ (0.1639)$</td>
<td>$-0.0281 \ (0.1742)$</td>
<td>$-0.0255 \ (0.1726)$</td>
</tr>
<tr>
<td>diagnose</td>
<td>$-1.3139 \ (0.1464)$</td>
<td>$-1.3139 \ (0.1460)$</td>
<td>$-1.3048 \ (0.1450)$</td>
</tr>
<tr>
<td>treat</td>
<td>$-0.0596 \ (0.2222)$</td>
<td>$-0.0593 \ (0.2286)$</td>
<td>$-0.0543 \ (0.2271)$</td>
</tr>
<tr>
<td>time</td>
<td>$0.4824 \ (0.1148)$</td>
<td>$0.4825 \ (0.1199)$</td>
<td>$0.4758 \ (0.1190)$</td>
</tr>
<tr>
<td>treat*time</td>
<td>$1.0174 \ (0.1888)$</td>
<td>$1.0172 \ (0.1877)$</td>
<td>$1.0129 \ (0.1865)$</td>
</tr>
</tbody>
</table>

Working correlation for exchangeable = $-0.0034$

Correlation Matrix for Unstructured:

<table>
<thead>
<tr>
<th></th>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row1</td>
<td>1.0000</td>
<td>0.0747</td>
<td>$-0.0277$</td>
</tr>
<tr>
<td>Row2</td>
<td>0.0747</td>
<td>1.0000</td>
<td>$-0.0573$</td>
</tr>
<tr>
<td>Row3</td>
<td>$-0.0277$</td>
<td>$-0.0573$</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

(Interpretation the same as when we ignored clustering.)
SAS and GEE

\begin{verbatim}
title 'GEE with Exchangeable';
proc genmod descending data=depress;
class case;
model outcome = diagnose treat time treat*time
   / dist=bin link=logit type3;
repeated subject=case / type=exch corrw;
run;

Other correlational structures

\begin{verbatim}
title 'GEE with AR(1)';
repeated subject=case / type=AR(1) corrw;
title 'GEE with Unstructured';
repeated subject=case / type=unstr corrw;
\end{verbatim}
\end{verbatim}
### R and GEE

**Input:**

```r
model.gee ← gee(y ~ severe + Rx + time + Rx*time, id, data=depress, family=binomial, corstr="exchangeable")
```

**Output:**

<table>
<thead>
<tr>
<th>Coefficients:</th>
<th>Naive Estimate</th>
<th>Naive S.E.</th>
<th>Naive z</th>
<th>Robust Estimate</th>
<th>Robust S.E.</th>
<th>Robust z</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.0280</td>
<td>0.1625</td>
<td>-0.1728</td>
<td>0.1741</td>
<td>-0.1613</td>
<td></td>
</tr>
<tr>
<td>severe</td>
<td>-1.3139</td>
<td>0.1448</td>
<td>-9.0700</td>
<td>0.1459</td>
<td>-9.0016</td>
<td></td>
</tr>
<tr>
<td>Rx</td>
<td>-0.0592</td>
<td>0.2205</td>
<td>-0.2687</td>
<td>0.2285</td>
<td>-0.2593</td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>0.4824</td>
<td>0.1141</td>
<td>4.2278</td>
<td>0.1199</td>
<td>4.0226</td>
<td></td>
</tr>
<tr>
<td>Rx:time</td>
<td>1.0171</td>
<td>0.1877</td>
<td>5.4191</td>
<td>0.1877</td>
<td>5.4192</td>
<td></td>
</tr>
</tbody>
</table>

**Estimated Scale Parameter:** 0.985392
R and GEE

Working Correlation

\[
\begin{bmatrix}
[1,] & [2,] & [3,] \\
1.00000 & -0.00343 & -0.0034 \\
-0.00343 & 1.00000 & -0.0034 \\
-0.00343 & -0.00343 & 1.0000 \\
\end{bmatrix}
\]
GEE Example 2: Respiratory Data

We’ll do simple (just time) and then complex (lots of predictors):
Exchangeable Working Correlation
Correlation 0.049991012

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

| Parameter  | Estimate | Error | 95% Confidence Limits | Pr > |Z| |
|------------|----------|-------|-----------------------|------|---|
| Intercept  | -2.3355  | 0.1134| -2.5577               | -2.1133| -20.60| < .0001 |
| age        | -0.0243  | 0.0051| -0.0344               | -0.0142| -4.72 | < .0001 |

Score Statistics For Type 3 GEE Analysis

<table>
<thead>
<tr>
<th>Chi-Square DF</th>
<th>1</th>
<th>18.24</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>1</td>
<td>18.24</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

Estimated odds ratio = \( \exp(-.0243) = 0.96 \) (or \( 1/0.96 = 1.02 \))
Note ignoring correlation, odds ratio = 0.98 or \( 1/0.98 = 1.03 \).
## Marginal Model: Complex Model

Exchangeable Working correlation = 0.04

...some model refinement needed ...

### Analysis Of GEE Parameter Estimates

| Parameter | Estimate | \( \exp(\beta) \) | std. Error | Z     | Pr > |Z| |
|-----------|----------|------------------|------------|------|------|---|
| Intercept | -2.42    | 0.89             | 0.18       | -13.61 | < .01 |
| age       | -0.03    | 0.97             | 0.01       | -5.14  | < .01 |
| xero 1    | 1        | 1.86             | 0.44       | 1.41   | .16  |
| xero 0    | 0        | 1.00             | 0.00       |       | .   |
| female 1  | -0.42    | 0.66             | 0.24       | -1.77  | .08  |
| female 0  | 0        | 1.00             | 0.00       |       | .   |
| cosine    | -0.57    | 0.57             | 0.17       | -3.36  | < .01 |
| sine      | -0.16    | 0.85             | 0.15       | -1.11  | .27  |
| height    | -0.05    | 0.95             | 0.03       | -1.55  | .12  |
| stunted 1 | 1        | 1.16             | 0.41       | 0.36   | .72  |
| stunted 0 | 0        | 1.00             | 0.00       |       | .   |
Miscellaneous Comments on Marginal Models

- With GEE
  - There is no likelihood being maximized $\implies$ no likelihood based tests. (Information criteria statistics: QIC & UQIC)
  - Can do Wald type tests and confidence intervals for parameters. Score tests are also available.

- There are other ways to model the marginal distribution(s) of discrete variables that depend on the number of observations per group (macro unit). e.g.,
  - For matched pairs of binary variables, MacNemars test.
  - Loglinear models of quasi-symmetry and symmetry to test marginal homogeneity in square tables.
  - Transition models.
  - Others.
Random Effects Model

- GLM with allow random parameters in the systematic component:

\[ \eta_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \ldots + \beta_{Kj}x_{Kij} \]

where \( i \) is index of level 1 and \( j \) is index of level 2.

- Level 1: Model conditional on \( x_{ij} \) and \( U_j \):

\[
P(Y_{ij} = 1| x_{ij}, U_j) = \frac{\exp[\beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \ldots + \beta_{Kj}x_{Kij}]}{1 + \exp[\beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \ldots + \beta_{Kj}x_{Kij}]}\]

where \( Y \) is binomial with \( n = 1 \) (i.e., Bernoulli).

- Level 2: Model for intercept and slopes:

\[
\begin{align*}
\beta_{0j} & = \gamma_{00} + U_{0j} \\
\beta_{1j} & = \gamma_{10} + \ldots + U_{1j} \\
& \vdots \\
\beta_{Kj} & = \gamma_{K0} + U_{Kj}
\end{align*}
\]
Putting Levels 1 & 2 Together

\[ P(Y_{ij} = 1 | x_{ij}, U_j) = \frac{\exp[\gamma_0 + \gamma_1 x_{1ij} + \ldots + \gamma_K x_{Kij} + U_{0j} + \ldots + U_{KJ} x_{KJ}]}{1 + \exp[\gamma_0 + \gamma_1 x_{1ij} + \ldots + \gamma_K x_{Kij} + U_{0j} + \ldots + U_{KJ} x_{KJ}]} \]

Marginalizing gives us the **Marginal Model**...

\[ P(Y_{ij} = 1 | x_{ij}) = \int_{U_0} \ldots \int_{U_K} \frac{\exp(\gamma_0 + \gamma_1 x_{1ij} + \ldots U_0 + \ldots + U_{KJ} x_{Kij})}{1 + \exp(\gamma_0 + \gamma_1 x_{1ij} + \ldots U_0 + \ldots + U_{KJ} x_{Kij})} f(U) dU \]
A Simple Random Intercept Model

Level 1:

\[ P(Y_{ij} = 1|x_{ij}) = \frac{\exp[\beta_0j + \beta_1j x_{1ij}]}{1 + \exp[\beta_0j x_{1ij}]} \]

where \( Y_{ij} \) is Binomial (Bernoulli).

Level 2:

\[
\begin{align*}
\beta_0j &= \gamma_00 + U_{0j} \\
\beta_1j &= \gamma_01
\end{align*}
\]

where \( U_{0j} \sim \mathcal{N}(0, \tau_0^2) \) i.i.d..

Random effects model for micro unit \( i \) and macro unit \( j \):

\[ P(Y_{ij} = 1|x_{ij}, U_{0j}) = \frac{\exp[\gamma_00 + \gamma_01 x_{1ij} + U_{0j}]}{1 + \exp[\gamma_00 + \gamma_01 x_{1ij} + U_{0j}]} \]
Example 1: A Simple Random Intercept Model

The respiratory data of children.

The NLMIXED Procedure
Specifications

Data Set WORK.RESPIRE
Dependent Variable resp
Distribution for Dependent Variable Binary
Random Effects u
Distribution for Random Effects Normal
Subject Variable id
Optimization Technique Dual Quasi-Newton
Integration Method Adaptive Gaussian Quadrature
# Example 1: Dimensions Table

<table>
<thead>
<tr>
<th>Dimensions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations Used</td>
<td>1200</td>
</tr>
<tr>
<td>Observations Not Used</td>
<td>0</td>
</tr>
<tr>
<td>Total Observations</td>
<td>1200</td>
</tr>
<tr>
<td>Subjects</td>
<td>275</td>
</tr>
<tr>
<td>Max Obs Per Subject</td>
<td>6</td>
</tr>
<tr>
<td>Parameters</td>
<td>3</td>
</tr>
<tr>
<td>Quadrature Points</td>
<td>10</td>
</tr>
</tbody>
</table>
### Example 1: Input and Iteration History

**Parameters**

<table>
<thead>
<tr>
<th>lam</th>
<th>bAge</th>
<th>sigma</th>
<th>NegLogLike</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.3</td>
<td>0.02</td>
<td>0.8</td>
<td>380.400779</td>
</tr>
</tbody>
</table>

**Iteration History**

<table>
<thead>
<tr>
<th>Iter</th>
<th>Calls</th>
<th>NegLogLike</th>
<th>Diff</th>
<th>MaxGrad</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>349.488037</td>
<td>30.91274</td>
<td>143.893</td>
<td>-21839.1</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>346.476536</td>
<td>3.011501</td>
<td>31.70864</td>
<td>-5.85557</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>346.35526</td>
<td>0.121276</td>
<td>15.28376</td>
<td>-0.07869</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>346.281004</td>
<td>0.074256</td>
<td>10.98611</td>
<td>-0.06659</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>346.277792</td>
<td>0.003212</td>
<td>1.785371</td>
<td>-0.00551</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>346.277696</td>
<td>0.000096</td>
<td>0.02428</td>
<td>-0.00019</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>346.277696</td>
<td>1.858E-8</td>
<td>0.000435</td>
<td>-3.45E-8</td>
</tr>
</tbody>
</table>

NOTE: GCONV convergence criterion satisfied.
Example 1: Fit Statistics & Parameter Estimates

Fit Statistics

-2 Log Likelihood 692.6
AIC (smaller is better) 698.6
AICC (smaller is better) 698.6
BIC (smaller is better) 709.4

Parameter Estimates

| Parameter | Est     | Error   | DF   | Value | Pr > |t|   | Gradient |
|-----------|---------|---------|------|-------|------|----|---------|
| gamma0    | -2.6130 | 0.1723  | 274  | -15.17| <.0001|    | 0.000063 |
| gAge      | -0.02676| 0.006592| 274  | -4.06 | <.0001|    | -0.00044 |
| tau0      | 0.8528  | 0.2087  | 274  | 4.09  | <.0001|    | 0.000041 |

Note: I cut out Alpha, Lower & Upper
Example 1: Additional Parameter Estimates

| Label             | Standard Est | Standard Error | DF  | Standard Value | Pr > |t|   |
|-------------------|--------------|----------------|-----|----------------|-------|---|
| Var(Uo)           | 0.7273       | 0.356          | 274 | 2.04           | 0.0420|
| odds ratio        | 0.9736       | 0.00641        | 274 | 151.70         | <.0001|

I requested these using an “Estimate” statement in SAS/NLMIXED.
R for Example 1: A Simple Random Intercept Model

```r
mod1.quad ← glmer(resp ∼ 1 + age + (1 | id), data=resp,
                    family=binomial, nAGQ=10 )
```

<table>
<thead>
<tr>
<th></th>
<th>mod1.quad</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.61*** (0.17)</td>
</tr>
<tr>
<td>age</td>
<td>-0.03*** (0.01)</td>
</tr>
<tr>
<td>AIC</td>
<td>698.56</td>
</tr>
<tr>
<td>BIC</td>
<td>713.83</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-346.28</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>1200</td>
</tr>
<tr>
<td>Num. groups: id</td>
<td>275</td>
</tr>
<tr>
<td>Var: id (Intercept)</td>
<td>0.73</td>
</tr>
</tbody>
</table>

***p < 0.001, **p < 0.01, *p < 0.05
Some other useful things:
For profile confidence intervals of effects:
\[
\text{round(confint(mod1.quad, level=.95),digits=4)}
\]

\begin{align*}
\text{2.5}\% & \quad \text{97.5}\% \\
\text{.sig01} & \quad 0.3993 & 1.2727 \\
(\text{Intercept}) & \quad -2.9887 & -2.3059 \\
\text{age} & \quad -0.0403 & -0.0142
\end{align*}

For odds ratios
\[
\text{odds} \leftarrow \exp(fixef(mod1.quad))
\]
\[
\text{round(odds,digits=2)}
\]
\[
\begin{align*}
(\text{Intercept} & \quad \text{age} \\
0.07 & 0.97
\end{align*}
\]
Example 1: Estimated Probabilities

![Graph showing the probability of infection over age in months with three lines representing different models: ignoring dependency, GEE (exchangable), and random effects.](image-url)
Example 1: Estimated Probabilities

Probability of Infection

Fitted Probability of Infection

Age in Months

-40 -30 -20 -10 0 10 20 30 40 50

0.0 0.1 0.2 0.3 0.4 0.5
SAS PROC NLMIXED & GLIMMIX input

title 'Random Intercept, Simple Model';
proc nlmixed data=respire qpoints=10;
   parms gamma0=-2.3 gAge=0.02 tau0= 0.8;
   eta = gamma0 + gAge*Age + u;
   p = exp(eta)/(1 + exp(eta));
   model resp ~ binary(p);
   random u ~ normal(0, tau0*tau0) subject=id out=ebuR;
   estimate 'Var(Uo)' tau0**2;
   estimate 'odds ratio' exp(gAge);
run;

title 'Easier Way to Fit Model';
proc glimmix data=respire method=quad ;
   class id ;
   model resp = age / solution link=logit dist=bin;
   random intercept / subject=id;
run;
# Default is LaPlace

```r
mod1.laplace ← glmer(resp ~ 1 + age + (1 | id), data=resp,
family=binomial)
summary(mod1.laplace)
```

# Gauss quadrature – MLE gold standard

```r
# (nAGQ= adaptive gauss quadrature and nAGQ=number quadrature points)

mod1.quad ← glmer(resp ~ 1 + age + (1 | id), data=resp,
family=binomial, nAGQ=10)
summary(mod1.quad)
```
## Different Estimation Methods $\rightarrow$ Different Results

### Some GLIMMIX Estimation Options

<table>
<thead>
<tr>
<th>Param</th>
<th>MMPL est</th>
<th>s.e.</th>
<th>RMPL est</th>
<th>s.e.</th>
<th>RSPL est</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{00}$</td>
<td>$-2.3381$</td>
<td>$(0.1163)$</td>
<td>$-2.3379$</td>
<td>$(0.1167)$</td>
<td>$-2.3722$</td>
<td>$(0.1160)$</td>
</tr>
<tr>
<td>$\gamma_{01}$</td>
<td>$-0.0254$</td>
<td>$(0.0061)$</td>
<td>$-0.0254$</td>
<td>$(0.0061)$</td>
<td>$-0.0249$</td>
<td>$(0.0061)$</td>
</tr>
<tr>
<td>$\tau_0^2$</td>
<td>$0.5734$</td>
<td>$(0.2775)$</td>
<td>$0.5967$</td>
<td>$(0.2810)$</td>
<td>$0.4996$</td>
<td>$(0.2292)$</td>
</tr>
</tbody>
</table>

### GLIMMIX

<table>
<thead>
<tr>
<th>Param</th>
<th>LaPlace est</th>
<th>s.e.</th>
<th>quad est</th>
<th>s.e.</th>
<th>NLMIXED gauss est</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{00}$</td>
<td>$-2.6768$</td>
<td>$(0.1844)$</td>
<td>$-2.6129$</td>
<td>$(0.1723)$</td>
<td>$-2.6130$</td>
<td>$(0.1723)$</td>
</tr>
<tr>
<td>$\gamma_{01}$</td>
<td>$-0.0267$</td>
<td>$(0.0067)$</td>
<td>$-0.0268$</td>
<td>$(0.0066)$</td>
<td>$-0.0268$</td>
<td>$(0.0066)$</td>
</tr>
<tr>
<td>$\tau_0^2$</td>
<td>$0.8950$</td>
<td>$(0.3961)$</td>
<td>$0.7272$</td>
<td>$(0.3559)$</td>
<td>$0.7273$</td>
<td>$(0.3560)$</td>
</tr>
</tbody>
</table>

What’s going on?
Estimation of GLIMMs

- **Pseudo-likelihood**
  - Turn into linear mixed model problem.
  - “pseudo-likelihood” Implemented in SAS PROC/GLIMMIX

- **Maximum likelihood**
  - LaPlace implemented in HLM6, GLIMMIX (SAS v9.2 & beyond), and the lme4 package in R
  - Approximate the integral (numerical integration)
    - Gaussian Quadrature
    - Adaptive quadrature
    - Implemented in SAS v9.2+: PROC NLMIXED, GLIMMIX, and the lme4 package in R
  - Bayesian: WinBugs, R, SAS v9.4 PROC MCMC
## Comparison of PLE and MLE

<table>
<thead>
<tr>
<th>Approx. integrand</th>
<th>Approx. integral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fits as wider range of models (e.g., 3+ levels, more than random intercept)</td>
<td>Narrower range of models (only 2 level for QUAD, but more with LaPlace)</td>
</tr>
<tr>
<td>Estimation approximates the integrand, “pseudo-likelihood”</td>
<td>Estimation uses numerical integration (Gaussian or adaptive quadrature)</td>
</tr>
<tr>
<td>Parameter estimates are downward</td>
<td>Parameter estimates aren’t biased</td>
</tr>
<tr>
<td>Estimation can be very poor for small ( n ) per macro-unit</td>
<td>Estimation can be fine for small ( n ) per macro-unit</td>
</tr>
<tr>
<td>Faster</td>
<td>Slower</td>
</tr>
<tr>
<td>Easier to use (pertains to SAS)</td>
<td>Harder to use</td>
</tr>
<tr>
<td>No LR testings</td>
<td>This is MLE</td>
</tr>
</tbody>
</table>
Cool Kid Example: The empty/null model

A good starting point... for the cool kid data:

**Level 1:**  \( \text{ideal}_{ij} = y_{ij} \sim \text{Binomial}(\pi_{ij}, n_{ij}^*) \) and

\[
\ln \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \eta_{ij} = \beta_{0j}
\]

**Level 2:**

\[
\beta_{0j} = \gamma_{00} + U_{0j}
\]

where \( U_{0j} \sim N(0, \tau_0^2) \) i.i.d.

**Linear Mixed Predictor:**  \[
\ln \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \eta_{ij} = \gamma_{00} + U_{0j}
\]

Useful information from this model:

- An estimate of the classroom-specific odds (& probability) of a student nominates an ideal student.
- Amount of between school variability in the odds (& probability).
Results and Interpretation

From adaptive quadrature,

\[ \hat{\gamma}_{00} = -0.4412 \ (s.e. = 0.2599) \quad \text{and} \quad \hat{\tau}^2_0 = 2.9903 \]

Interpretation:

- Based on our model, the odds that a student in classroom \( j \) nominates an ideal student equals

\[ \exp[\gamma_{00} + U_{0j}] \]

- For a classroom with \( U_{0j} = 0 \), the estimated odds of nominating an ideal student equals

\[ \exp[\hat{\gamma}_{00}] = \exp[-0.4412] = .64. \]
Results and Interpretation (continued)

- The 95% confidence of classroom-specific odds equals when $U_{0j} = 0$
  \[ \exp[\hat{\gamma}_{00} - 1.96(s.e.)], \exp[\hat{\gamma}_{00} + 1.96(s.e.)] \rightarrow (.63, .65). \]

- The 95% of the estimated variability in odds over classrooms equals
  \[ \exp[\hat{\gamma}_{00} - 1.96\sqrt{\hat{\tau}_{00}}], \exp[\hat{\gamma}_{00} - 1.96\sqrt{\hat{\tau}_{00}}] \rightarrow (0.01, 8.93). \]

What does this imply?
and Probability Estimates

\[ \hat{\gamma}_{00} = -0.4412 \text{ (s.e. } = 0.2599) \text{ and } \hat{\tau}_{00} = 2.9903 \]

- We can also compute estimated probabilities using the estimated linear predictor by using the inverse of the logit:
  \[ \pi_{ij} = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})}. \]

- For a classroom with \( U_{0j} = 0 \), the probability that a student nominates an ideal student is
  \[ \hat{\pi}_{ij} = \frac{\exp(-0.4412)}{1 + \exp(-0.4412)} = .39 \]

- A 95\% confidence interval for this classroom-specific probability (i.e., \( U_{0j} = 0 \)) is
  \[ (\logit^{-1}(0.63), \logit^{-1}(0.65)) \rightarrow (.28, .52) \]

- 95\% of the classrooms have probabilities ranging from .01 to .90.
Intraclass Correlations

For the empty/null random intercept model there are at least two ways to define an interclass correlation. This definition will extend to residual interclass correlation case:

\[
ICC = \frac{\tau_{00}}{\tau_{00} + \pi^2 / 3}
\]

where \( \pi = 3.141593 \ldots \)

\[
ICC = \frac{2.9903}{2.9903 + 3.141593^2 / 3} = .48
\]

- Lots of variability between classrooms.
- Lots of dependency within classrooms.
Random Intercept Model

Level 1: \( y_{ij} \sim \text{Binomial}(\pi_{ij}, n_{ij}^*) \) where

\[
\logit(\pi_{ij}) = \eta_{ij} = \beta_{0j} + \beta_{1j}x_{1ij}
\]

Level 2:

\[
\begin{align*}
\beta_{0j} & = \gamma_{00} + U_{0j} & \text{where} & U_{0j} \sim N(0, \tau_{00}) \\
\beta_{1j} & = \gamma_{10}
\end{align*}
\]

For interpretation:

\[
\frac{(\pi_{ij}|x_{ij}, U_{0j})}{1 - (\pi_{ij}|x_{ij}, U_{0j})} = \exp[\gamma_{00} + \gamma_{10}x_{ij} + U_{0j}]
\]

The intercept:

- When \( x_{ij} = 0 \), the odds in cluster \( j \) equals \( \exp(\gamma_{00} + U_{0j}) \).
- When \( x_{ij} = 0 \), the odds within an average cluster (i.e., \( U_{0j} = 0 \)) equals \( \exp(\gamma_{00}) \).

The slope: The odds ratio within a cluster for a 1 unit change in \( x_{ij} \) equals

\[
\frac{\exp(\gamma_{00})\exp(\gamma_{10}(x_{ij} + 1))\exp(U_{0j})}{\exp(\gamma_{00})\exp(\gamma_{10}x_{ij})\exp(U_{0j})} = \exp(\gamma_{10})
\]
Example of Random Intercept Model

We fit a random intercept model to the “cool” kid data set with only Level 1 predictors. The estimated model is

\[
\frac{\hat{\pi}_{ij}}{1 - \hat{\pi}_{ij}} = \exp [0.3240 + 0.1080\text{Popularity}_{ij} - 0.6486\text{Gender}_{ij} - 1.3096\text{Race}_{ij}]
\]

Holding other predictors constant,

- **Popularity: ****Within a cluster,** the odds that a highly popular student nominates an ideal student is \(\exp(0.1080) = 1.11\) times the odds for a low popular student.

- **Gender: ****Within a cluster,** the odds that a girl nominates an ideal student is \(\exp(0.6486) = 1.92\) times the odds for a boy.

- **Race: ****Within a cluster,** the odds that a white student nominates an ideal student is \(\exp(1.3096) = 3.70\) times the odds for a black student.
Estimated Probabilities within a Cluster

\( \hat{\pi}(\text{pop, gender, race, } U_{0j}) \times 100\%: \)

<table>
<thead>
<tr>
<th>Popular</th>
<th>Gender</th>
<th>Race</th>
<th>Random Classroom Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>( U_{0j} = -2 )</td>
</tr>
<tr>
<td>Yes</td>
<td>female</td>
<td>white</td>
<td>27.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>black</td>
<td>9.88</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>white</td>
<td>14.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>black</td>
<td>3.56</td>
</tr>
<tr>
<td>No</td>
<td>female</td>
<td>white</td>
<td>26.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>black</td>
<td>3.57</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>white</td>
<td>12.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>black</td>
<td>1.79</td>
</tr>
</tbody>
</table>
Random Intercept with Predictors

Level 1: \( y_{ij} \sim \text{Binomial}(\pi_{ij}, n_{ij}^*) \) where

\[
\frac{\pi_{ij}}{1 - \pi_{ij}} = \exp \left[ \beta_{0j} + \beta_{1j} \text{Popularity}_{ij} + \beta_{2j} \text{Gender}_{ij} + \beta_{3j} \text{Race}_{ij} \right]
\]

Level 2:

\[
\begin{align*}
\beta_{0j} &= \gamma_{00} + \gamma_{01} \text{ClassAggress}_j + U_{0j} \\
\beta_{1j} &= \gamma_{10} \\
\beta_{2j} &= \gamma_{20} \\
\beta_{3j} &= \gamma_{30}
\end{align*}
\]

where \((U_{0j} \sim N(0, \tau^2_0)).\)

For interpretation:

\[
\frac{(\pi_{ij} | U_{0j})}{1 - (\pi_{ij} | U_{0j})} = \exp \left[ \gamma_{00} + \gamma_{10} \text{Popularity}_{ij} + \gamma_{20} \text{Gender}_{ij} + \gamma_{30} \text{Race}_{ij} + \gamma_{01} \text{ClassAggress}_j + U_{0j} \right]
\]
## Results

### Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>CLASSID</td>
<td>1.9907</td>
<td>0.6365</td>
</tr>
</tbody>
</table>

### Solutions for Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>Error</th>
<th>DF</th>
<th>Value</th>
<th>Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.2392</td>
<td>0.2805</td>
<td>54</td>
<td>0.85</td>
<td>0.3976</td>
</tr>
<tr>
<td>DPOP</td>
<td>0.07219</td>
<td>0.2101</td>
<td>431</td>
<td>0.34</td>
<td>0.7314</td>
</tr>
<tr>
<td>SEX</td>
<td>-0.6366</td>
<td>0.2138</td>
<td>431</td>
<td>-2.98</td>
<td>0.0031</td>
</tr>
<tr>
<td>black</td>
<td>-1.0709</td>
<td>0.3268</td>
<td>431</td>
<td>-3.28</td>
<td>0.0011</td>
</tr>
<tr>
<td>classAgg</td>
<td>-0.6390</td>
<td>0.2374</td>
<td>54</td>
<td>-2.69</td>
<td>0.0095</td>
</tr>
</tbody>
</table>
Results and Interpretation

\[
\frac{\hat{\pi}_{ij}}{1 - \hat{\pi}_{ij}} = \exp\left[0.2392 + 0.0722\text{Popularity}_{ij} - 0.6366\text{Gender}_{ij}
- 1.0709\text{Race}_{ij} - 0.6390\text{ClassAggress}_j\right]
\]

Classroom aggression helps to explain the differences between cluster intercepts. Within in class \(j\), for students with the same popularity, gender and race, the odds of a student choosing an ideal student is

\[
\frac{\exp[0.2392 - 0.6390(\text{ClassAggress}_j) + u_{0j}]}{\exp[0.2392 - 0.6390(\text{ClassAggress}_k) + u_{0k}]}
= \exp[-0.6390(\text{ClassAggress}_j - \text{ClassAggress}_k) + (u_{0j} - u_{0k})]
\]

times those of a student in class \(k\).

\ldots So the systematic differences between classrooms can be in part explained by mean classroom aggression such that the lower classroom aggression, the greater the tendency for ideal students to be nominated as “cool”.

For students with the same popularity, gender and race, from two different schools where $u_{0j} = u_{1j}$ but the schools differ by one unit of classroom aggression, the odds ratio of nominating an ideal student equals $\exp[-0.6390] = 0.52$.

Interpretation of Popularity, Gender and Race are basically the same, but for the sake of completeness,

- **Popularity:** Within a classroom and holding other variables constant, the odds that a highly popular student nominates an ideal student is $\exp(0.0722) = 1.07$ times the odds of a low popular student.
- **Gender:** Within a classroom and holding other variables constant, the odds that a girl nominates an ideal student is $\exp(0.6366) = 1.89$ times the odds for a boy.
- **Race:** Within a classroom and holding other variables constant, the odds that a white student nominates an ideal student is $\exp(1.0709) = 2.92$ times the odds for a black student.
Residual Intraclass Correlation

We can use our estimate of $\tau_{00}$ to see what this now equals given that we have both Level 1 and Level 2 predictors in the model using

$$ICC = \frac{\hat{\tau}_{00}}{\tau_{00} + \pi^2/3}$$

where $\pi = 3.141593\ldots$ (i.e., “pi” and not probability).

For three random intercept models we have fit so far:

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{\tau}_{00}^2$</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null/Empty</td>
<td>2.9903</td>
<td>.48</td>
</tr>
<tr>
<td>+ Popularity + Gender + Minority</td>
<td>2.3129</td>
<td>.41</td>
</tr>
<tr>
<td>+ Class Aggression</td>
<td>1.9907</td>
<td>.38</td>
</tr>
</tbody>
</table>
I Random Intercept and Slope

Level 1: \( y_{ij} \sim \text{Bionmial}(\pi_{ij}, n_{ij}^*) \) where

\[
\frac{\pi_{ij}}{1 - \pi_{ij}} = \exp \left[ \beta_{0j} + \beta_{1j}\text{Popularity}_{ij} + \beta_{2j}\text{Gender}_{ij} + \beta_{3j}\text{Race}_{ij} \right]
\]

Level 2:

\[
\begin{align*}
\beta_{0j} &= \gamma_{00} + \gamma_{01}\text{ClassAggress}_j + U_{0j} \\
\beta_{1j} &= \gamma_{10} \\
\beta_{2j} &= \gamma_{20} + U_{2j} \\
\beta_{3j} &= \gamma_{30}
\end{align*}
\]

To help interpretation:

\[
\frac{(\pi_{ij}|U_{0j})}{1 - (\pi_{ij}|U_{0j})} = \exp \left[ \gamma_{00} + \gamma_{10}\text{Popularity}_{ij} + \gamma_{20}\text{Gender}_{ij} + \gamma_{3}\text{Race}_{ij} + \gamma_{01}\text{ClassAggress}_j + U_{0j} + U_{2j}\text{Gender}_{ij} \right]
\]
## Results and Comparisons

<table>
<thead>
<tr>
<th>Effects</th>
<th>Empty</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>s.e.</td>
<td>Est.</td>
<td>s.e.</td>
</tr>
<tr>
<td>Intercept</td>
<td>−0.44 (0.26)</td>
<td>0.31 (0.29)</td>
<td>0.24 (0.28)</td>
<td>0.26 (0.36)</td>
</tr>
<tr>
<td>Popularity</td>
<td>0.10 (0.21)</td>
<td>0.07 (0.21)</td>
<td>−0.10 (0.24)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>−0.65 (0.21)</td>
<td>−0.64 (0.21)</td>
<td>−0.48 (0.42)</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>−1.30 (0.33)</td>
<td>−1.07 (0.33)</td>
<td>−1.14 (0.36)</td>
<td></td>
</tr>
<tr>
<td>ClassAgg</td>
<td>−0.64 (0.24)</td>
<td>−0.71 (0.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_{00}$</td>
<td>2.99 (0.85)</td>
<td>2.31 (0.73)</td>
<td>1.99 (0.64)</td>
<td>3.74 (1.41)</td>
</tr>
<tr>
<td>$\tau_{10}$</td>
<td></td>
<td></td>
<td>−2.80 (1.42)</td>
<td></td>
</tr>
<tr>
<td>$\tau_{11}$</td>
<td></td>
<td></td>
<td></td>
<td>4.87 (1.96)</td>
</tr>
<tr>
<td># param</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>$-2\ln\text{Like}$</td>
<td>720.75</td>
<td>694.38</td>
<td>687.07</td>
<td>656.20</td>
</tr>
<tr>
<td>$AIC$</td>
<td>724.75</td>
<td>704.38</td>
<td>699.07</td>
<td>672.20</td>
</tr>
<tr>
<td>$BIC$</td>
<td>728.80</td>
<td>714.51</td>
<td>711.23</td>
<td>688.41</td>
</tr>
</tbody>
</table>
Some Model Refinements

- Popularity is clearly not significant with a $t = -0.10/0.24 = -0.42$.
- Gender is no longer significant with a $t = -0.48/0.42 = -1.16$. Should we drop gender?
- Test $H_0: \tau_1^2 = \tau_{01} = 0$ versus $H_a$: not $H_0$. Use same method as we did for HLM: compute $LR^*$ and compare to a mixture of chi-square distributions.

<table>
<thead>
<tr>
<th>Model</th>
<th>$-2\ln\text{Like}$</th>
<th>$LR^*$</th>
<th>$\chi_2$</th>
<th>$\chi_1$</th>
<th>$p$-value</th>
<th>$p$-value</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null</td>
<td>687.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_a$</td>
<td>656.20</td>
<td>30.87</td>
<td>tiny</td>
<td>tiny</td>
<td>&lt;.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Drop Popularity from the model but keep random Gender effect.
## Results and Comparisons

<table>
<thead>
<tr>
<th>Effects</th>
<th>Model 2</th>
<th>Est.</th>
<th>s.e.</th>
<th>Model 3</th>
<th>Est.</th>
<th>s.e.</th>
<th>Model 4</th>
<th>Est.</th>
<th>s.e.</th>
<th>Refined</th>
<th>Est.</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.31</td>
<td>(0.29)</td>
<td></td>
<td>0.24</td>
<td>(0.28)</td>
<td></td>
<td>0.26</td>
<td>(0.36)</td>
<td></td>
<td>0.22</td>
<td>(0.35)</td>
<td></td>
</tr>
<tr>
<td>Popularity</td>
<td>0.10</td>
<td>(0.21)</td>
<td></td>
<td>0.07</td>
<td>(0.21)</td>
<td></td>
<td>−0.10</td>
<td>(0.24)</td>
<td></td>
<td>−0.48</td>
<td>(0.41)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>−0.65</td>
<td>(0.21)</td>
<td></td>
<td>−0.64</td>
<td>(0.21)</td>
<td></td>
<td>−0.48</td>
<td>(0.42)</td>
<td></td>
<td>−0.48</td>
<td>(0.41)</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>−1.30</td>
<td>(0.33)</td>
<td></td>
<td>−1.07</td>
<td>(0.33)</td>
<td></td>
<td>−1.14</td>
<td>(0.36)</td>
<td></td>
<td>−1.14</td>
<td>(0.36)</td>
<td></td>
</tr>
<tr>
<td>ClassAgg</td>
<td></td>
<td></td>
<td></td>
<td>−0.64</td>
<td>(0.24)</td>
<td></td>
<td>−0.71</td>
<td>(0.27)</td>
<td></td>
<td>−0.69</td>
<td>(0.27)</td>
<td></td>
</tr>
<tr>
<td>( \tau_{00} )</td>
<td>2.31</td>
<td>(0.73)</td>
<td></td>
<td>1.99</td>
<td>(0.64)</td>
<td></td>
<td>3.74</td>
<td>(1.41)</td>
<td></td>
<td>3.63</td>
<td>(1.35)</td>
<td></td>
</tr>
<tr>
<td>( \tau_{20} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−2.80</td>
<td>(1.42)</td>
<td></td>
<td>−2.73</td>
<td>(1.38)</td>
<td></td>
</tr>
<tr>
<td>( \tau_{22} ) gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.87</td>
<td>(1.96)</td>
<td></td>
<td>4.76</td>
<td>(1.91)</td>
<td></td>
</tr>
<tr>
<td># param</td>
<td>5</td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-2\ln\text{Like})</td>
<td>694.38</td>
<td></td>
<td></td>
<td>687.07</td>
<td></td>
<td></td>
<td>656.20</td>
<td></td>
<td></td>
<td>656.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( AIC )</td>
<td>704.38</td>
<td></td>
<td></td>
<td>699.07</td>
<td></td>
<td></td>
<td>672.20</td>
<td></td>
<td></td>
<td>670.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( BIC )</td>
<td>714.51</td>
<td></td>
<td></td>
<td>711.23</td>
<td></td>
<td></td>
<td>688.41</td>
<td></td>
<td></td>
<td>684.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C.J. Anderson  (Illinois)  Multilevel Logistic Regression  Spring 2020  80.80/130
Comments on Results

- A likelihood ratio test for popularity,

\[ LR = 656.38 - 656.20 = 0.18 \]

compared to a \( \chi^2_1 \) has \( p = .67 \).

- Fixed parameter estimates and their standard errors are the same.

- Estimates variance and their standard errors changed a little.

- Empirical standard errors are very similar to the model based ones reported on the previous slide.

- Before getting serious about this model, we should consider 3 level models because students are nested within peer groups nested within classrooms.
Three Level Models

Regardless of the type of response variable (e.g., normal, binomial, etc), additional levels of nesting can be included.

A very simple example:

**Level 1** \( y_{ijk} \sim \text{Binomial}(\pi_{ijk}, n^*_{ijk}) \) and \( \text{logit}(\pi_{ijk}) = \beta_{0jk} \)

**Level 2**

\[
\beta_{0jk} = \gamma_{00k} + U_{0jk},
\]

where \( U_{0jk} \sim N(0, \tau^2_0) \) i.i.d.

**Level 3:**

\[
\gamma_{00k} = \xi_{00} + W_{0k}
\]

where \( W_{0k} \sim N(0, \psi^2) \) i.i.d and independent of \( U_{0jk} \).

**Linear Mixed Predictor:**

\[
\text{logit}(\pi_{ijk}) = \underbrace{\xi_{00}}_{\text{fixed}} + \underbrace{U_{0jk} + W_{0k}}_{\text{random}}
\]
Adding Predictors

- Predictors can be added at every level.
- Predictors at lower levels can have random coefficients that are modeled at a high level.
- Can have cross-level interactions.

Predictors for the “cool” kid data:

- **Level 1**
  - Black$_{ijk}$ = 1 if or black student, 0 for white
  - Zpop$_{ijk}$ = standardized popularity score
  - Zagg$_{ijk}$ = standardized aggression score

- **Level 2**
  - Gnom$_{jk}$ = Peer group centrality
  - Gagg$_{jk}$ = Peer group aggression score
  - gender$_{jk}$ = 1 boy group and 0 girl group

- **Level 3**
  - ClassAgg$_{k}$ = Mean class aggression score
  - Majority$_{k}$ = 1 more white and 0 more Black
To show what happens when we enter variable, I’ll do this one set at a time.

**Level 1** \( y_{ijk} \sim \text{Binomial}(\pi_{ijk}, n_{ijk}^*) \) and

\[
\text{logit}(\pi_{ijk}) = \beta_{0jk} + \beta_{1jk} Z_{\text{pop}ijk} + \beta_{2jk} Z_{\text{agg}ijk}
\]

**Level 2**

\[
\begin{align*}
\beta_{0jk} &= \gamma_{00k} + U_{0jk} \\
\beta_{1jk} &= \gamma_{10k} \\
\beta_{2jk} &= \gamma_{20k}
\end{align*}
\]

where \( U_{0jk} \sim N(0, \tau_0^2) \ i.i.d. \)

**Level 3:**

\[
\begin{align*}
\gamma_{00k} &= \xi_{00} + W_{0k} \\
\gamma_{10k} &= \xi_{10} \\
\gamma_{20k} &= \xi_{20}
\end{align*}
\]

where \( W_{0k} \sim N(0, \psi^2) \ i.i.d \) and independent of \( U_{0jk} \).

What’s the **Linear Mixed Predictor**?
Adding Predictors of Intercepts

To get convergence, I switched from Method=quad to Method=LaPlace

<table>
<thead>
<tr>
<th>Model</th>
<th>#</th>
<th>(-2\ln\text{like})</th>
<th>AIC</th>
<th>BIC</th>
<th>(\hat{\tau}_0^2)</th>
<th>(\hat{\psi}_0^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empty</td>
<td>3</td>
<td>701.14</td>
<td>707.14</td>
<td>713.21</td>
<td>1.28</td>
<td>2.84</td>
</tr>
<tr>
<td>+ Level 1</td>
<td>5</td>
<td>682.21</td>
<td>700.21</td>
<td>718.44</td>
<td>1.32</td>
<td>2.09</td>
</tr>
<tr>
<td>+ Level 2</td>
<td>8</td>
<td>673.06</td>
<td>691.06</td>
<td>709.29</td>
<td>1.02</td>
<td>1.90</td>
</tr>
<tr>
<td>+ Level 3</td>
<td>10</td>
<td>664.44</td>
<td>686.44</td>
<td>708.99</td>
<td>1.06</td>
<td>1.53</td>
</tr>
</tbody>
</table>

- Adding Level 1 predictors improves the fit of the model but has little effect on \(\hat{\tau}_0^2\), but some on 
- Adding Level 2 predictors improves the fit of the models, have an effect on \(\hat{\tau}_0^2\) but little effect \(\hat{\psi}^2\).
- Adding Level 3 predictors improves the fit of the models, has an effect on \(\hat{\psi}^2\) but little effect on \(\hat{\tau}_0^2\).
What the Last Model Looks Like

**Level 1** \( y_{ijk} \sim \text{Binomial}(\pi_{ijk}, n^*_{ijk}) \) and

\[
\text{logit}(\pi_{ijk}) = \beta_{0jk} + \beta_{1jk} \text{Black}_{ijk} + \beta_{2jk} \text{Zpop}_{ijk} + \beta_{3jk} \text{Zagg}_{ijk}
\]

**Level 2**

\[
\begin{align*}
\beta_{0jk} &= \gamma_{00k} + \gamma_{01k} \text{Gnom}_{jk} + \gamma_{02k} \text{Gagg}_{jk} + \gamma_{03k} \text{Sex}_{jk} + U_{0jk} \\
\beta_{1jk} &= \gamma_{10k} \\
\beta_{2jk} &= \gamma_{20k} \\
\beta_{3jk} &= \gamma_{30k}
\end{align*}
\]

where \( U_{0jk} \sim N(0, \tau_0^2) \) i.i.d..

**Level 3:**

\[
\begin{align*}
\gamma_{00k} &= \xi_{000} + \xi_{001} \text{ClassAgg}_k + \xi_{002} \text{Majority}_k + W_{0k} \\
\gamma_{01k} &= \xi_{010} \\
\gamma_{02k} &= \xi_{020} \\
\gamma_{03k} &= \xi_{030} \\
\gamma_{10k} &= \xi_{100} \\
\gamma_{20k} &= \xi_{200} \\
\gamma_{30k} &= \xi_{300}
\end{align*}
\]
Linear Mixed Predictor

\[
\text{logit}(\pi_{ijk}) = \xi_{000} + \xi_{010} \text{Gnom}_{jk} + \xi_{020} \text{Gagg}_{jk} + \xi_{001} \text{ClassAgg}_j \\
\xi_{002} \text{Majority}_k + \xi_{100} \text{Zpop}_{ijk} + \xi_{200} \text{Zagg}_{ijk} \\
+ U_{0jk} + W_{0k}
\]

By adding all of these variable to model the intercept, total variance of the intercept has decreased from

\[1.28 + 2.84 = 4.12\] to \[1.06 + 1.53 = 2.59\]

(about a 63% decrease).

We can also add additional fixed effects and random effects to the Level 2 and Level 3 regressions.
## Adding More Random and Fixed

<table>
<thead>
<tr>
<th>Model</th>
<th>#</th>
<th>$-2\ln$like</th>
<th>AIC</th>
<th>BIC</th>
<th>$\hat{\tau}_0^2$ (s.e.)</th>
<th>$\hat{\psi}_0^2$ (se.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empty</td>
<td>3</td>
<td>701.14</td>
<td>707.14</td>
<td>713.21</td>
<td>1.28(0.51)</td>
<td>2.84(0.95)</td>
</tr>
<tr>
<td>+ Level 1</td>
<td>5</td>
<td>682.21</td>
<td>700.21</td>
<td>718.44</td>
<td>1.32(0.55)</td>
<td>2.09(0.81)</td>
</tr>
<tr>
<td>+ Level 2</td>
<td>8</td>
<td>673.06</td>
<td>691.06</td>
<td>709.29</td>
<td>1.02(0.47)</td>
<td>1.90(0.73)</td>
</tr>
<tr>
<td>+ Level 3</td>
<td>10</td>
<td>664.44</td>
<td>686.44</td>
<td>708.99</td>
<td>1.06(0.48)</td>
<td>1.53(0.63)</td>
</tr>
<tr>
<td>$+w_{3k}$</td>
<td>11</td>
<td>648.20</td>
<td>674.20</td>
<td>700.53</td>
<td>0.05(0.25)</td>
<td>3.25(1.34)</td>
</tr>
<tr>
<td>(gender)$_{jk}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-Z_{\text{pop}}_{ijk}$</td>
<td>7</td>
<td>650.18</td>
<td>670.18</td>
<td>690.44</td>
<td>0.10(0.25)</td>
<td>3.61(1.43)</td>
</tr>
<tr>
<td>$-G_{\text{nom}}_{jk}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-\text{Majority}_k$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$+B_{\text{black}}_{ijk}$</td>
<td>8</td>
<td>646.08</td>
<td>668.08</td>
<td>690.36</td>
<td>0.04(.22)</td>
<td>3.25(1.22)</td>
</tr>
<tr>
<td>$\times C_{\text{lassagg}}_k$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C.J. Anderson (Illinois)  Multilevel Logistic Regression  Spring 2020  88.88/ 130
One More Model

<table>
<thead>
<tr>
<th>Model</th>
<th>#</th>
<th>−2lnlike</th>
<th>AIC</th>
<th>BIC</th>
<th>(\hat{\tau}_{00}) (s.e.)</th>
<th>(\hat{\psi}_{00}) (se.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(−U_{0jk})</td>
<td>7</td>
<td>646.12</td>
<td>666.12</td>
<td>686.37</td>
<td>3.26 (1.22)</td>
<td>−2.69 (1.33)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.77 (1.93)</td>
<td></td>
</tr>
</tbody>
</table>

Test for \(H_0 : \tau_{00} = 0\) versus \(H_a : \tau_{00} > 0\), \(LR = 646.12 - 646.08 = .04\). This would be compared to a mixture of \(\chi_1^2\) and \(\chi_0^2\)—clearly not significant.

So what’s the final model?
What the Final (?) Model Looks Like

**Level 1**

\[ y_{ijk} \sim \text{Binomial}(\pi_{ijk}, n_{ijk}) \text{ and} \]

\[
\logit(\pi_{ijk}) = \beta_{0jk} + \beta_{1jk} \text{Black}_{ijk} + \beta_{2jk} \text{Zagg}_{ijk}
\]

**Level 2**

\[
\beta_{0jk} = \gamma_{00k} + \gamma_{02k} \text{Gagg}_{jk} + \gamma_{03k} \text{Gender}_{jk}
\]

\[
\beta_{1jk} = \gamma_{10k}
\]

\[
\beta_{2jk} = \gamma_{20k}
\]

**Level 3:**

\[
\gamma_{00k} = \xi_{000} + \xi_{001} \text{ClassAgg}_{k} + W_{00k}
\]

\[
\gamma_{02k} = \xi_{020}
\]

\[
\gamma_{03k} = \xi_{030} + W_{03k}
\]

\[
\gamma_{10k} = \xi_{100} + \xi_{101} \text{ClassAgg}_{k}
\]

\[
\gamma_{20k} = \xi_{200}
\]
Linear Mixed for $\eta_{ijk}$

$$\eta_{ijk} = \xi_{000} + \xi_{001} \text{ClassAgg}_k + \xi_{020} \text{Gagg}_k + \xi_{030} \text{Sex}_{jk} + \xi_{100} \text{Black}_{ijk}$$

$$+ \xi_{101} \text{ClassAgg}_k \text{Black}_{ijk} + \xi_{200} \text{Zagg}_{ijk} + W_{00k} + W_{03k} \text{gender}_{jk}$$
# Parameter Estimates

## Solutions for Fixed Effects

| Effect           | Estimate | Error  | DF   | Value | Pr>|t| |
|------------------|----------|--------|------|-------|-----|
| Intercept        | 0.6026   | 0.3535 | 54   | 1.70  | .0940|
| ZAGG             | 0.3069   | 0.1349 | 386  | 2.28  | .0234|
| black            | -1.1183  | 0.3787 | 386  | -2.95 | .0033|
| GAGG             | -0.8538  | 0.3268 | 386  | -2.61 | .0093|
| SEX              | -0.3073  | 0.4191 | 43   | -0.73 | .4674|
| classAgg         | -0.3718  | 0.3056 | 386  | -1.22 | .2246|
| black*classAgg   | -0.8255  | 0.4108 | 386  | -2.01 | .0452|

Note: This is not a publicly available data set, but I put code without data online.
IRT: A Rasch Model

- Suppose we have 4 items where $Y_{ij} = 1$ for a correct response and $Y_{ij} = 0$ for incorrect.
- Explanatory Variables are indicator variables,

$$x_{1j} = \begin{cases} 1 & \text{if person } j \text{ responds to item } 1 \\ 0 & \text{otherwise} \end{cases} \quad \ldots \quad x_{4j} = \begin{cases} 1 & \text{if person } j \text{ responds to item } 4 \\ 0 & \text{otherwise} \end{cases}$$

- **Level 1**: The linear predictor

$$\eta_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \beta_{3j}x_{3ij} + \beta_{4j}x_{4ij}$$

and the link is the logit:

$$P(Y_{ij} | x_{1ij}, x_{2ij}, x_{3ij}, x_{4ij}) = \frac{\exp[\eta_{ij}]}{1 + \exp[\eta_{ij}]}$$
To get Rasch Model (continued)

- **Level 2:**
  
  $\beta_{0j} = U_{0j} \sim N(0, \tau_{00})$
  
  $\beta_{1j} = \gamma_{10}$
  
  $\beta_{2j} = \gamma_{20}$
  
  $\beta_{3j} = \gamma_{30}$
  
  $\beta_{4j} = \gamma_{40}$

- **Our model for each item**

  $P(Y_{1j}|x_{1i}, \ldots, x_{5ij}, U_{0j}) = \frac{\exp[\gamma_{10} + U_{0j}]}{1 + \exp[\gamma_{10} + U_{0j}]}$
  
  $P(Y_{2j}|x_{1i}, \ldots, x_{5ij}, U_{0j}) = \frac{\exp[\gamma_{20} + U_{0j}]}{1 + \exp[\gamma_{20} + U_{0j}]}$
  
  $P(Y_{3j}|x_{1i}, \ldots, x_{5ij}, U_{0j}) = \frac{\exp[\gamma_{30} + U_{0j}]}{1 + \exp[\gamma_{30} + U_{0j}]}$
  
  $P(Y_{4j}|x_{1i}, \ldots, x_{5ij}, U_{0j}) = \frac{\exp[\gamma_{40} + U_{0j}]}{1 + \exp[\gamma_{40} + U_{0j}]}$
The IRT Connection

Set

- $\gamma_{i0} = b_i$, the difficulty for item $i$
- $U_{0j} = \theta_j$, value on latent variable for examinee $j$.

For item $i$,

$$P(Y_{ij} = 1|x_{1ij}, \ldots, x_{4ij}, U_{0j}) = \frac{\exp(\gamma_{i0} + U_{0j})}{1 + \exp(\gamma_{i0} + U_{0j})} = \frac{\exp(b_i + \theta_j)}{1 + \exp(b_j + \theta_j)}$$

“One Parameter Logistic Regression Model” or the Rasch model
Example 2: LSAT data

For this, we’ll use the LSAT6 Data: \( N = 1000, n = 5 \).

- Responses (items) are nested within examinees.
- Response \( Y_{ij} \) is correct (\( Y = 1 \)) or not (\( Y = 0 \)) and assumed to be binomial.
- Explanatory Variables are dummy variables indicating the item the examinee is responding to

\[
x_{1ij} = \begin{cases} 
1 & \text{if item 1} \\ 
0 & \text{otherwise} 
\end{cases} \quad \ldots \quad x_{5ij} = \begin{cases} 
1 & \text{if item 5} \\ 
0 & \text{otherwise} 
\end{cases}
\]

- Level 1: The “linear predictor”

\[
\eta_{ij} = \beta_0j + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \beta_{3j}x_{3ij} + \beta_{4j}x_{4ij} + \beta_{5j}x_{5ij}
\]

and the link is the logit:

\[
P(Y_{ij} \mid x_{1ij}, x_{2ij}, x_{3ij}, x_{4ij}, x_{5ij}) = \frac{\exp[\eta_{ij}]}{1 + \exp[\eta_{ij}]}
\]

Observations are independent at level 1.
Example: LSAT6

Level 2:

\[
\begin{align*}
\beta_{0j} &= U_{0j} & \text{← on average equals 0} \\
\beta_{1j} &= \gamma_{10} \\
\beta_{2j} &= \gamma_{20} \\
\beta_{3j} &= \gamma_{30} \\
\beta_{4j} &= \gamma_{40} \\
\beta_{5j} &= \gamma_{50}
\end{align*}
\]
Example: LSAT6

Our model for each item

\[
P(Y_{1j}|x_{1ij}, \ldots, x_{5ij}, U_{0j}) = \frac{\exp[\gamma_{10} + U_{0j}]}{1 + \exp[\gamma_{10} + U_{0j}]} \\
P(Y_{2j}|x_{1ij}, \ldots, x_{5ij}, U_{0j}) = \frac{\exp[\gamma_{20} + U_{0j}]}{1 + \exp[\gamma_{20} + U_{0j}]} \\
P(Y_{3j}|x_{1ij}, \ldots, x_{5ij}, U_{0j}) = \frac{\exp[\gamma_{30} + U_{0j}]}{1 + \exp[\gamma_{30} + U_{0j}]} \\
P(Y_{4j}|x_{1ij}, \ldots, x_{5ij}, U_{0j}) = \frac{\exp[\gamma_{40} + U_{0j}]}{1 + \exp[\gamma_{40} + U_{0j}]} \\
P(Y_{5j}|x_{1ij}, \ldots, x_{5ij}, U_{0j}) = \frac{\exp[\gamma_{50} + U_{0j}]}{1 + \exp[\gamma_{50} + U_{0j}]} 
\]

What very well known model is this?
The Model for the LSAT6

- Set $\gamma_{i0} = -b_i$ and $U_{0j} = \theta_j$.

- This is an example of a model that can be fit without numerical integration (by conditional MLE).

- Implications for applications (i.e., can add individual and/or item level predictor variables as was done with the GSS vocabulary items).
## LSAT6: Estimated Parameters

### Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>DF</th>
<th>Value</th>
<th>&gt;</th>
<th>Pr</th>
<th>Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td>b1</td>
<td>-2.73</td>
<td>0.13</td>
<td>999</td>
<td>-20.93</td>
<td>&lt;</td>
<td>.01</td>
<td>0.000002</td>
</tr>
<tr>
<td>b2</td>
<td>-1.00</td>
<td>0.08</td>
<td>999</td>
<td>-12.61</td>
<td>&lt;</td>
<td>.01</td>
<td>0.000175</td>
</tr>
<tr>
<td>b3</td>
<td>-0.24</td>
<td>0.07</td>
<td>999</td>
<td>-3.34</td>
<td>0.01</td>
<td>-0.00016</td>
<td></td>
</tr>
<tr>
<td>b4</td>
<td>-1.31</td>
<td>0.08</td>
<td>999</td>
<td>-15.44</td>
<td>&lt;</td>
<td>.01</td>
<td>0.000107</td>
</tr>
<tr>
<td>b5</td>
<td>-2.10</td>
<td>0.11</td>
<td>999</td>
<td>-19.91</td>
<td>&lt;</td>
<td>.01</td>
<td>-0.00008</td>
</tr>
<tr>
<td>std</td>
<td>-0.76</td>
<td>0.07</td>
<td>999</td>
<td>-10.88</td>
<td>&lt;</td>
<td>.01</td>
<td>-0.00027</td>
</tr>
</tbody>
</table>

### Additional Estimates

<table>
<thead>
<tr>
<th>Label</th>
<th>Estimate</th>
<th>Error</th>
<th>DF</th>
<th>Value</th>
<th>&gt;</th>
<th>Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>var(theta)</td>
<td>0.57</td>
<td>0.10</td>
<td>999</td>
<td>5.44</td>
<td>&lt;</td>
<td>.01</td>
</tr>
</tbody>
</table>
LSAT6: Fitted Probabilities

LSAT6: Rasch Model Fitted Item Response Functions

Fitted Probability

Theta (ability, knowledge, etc)
# How Model was Fit: Data file/Design Matrix

<table>
<thead>
<tr>
<th>ID</th>
<th>$Y_{ij}$</th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
<th>x4</th>
<th>x5</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>298</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>298</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>298</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>298</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>298</td>
</tr>
</tbody>
</table>
Preparing the Data

To reformat data from tabled data to format on previous page (note: \(x_i = 1, 2\)):

```plaintext
data vector;
  set table;
  id = __n__;
  i1=1; i2=0; i3=0; i4=0; i5=0; y=(x1-1); output;
  i1=0; i2=1; i3=0; i4=0; i5=0; y=(x2-1); output;
  i1=0; i2=0; i3=1; i4=0; i5=0; y=(x3-1); output;
  i1=0; i2=0; i3=0; i4=1; i5=0; y=(x4-1); output;
  i1=0; i2=0; i3=0; i4=0; i5=1; y=(x5-1); output;
drop x1-x5;
```
### Proc NLMIXED for Rasch Model

```plaintext
title 'Rasch model as a random intercept model';
proc nlmixed data=vector qpoints=20;
   parms b1-b5=.2 std=.1;
   eta = theta -(b1*i1 + b2*i2 + b3*i3 + b4*i4 + b5*i5);
   p = exp(eta)/(1 + exp(eta));
   model y ~ binary(p);
   random theta ~ normal(0,std*std) subject = id;
   replicate count;
   estimate 'var(theta)' std**2;
OR
proc glimmix data= vector method=quad ;
   class id;
   model count = i1 i2 i3 i4 i5 / link=logit dist=binomial
      solution noint;
   random intercept / subject=id;
```
R for Rasch Model: glmer

- Same basic data set up as used by SAS:
  ```
  id   i1  i2  i3  i4  i5  y
  1    1   0   0   0   0   0
  1    0   1   0   0   0   0
  1    0   0   1   0   0   0
  1    0   0   0   1   0   0
  1    0   0   0   0   1   0
  1    1   0   0   0   0   0
  ...
  12   1   0   0   0   0   0
  12   0   1   0   0   0   0
  12   0   0   1   0   0   0
  12   0   0   0   1   0   1
  12   0   0   0   1   1
  ```

- By Gauss quadrature:
  ```r
  rasch.quad <- glmer(y ~ -1 + i1 + i2 + i3 + i4 + i5 + (1 | id), data=lsat, family=binomial, nAGQ=10)
  ```

- Defaults method is LaPlace:
  ```r
  rasch.laplace <- glmer(y ~ -1 + i1 + i2 + i3 + i4 + i5 + (1 | id), data=lsat, family=binomial)
  ```
R for Rasch Model: nlme

A more complex but flexible way to fit Rasch model:

```r
onePL <- function(b1,b2,b3,b4,b5,theta) {
  b=b1*i1 + b2*i2 + b3*i3 + b4*i4 + b5*i5
  exp(theta-b)/( 1 + exp(theta-b) )
}

rasch3 <- nlme(y ~ onePL(b1,b2,b3,b4,b5,theta),
               data=lsat,
               fixed= b1+b2+b3+b4+b5 ~ 1,
               random = theta ~ 1 | id,
               start=c(b1=1, b2=1,b3=1, b4=1, b5=1) )
```
Another Random Effects Model

For the LSAT6 data

- Explanatory variables are dummy variables indicating the item the examinee is responding to.

- Random Effects Model: the linear predictor

\[
\text{logit}(P(Y_{ij} = 1|\theta_j)) = b_1 x_{1ij} + b_2 x_{2ij} + b_3 x_{3ij} + b_4 x_{4ij} + b_5 x_{5ij} \\
+ (a_1 x_{1ij} + a_2 x_{2ij} + a_3 x_{3ij} + a_4 x_{4ij} + a_5 x_{5ij})\theta_j
\]

- The model for item $i$ is

\[
P(Y_{ij} = 1|\theta_j) = \frac{\exp[b_i + a_i\theta_j]}{1 + \exp[b_i + a_1\theta_j]}
\]

- The model is often written as

\[
P(Y_{ij} = 1|\theta_j) = \frac{\exp[a_i(\theta_j - b_i^*)]}{1 + \exp[a_i(\theta_j - b_i^*)]}
\]

where $b_i^* = -b_i/a_i$. 

# Two Parameter Logistic Model Estimates

| Parameter | Estimate | Error | DF  | $t$-Value | Pr $>|t|$ | Gradient |
|-----------|----------|-------|-----|-----------|-----------|----------|
| $b_1$     | $-2.77$  | 0.21  | 999 | $-13.48$  | $<.01$    | $-9.64E-6$ |
| $b_2$     | $-0.99$  | 0.09  | 999 | $-11.00$  | $<.01$    | 0.000031  |
| $b_3$     | $-0.25$  | 0.08  | 999 | $-3.27$   | .01       | 0.000011  |
| $b_4$     | $-1.28$  | 0.10  | 999 | $-12.98$  | $<.01$    | $-0.00004$|
| $b_5$     | $-2.05$  | 0.14  | 999 | $-15.17$  | $<.01$    | $4.593E-6$|
| $a_1$     | 0.83     | 0.26  | 999 | 3.20      | $<.01$    | $-7.23E-6$|
| $a_2$     | 0.72     | 0.19  | 999 | 3.87      | $<.01$    | 0.000018  |
| $a_3$     | 0.89     | 0.23  | 999 | 3.83      | $<.01$    | $-2.7E-6$ |
| $a_4$     | 0.69     | 0.19  | 999 | 3.72      | $<.01$    | $-3.65E-6$|
| $a_5$     | 0.66     | 0.21  | 999 | 3.13      | $<.01$    | 0.00001   |

$\tau^2_{00}$ set to 1 for identification and Converged (i.e., $\theta \sim N(0, 1)$)
LSAT6: Fitted Probabilities

LSAT6: 2pl Model Fitted Item Response Functions

Fitted Probability

Theta (ability, knowledge, etc)

C.J. Anderson (Illinois)
### Two Parameter Logistic Model Estimates

2PL model is too complex for the data? Should we go with the simpler one? (i.e., the Rasch model)

Need to test $H_0 : a_i = 1$.

Wald statistics are very small (i.e. $t$-statistics $= (\hat{a}_i - 1)/se(a_i)$):

- $(0.83 - 1)/.26 = -0.68 \quad (p = .50)$
- $(0.72 - 1)/.19 = -1.49 \quad (p = .14)$
- $(0.89 - 1)/.23 = -0.47 \quad (p = .64)$
- $(0.69 - 1)/.19 = -1.68 \quad (p = .09)$
- $(0.66 - 1)/.21 = -1.63 \quad (p = .10)$

Retain $H_0$ for all items.
**Better Test of** \( H_0 : a_i = 1 \)

The likelihood ratio test:

<table>
<thead>
<tr>
<th>Model</th>
<th># parameters</th>
<th>(-\text{logLike})</th>
<th>(LR)</th>
<th>(df)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rasch</td>
<td>6</td>
<td>4933.875</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2PL</td>
<td>10</td>
<td>4933.307</td>
<td>0.548</td>
<td>4</td>
<td>.97</td>
</tr>
</tbody>
</table>

Retain \( H_0 : a_1 = a_2 = a_3 = a_4 = a_5 = 1 \)
SAS and the 2PL Model

```sas
proc nlmixed data=vector method=gauss qpoints=15 noad;
    parms b1-b5=.2 a1-a5=1 ;
    eta = (a1*i1 + a2*i2 + a3*i3 + a4*i4 + a5*i5)*theta
        -(b1*i1 + b2*i2 + b3*i3 + b4*i4 +b5*i5);
    p = exp(eta)/(1 + exp(eta));
    model y ~ binary(p);
    random theta ~ normal(0,1) subject = id; * $\tau_0^2 = 1$ for ID;
    replicate count;
    estimate 'Ho: a1=1' a1-1;  * For tests on each slope;
    estimate 'Ho: a2=1' a2-1;
    estimate 'Ho: a3=1' a3-1;
    estimate 'Ho: a4=1' a4-1;
    estimate 'Ho: a5=1' a5-1;
    estimate 'b1/a1' b1/a1; * More standard IRT parametrization;
```
Rasch Example

Data are response to 10 vocabulary items from the 2004 General Social Survey from $n = 1155$ respondents. . . . need $x_{1ij}$ through $x_{10ij}$ for this model.

The model was fit using SAS PROC NLMIXED. Edited output:

**NOTE:** GCONV convergence criterion satisfied.

**Fit Statistics**

-2 Log Likelihood 10197
AIC (smaller is better) 10219
AICC (smaller is better) 10220
BIC (smaller is better) 10275
## Rasch Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>DF</th>
<th>Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th>Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td>b1</td>
<td>-2.0756</td>
<td>0.09925</td>
<td>1154</td>
<td>-20.91</td>
<td>&lt; .0001</td>
<td>-0.000035</td>
<td></td>
</tr>
<tr>
<td>b2</td>
<td>-3.4436</td>
<td>0.1435</td>
<td>1154</td>
<td>-23.99</td>
<td>&lt; .0001</td>
<td>2.484E-7</td>
<td></td>
</tr>
<tr>
<td>b3</td>
<td>1.4330</td>
<td>0.08806</td>
<td>1154</td>
<td>16.27</td>
<td>&lt; .0001</td>
<td>-4.6E-6</td>
<td></td>
</tr>
<tr>
<td>b4</td>
<td>-3.6519</td>
<td>0.1537</td>
<td>1154</td>
<td>-23.76</td>
<td>&lt; .0001</td>
<td>6.38E-7</td>
<td></td>
</tr>
<tr>
<td>b5</td>
<td>-2.2280</td>
<td>0.1026</td>
<td>1154</td>
<td>-21.71</td>
<td>&lt; .0001</td>
<td>7.15E-6</td>
<td></td>
</tr>
<tr>
<td>b6</td>
<td>-2.1304</td>
<td>0.1004</td>
<td>1154</td>
<td>-21.21</td>
<td>&lt; .0001</td>
<td>0.00039</td>
<td></td>
</tr>
<tr>
<td>b7</td>
<td>0.6455</td>
<td>0.08044</td>
<td>1154</td>
<td>8.02</td>
<td>&lt; .0001</td>
<td>-6.66E-6</td>
<td></td>
</tr>
<tr>
<td>b8</td>
<td>0.6601</td>
<td>0.08053</td>
<td>1154</td>
<td>8.20</td>
<td>&lt; .0001</td>
<td>3.56E-6</td>
<td></td>
</tr>
<tr>
<td>b9</td>
<td>-1.6749</td>
<td>0.09176</td>
<td>1154</td>
<td>-18.25</td>
<td>&lt; .0001</td>
<td>4.163E-6</td>
<td></td>
</tr>
<tr>
<td>b10</td>
<td>1.0332</td>
<td>0.08342</td>
<td>1154</td>
<td>12.39</td>
<td>&lt; .0001</td>
<td>0.000015</td>
<td></td>
</tr>
<tr>
<td>std</td>
<td>1.3303</td>
<td>0.04981</td>
<td>1154</td>
<td>26.71</td>
<td>&lt; .0001</td>
<td>2.979E-6</td>
<td></td>
</tr>
<tr>
<td>var(theta)</td>
<td>1.7697</td>
<td>0.1325</td>
<td>1154</td>
<td>13.35</td>
<td>&lt; .0001</td>
<td>* n.a.</td>
<td></td>
</tr>
</tbody>
</table>
1pl: Item Curves

Rasch (1PL) fit to 10 Vocabulary Items

- Fitted Probability of Correct
- Theta

C.J. Anderson (Illinois)
2 Parameter Logistic Model

- For 2PL, we allow different slope (discrimination parameter) for each item.
- This is a generalization **non-linear** mixed model.
- Change the model for Level 1 intercept to

  \[
  \beta_{0j} = (a_1 x_{1ij} + a_2 x_{2ij} + \ldots + a_{10} x_{10ij}) U_{0j}
  \]

- Model for item \(i\) and person \(j\) becomes

  \[
  P(Y_{ij} = 1) = \frac{\exp(\gamma_{i0} + a_i U_{0j})}{(1 + \exp(\gamma_{i0} + a_i U_{0j}))} = \frac{\exp(b_i + a_i \theta_j)}{(1 + \exp(b_i + a_i \theta_j))}
  \]
Example of 2PL: 10 vocabulary items

NOTE: GCONV convergence criterion satisfied.

Fit Statistics

-2 Log Likelihood 10084
AIC (smaller is better) 10124
AICC (smaller is better) 10124
BIC (smaller is better) 10225

\[ LR = -2(\text{LogLike}_{1\text{PL}} - \text{LogLike}_{2\text{PL}}) \]
\[ = 10197 - 10084 = 113 \]

\[ df = 9, \ p < .01. \]
### Parameter Estimates

| Parameter | Estimate | Standard Error | DF  | Value | Pr > |t| | Gradient |
|-----------|----------|----------------|-----|-------|------|---|----------|
| b1        | -1.8619  | 0.1064         | 1154| -17.49| < .0001 | -3.73E-6 |
| b2        | -5.1574  | 0.5489         | 1154| -9.40  | < .0001 | 5.538E-7  |
| b3        | 1.2855   | 0.08804        | 1154| 14.60  | < .0001 | -2.29E-6  |
| b4        | -6.9405  | 1.1282         | 1154| -6.15  | < .0001 | 6.59E-7   |
| b5        | -2.4447  | 0.1668         | 1154| -14.66 | < .0001 | 1.077E-7  |
| b6        | -2.5777  | 0.1965         | 1154| -13.12 | < .0001 | 8.384E-7  |
| b7        | 0.5891   | 0.07594        | 1154| 7.76   | < .0001 | 1.082E-6  |
| b8        | 0.6351   | 0.08126        | 1154| 7.82   | < .0001 | 3.174E-6  |
| b9        | -1.4837  | 0.09179        | 1154| -16.16 | < .0001 | 2.323E-6  |
| b10       | 1.1351   | 0.1071         | 1154| 10.60  | < .0001 | 2.414E-6  |

Looks good so far...
### $a_i$ Parameter Estimates

| Parameter | Estimate | Standard Error | DF  | Value | Pr $>|t|$ | Gradient |
|-----------|----------|----------------|-----|-------|-----------|----------|
| a1        | 0.9406   | 0.1213         | 1154| 7.75  | $< .0001$| 3.191E-6 |
| a2        | 2.8594   | 0.4183         | 1154| 6.84  | $< .0001$| 1.015E-7 |
| a3        | 0.9606   | 0.1117         | 1154| 8.60  | $< .0001$| -3.01E-6 |
| a4        | 3.9334   | 0.7862         | 1154| 5.00  | $< .0001$| 1.361E-6 |
| a5        | 1.6386   | 0.1821         | 1154| 9.00  | $< .0001$| -1.18E-6 |
| a6        | 1.9667   | 0.2202         | 1154| 8.93  | $< .0001$| 1.169E-6 |
| a7        | 1.0491   | 0.1094         | 1154| 9.59  | $< .0001$| -4.48E-6 |
| a8        | 1.2324   | 0.1247         | 1154| 9.88  | $< .0001$| -3.34E-6 |
| a9        | 0.8965   | 0.1111         | 1154| 8.07  | $< .0001$| -2.82E-6 |
| a10       | 1.6562   | 0.1697         | 1154| 9.76  | $< .0001$| -6.75E-7 |

All $a_i$ are significant for the hypothesis $H_0: a_i = 0$, but there are other hypotheses that we may be interested in (e.g., $H_0: a_1 = a_2$, or $H_0: a_i = 1$).
Parameter Estimates

To test whether 1 PL is sufficient, we could perform 10 tests of $H_{oi} : a_i = 1$, so do one likelihood ratio test.

- 1PL is a special case of 2PL where $H_o : a_1 = a_2 = \ldots = a_{10} = 1$.
- Likelihood ratio test

$$LR = -2(\text{LogLike}_{1PL} - \text{LogLike}_{2PL})$$

$$= 10197 - 10084 = 113$$

$df = 9$, $p < .01$.

- Data support conclusion that 2PL is the better model (i.e, slopes differ).
2pl: Item Curves

2PL Item Characteristic Curves for 10 Vocabulary Items
Often we wish to use $\theta$ as a response variable in a regression model (i.e., see what influences or explains variability in $\theta$).

This strategy is problematic because it introduces additional error — error due to estimation of $\theta$.

Solution: Put predictors of $\theta$ into the IRT model, e.g.,

$$\beta_{0j} = \theta_j = \nu_1 \text{age}_j + \nu_2 \text{HS degree}_j + \nu_3 \text{Primary}_j + e_j.$$ 

Can also put predictor for difficulty parameters into the IRT model (any IRT model).
## 2PL with Predictors of Vocabulary

| Parm | Est.   | Error  | DF   | Value | Pr > | Pr > $|t|$ | Gradient |
|------|--------|--------|------|-------|-------|--------|--------|----------|
| b1   | -0.3211| 0.2001 | 1154 | -1.60 | 0.1089| 0.000058|
| b2   | -0.4985| 0.4330 | 1154 | -1.15 | 0.2498| -0.00002|
| b3   | 3.0137 | 0.2804 | 1154 | 10.75 | < .0001| 0.00001|
| b4   | -0.6312| 0.5071 | 1154 | -1.24 | 0.2135| 0.000036|
| b5   | 0.2717 | 0.2954 | 1154 | 0.92  | 0.3579| 0.000136|
| b6   | 0.6106 | 0.3291 | 1154 | 1.86  | 0.0638| 0.000021|
| b7   | 2.4326 | 0.2651 | 1154 | 9.18  | < .0001| -0.00003|
| b8   | 2.7968 | 0.2995 | 1154 | 9.34  | < .0001| -0.00005|
| b9   | -0.0960| 0.1836 | 1154 | -0.52 | 0.6013| -0.00013|
| b10  | 3.6980 | 0.3584 | 1154 | 10.32 | < .0001| -1.57E-6|
## 2PL with Predictors of Vocabulary

<table>
<thead>
<tr>
<th>Parm</th>
<th>Est.</th>
<th>Error</th>
<th>DF</th>
<th>Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
<th>Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>0.8166</td>
<td>0.1033</td>
<td>1154</td>
<td>7.91</td>
<td>&lt; .0001</td>
<td>0.000147</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a2</td>
<td>2.4461</td>
<td>0.3552</td>
<td>1154</td>
<td>6.89</td>
<td>&lt; .0001</td>
<td>-0.00003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a3</td>
<td>0.9020</td>
<td>0.0990</td>
<td>1154</td>
<td>9.11</td>
<td>&lt; .0001</td>
<td>-2.01E-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a4</td>
<td>2.9506</td>
<td>0.5177</td>
<td>1154</td>
<td>5.70</td>
<td>&lt; .0001</td>
<td>-0.00007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a5</td>
<td>1.4529</td>
<td>0.1573</td>
<td>1154</td>
<td>9.23</td>
<td>&lt; .0001</td>
<td>6.616E-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a6</td>
<td>1.6744</td>
<td>0.1825</td>
<td>1154</td>
<td>9.18</td>
<td>&lt; .0001</td>
<td>0.000026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a7</td>
<td>0.9712</td>
<td>0.0968</td>
<td>1154</td>
<td>10.03</td>
<td>&lt; .0001</td>
<td>0.000011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a8</td>
<td>1.1382</td>
<td>0.1099</td>
<td>1154</td>
<td>10.36</td>
<td>&lt; .0001</td>
<td>-0.00001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a9</td>
<td>0.7217</td>
<td>0.0924</td>
<td>1154</td>
<td>7.81</td>
<td>&lt; .0001</td>
<td>-0.00012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a10</td>
<td>1.3724</td>
<td>0.1375</td>
<td>1154</td>
<td>9.98</td>
<td>&lt; .0001</td>
<td>-0.00002</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2PL with Predictors of Vocabulary

<table>
<thead>
<tr>
<th>Parm</th>
<th>Est.</th>
<th>Error</th>
<th>DF</th>
<th>Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_1$ (age)</td>
<td>0.0154</td>
<td>0.0024</td>
<td>1154</td>
<td>6.47</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu_2$ (elementary)</td>
<td>0.9891</td>
<td>0.1310</td>
<td>1154</td>
<td>7.55</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu_3$ (high school)</td>
<td>1.7075</td>
<td>0.1412</td>
<td>1154</td>
<td>12.09</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parm Gradient

$\nu_1$ (age) -0.00014
$\nu_2$ (elementary) -0.00008
$\nu_3$ (high school) -0.0001

Interpretation?
Figure of Item Curves
Summary of Models for GSS Vocabulary

<table>
<thead>
<tr>
<th>Model</th>
<th>Number parameters</th>
<th>$-2\text{(LogLike)}$</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rasch (1PL)</td>
<td>11</td>
<td>10197</td>
<td>10219</td>
<td>10275</td>
</tr>
<tr>
<td>2PL</td>
<td>20</td>
<td>10084</td>
<td>10124</td>
<td>10225</td>
</tr>
<tr>
<td>2PL covariates $\theta$</td>
<td>23</td>
<td>9871</td>
<td>9917</td>
<td>10033</td>
</tr>
<tr>
<td>2PL drop age</td>
<td>22</td>
<td>9914</td>
<td>9958</td>
<td>10069</td>
</tr>
</tbody>
</table>

Improvement due to addition of 3 predictors in 2PL model:

$$LR = 10084 - 9871 = 213$$

compare to $\chi^2_3$, $p$-value is “vanishingly small”.
Concluding Comments

- The tests for fixed and random effects are the same as they were in HLM.
- Regression diagnostics that were used in HLM really don’t work well with logistic regression (whether multilevel or not). This is an area of active research.
- $R^2$ concept really doesn’t apply to logistic regression.
- Three levels models can be fit to multilevel models for different types of response variables.
- For all 3-level MLM models, I switched from quad to LaPlace.
If you Have Clustered Discrete Data . . .

(Figure idea from Molenberghs & Verbeke (2004))
Some Things I Didn’t have Time for

Hierarchical models (in general)

- Other cases of GLMMs (e.g., counts, skewed, multi-category (nominal), ordinal, rating data, . . .)
- Computing Power
- More model diagnostics and assumption checking.
- Other kinds of applications (go back to introduction).
- Software alternatives.
- Lots of other things.