

Multilevel Modeling of Categorical Data

Sophia Rabe-Hesketh

Graduate School of Education & Graduate Group in Biostatistics
University of California, Berkeley
Institute of Education, University of London

Joint work with Anders Skrondal
Division of Epidemiology
Norwegian Institute of Public Health

Methods of Analysis Program in the Social Sciences
Stanford University, November 2008

GLLAMM – p.1

Outline

- ▶ I. Ordered categorical responses
 - Latent response and generalized linear model formulations
 - Two-level random intercept models
 - Conditional versus marginal effects
 - Multilevel random coefficient models
 - Estimation methods
 - Example: Cluster randomized trial of sex education in Norway
- ▶ II. Unordered categorical responses
 - Random utility formulation
 - Multilevel random utility models
 - Example: Abuse of antibiotics in China

GLLAMM – p.2

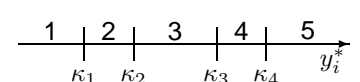
I. Ordered categorical responses

- ▶ Small number of mutually exclusive 'categories', $y_i = 1, \dots, S$
- ▶ Categories are ordered, so y_i is **ordinal**
- ▶ Examples:
 - Severity of symptom (e.g. pain): none, moderate, severe
 - Frequency of behavior: never, occasionally, daily
 - Agreement (Likert scale): disagree strongly, disagree, agree, agree strongly
 - Educational attainment: high school, college degree, graduate or professional degree

GLLAMM – p.3

Latent response models

- ▶ Latent (unobserved) continuous response y_i^* underlies observed ordinal response y_i
- ▶ Threshold model determines observed response:

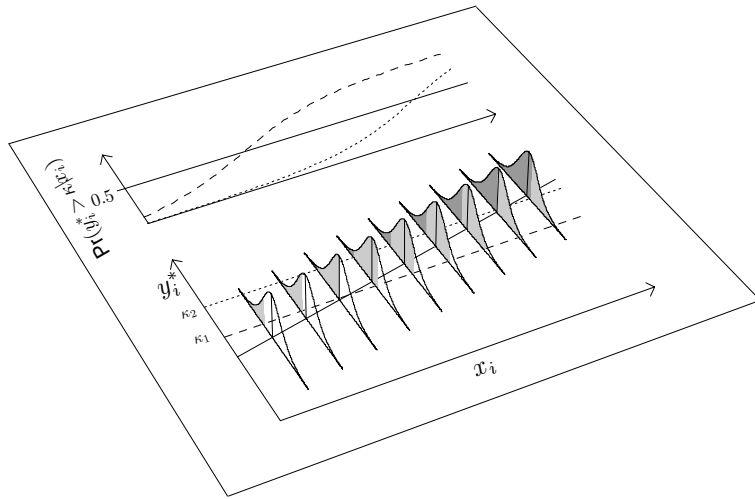
$$y_i = \begin{cases} 1 & \text{if } y_i^* \leq \kappa_1 \\ 2 & \text{if } \kappa_1 < y_i^* \leq \kappa_2 \\ \vdots & \vdots \\ S & \text{if } \kappa_{S-1} < y_i^* \end{cases}$$


- ▶ Latent response modeled as linear regression without intercept (for identification)

$$\begin{aligned} y_i^* &= \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \epsilon_i \\ &= \mathbf{x}'_i \boldsymbol{\beta} + \epsilon_i \end{aligned}$$

GLLAMM – p.4

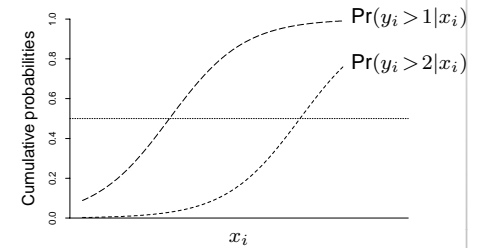
Latent response modeled as linear regression



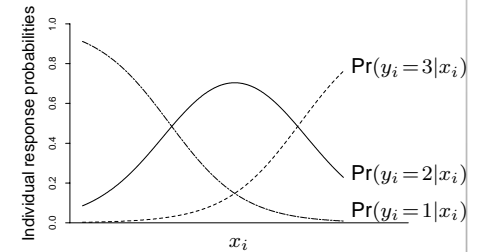
GLLAMM – p.5

Cumulative versus individual response probabilities

$$\begin{aligned} \Pr(y_i > 1) &= \Pr(y_i = 2) + \Pr(y_i = 3) \\ \Pr(y_i > 2) &= \Pr(y_i = 3) \end{aligned}$$



$$\begin{aligned} \Pr(y_i = 3) &= \Pr(y_i > 2) \\ \Pr(y_i = 2) &= \Pr(y_i > 1) - \Pr(y_i > 2) \\ \Pr(y_i = 1) &= 1 - \Pr(y_i > 1) \end{aligned}$$



GLLAMM – p.6

From latent response to generalized linear model formulations

- Cumulative probabilities:

$$\begin{aligned} \Pr(y_i > s | x_i) &= \Pr(y_i^* > \kappa_s | x_i) = \Pr(\mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i > \kappa_s | x_i) \\ &= \Pr(-\epsilon_i \leq \mathbf{x}_i' \boldsymbol{\beta} - \kappa_s | x_i) = F(\mathbf{x}_i' \boldsymbol{\beta} - \kappa_s) \end{aligned}$$

- $F(\cdot)$ is the cumulative density function of $-\epsilon_i$
- This is a generalized linear model where
 - $F(\cdot)$ is inverse **link function** and $\mathbf{x}_i' \boldsymbol{\beta} - \kappa_s$ is **linear predictor**

Distribution of $-\epsilon_i$	$\theta \equiv \text{Var}(\epsilon_i)$	Link $g = F^{-1}$	Model
Logistic	$\pi^2/3$	Logit	Prop. odds
Standard normal	1	Probit	Ordinal probit
*Gumbel	$\pi^2/6$	Compl. log-log	Compl. log-log

*Asymmetric (non-equivalent model if scale reversed)

GLLAMM – p.7

Logit link and odds ratios

- With a logistic distribution for ϵ_i , a **logit link** (one covariate x_i):

$$\ln \left\{ \frac{\Pr(y_i > s | x_i)}{1 - \Pr(y_i > s | x_i)} \right\} = \beta x_i - \kappa_s$$

odds $(y_i > s | x_i)$

$$\begin{aligned} \ln\{\text{odds}(y_i > s | x_i = a + 1)\} - \ln\{\text{odds}(y_i > s | x_i = a)\} &= \beta \\ &= \ln \left\{ \frac{\text{odds}(y_i > s | x_i = a + 1)}{\text{odds}(y_i > s | x_i = a)} \right\} \end{aligned}$$

- Coefficient β represents increase in log-odds associated with unit increase in x_i
- Exponentiated coefficient $\exp(\beta)$ represents odds ratio associated with unit increase in x_i

GLLAMM – p.8

Two-level random intercept models

- ▶ Subjects i nested in clusters j (e.g. schools)
- ▶ Include random intercept ζ_j for clusters

$$y_{ij}^* = \mathbf{x}_{ij}'\boldsymbol{\beta} + \zeta_j + \epsilon_{ij}, \quad \zeta_j | \mathbf{x}_{ij} \sim N(0, \psi)$$

- ζ_j independent of ϵ_{ij}
- ζ_j and ϵ_{ij} independent across clusters
- ϵ_{ij} independent across units within clusters

Intraclass correlation of latent responses

- ▶ Total residual $\xi_{ij} = \zeta_j + \epsilon_{ij}$ has variance

$$\text{Var}(\xi_{ij} | \mathbf{x}_{ij}) = \begin{cases} \psi + 1 & \text{for probit models} \\ \psi + \pi^2/3 & \text{for logit models} \end{cases}$$

- ▶ Covariance between total residuals ξ_{ij} and $\xi_{i'j}$ of two subjects in same cluster is ψ and **intraclass correlation** is

$$\text{Cor}(\xi_{ij}, \xi_{i'j} | \mathbf{x}_{ij}, \mathbf{x}_{i'j}) = \begin{cases} \psi / (\psi + 1) & \text{for probit models} \\ \psi / (\psi + \pi^2/3) & \text{for logit models} \end{cases}$$

Conditional versus marginal effects

- ▶ For probit model, **conditional** or cluster-specific probabilities are

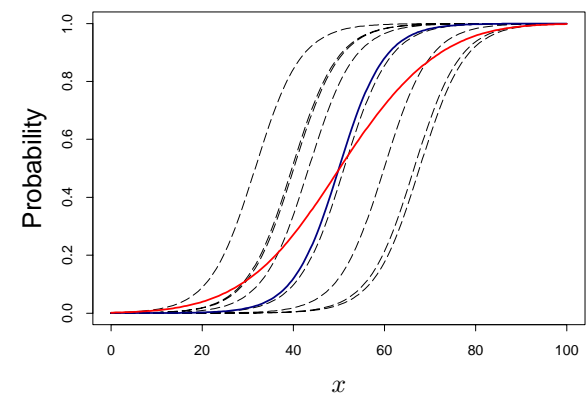
$$\Pr(y_{ij} > s | \zeta_j, \mathbf{x}_{ij}) = \Phi(\mathbf{x}_{ij}'\boldsymbol{\beta} + \zeta_j - \kappa_s)$$

- ▶ **Marginal** or population-averaged response probabilities are

$$\begin{aligned} \Pr(y_{ij} > s | \mathbf{x}_{ij}) &= \Pr(y_{ij}^* > \kappa_s) = \Pr(\mathbf{x}_{ij}'\boldsymbol{\beta} + \xi_{ij} > \kappa_s) \\ &= \Pr(-\xi_{ij} \leq \mathbf{x}_{ij}'\boldsymbol{\beta} - \kappa_s) \\ &= \Pr\left(\frac{\xi_{ij}}{\sqrt{\psi + 1}} \leq \frac{\mathbf{x}_{ij}'\boldsymbol{\beta} - \kappa_s}{\sqrt{\psi + 1}}\right) \\ &= \Phi\left(\frac{\mathbf{x}_{ij}'\boldsymbol{\beta} - \kappa_s}{\sqrt{\psi + 1}}\right) \end{aligned}$$

- ▶ Therefore, marginal effects $\beta / \sqrt{\psi + 1}$ attenuated compared with conditional effects β

Illustration: Conditional versus marginal relationship



- cluster-specific (random sample)
- median
- marginal or population-averaged

Multilevel random coefficient models

- ▶ Consider clustered longitudinal data with occasions i (level 1) nested in student j (level 2) in schools k (level 3)
- ▶ Example of three-level random coefficient model:

$$y_{ijk}^* = [\zeta_{0jk}^{(2)} + \zeta_{0k}^{(3)}] + [\beta_1 + \zeta_{1jk}^{(2)} + \zeta_{1k}^{(3)}]x_{1ijk} + \beta_2 x_{2ijk} + \epsilon_{ijk}$$

- $\zeta_{0jk}^{(2)}$ and $\zeta_{0k}^{(3)}$ are random intercepts at levels 2 and 3
- $\zeta_{1jk}^{(2)}$ and $\zeta_{1k}^{(3)}$ are random coefficients of x_{1ijk} at levels 2 and 3
- Random effects at the same level, e.g. $(\zeta_{0jk}^{(2)}, \zeta_{1jk}^{(2)})$, bivariate normal with zero means, independent across units jk
- ▶ General three-level random coefficient model

$$y_{ijk}^* = \mathbf{x}'_{ijk}\boldsymbol{\beta} + \mathbf{z}'_{ijk}\boldsymbol{\zeta}_{jk}^{(2)} + \mathbf{z}'_{ijk}\boldsymbol{\zeta}_k^{(3)} + \epsilon_{ijk}$$

GLLAMM – p.13

Estimation: Approximate methods

- ▶ Estimation of multilevel models with categorical responses difficult because likelihood does not generally have closed form
- ▶ Penalized Quasilikelihood (PQL)
 - Two versions: First and second order (PQL-1, PQL-2), the latter being better
 - ◊ PQL-1 in GLIMMIX in SAS, MLwiN and HLM
 - ◊ PQL-2 in MLwiN
 - ◊ Even PQL-2 produces biased estimates for small clusters and high intraclass correlations
- ▶ Sixth order Laplace in HLM
- ▶ H-likelihood in Genstat
- ▶ Methods do not provide a likelihood

GLLAMM – p.14

Estimation: Maximum likelihood

- ▶ Numerical integration
 - Gauss-Hermite (ordinary) quadrature used in MIXOR/MIXNO (two-level only), aML, NLMIXED in SAS (two-level only) and gllamm in Stata
 - Adaptive quadrature superior to 'ordinary quadrature', particularly for large clusters and high intraclass correlations and available in NLMIXED in SAS (two-level only), gllamm in Stata, glme in S-PLUS
- ▶ Monte Carlo integration
 - Simulated maximum likelihood in nlogit
 - Monte Carlo EM - no software?
- ▶ MCMC with vague priors approximates maximum likelihood and available in MLwiN and WinBUGS

GLLAMM – p.15

Example: Cluster randomized trial of sex education in Norway

- ▶ Schools randomized to receive special sex education or not
- ▶ Assessments before, 6 and 18 months after randomization
- ▶ One outcome is question relating to 'Contraceptive self-efficacy':
 - "If my partner and I were about to have intercourse without either of us having mentioned contraception, it would be easy for me to produce a condom (if I brought one)"
 - Question answered in terms of five ordinal categories: not at all true of me, slightly true of me, somewhat true of me, mostly true of me, completely true of me
- ▶ Multilevel data with responses at occasions (level 1) from 1184 students (level 2) in 46 schools (level 3)
- ▶ Only 570 students always responded, 400 responded on some occasions and 114 never responded

GLLAMM – p.16

Models and estimation using gllamm

► Occasions t , students j , schools k

► Covariates

- x_{1t} [Time] (0, 1, 3)
- x_{2jk} [Treat] (yes=1, no=0)
- x_{3tjk} [Treat] \times [Time]

► Model the probability of exceeding a category s , $s = 1, 2, 3, 4$

$$\text{logit}[\Pr(y_{tjk} > s | \zeta_{jk}^{(2)}, \zeta_k^{(3)}, \mathbf{x}_{tjk})] = \beta_1 x_{1t} + \beta_2 x_{2jk} + \beta_3 x_{3tjk} + \zeta_{jk}^{(2)} + \zeta_k^{(3)} - \kappa_s$$

► Estimation using adaptive quadrature in gllamm:

```
gllamm use treat time treat_time, i(id school) ///
link(ologit) family(binom) adapt
```

GLLAMM – p.17

Estimates

Parameter	Single-level model		Two-level model		Three-level model	
	Est	(SE)	Est	(SE)	Est	(SE)
β_1 [Time]	-0.12	(0.06)	-0.13	(0.06)	-0.13	(0.06)
β_2 [Treat]	-0.05	(0.14)	-0.02	(0.19)	-0.02	(0.19)
β_3 [Time] \times [Treat]	0.17	(0.08)	0.17	(0.09)	0.17	(0.09)
$\text{Var}(\zeta_{jk}^{(2)})$	–		2.03	(0.31)	2.03	(0.31)
$\text{Var}(\zeta_k^{(3)})$	–		–		0.00	(0.00)
κ_1	-3.54	(0.17)	-4.41	(0.23)	-4.41	(0.23)
κ_2	-2.43	(0.13)	-3.15	(0.19)	-3.15	(0.19)
κ_3	-1.18	(0.12)	-1.58	(0.16)	-1.58	(0.16)
κ_4	0.16	(0.12)	0.25	(0.15)	0.25	(0.15)
Log-likelihood	-2531		-2471		-2471	

GLLAMM – p.18

Predicted probabilities

► Conditional probabilities:

$$\Pr(y_{tjk} > 2 | \zeta_{jk}^{(2)} = 0, \zeta_k^{(3)} = 0, \mathbf{x}_{tjk}) = \frac{\exp(\mathbf{x}'_{tjk} \hat{\boldsymbol{\beta}} + 0 + 0 - \hat{\kappa}_2)}{1 + \exp(\mathbf{x}'_{tjk} \hat{\boldsymbol{\beta}} + 0 + 0 - \hat{\kappa}_2)}$$

```
gen u1 = 0
gen u2 = 0
gllapred p_cond, mu us(u) above(2)
```

► Marginal probabilities:

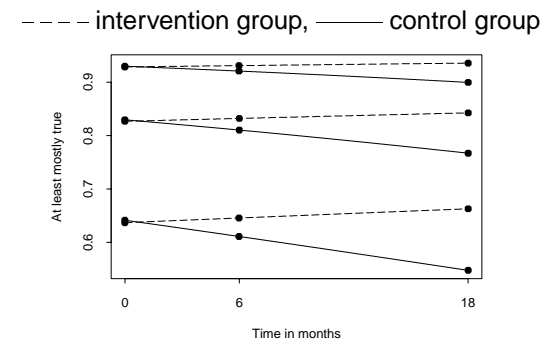
$$\Pr(y_{tjk} > 2 | \mathbf{x}_{tjk}) = \iint \frac{\exp(\mathbf{x}'_{tjk} \hat{\boldsymbol{\beta}} + \zeta_{jk}^{(2)} + \zeta_k^{(3)} - \hat{\kappa}_2)}{1 + \exp(\mathbf{x}'_{tjk} \hat{\boldsymbol{\beta}} + \zeta_{jk}^{(2)} + \zeta_k^{(3)} - \hat{\kappa}_2)} g(\zeta_{jk}^{(2)}) g(\zeta_k^{(3)}) d\zeta_{jk}^{(2)} d\zeta_k^{(3)}$$

```
gllapred p_marg, mu marg above(2)
```

GLLAMM – p.19

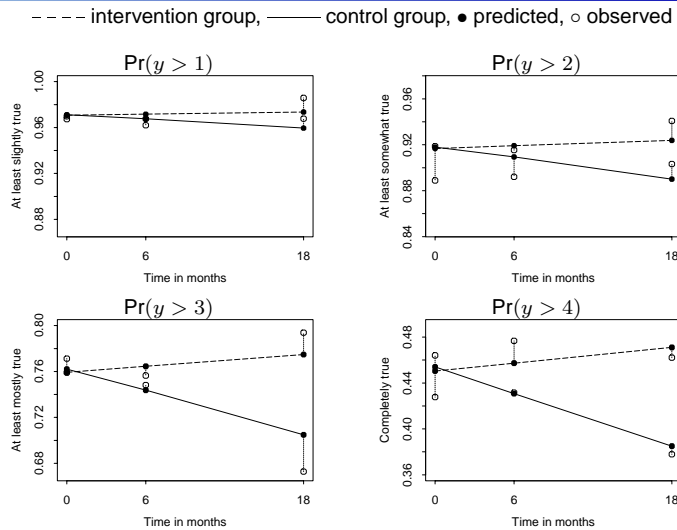
Predicted conditional probabilities

- Probability of responding at least 'mostly true of me' (category 3)
- Relationship between probability and occasion for two treatment groups, when $\zeta_k^{(3)} = 0$ and $\zeta_{jk}^{(2)}$ is -1, 0 and 1:



GLLAMM – p.20

Observed proportions and predicted marginal probabilities



GLLAMM – p.21

II. Unordered categorical responses or discrete choice

- ▶ Small number of mutually exclusive categories a , $a = 1, \dots, A$.
- ▶ Categories cannot be ordered a priori
- ▶ Examples:
 - Candidate voted for: Obama, McCain, Nader, other
 - Brand of cola preferred: Coca Cola, Pepsi, other
 - Method of birth control used: pill, condom, etc.
 - Diagnosis: Autism, Asperger's syndrome, pervasive developmental disorder, other
- ▶ Responses often correspond to discrete choices among alternatives (categories)

GLLAMM – p.22

Random utility models

- ▶ Unobserved 'utility' U_i^a associated with each alternative $a = 1, \dots, A$ for unit $i = 1, \dots, N$

- ▶ Random utility models composed as

$$U_i^a = V_i^a + \epsilon_i^a$$

- V_i^a is linear predictor
- ϵ_i^a is residual term (independent over i and a)

- ▶ Alternative f chosen if

$$U_i^f > U_i^g \text{ for all } g \neq f$$

ϵ_i^a independent Gumbel distributed



$$\Pr(f_i) = \frac{\exp(V_i^f)}{\sum_{a=1}^A \exp(V_i^a)} \quad \text{multinomial logit}$$

GLLAMM – p.23

Covariate effects on utilities

Linear predictor for unit i and alternative a :

$$V_i^a = m^a + \mathbf{x}_i' \mathbf{g}^a + \mathbf{x}_i^{a'} \mathbf{b}$$

- ▶ Covariates and parameters:

- m^a alternative-specific constants
- \mathbf{x}_i varies over subjects (but not alternatives) and has fixed alternative-specific effects \mathbf{g}^a
Example: Age of subject
- \mathbf{x}_i^a varies over alternatives (and possibly subjects) and has fixed effects \mathbf{b} , constant across alternatives
Example: Cost of treatment alternative

GLLAMM – p.24

Identification

- ▶ Probability of choosing alternative 1 among alternatives 1, 2 and 3 can be expressed in terms of utility differences

$$\Pr(U^1 - U^2 > 0 \text{ and } U^1 - U^3 > 0)$$

- ▶ Therefore location of V_i^a arbitrary:

$$\frac{\exp(V_i^1)}{\sum_a \exp(V_i^a)} = \frac{\exp(V_i^1 + c_i)}{\sum_a \exp(V_i^a + c_i)}$$

- ▶ Solution: Last alternative S serves as reference alternative,

$$\text{set } m^S = 0 \text{ and } \mathbf{g}^S = \mathbf{0}$$

Multilevel random utility models

- ▶ Consider three-level data with patients i (level 1) treated by doctors j (level 2) working in hospitals k (level 3)

- ▶ Can include random effects in linear predictor:

$$\begin{aligned} V_{ijk}^a = & [m^a + \gamma_{0jk}^{a(2)} + \gamma_{0k}^{a(3)}] \\ & + \mathbf{x}'_{ijk} [\mathbf{g}^a + \gamma_{jk}^{a(2)} + \gamma_k^{a(3)}] \\ & + \mathbf{x}'_{ijk} [\mathbf{b} + \beta_{jk}^{(2)} + \beta_k^{(3)}] \end{aligned}$$

- ▶ Random intercepts: $\gamma_{0jk}^{a(2)}$ and $\gamma_{0k}^{a(3)}$
- ▶ Random Coefficients I: $\gamma_{jk}^{a(2)}$ and $\gamma_k^{a(3)}$ are **alternative-specific** random coefficients for **subject-specific** covariates \mathbf{x}_{ijk}
- ▶ Random Coefficients II: $\beta_{jk}^{(2)}$ and $\beta_k^{(3)}$ are random coefficients for **alternative-specific** covariates \mathbf{x}_{ijk}^a

Abuse of antibiotics in China

- ▶ Acute respiratory tract infection (ARI) can lead to pneumonia and death if not properly treated
- ▶ Inappropriate frequent use of antibiotics common in China in 1990's, leading to drug resistance
- ▶ In the 1990's WHO introduced program of case management for children under 5 with ARI in China
- ▶ Data collected on 855 children i (level 1) treated by 134 doctors j (level 2) in 36 hospitals k (level 3) in two counties (one of which was in WHO program)
- ▶ **Response variable**: Abuse defined as prescription of antibiotics when there were no clinical indications based on medical files
 1. Abuse of several antibiotics
 2. Abuse of one antibiotic
 3. Correct use of antibiotics (reference category)

Covariates

- ▶ 7 covariates \mathbf{x}_{ijk}
 - Patient level i
 - ◇ [Age] Age in years (0-4)
 - ◇ [Temp] Body temperature, centered at 36°C
 - ◇ [Paymed] Pay for medication (yes=1, no=0)
 - ◇ [Selfmed] Self medication (yes=1, no=0)
 - ◇ [Wrdiag] Failure to diagnose ARI early (yes=1, no=0)
 - Doctor level j
 - ◇ [DRed] Doctor's education (6 categories from self-taught to medical school)
 - Hospital level k
 - ◇ [WHO] Hospital in WHO program (yes=1, no=0)

Reference: Min Yang (2001). *Multinomial Regression*. In Goldstein and Leyland (Eds). *Multilevel Modelling of Health Statistics*, pages 107-123.

Model

- ▶ No alternative-specific covariates

$$V_{ijk}^a = [m^a + \gamma_{0jk}^{a(2)} + \gamma_{0k}^{a(3)}] + \mathbf{x}'_{ijk} \mathbf{g}^a$$

doc child alt choice

1 1 1 0

1 1 2 1

▶ Data: 1 1 3 0

1 2 1 0

1 2 2 0

1 2 3 1

- ▶ Estimation in gllamm

```
gen categ1 = alt == 1
gen categ2 = alt == 2
eq c1: categ1
eq c2: categ2
gllamm alt age temp ... , i(doc hosp) nrf(2 2) eqs(c1 c2 c1 c2) ///
link(mlogit) expanded(child choice m) basecat(3) adapt
```

Maximum likelihood estimates


Parameter	Abuse several		Abuse one	
	Est	(SE)	Est	(SE)
g_0^a [Cons]	-5.72	(0.99)	-0.23	(0.55)
g_1^a [Age]	0.07	(0.09)	0.17	(0.08)
g_2^a [Temp]	-0.27	(0.13)	-0.96	(0.12)
g_3^a [Paymed]	0.92	(0.40)	0.12	(0.32)
g_4^a [Selfmed]	-0.86	(0.29)	-0.49	(0.24)
g_5^a [Wrdiag]	1.85	(0.26)	2.08	(0.23)
g_6^a [DRed]	-0.62	(0.17)	0.08	(0.11)
g_7^a [WHO]	-2.40	(0.62)	-0.88	(0.33)
Doctor-level variances				
$\text{Var}(\gamma_{0jk}^{a(2)})$	0.46	(0.28)	0.43	(0.22)
$\text{Cov}(\gamma_{0jk}^{1(2)}, \gamma_{0jk}^{2(2)})$			-0.44	(0.13)
Hospital-level variances				
$\text{Var}(\gamma_{0k}^{a(3)})$	0.88	(0.45)	0.11	(0.12)
$\text{Cov}(\gamma_{0k}^{1(3)}, \gamma_{0k}^{2(3)})$			0.31	(0.20)

References

- ▶ “Multilevel logistic regression for polytomous data and rankings”. *Psychometrika* 68, 267-287, 2003. (A. Skrondal & S. Rabe-Hesketh)
- ▶ “Generalized linear mixed effects models”. In G. Fitzmaurice et al. (Eds.) *Longitudinal Data Analysis*. Chapman & Hall/CRC, pp. 79-106, 2008. (S. Rabe-Hesketh & A. Skrondal)
- ▶ “Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects”. *Journal of Econometrics* 128, 301-323, 2005. (S. Rabe-Hesketh, A. Skrondal & A. Pickles)
- ▶ “Prediction in multilevel generalized linear models”. *Journal of the Royal Statistical Society, Series A*, in press. (A. Skrondal & S. Rabe-Hesketh)
- ▶ *GLLAMM Manual*. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 160, 2004. <http://www.bepress.com/ucbbiostat/paper160/> (S. Rabe-Hesketh, A. Skrondal & A. Pickles)
- ▶ See also <http://www.gllamm.org>



Further reading

- ▶ Snijders & Bosker (1999): Excellent introduction to multilevel modelling (MLM)
- ▶ Rabe-Hesketh & Skrondal (2008): MLM using Stata 
- ▶ Raudenbush & Bryk (2002): General book on MLM
- ▶ Goldstein (2003): General book on MLM
- ▶ McCulloch, Searle & Neuhaus (2008): Theoretical treatment of GLMMs
- ▶ Skrondal & Rabe-Hesketh (2004): Generalized latent variable modeling 