Multilevel Logistic Regression

Edps/Psych/Soc 587

Carolyn J. Anderson

Department of Educational Psychology

ILLINOIS

Spring 2020

Outline

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
- Molenberghs, G. & Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer.
- Agresti, A. (2013). *Categorical Data Analsysi*, 3rd Edition. Wiley.
- Agresti, A. (2019). *Introduction to Categorical Data Analysis*, 3rd edition. Wiley. (included R and SAS code).

More References

- Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling*. NY: Chapman & Hall/CRC.
- de Boeck, P. & Wilson, M. (editors) (2004). *Explanatory Item Response Models*. Springer.
- Molenbergahs, G. & Verbeke, G. (2004). An introduction to (Generalized Non) Linear Mixed Models, Chapter 3, pp 11-148. In de Boeck, P. & Wilson, M. (Eds.) *Explanatory Item Response Models*. Springer.

Data

- Clustered, nested, hierarchial, 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

- From Skrondal & Rabe-Hesketh (2004) also analyzed by Zeger & Karim (1991), Diggle et. al (2002), but originally from Sommer et al (1983)
- Preschool children from Indonesia who were examined up to 6 consecutive quarters for respiratory infection.
- Predictors/explanatory/covariates:
  - Age in months
  - Xeropthalmia as indicator of chronic vitamin A deficiency (dummy variable)— night blindness & dryness of membranes $\rightarrow$ dryness of cornea $\rightarrow$ 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

- From Agresti (2002) who got it from Koch et al (1977)
- Comparison of new drug with a standard drug for treating depression.
- Classified as N= Normal and A= Abnormal at 1, 2 and 4 weeks.

Diagnosis	R_x	Response at Each of 3 Time Points							
		NNN	NNA	NAN	NAA	ANN	ANA	AAN	AAA
Mild	std	16	13	9	3	14	4	15	6
	new	31	0	6	0	22	2	9	0
Severe	std	2	2	8	9	9	15	27	28
	new	7	2	5	2	31	5	32	6

“Cool” Kids

- Rodkin, P.C., Farmer, T.W, Pearl, R. & Acker, R.V. (2006). They’re cool: social status and peer group for aggressive boys and girls. *Social Development, 15*, 175–204.
- 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

LSAT6

Law School Admissions data: 5 items, $N = 1000$

Y_1	Y_2	Y_3	Y_4	Y_5	Freqency
1	1	1	1	1	3
1	1	1	1	2	6
1	1	1	2	1	2
1	1	1	2	2	11
1	1	2	1	1	1
1	1	2	1	2	1
1	1	2	2	1	3
⋮	⋮	⋮	⋮	⋮	
2	2	2	1	2	61
2	2	2	2	1	28
2	2	2	2	2	298

General Social Survey

- Data are responses to 10 vocabulary items from the 2004 General Social Survey from $n = 1155$ respondents.
- 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	1	1	0	1	1	0	1	0
34	17	3	2	1	1	1	1	1	0	0	1	1	1	0	0	1
26	14	2	1	1	1	0	1	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	0
29	11	1	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 interesting 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 π
- The probability that $X = 0$ is $(1 - \pi)$

Such variables are called Bernoulli random variables.

Bernoulli Random Variable

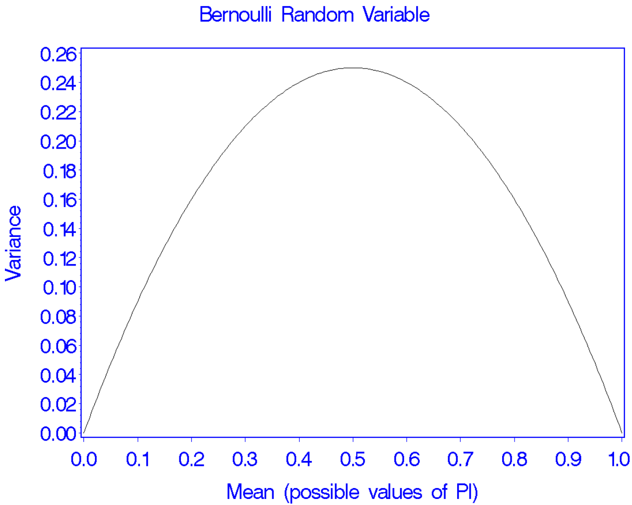
The mean of a Bernoulli random variable is

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

The variance of X is

$$\begin{aligned} \text{var}(X) = \sigma^2_X &= E[(X-\mu_X)^2] \\ &= (1-\pi)^2\pi + (0-\pi)^2(1-\pi) \\ &= \pi(1-\pi) \end{aligned}$$

Bernoulli Variance vs Mean



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 π, 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

$$\begin{aligned}
\mu_y = E(Y) &= E(\sum_{i=1}^{n} X_i) \\
&= E(X_1) + E(X_2) + \ldots + E(X_n) \\
&= \overbrace{\mu_x + \mu_x + \ldots + \mu_x}^{n} \\
&= \overbrace{\pi + \pi + \ldots + \pi}^{n} \\
&= n\pi
\end{aligned}$$

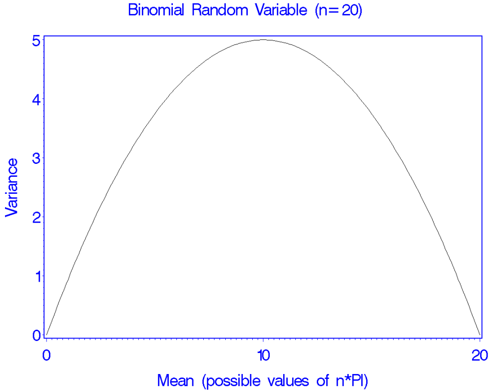
Variance of Binomial Random Variable

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

$$\begin{aligned}
\text{var}(Y) = \sigma_y^2 &= \text{var}(X_1 + X_2 + \ldots + X_n) \\
&= \overbrace{\text{var}(X) + \text{var}(X) + \ldots + \text{var}(X)}^{n} \\
&= \overbrace{\pi(1-\pi) + \pi(1-\pi) + \ldots + \pi(1-\pi)}^{n} \\
&= n\pi(1-\pi)
\end{aligned}$$

Note: Once you know π and n, you know the mean and variance of the Binomial distribution.

Variance vs Mean



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.

Possible Outcomes $X_1 + X_2 + X_3 = Y$	Probability of a Sequence $P(X_1, X_2, X_3)$	Prob(Y) $P(Y)$
$1+1+1=3$	$(.4)(.4)(.4) = (.4)^3(.6)^0 = .064$	$.064$
$1+1+0=2$	$(.4)(.4)(.6) = (.4)^2(.6)^1 = .096$	
$1+0+1=2$	$(.4)(.6)(.4) = (.4)^2(.6)^1 = .096$	$3(.096) = .288$
$0+1+1=2$	$(.6)(.4)(.4) = (.4)^2(.6)^1 = .096$	
$1+0+0=1$	$(.4)(.6)(.6) = (.4)^1(.6)^2 = .144$	
$0+1+0=1$	$(.6)(.4)(.6) = (.4)^1(.6)^2 = .144$	$3(.144) = .432$
$0+0+1=1$	$(.6)(.6)(.4) = (.4)^1(.6)^2 = .144$	
$0+0+0=0$	$(.6)(.6)(.6) = (.4)^0(.6)^3 = .216$	$.216$
	1.000	1.000

Binomial Distribution Function

The formula for the probability of a Binomial random variable is

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

where

$$\left(\begin{array}{c} n \\ a \end{array}\right) = \frac{n!}{a!(n-a)!} = \frac{n(n-1)(n-2)\dots 1}{a(a-1)\dots 1((n-a)(n-a-1)\dots 1)}$$

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:

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

where i is index of level 1 and j is index of level 2.

The Link Function:

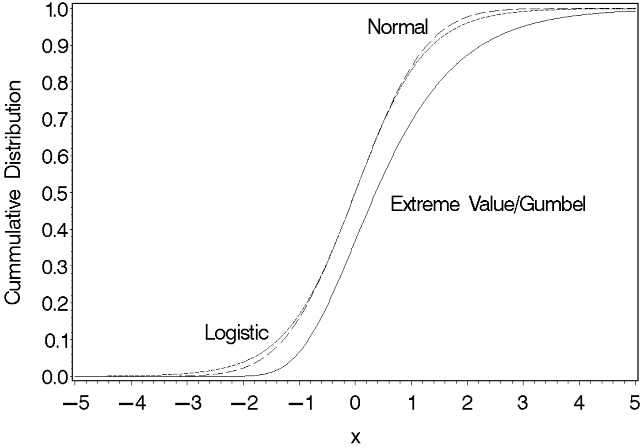
Problem:

- Probabilities must be between 0 and 1.
- η_i could be between $-\infty$ to ∞.

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 $\longrightarrow$ "probit" link
  - Logistic $\longrightarrow$ "logit" link
  - Gumbel (extreme value) $\longrightarrow$ Complementary log-log link $\log[-\log(1-\pi)]$

Some Example cdf's



Putting All the Components Together

$$\begin{aligned}\log\left(\frac{P(Y_i=1|\boldsymbol{x}_i)}{P(Y_i=0|\boldsymbol{x}_i)}\right) &= \text{logit}(P(Y_i=1|\boldsymbol{x}_i)) \\ &= \beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\ldots+\beta_K x_{Ki}\end{aligned}$$

where $\boldsymbol{x}_i=(x_{0i},x_{1i},\ldots,x_{Ki})$.

or in-terms of probabilities

$$\begin{aligned}E(Y_i|\boldsymbol{x}_i) &= P(Y_i=1|\boldsymbol{x}_i) \\ &= \frac{\exp[\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\ldots+\beta_K x_{Ki}]}{1+\exp[\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\ldots+\beta_K x_{Ki}]}\end{aligned}$$

Implicit assumption (for identification):

For $P(Y_i=0|\boldsymbol{x}_i)$: $\beta_0=\beta_1=\ldots=\beta_K=0$.

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

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

Example 1: Respiratory Data

One with a continuous explanatory variable (for now)

- Response variable
  - Y = 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

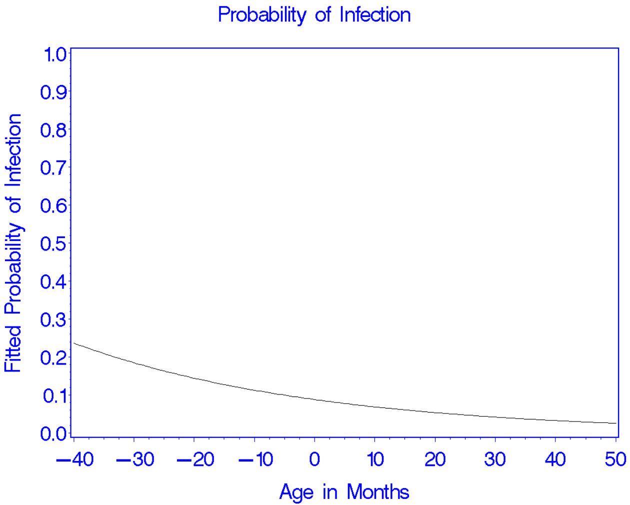
Analysis Of Parameter Estimates

Parameter	Estimate	Standard Error	95% Conf. Limits		Chi-Square	Pr > ChiSq
Intercept	-2.3436	0.1053	-2.55	-2.14	495.34	$< .0001$
age	-0.0248	0.0056	-0.04	-0.01	19.90	$< .0001$

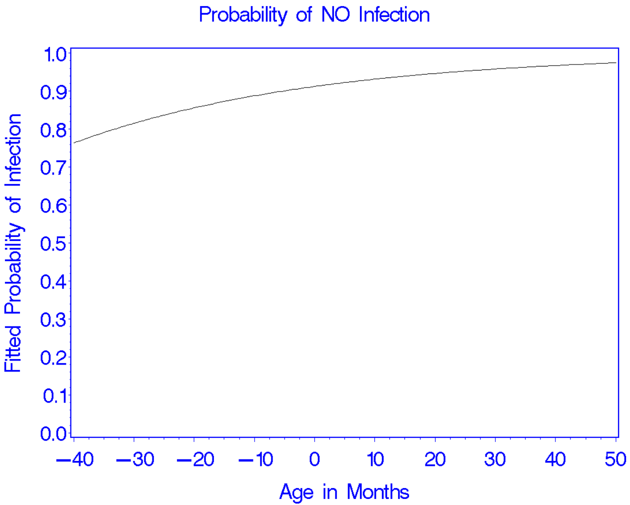
Interpretation: The odds of an infection equals $\exp(-.0248) = 0.98$ times that for a person one year younger.

OR The odds of *no* infection equals $\exp(0.0248) = 1/.98 = 1.03$ times the odds for a person one year older.

Probability of Infection



Probability of NO infection



Example 2: Longitudinal Depression Data

- From Agresti (2002) who got it from Koch et al (1977)
- 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).

Parameter	DF	Estimate	$\exp \hat{\beta}$	Std. Error	X^2	Pr> χ^2_1
Diagnose	1	-1.3139	0.27	0.1464	80.53	$< .0001$
Drug	1	-0.0596	0.94	0.2222	0.07	0.7885
Time	1	0.4824	1.62	0.1148	17.67	$< .0001$
Drug*Time	1	1.0174	2.77	0.1888	29.04	$< .0001$

- The odds of normal when diagnosis is severe is 0.27 times the odds when diagnosis is mild (or $1/.27 = 3.72$).
- 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"

$$P(Y_{ij} = 1|\boldsymbol{x}_{ij}) = \frac{\exp(\beta_0 + \beta_1 x_{1ij} + \ldots + \beta_K x_{Kij})}{1 + \exp(\beta_0 + \beta_1 x_{1ij} + \ldots + \beta_K x_{Kij})}$$

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

- Random Effects: "subject-specific"

$$P(Y_{ij} = 1|\boldsymbol{x}_{ij}, \boldsymbol{U}_j) = \frac{\exp(\beta_{0j} + \beta_{1j} x_{1ij} + \ldots + \beta_{Kj} x_{Kij})}{1 + \exp(\beta_{0j} + \beta_{1j} x_{1ij} + \ldots + \beta_{Kj} x_{Kij})}$$

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

$$P(Y_{ij} = 1|\boldsymbol{x}_{ij}) = \int_{U_0} \frac{\exp(\gamma_{00} + \gamma_{10} x_{1ij} + \ldots + \gamma_{K0} x_{Kij} + U_0)}{1 + \exp(\gamma_{00} + \gamma_{10} x_{1ij} + \ldots + \gamma_{K0} x_{Kij} + U_0)} f(U_0) dU_0$$

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 $\epsilon_{ij} \sim \mathcal{N}(0, .01)$.
- $U_O \sim \mathcal{N}(0, 4) i.i.d.$
- x_i^*, ϵ_{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

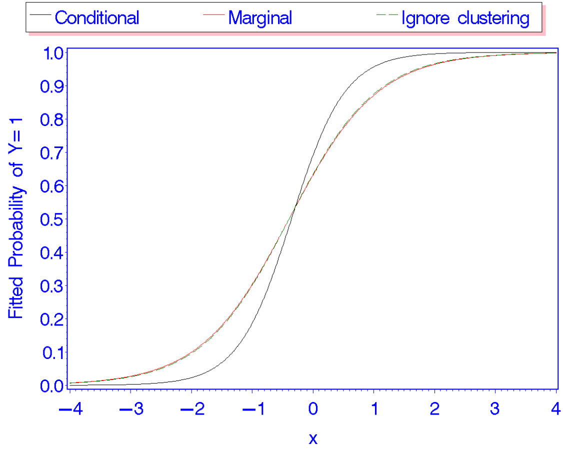
	MLE Ignoring clustering		GEE (exchangeable)		MLE Random Effects	
Parameter	Estimate	Std Error	Estimate	Std Error	Estimate	Std Error
Intercept	0.545	0.205	0.535	0.314	0.801	0.483
x	1.396	0.206	1.370	0.293	2.278	0.543

From GEE: correlation $= .42$

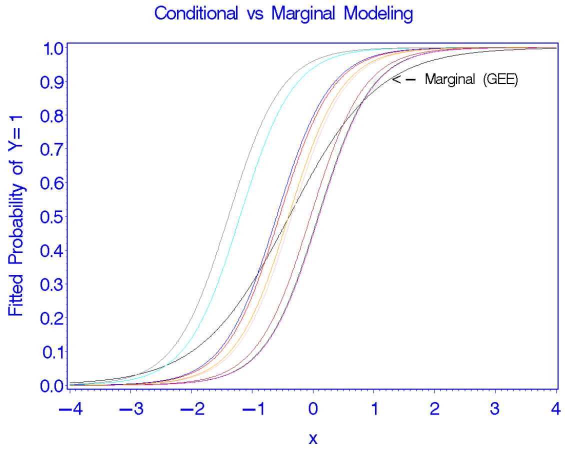
From Random effects :$\sqrt{\hat{\tau}_o^2} = 2.1636$ (s.e. $= .6018$) and $\hat{\tau}_o^2 = 4.6811$

What do you notice?

Simulation: Fitted Values



Conditional vs Marginal Models



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 (& 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 τ_0^2 and larger correlation within cluster), the greater the difference.

Population Averaged Model

- Have repeated measures data or nested data $\longrightarrow$ 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 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 requires (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

	Initial	Exchangeable	Unstructured
Intercept	-0.0280 (0.1639)	-0.0281 (0.1742)	-0.0255 (0.1726)
diagnose	-1.3139 (0.1464)	-1.3139 (0.1460)	-1.3048 (0.1450)
treat	-0.0596 (0.2222)	-0.0593 (0.2286)	-0.0543 (0.2271)
time	0.4824 (0.1148)	0.4825 (0.1199)	0.4758 (0.1190)
treat*time	1.0174 (0.1888)	1.0172 (0.1877)	1.0129 (0.1865)

Working correlation for exchangeable $= -.0034$

Correlation Matrix for Unstructured:

Working Correlation Matrix

	Col1	Col2	Col3
Row1	1.0000	0.0747	-0.0277
Row2	0.0747	1.0000	-0.0573
Row3	-0.0277	-0.0573	1.0000

(Interpretation the same as when we ignored clustering.)

SAS and GEE

```
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

```
title 'GEE with AR(1)';
repeated subject=case / type=AR(1) corrw;
```

```
title 'GEE with Unstructured';
repeated subject=case / type=unstr corrw;
```

R and GEE

Input:

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

Output:

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	-0.0280	0.1625	-0.1728	0.1741	-0.1613
severe	-1.3139	0.1448	-9.0700	0.1459	-9.0016
Rx	-0.0592	0.2205	-0.2687	0.2285	-0.2593
time	0.4824	0.1141	4.2278	0.1199	4.0226
Rx:time	1.0171	0.1877	5.4191	0.1877	5.4192

Estimated Scale Parameter: 0.985392

R and GEE

Working Correlation

	[,1]	[,2]	[,3]
[1,]	1.00000	-0.00343	-0.0034
[2,]	-0.00343	1.00000	-0.0034
[3,]	-0.00343	-0.00343	1.0000

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	Standard Error	95% Confidence Limits		Z	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

Source	DF	Chi-Square	Pr$>$ ChiSq
age	1	18.24	$<.0001$

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
Empirical Standard Error 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	0.62	1.86	0.44	1.41	$.16$
xero	0	0.00	1.00	0.00	.	.
female	1	-0.42	0.66	0.24	-1.77	$.08$
female	0	0.00	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	0.15	1.16	0.41	0.36	$.72$
stunted	0	0.00	1.00	0.00	.	.

Miscellaneous Comments on Marginal Models

- With GEE
  - There is no likelihood being maximized $\Longrightarrow$ 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 $\boldsymbol{x}_{ij}$ and $\boldsymbol{U}_j$:

$$P(Y_{ij} = 1|\boldsymbol{x}_{ij}, \boldsymbol{U}_j) = \frac{\exp[\beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \ldots + \beta_{Kj}x_{Kij}]}{1 + \exp[\beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \ldots + \beta_{Kj}x_{Kij}]}$$

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

- Level 2: Model for intercept and slopes:

$$\begin{array}{rcl} \beta_{0j} & = & \gamma_{00} + U_{0j} \\ \beta_{1j} & = & \gamma_{10} + \ldots + U_{1j} \\ \vdots & \vdots & \vdots \\ \beta_{Kj} & = & \gamma_{K0} + U_{Kj} \end{array}$$

Putting Levels 1 & 2 Together

$$P(Y_{ij}=1|\boldsymbol{x}_{ij},\boldsymbol{U}_j)=\frac{\exp[\gamma_{00}+\gamma_1 x_{1ij}+\ldots+\gamma_K x_{Kij}+U_{0j}+\ldots+U_{KJ}x_{KJ}]}{1+\exp[\gamma_0+\gamma_1 x_{1ij}+\ldots+\gamma_K x_{Kij}+U_{0j}+\ldots+U_{KJ}x_{KJ}]}$$

Marginalizing gives us the Marginal Model...

$$P(Y_{ij}=1|\boldsymbol{x}_{ij})=\int_{U_0}\ldots\int_{U_K}\frac{\exp(\gamma_{00}+\gamma_{10}x_{1ij}+\ldots U_0+\ldots+U_K x_{Jij})}{1+\exp(\gamma_{00}+\gamma_{10}x_{1ij}\ldots+U_0+\ldots+U_K x_{Jij})}f(\boldsymbol{U})d\boldsymbol{U}$$

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:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + U_{0j} \\ \beta_{1j} &= \gamma_{01} \end{aligned}$$

  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

Dimensions	
Observations Used	1200
Observations Not Used	0
Total Observations	1200
Subjects	275
Max Obs Per Subject	6
Parameters	3
Quadrature Points	10

Example 1: Input and Iteration History

Parameters

lam	bAge	sigma	NegLogLike
-2.3	0.02	0.8	380.400779

Iteration History

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

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	Standard Error	DF	t 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

Additional Estimates

Label	Est	Standard Error	DF	t Value	$\Pr > \lvert t \rvert$
Var(Uo)	0.7273	0.356	0 274	2.04	0.0420
odds ratio	0.9736	0.00641	8 274	151.70	$< .0001$

Alpha	Lower	Upper
0.05	0.02658	1.4281
0.05	0.9610	0.9862

I requested these using an "Estimate" statement in SAS/NLMIXED.

R for Example 1: A Simple Random Intercept Model

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

	mod1.quad	
(Intercept)	-2.61^{***}	(0.17)
age	-0.03^{***}	(0.01)
AIC	698.56	
BIC	713.83	
Log Likelihood	-346.28	
Num. obs.	1200	
Num. groups: id	275	
Var: id (Intercept)	0.73	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

R for Example 1: A Simple Random Intercept Model

Some other useful things:

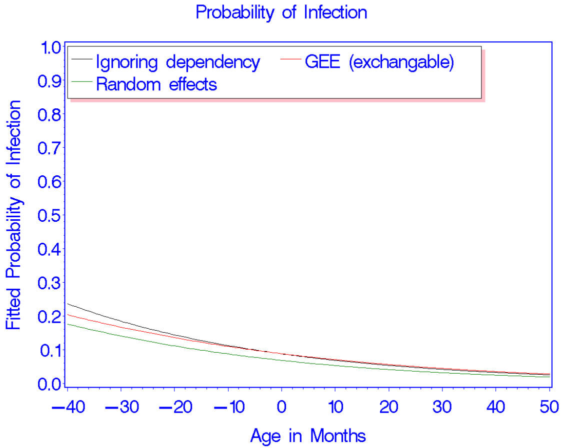
For profile confidence intervals of effects:

```
round(confint(mod1.quad, level=.95),digits=4)
                 2.5%    97.5%
 .sig01         0.3993   1.2727
 (Intercept)   -2.9887  -2.3059
 age           -0.0403  -0.0142
```

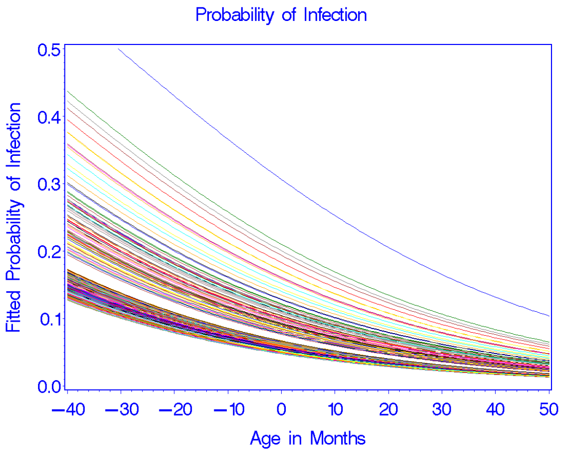
For odds ratios

```
odds ← exp(fixef(mod1.quad))
round(odds,digits=2)
 (Intercept)   age
     0.07      0.97
```

Example 1: Estimated Probabilities



Example 1: Estimated Probabilities



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;
```

R and glmer

```
# Default is LaPlace

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


# Gauss quadrature – MLE gold standard
#(nAGQ= adaptive gauss quadrature and nAGQ=numnber quadrature
points)

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

Different Estimation Methods → Different Results

Some GLIMMIX Estimation Options

	MMPL		RMPL		RSPL	
Param	est	s.e.	est	s.e.	est	s.e.
$\hat{\gamma}_{00}$	-2.3381	(0.1163)	-2.3379	(0.1167)	-2.3722	(0.1160)
$\hat{\gamma}_{01}$	-0.0254	(0.0061)	-0.0254	(0.0061)	-0.0249	(0.0061)
$\hat{\tau}_0^2$	0.5734	(0.2775)	0.5967	(0.2810)	0.4996	(0.2292)

	GLIMMIX				NLMIXED	
	LaPlace		quad		gauss	
Param	est	s.e.	est	s.e.	est	s.e.
$\hat{\gamma}_{00}$	-2.6768	(0.1844)	-2.6129	(0.1723)	-2.6130	(0.1723)
$\hat{\gamma}_{01}$	-0.0267	(0.0067)	-0.0268	(0.0066)	-0.0268	(0.0066)
$\hat{\tau}_0^2$	0.8950	(0.3961)	0.7272	(0.3559)	0.7273	(0.3560)

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

Appox. integrand	Approx. integral
Fits as wider range of models (e.g., $3+$ levels, more than random intercept)	Narrower range of models (only 2 level for QUAD, but more with LaPlace)
Estimation approximates the integrand, "pseudo-likelihood"	Estimation uses numerical integration (Gaussian or adaptive quadrature)
Parameter estimates are downward	Parameter estimates aren't biased
Estimation can be very poor for small n per macro-unit	Estimation can be fine for small n per macro-unit
Faster	Slower
Easier to use (pertains to SAS)	Harder to use
No LR testings	This is MLE

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,

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

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 $U_{0j} = 0$

$$\exp[\hat{\gamma}_{00} - 1.96(s.e.)], \exp[\hat{\gamma}_{00} + 1.96(s.e.)] \longrightarrow (.63, .65).$$

- The 95% of the estimated variability in odds over classrooms equals

$$\exp[\hat{\gamma}_{00} - 1.96\sqrt{\hat{\tau}_{00}}], \exp[\hat{\gamma}_{00} - 1.96\sqrt{\hat{\tau}_{00}}] \longrightarrow (0.01, 8.93).$$

What does this imply?

and Probability Estimates

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

- 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

$$(\text{logit}^{-1}(.63), \text{logit}^{-1}(.65)) \longrightarrow (.28, .52)$$

- 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: $y_{ij} \sim$ Bionmial(π_{ij}, n^*_{ij}) where

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

Level 2:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + U_{0j} \quad \text{where} \quad \underline{U}_{0j} \sim N(0, \tau_{00}) \\ \beta_{1j} &= \gamma_{10} \end{aligned}$$

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

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

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

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.

Esttimated Probabilities within a Cluster

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

			Random Classroom Effect		
Popular	Gender	Race	$U_{0j} = -2$	$U_{0j} = 0$	$U_{0j} = 2$
Yes	female	white	27.50	23.56	18.26
		black	9.88	12.98	17.32
	male	white	14.94	16.36	16.22
		black	3.56	5.25	9.38
No	female	white	26.24	21.47	16.06
		black	3.57	4.57	5.81
	male	white	12.52	13.22	12.53
		black	1.79	2.59	4.42

Random Intercept with Predictors

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:

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

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

For interpretation:

$$\frac{(\pi_{ij}|U_{0j})}{1-(\pi_{ij}|U_{0j})} = \exp\left[\gamma_{00} + \gamma_{10}\text{Popularity}_{ij} + \gamma_{20}\text{Gender}_{ij} + \gamma_3\text{Race}_{ij} + \gamma_{01}\text{ClassAggress}_j + U_{0j}\right]$$

Results

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Intercept	CLASSID	1.9907	0.6365

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr $> \|t\|$
Intercept	0.2392	0.2805	54	0.85	0.3976
DPOP	0.07219	0.2101	431	0.34	0.7314
SEX	-0.6366	0.2138	431	-2.98	0.0031
black	-1.0709	0.3268	431	-3.28	0.0011
classAgg	-0.6390	0.2374	54	-2.69	0.0095

Results and Interpretation

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

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.

. . . 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".

Results and Interpretation (continued)

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 $\exp[-0.6390] = .52$.

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 τ_{00} to see what this now equals give that we have both Level 1 and Level 2 predictors in the model using

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

where $\pi = 3.141593\ldots$ (i.e.,"pi" and not probability).
For three random intercept models we have fit so far:

Model	$\hat{\tau}_0^2$	ICC
Null/Empty	2.9903	.48
+Popularity + Gender + Minority	2.3129	.41
+ Class Aggression	1.9907	.38

Random Intercept and Slope

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

$$\begin{aligned} \frac{\pi_{ij}}{1-\pi_{ij}} &= \exp\left[\beta_{0j} + \beta_{1j}\text{Popularity}_{ij} + \beta_{2j}\text{Gender}_{ij}\right. \\ &\quad \left. + \beta_{3j}\text{Race}_{ij}\right] \end{aligned}$$

Level 2:

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

To help interpretation:

$$\begin{aligned} \frac{(\pi_{ij}|U_{0j})}{1-(\pi_{ij}|U_{0j})} &= \exp\left[\gamma_{00} + \gamma_{10}\text{Popularity}_{ij} + \gamma_{20}\text{Gender}_{ij} + \gamma_{3}\text{Race}_{ij}\right. \\ &\quad \left. + \gamma_{01}\text{ClassAggress}_j + U_{0j} + U_{2j}\text{Gender}_{ij}\right] \end{aligned}$$

Results and Comparisons

	Empty		Model 2		Model 3		Model 4	
Effects	Est.	s.e.	Est.	s.e.	Est.	s.e.	Est.	s.e.
Intercept	−0.44	(0.26)	0.31	(0.29)	0.24	(0.28)	0.26	(0.36)
Popularity			0.10	(0.21)	0.07	(0.21)	−0.10	(0.24)
Gender			−0.65	(0.21)	−0.64	(0.21)	−0.48	(0.42)
Race			−1.30	(0.33)	−1.07	(0.33)	−1.14	(0.36)
ClassAgg					−0.64	(0.24)	−0.71	(0.27)
τ_{00}	2.99	(0.85)	2.31	(0.73)	1.99	(0.64)	3.74	(1.41)
τ_{10}							−2.80	(1.42)
τ_{11} gender							4.87	(1.96)
# param	2		5		6		8	
−2lnLike	720.75		694.38		687.07		656.20	
AIC	724.75		704.38		699.07		672.20	
BIC	728.80		714.51		711.23		688.41	

Some Model Refinements

- Popularity is clearly not significant with a $t = -.10/0.24 = -0.42$.
- Gender is no longer significant with a $t = -.48/0.42 = -1.16$. Should we drop gender?
- Test $H_o : \tau_1^2 = \tau_{01} = 0$ versus H_a: not H_o. Use same method as we did for HLM: compute LR^* and compare to a mixture of chi-square distributions.

Model	-2lnLike	LR^*	p-value χ_2	p-value χ_1	p-value for test
Null	687.07				
H_a	656.20	30.87	tiny	tiny	$< .01$

- Drop Popularity from the model but keep random Gender effect.

Results and Comparisons

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

Comments on Results

- A likelihood ratio test for popularity,

$$LR = 656.38 - 656.20 = 0.18$$

  compared to a χ^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 $y_{ijk} \sim$ Binomial(π_{ijk}, n^*_{ijk}) and logit$(\pi_{ijk}) = \beta_{0jk}$ Level 2

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

where $U_{0jk} \sim N(0, \tau_0^2)$ $i.i.d.$.
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 a high level.
- Can have cross-level interactions.

Predictors for the "cool" kid data:

Level 1	Black_{ijk}	$=$	1f 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

Three Level Random Intercept

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

$$\text{logit}(\pi_{ijk}) = \beta_{0jk} + \beta_{1jk}\text{Zpop}_{ijk} + \beta_{2jk}\text{Zagg}_{ijk}$$

Level 2

$$\begin{aligned} \beta_{0jk} &= \gamma_{00k} + U_{0jk} \\ \beta_{1jk} &= \gamma_{10k} \\ \beta_{2jk} &= \gamma_{20k} \end{aligned}$$

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

Level 3:

$$\begin{aligned} \gamma_{00k} &= \xi_{00} + W_{0k} \\ \gamma_{10k} &= \xi_{10} \\ \gamma_{20k} &= \xi_{20} \end{aligned}$$

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

Model	#	-2lnlike	AIC	BIC	$\hat{\tau}_0^2$	$\hat{\psi}_0^2$
Empty	3	701.14	707.14	713.21	1.28	2.84
+ Level 1	5	682.21	700.21	718.44	1.32	2.09
+ Level 2	8	673.06	691.06	709.29	1.02	1.90
+ Level 3	10	664.44	686.44	708.99	1.06	1.53

- 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

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

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

Level 3:

$$\begin{aligned}
\gamma_{00k} &= \xi_{000} + \xi_{001}\text{ClassAgg}_k + \xi_{002}\text{Majority}_k + W_{0k} \\
\gamma_{01k} &= \xi_{010} \\
\gamma_{02k} &= \xi_{020} \\
\gamma_{03k} &= \xi_{030} \\
\gamma_{10k} &= \xi_{100} \\
\gamma_{20k} &= \xi_{200} \\
\gamma_{30k} &= \xi_{300}
\end{aligned}$$

Linear Mixed Predictor

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

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

$$1.28 + 2.84 = 4.12 \qquad \text{to} \qquad 1.06 + 1.53 = 2.59$$

(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

Model	#	−2lnlike	AIC	BIC	$\hat{\tau}_0^2$ (s.e.)	$\hat{\psi}_0^2$ (se.)
Empty	3	701.14	707.14	713.21	1.28(0.51)	2.84(0.95)
+ Level 1	5	682.21	700.21	718.44	1.32(0.55)	2.09(0.81)
+ Level 2	8	673.06	691.06	709.29	1.02(0.47)	1.90(0.73)
+ Level 3	10	664.44	686.44	708.99	1.06(0.48)	1.53(0.63)
$+\underline{w}_{3k}$	11	648.20	674.20	700.53	0.05(0.25)	3.25(1.34)
(gender)$_{jk}$						−2.27(1.52)
						−4.49(1.83)
−Zpop$_{ijk}$	7	650.18	670.18	690.44	0.10(0.25)	3.61(1.43)
−Gnom$_{jk}$						−2.99(1.41)
−Majority$_{k}$						4.64(1.91)
+Black$_{ijk}$	8	646.08	668.08	690.36	0.04(.22)	3.25(1.22)
×ClassAgg$_{k}$						−2.69(1.34)
						4.76(1.94)

One More Model

Model	#	-2lnlike	AIC	BIC	$\hat{\tau}_{00}$ (s.e.)	$\hat{\psi}_{00}$ (se.)
$-U_{0jk}$	7	646.12	666.12	686.37		3.26 (1.22)
						-2.69 (1.33)
						4.77 (1.93)

Test for $H_o : \tau_{00} = 0$ versus $H_a : \tau_{00} > 0$, $LR = 646.12 - 646.08 = .04$.
This would be compared to a mixture of χ^2_1 and χ^2_0—clearly not significant.
So what's the final model?

What the Final (?) Model Looks Like

Level 1 $\underline{y}_{ijk} \sim \text{Binomial}(\underline{\pi}_{ijk}, n^*_{ijk})$ and

$$\text{logit}(\underline{\pi}_{ijk}) = \underline{\beta}_{0jk} + \beta_{1jk}\text{Black}_{ijk} + \beta_{2jk}\text{Zagg}_{ijk}$$

Level 2

$$\begin{aligned}
\underline{\beta}_{0jk} &= \underline{\gamma}_{00k} + \gamma_{02k}\text{Gagg}_{jk} + \underline{\gamma}_{03k}\text{Gender}_{jk} \\
\beta_{1jk} &= \gamma_{10k} \\
\beta_{2jk} &= \gamma_{20k}
\end{aligned}$$

Level 3:

$$\begin{aligned}
\underline{\gamma}_{00k} &= \xi_{000} + \xi_{001}\text{ClassAgg}_k + \underline{W}_{00k} \\
\gamma_{02k} &= \xi_{020} \\
\underline{\gamma}_{03k} &= \xi_{030} + \underline{W}_{03k} \\
\gamma_{10k} &= \xi_{100} + \xi_{101}\text{ClassAgg}_k \\
\gamma_{20k} &= \xi_{200}
\end{aligned}$$

Linear Mixed for η_{ijk}

$$
\begin{aligned}
\underline{\eta}_{ijk} \quad = \quad & \xi_{000} + \xi_{001}\mathsf{ClassAgg}_k + \xi_{020}\mathsf{Gagg}_k + \xi_{030}\mathsf{Sex}_{jk} + \xi_{100}\mathsf{Black}_{ijk} \\
& + \xi_{101}\mathsf{ClassAgg}_k\mathsf{Black}_{ijk} + \xi_{200}\mathsf{Zagg}_{ijk} + \underline{W}_{00k} + \underline{W}_{03k}\mathsf{gender}_{jk}
\end{aligned}
$$

Parameter Estimates

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t 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:

$$\begin{aligned}\beta_{0j} &= U_{0j} \quad \sim N(0,\tau_{00})\\ \beta_{1j} &= \gamma_{10}\\ \beta_{2j} &= \gamma_{20}\\ \beta_{3j} &= \gamma_{30}\\ \beta_{4j} &= \gamma_{40}\end{aligned}$$

- Our model for each item

$$\begin{aligned}P(Y_{1j}|x_{1i},\ldots,x_{5ij},U_{0j}) &= \exp[\gamma_{10}+U_{0j}]/(1+\exp[\gamma_{10}+U_{0j}])\\ P(Y_{2j}|x_{1i},\ldots,x_{5ij},U_{0j}) &= \exp[\gamma_{20}+U_{0j}]/(1+\exp[\gamma_{20}+U_{0j}])\\ P(Y_{3j}|x_{1i},\ldots,x_{5ij},U_{0j}) &= \exp[\gamma_{30}+U_{0j}]/(1+\exp[\gamma_{30}+U_{0j}])\\ P(Y_{4j}|x_{1i},\ldots,x_{5ij},U_{0j}) &= \exp[\gamma_{40}+U_{0j}]/(1+\exp[\gamma_{40}+U_{0j}])\end{aligned}$$

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,

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

"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:

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

Observations are independent at level 1.

Example: LSAT6

Level 2:

$$\begin{array}{rcll} \beta_{0j} & = & U_{0j} & \leftarrow \text{ on average equals } 0 \\ \beta_{1j} & = & \gamma_{10} & \\ \beta_{2j} & = & \gamma_{20} & \\ \beta_{3j} & = & \gamma_{30} & \\ \beta_{4j} & = & \gamma_{40} & \\ \beta_{5j} & = & \gamma_{50} & \end{array}$$

Example: LSAT6

Our model for each item

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

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

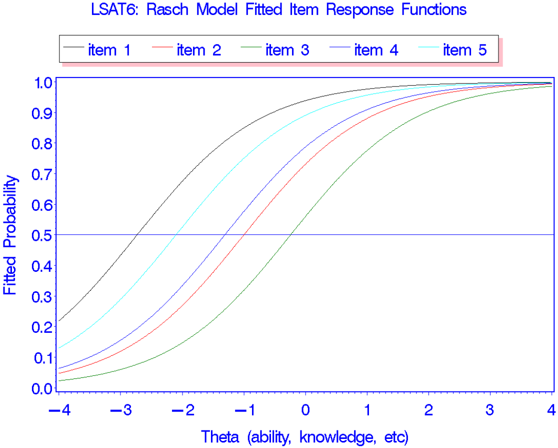
Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr $> \|t\|$	Gradient
b1	-2.73	0.13	999	-20.93	$< .01$	0.00002
b2	-1.00	0.08	999	-12.61	$< .01$	0.000175
b3	-0.24	0.07	999	-3.34	0.01	-0.00016
b4	-1.31	0.08	999	-15.44	$< .01$	0.000107
b5	-2.10	0.11	999	-19.91	$< .01$	-0.00008
std	-0.76	0.07	999	-10.88	$< .01$	-0.00027

Additional Estimates

Label	Estimate	Standard Error	DF	t Value	Pr $> \|t\|$
var(theta)	0.57	0.10	999	5.44	$< .01$

LSAT6: Fitted Probabilities



How Model was Fit: Data file/Design Matrix

ID	Y_{ij}	x1	x2	x3	x4	x5	Count
1	0	1	0	0	0	0	3
⋮	⋮						⋮
1	0	0	0	0	0	1	3
2	0	1	0	0	0	0	6
2	0	0	1	0	0	0	6
2	0	0	0	1	0	0	6
2	0	0	0	0	1	0	6
2	1	0	0	0	0	1	6
⋮							
32	1	1	0	0	0	0	298
32	1	0	1	0	0	0	298
32	1	0	0	1	0	0	298
32	1	0	0	0	1	0	298
32	1	0	0	0	0	1	298

Preparing the Data

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

```
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

```
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	0	1	1
⋮						

- By Guass quadrature:

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

- Defaults method is LaPlace:

```
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:

```
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

$$\begin{aligned}\text{logit}(P(Y_{ij}=1|\theta_j)) &= b_1x_{1ij}+b_2x_{2ij}+b_3x_{3ij}+b_4x_{4ij}+b_5x_{5ij}\\ &\quad +(a_1x_{1ij}+a_2x_{2ij}+a_3x_{3ij}+a_4x_{4ij}+a_5x_{5ij})\theta_j\end{aligned}$$

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

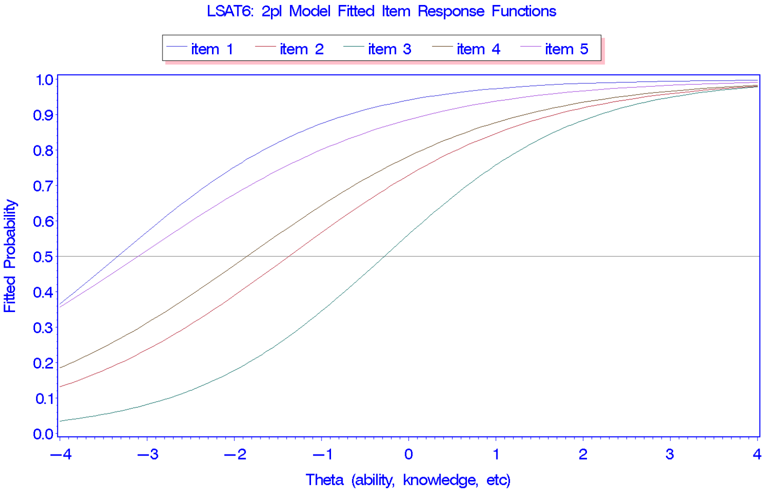
Two Parameter Logistic Model Estimates

Parameter Estimates from 2PL

Parameter	Estimate	Error	DF	t-Value	Pr $> \|t\|$	Gradient
b1	-2.77	0.21	999	-13.48	$< .01$	$-9.64E-6$
b2	-0.99	0.09	999	-11.00	$< .01$	0.000031
b3	-0.25	0.08	999	-3.27	$.01$	0.000011
b4	-1.28	0.10	999	-12.98	$< .01$	-0.00004
b5	-2.05	0.14	999	-15.17	$< .01$	$4.593E-6$
a1	0.83	0.26	999	3.20	$< .01$	$-7.23E-6$
a2	0.72	0.19	999	3.87	$< .01$	0.000018
a3	0.89	0.23	999	3.83	$< .01$	$-2.7E-6$
a4	0.69	0.19	999	3.72	$< .01$	$-3.65E-6$
a5	0.66	0.21	999	3.13	$< .01$	0.00001

τ_{00}^2 set to 1 for identification and Converged (i.e., $\theta \sim N(0,1)$)

LSAT6: Fitted Probabilities



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

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

Retain H_o for all items.

Better Test of $H_o : a_i = 1$

The likelihood ratio test:

Model	# parameters	−logLike	LR	df	p
Rasch	6	4933.875	—		
2PL	10	4933.307	0.548	4	.97

Retain $\text{H}_o : \text{a}_1 = a_2 = a_3 = a_4 = a_5 = 1$

SAS and the 2PL Model

```
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; * τ0^2 = 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.

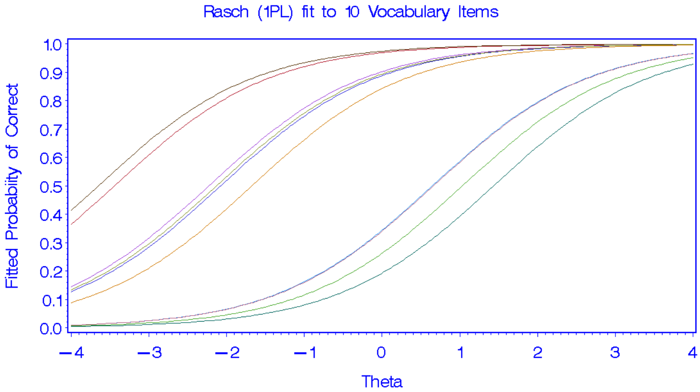
The model was fit using SAS PROC NLMIXED. Edited output:

```
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

Parameter	Estimate	Standard Error	DF	Value	Pr $> \|t\|$	t Gradient
b1	-2.0756	0.09925	1154	-20.91	$< .0001$	-0.00003
b2	-3.4436	0.1435	1154	-23.99	$< .0001$	2.484E-7
b3	1.4330	0.08806	1154	16.27	$< .0001$	-4.6E-6
b4	-3.6519	0.1537	1154	-23.76	$< .0001$	-6.38E-7
b5	-2.2280	0.1026	1154	-21.71	$< .0001$	-7.15E-6
b6	-2.1304	0.1004	1154	-21.21	$< .0001$	0.000039
b7	0.6455	0.08044	1154	8.02	$< .0001$	-6.66E-6
b8	0.6601	0.08053	1154	8.20	$< .0001$	-3.56E-6
b9	-1.6749	0.09176	1154	-18.25	$< .0001$	4.163E-6
b10	1.0332	0.08342	1154	12.39	$< .0001$	0.000015
std	1.3303	0.04981	1154	26.71	$< .0001$	2.979E-6
var(theta)	1.7697	0.1325	1154	13.35	$< .0001$ * n.a.	

1pl: Item Curves



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

$$\begin{aligned} P(Y_{ij} = 1) &= \frac{\exp(\gamma_{i0} + a_i U_{0j})}{(1 + \exp(\gamma_{i0} + a_i U_{0j}))} \\ &= \frac{\exp(b_i + a_i \theta_j)}{(1 + \exp(b_i + a_i \theta_j))} \end{aligned}$$

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

$$\begin{aligned} LR &= -2(\text{LogLike}_{1PL} - \text{LogLike}_{2PL}) \\ &= 10197 - 10084 = 113 \end{aligned}$$

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

b_i Parameter Estimates

Parameter	Estimate	Standard Error	DF	t 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	t 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 $H_o : a_i = 0$, but there are other hypotheses that we may be interested in (e.g., $H_o : a_1 = a_2$, or $H_o : a_i = 1$).

a_i 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.

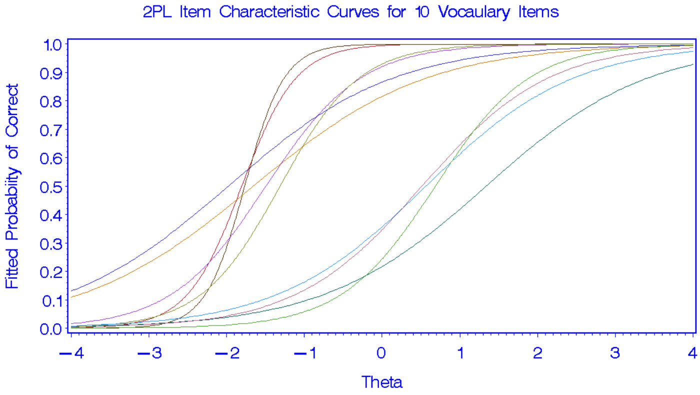
- 1PL is a special case of 2PL where $H_o: a_1 = a_2 = \ldots = a_{10} = 1$.
- Likelihood ratio test

$$\begin{aligned} LR &= -2(\text{LogLike}_{\text{1PL}} - \text{LogLike}_{\text{2PL}}) \\ &= 10197 - 10084 = 113 \end{aligned}$$

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

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

2pl: Item Curves



2PL with Predictors of theta

- Often we wish to use θ as a response variable in a regression model (i.e., see what influences or explains variability in θ).
- This strategy is problematic because it introduces additional error — error due to estimation of θ.
- Solution: Put predictors of θ 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.	Std Error	DF	t Value	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

Parm	Est.	Std Error	DF	t Value	Pr $> \|t\|$	Gradient
a1	0.8166	0.1033	1154	7.91	$<.0001$	0.000147
a2	2.4461	0.3552	1154	6.89	$<.0001$	-0.00003
a3	0.9020	0.0990	1154	9.11	$<.0001$	-2.01E-6
a4	2.9506	0.5177	1154	5.70	$<.0001$	-0.00007
a5	1.4529	0.1573	1154	9.23	$<.0001$	6.616E-6
a6	1.6744	0.1825	1154	9.18	$<.0001$	0.000026
a7	0.9712	0.0968	1154	10.03	$<.0001$	0.000011
a8	1.1382	0.1099	1154	10.36	$<.0001$	-0.00001
a9	0.7217	0.0924	1154	7.81	$<.0001$	-0.00012
a10	1.3724	0.1375	1154	9.98	$<.0001$	-0.00002

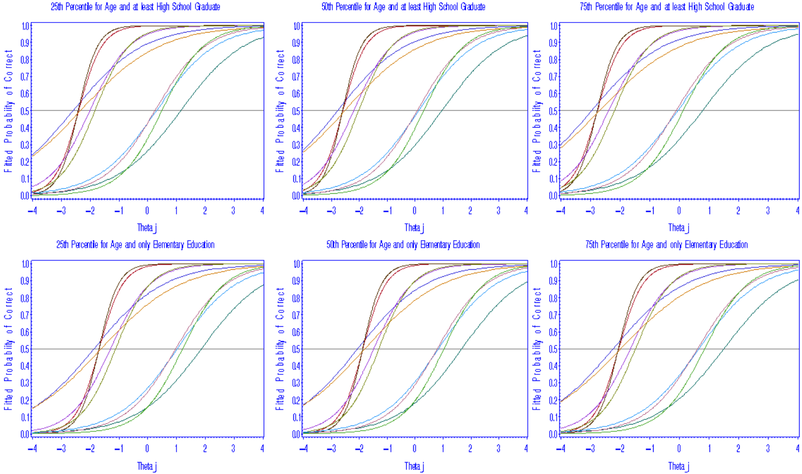
2PL with Predictors of Vocabulary

Parm	Est.	Std Error	DF	t Value	Pr $> \|t\|$
ν_1 (age)	0.0154	0.0024	1154	6.47	$< .0001$
ν_2 (elementary)	0.9891	0.1310	1154	7.55	$< .0001$
ν_3 (high school)	1.7075	0.1412	1154	12.09	$< .0001$

Parm	Gradient
ν_1 (age)	-0.00014
ν_2 (elementary)	-0.00008
ν_3 (high school)	-0.0001

Interpretation?

Figure of Item Curves



Summary of Models for GSS Vocabulary

Model	Number parameters	-2(LogLike)	AIC	BIC
Rasch (1PL)	11	10197	10219	10275
2PL	20	10084	10124	10225
2PL covariates θ	23	9871	9917	10033
2PL drop age	22	9914	9958	10069

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

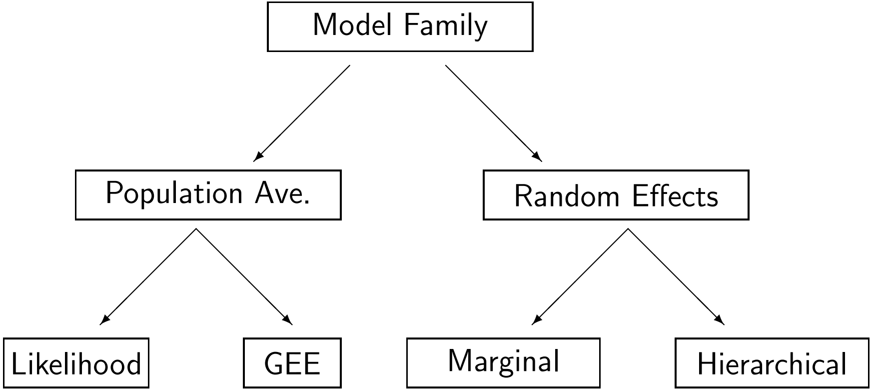
$$LR = 10084 - 9871 = 213$$

compare to χ^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 are 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.