a0005 # Generalized Linear Mixed Models

S Rabe-Hesketh, University of California, Berkeley, CA, USA; University of London, London, UK
A Skrondal, Norwegian Institute of Public Health, Oslo, Norway

Glossary

g0005 **Linear predictor** – A linear combination of explanatory variables that is part of a regression model or generalized linear mixed model.

g0010 **Link function** – A function applied to the conditional expectation of the response variable before this is equated to the linear predictor (in a generalized linear model). Examples are the identity, log, and logit link functions.

g0015 **Random coefficient** – Cluster-specific coefficient of a covariate in a regression model for clustered data that varies randomly between clusters.

g0020 **Random intercept** – Cluster-specific intercept in a regression model for clustered data that varies randomly between clusters.

s0005 ## Logistic Random-Intercept Model

p0005 To introduce the idea of generalized linear mixed models, we consider the following example. The Program for International Student Assessment (PISA) is an international educational survey funded by the Organisation for Economic Co-operation and Development (OECD) that measures attainment in reading, mathematics, and science among 15-year-old students. Using the United States sample of PISA 2000, we estimate the relationship between reading proficiency (Prof_{ij}) of student i in school j and the following covariates:

- Fem_{ij}: indicator for student being female;
- SES_{ij}: international socioeconomic index (continuous), grand-mean centered and divided by 20;
- HS_{ij}: indicator for highest education level of either parent being high school;
- Coll_{ij}: indicator for highest education level of either parent being college; and
- Eng_{ij}: indicator for test language (English) spoken at home.

p0010 Prof_{ij} is a binary variable (1 = Yes, 0 = No); therefore, an appropriate level-1 model is a logistic regression model. Specifically, the log of the odds that the student is proficient versus not proficient is specified as a linear function of the covariates,

$$\log\left\{\frac{P(\text{Prof}_{ij}=1)}{P(\text{Prof}_{ij}=0)}\right\} = \beta_{0j} + \beta_1\text{Fem}_{ij} + \beta_2\text{SES}_{ij} + \beta_3\text{HS}_{ij} + \beta_4\text{Coll}_{ij} + \beta_5\text{Eng}_{ij}$$

(Here $P(\cdot)$ refers to the conditional probability, given β_{0j} and the covariates.) The school-specific intercept β_{0j} represents the log-odds of being proficient for students in school j when all the covariates take the value zero, that is for males with average Socioeconomic status (SES) whose parents' highest education level is less than high school and who do not speak English at home. The regression coefficient of a covariate represents the increase in log-odds per unit increase in the corresponding covariate when the other covariates and β_{0j} remain constant. These coefficients are sometimes referred to as conditional or school-specific effects because β_{0j} is held constant. Exponentiating the coefficients yields school-specific odds ratios.

The level-2 model for school j is p0015

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\overline{\text{SES}}_{\cdot j} + \zeta_{0j}$$

where $\overline{\text{SES}}_{\cdot j}$ is the mean SES for school j, and ζ_{0j} is a school-specific random intercept assumed to be independent of the covariates and independent across schools with $\zeta_{0j} \sim N(0, \psi)$.

Substituting the school-level model into the student- p0020
level model and writing η_{ij} for the log-odds, we obtain the so-called reduced form

$$\eta_{ij} = \gamma_{00} + \gamma_{01}\overline{\text{SES}}_{\cdot j} + \beta_1\text{Fem}_{ij} + \beta_2\text{SES}_{ij} + \beta_3\text{HS}_{ij} + \beta_4\text{Coll}_{ij} + \beta_5\text{Eng}_{ij} + \zeta_{0j}$$

a logistic regression model with both student-level (or level 1) and school-level (or level 2) covariates and both a fixed intercept γ_{00} and a random intercept ζ_{0j}.

Maximum likelihood estimates for the logistic random p0025
intercept model are presented in **Table 1**. Higher school mean SES, being female, higher individual SES, having at least one parent with a college education, and speaking English at home are all associated with a significantly greater odds of reading proficiency at the 5% level. For an indicator variable, the odds ratio represents the ratio of the odds for two groups, holding constant the other covariates. For instance, the estimated odds ratio of 1.9 for Eng_{ij} means that students who speak English at home have 1.9 times the odds of being proficient in comparison to students who do not speak English at home for a given

school mean SES, gender, parents' education, and random intercept ζ_{0j}.

p0030 For a continuous covariate, the odds ratio can be interpreted as the effect of increasing the covariate by a unit; therefore, it is important to understand the scale of the covariate. Individual SES has a standard deviation of 0.88 and school mean SES has a standard deviation of 0.45. Each unit increase in school mean SES is associated with a quadrupling in the estimated odds of being proficient, controlling for individual SES, the other covariates, and the random intercept. Such an additional effect of the school mean of a variable after controlling for the individual-level variable is sometimes referred to as a contextual effect. Here it could be due to the effect of the peers' SES, as well as any omitted variables that are correlated with school SES, such as school resources, parent involvement, and teacher qualifications.

p0035 If we had omitted school mean SES from the model, the estimated coefficient of individual SES would have been greater (0.41) due to absorbing some of the contextual effect. Using econometric terminology, we would have had an endogeneity problem because the random intercept, representing the combined effects of all omitted covariates at the cluster level (including school mean SES), would be correlated with individual SES. Another way of thinking about the problem is by remembering that the random intercept is assumed to be uncorrelated with the covariates; therefore, the effects of included covariates are not controlled for any possible school-level confounders. Such control can be achieved for an individual level-1 covariate by mean centering it and/or including its cluster mean as an additional covariate. Alternatively, we can eliminate cluster-level confounding for all level-1 covariates by specifying fixed effects for schools. In linear models, this could be accomplished by including indicator variables for schools. In logistic regression models, an appropriate approach is conditional maximum likelihood estimation. Using this method gives similar estimated coefficients except for a lower estimate of 0.46 for Eng_{ij}, suggesting that there might be an endogeneity problem or contextual effect for this variable.

t0005 **Table 1** Maximum likelihood estimates for logistic random intercept model

Parameter	*Est*	*(SE)*	*OR*	*(95% CI)*
γ_{00}	−2.02	(0.29)		
$\gamma_{01}\,[\overline{\text{SES}}_{\cdot j}]$	1.38	(0.18)	4.0	(2.8,5.7)
β_1 [Fem_{ij}]	0.56	(0.10)	1.7	(1.4,2.1)
β_2 [SES_{ij}]	0.29	(0.07)	1.3	(1.2,1.5)
β_3 [HS_{ij}]	0.39	(0.25)	1.5	(0.9,2.4)
β_4 [Coll_{ij}]	0.71	(0.23)	2.0	(1.2,3.3)
β_5 [Eng_{ij}]	0.62	(0.29)	1.9	(1.2,2.9)
ψ	0.27	(0.09)		

It is useful to visualize the magnitude of the effects of some variables on the probability of reading proficiency, holding constant other variables. For instance, we can consider boys whose parents' highest level of education is high school and who have SES equal to the overall mean, and obtain predicted probabilities as a function of school mean SES and the indicator variable for speaking English at home. A graph of these predicted probabilities (for the range of school mean SES in the data) is shown in **Figure 1**, where the dashed curves are for students who speak English at home and the solid curves are for students who do not speak English at home. For each line pattern, the curve that is closer to 0.5 represents the population averaged or marginal probability, after integrating or averaging out the random intercept, whereas the other curve represents the conditional or school-specific probability, with the random intercept set to 0, which is also the median probability. p0040

We can see that school mean SES has a large effect, with predicted probabilities differing by more than 0.4 between the lowest and highest SES schools. Speaking English at home increases the probability by about 0.1 for average and above average SES schools. In the very low SES schools, boys with average SES, but whose parents' highest level of education is high school, have very low probabilities of reading proficiency. p0045

Random Coefficient Models s0010

The random intercept model for the log-odds considered in the previous section had the form p0050

$$\eta_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \ldots + \beta_P x_{Pij}$$

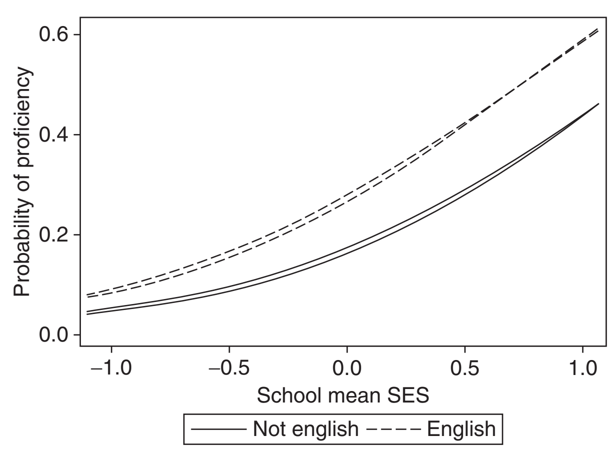


Figure 1 Predicted probability of reading proficiency as a function of school mean SES for students who do (dashed curves) and do not (solid curves) speak English at home (with $\text{Fem}_{ij} = 0$, $\text{SES}_{ij} = 0$, $\text{HS}_{ij} = 1$, and $\text{Coll}_{ij} = 0$). For each line pattern, the curves closer to 0.5 are marginal probabilities and the other curves, median probabilities. f0005

where β_{0j} is a school-specific random intercept and $x_{pij}(p = 1, \ldots, P)$ are level-1 covariates with fixed regression coefficients β_p.

p0055 In random coefficient models, not only is the intercept school specific, but the effects of at least some level-1 covariates also vary between schools. The simplest example is a model with a random intercept and a random slope for a covariate x_{1ij}. The level-1 model for the log-odds then has the form

$$\eta_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_2 x_{2ij} + \ldots + \beta_P x_{Pij}$$

where x_{1ij} has a school-specific regression coefficient or slope β_{1j}, whereas the other level-1 covariates have fixed regression coefficients $\beta_2, \ldots, \beta_P$.

p0060 The school-specific intercept and slope each, has its own level-2 model of the form

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\omega_{1j} + \ldots \gamma_{0Q}\,\omega_{Q\,j} + \zeta_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}\omega_{1j} + \ldots \gamma_{1Q}\,\omega_{Q\,j} + \zeta_{1j}$$

where $\omega_{1j}, \ldots, \omega_{Q\,j}$ are school-level covariates. The school-specific random intercept ζ_{0j} and random slope ζ_{1j} are typically assumed to have a bivariate normal distribution with zero means for given covariate values.

p0065 Substituting the level-2 models into the level-1 model produces a reduced form with cross-level interaction terms. For instance, the model for β_{1j} includes the term $\gamma_{11}\,\omega_{1j}$, and, since β_{1j} multiplies x_{1ij}, we obtain the cross-level interaction $\gamma_{11}\,\omega_{1j}x_{1ij}$. Random coefficient models could, of course, also include random effects of more than one covariate varying at level-1.

s0015 Different Response Types

p0070 All generalized linear mixed models have a so-called linear predictor η_{ij} of the form shown in the previous section. However, the relationship between the linear predictor and the observed response y_{ij} can be specified in a number of different ways depending on the type of response variable. First, a conditional distribution for y_{ij} is specified as a function of the conditional expectation μ_{ij}, given the covariates and random effects. Second, a link function $g(\cdot)$ is specified so that

$$g(\mu_{ij}) = \eta_{ij}$$

p0075 The conditional response distribution is from the exponential family and is characterized by the conditional expectation μ_{ij} as well as a dispersion parameter ϕ that affects the conditional variance

$$\mathrm{Var}(y_{ij}|\mu_{ij}) = \phi V(\mu_{ij})$$

where the variance function $V(\mu_{ij})$ is determined by the chosen distribution.

We briefly discuss the most common generalized linear mixed models for continuous responses, binary responses, and counts. p0080

Continuous Responses s0020

In the continuous case, a linear mixed model is typically assumed, in which the conditional response distribution is normal, p0085

$$y_{ij}|\mu_{ij} \sim N(\mu_{ij}, \sigma_e^2)$$

and the link function is the identity link

$$\mu_{ij} = \eta_{ij}$$

The variance function is $V(\mu_{ij}) = 1$ and the dispersion parameter is $\phi = \sigma^2$.

Binary Responses s0025

Binary responses are assumed to be independently Bernoulli distributed as p0090

$$y_{ij}|\mu_{ij} \sim \mathrm{Bernoulli}(\mu_{ij})$$

for given μ_{ij}. Here the conditional expectation μ_{ij} is also the conditional probability $\mathrm{P}(y_{ij} = 1)$ for given values of the covariates and random effects. The most common link function is the logit link

$$\eta_{ij} = \mathrm{logit}(\mu_{ij}) \equiv \log\left\{\frac{\mu_{ij}}{1-\mu_{ij}}\right\} = \log\left\{\frac{\mathrm{P}(y_{ij}=1)}{\mathrm{P}(y_{ij}=0)}\right\}$$

which can be interpreted as the log of the odds that y_{ij} is 1. An alternative link function is the probit link $\Phi^{-1}(\mu_{ij})$, the inverse standard normal cumulative distribution function. The variance function in either case is

$$V(\mu_{ij}) = \mu_{ij}(1-\mu_{ij})$$

and the dispersion parameter is equal to 1.

If there are several $(n_{ij} > 1)$ independent binary responses per unit ij, the conditional distribution of the number of responses that are 1 is binomial with probability of success μ_{ij} and number of trials n_{ij}. In this case, the variance function is $n_{ij}\,\mu_{ij}(1-\mu_{ij})$ and $\phi = 1$. However, if $n_{ij} > 1$, it is possible that the empirical variances are larger or smaller than the model-implied variances. Such overdispersion or underdisperson can be accommodated by estimating ϕ as a free parameter in a so-called quasi-likelihood approach. In likelihood or Bayesian methods, overdispersion is instead induced by including additional random effects varying over level-1 units. Note that there is no such thing as overdispersion or underdispersion if $n_{ij} = 1$, although researchers sometimes attempt to model it. p0095

s0030 *Latent response formulation*

p0100 In econometrics and psychometrics, models for binary responses are often specified by imaging an underlying or latent continuous response y^*_{ij} such that the observed response y_{ij} is 1 if the latent response exceeds 0 and y_{ij} is 0 otherwise. An obvious interpretation of y^*_{ij} in the proficiency example is as a continuous measure of reading achievement that must exceed a threshold for the student to be proficient. A linear mixed model is then specified for the latent response

$$y^*_{ij} = \eta_{ij} + e_{ij}$$

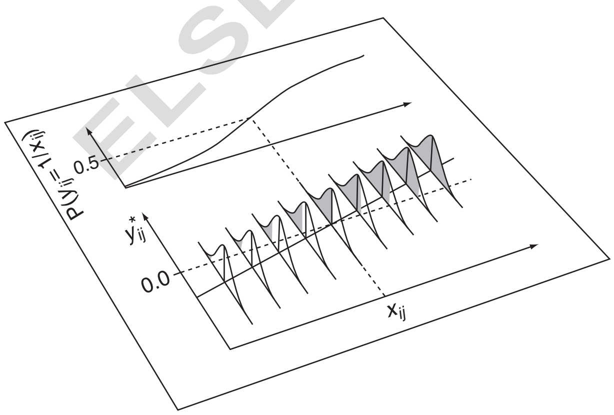
p0105 If a logistic distribution is assumed for e_{ij}, we obtain a logistic regression model, and, if a standard normal distribution is assumed, we obtain a probit regression model.

p0110 A simple latent response model with $\eta_{ij} = \beta_0 + \beta_1 x_{ij}$ is shown in the lower portion of **Figure 2**. The density curves represent the distribution of y^*_{ij} for different values of x_{ij}, with means falling on the upward sloping regression line. The threshold 0 is indicated by the dashed horizontal line, and the areas under the density curves exceeding this threshold are shaded. These probabilities that y^*_{ij} is greater than 0 represent the corresponding probabilities that $y_{ij} = 1$. The upper portion of the figure is a graph of these probabilities as a function of x_{ij}, and we can recognize the familiar logistic curve. The logistic regression model can thus either be specified through a logit link or using a latent response formulation.

p0115 The latent response formulation can be useful for interpreting the random part of the model. For a random intercept model, we can express the within-cluster dependence as the residual intraclass correlation among the latent responses, given the covariates

$$\text{cor}(y^*_{ij} \cdot y^*_{ij'} | \mu_{ij}) = \frac{\psi}{\psi + \text{Var}(e_{ij})}$$

where $\text{Var}(e_{ij})$ is $\pi^2/3$ in logistic models and 1 in probit models. For the estimates in **Table 1**, the residual intraclass correlation of the latent responses is estimated as 0.08. Hence, 8% of the residual variance in underlying reading achievement is due to schools. Although this intraclass correlation relies on the concept of a latent response, it is preferable to the correlation among the observed responses which depends on the covariate values.



f0010 **Figure 2** Illustration of the relationship between latent-response model and probability of observed response being 1. (Source: Rabe-Hesketh and Skrondal, 2008).

Counts s0035

Sometimes a response variable is a count of some event such as the number of days a student is absent from school in a year or the number of times a teacher shouts in an hour. The conditional response distribution is typically specified as Poisson, p0120

$$y_{ij}|\mu_{ij} \sim \text{Poisson}(\mu_{ij})$$

and a log link is used

$$\log(\mu_{ij}) = \eta_{ij}$$

The log link ensures that the expected count is nonnegative and produces a multiplicative model for the expected count.

The variance function is $V(\mu_{ij}) = \mu_{ij}$ and $\phi = 1$. Overdispersion or underdisperson can be accommodated by estimating ϕ as a free parameter, and overdispersion can be modeled by including additional random effects varying over level-1 units. The Poisson model cannot be specified through a latent response formulation. p0125

Other Response Types s0040

Ordinal responses include Likert scales for agreement with attitude statements (e.g., disagree, neither agree nor disagree, and agree) and reported frequencies of doing something such as helping children with homework (e.g., daily, several times per week, occasionally, and never). The most common models are so-called cumulative logit or probit models, and these can be specified as logit or probit models for the probabilities of exceeding each of the ordered categories (except the last). A parallel regression assumption is usually made by allowing only the intercept to take different values for different categories. The models can alternatively be specified using a latent response formulation with several, freely estimated thresholds. p0130

Discrete time durations are often of interest in education. Examples include number of semesters to dropout from college, age at which children first enter preschool, and number of years teachers teach at a school. A popular model for analyzing discrete time durations is the continuation ratio logit model. After expanding the data appropriately, this model can be estimated using logistic regression for binary responses. p0135

Unordered categorical data arise when people choose among different alternatives, such as community college, p0140

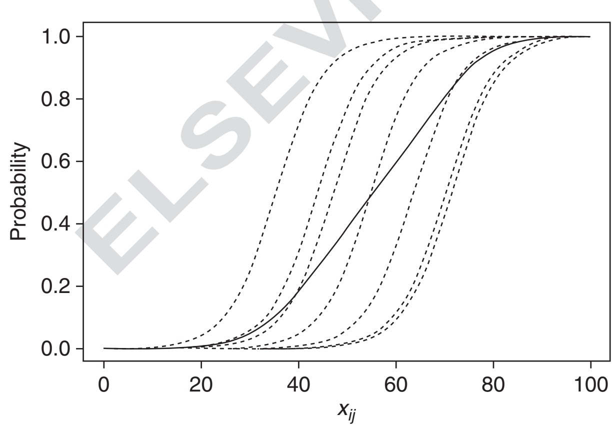
4-year college, or no higher education after graduation from high school. Multinomial logit models are typically used for such data.

s0045 Conditional and Marginal Relationships

p0145 The regression coefficients in generalized linear mixed models represent conditional effects in the sense that they express comparisons holding the cluster-specific random effects (and covariates) constant. For this reason, conditional effects are sometimes referred to as cluster-specific effects. In contrast, marginal effects can be obtained by averaging the conditional expectation μ_{ij} over the random effects distribution. Marginal effects express comparisons of entire sub-population strata defined by covariate values and are sometimes referred to as population-averaged effects.

p0150 In linear mixed models (identity link), the regression coefficents can be interpreted as either conditional or marginal effects. However, conditional and marginal effects differ for most other link functions. This can easily be seen for a random intercept logistic regression model with a single covariate in **Figure 3**. The cluster-specific, conditional relationships are shown as dotted curves with horizontal shifts due to different values of the random intercept. The population-averaged, marginal curve is obtained by averaging the conditional curves at each value of x_{ij}. We see that the marginal curve resembles a logistic curve with a smaller regression coefficient. Hence, the marginal effect of x_{ij} is smaller than the conditional effect.

p0155 The difference between conditional and marginal relationships is also visible in **Figure 1**, but it is much less pronounced due to the relatively small estimated random intercept variance.



f0015 **Figure 3** Conditional relationships (dotted curves) and marginal relationship (solid curve) for a random intercept logistic model. From Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latest Variable Modeling: Multilevel, Longitudinal, Structural Equation Models*, Boca Raton, FL: Chapman and Hall/CRC.

Estimation and Software s0050

The model parameters to be estimated in most generalized linear mixed models are the fixed regression coefficients and the covariance matrix of the random effects (the variance ψ for a random-intercept model). The most commonly used estimation methods are maximum likelihood, penalized quasi-likelihood (PQL-1), and Markov chain Monte Carlo (MCMC). p0160

Maximum likelihood estimation is not straightforward for generalized linear mixed models because the likelihood involves integrals that cannot be solved analytically. Many software packages therefore use numerical integration, typically adaptive or ordinary Gauss–Hermite quadrature. Examples include the Stata programs xtlogit, xtpoisson, xtmelogit, xtmepoisson, and gllamm; the SAS procedure NLMIXED; and the R-program lmer. The estimates in **Table 1** and the predictions for **Figure 1** were obtained using gllamm (see multimedia). p0165

Numerical integration is time consuming, particularly if the model includes several random effects. Approximate methods have therefore been suggested, including PQL-1 which is implemented in the SAS procedure GLIMMIX, the S-PLUS or R function glmmPQL, and the standalone programs HLM and MLwiN. Unfortunately, these methods sometimes produce biased estimates, in particular for binary responses, small cluster sizes and large intraclass correlations of the latent responses. More accurate approximations have therefore also been implemented (PQL-2 in MLwiN and Laplace6 in HLM). Software implementing MCMC for Bayesian estimation includes the general program BUGS or WINBUGS and the program MLwiN which is custom made for generalized linear mixed models. p0170

Assigning Values to Random Effects s0055

It is sometimes required to assign values to the random effects for individual clusters. For instance, random intercepts for schools or teachers in models for student outcomes can sometimes be viewed as measures of effectiveness since they represent the school- or teacher-specific value added taking into account after having taken observed covariates such as the prior achievement of students at intake. p0175

If MCMC is used for estimation, there is no real distinction between parameters and random effects, and estimates of the latter are obtained in a straightforward manner. If maximum likelihood or approximate methods are used for estimation, the parameters (typically the fixed regression coefficients and the covariance matrix of the random effects) are treated as known and replaced by their estimates for the purpose of assigning values to p0180 the random effects. The random effects are then either estimated by maximum likelihood or predicted using empirical Bayes.

p0185 Viewing the random effects as the only unknown parameters, they can be estimated by maximizing the joint probability distribution (or likelihood) of the responses, given the random effects and covariates, with respect to the random effects. This maximum likelihood estimation is performed independently for each cluster. In linear mixed models, such estimates are also known as ordinary least squares (OLS) estimates.

p0190 In empirical Bayes prediction of the random effects, we exploit the information that we have about the random effects before seeing the data for a cluster, namely the estimated random effects distribution, known as the prior distribution. The posterior distribution of the random effects for a cluster is proportional to the product of the prior distribution and the likelihood for the cluster. The mean of the posterior distribution is called the empirical Bayes predictor. In linear mixed models, these predictions are also known as best linear unbiased predictors (BLUPs).

p0195 The empirical Bayes predictions are shrunken toward zero compared with the maximum likelihood estimates. This shrinkage occurs because the prior distribution has its mean at zero. Shrinkage is negligible when the likelihood dominates the prior, for instance due to a large cluster size, but can be pronounced for small cluster sizes. Empirical Bayes predictors have lower mean squared prediction errors than maximum likelihood estimators and are, therefore, usually preferred.

p0200 There are generally no closed-form expressions for the empirical Bayes predictor except for linear mixed models. For other generalized linear mixed models, empirical Bayes predictions can be obtained by numerical integration.

s0060 Some Extensions

p0205 We have considered two-level models in this article which are applicable for typical educational datasets where students are nested in schools or where repeated observations on students over time are nested in students. There are often further levels of nesting. For instance, we may have repeated observations nested within students who are nested in schools. Alternatively, students may be nested in classrooms nested in schools, or schools may be nested in school districts or countries. In these cases, we can include random effects at each of the nested levels. More complex designs require crossed random effects. For instance, in longitudinal studies, students may be in middle school in the first two waves and, subsequently, high school in the next two waves. Middle schools are typically crossed with high schools in the sense that all students from a given middle school do not attend the same high school (or vice versa).

Estimation of models with nested random effects is not p0210
much more complex than estimation of standard generalized linear mixed models. In contrast, estimation of models with crossed random effects remains a challenge, except in linear mixed models.

The random part of the linear predictor in generalized p0215
linear mixed models is quite restrictive because each random effect ζ_{pj} merely multiplies an observed variable x_{pij} in the reduced form. Generalized linear latent and mixed models (GLLAMMs) extend generalized linear mixed models by allowing the random effects to be multiplied by different parameters for different responses. Furthermore, the random effects can be regressed on observed covariates and other random effects (at the same or higher hierarchical levels). This is useful for incorporating measurement models, such as IRT models, within generalized linear mixed models.

See also: Categorical data analysis (01311); Empirical Bayes Methods (01325); Generalized Linear Models (01331); Growth modeling (01335); Hierarchical Linear Model (01336); Value-added models (01374).

Bibliography

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman and Hall/CRC.

Further Reading

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.

Clayton, D. G. (1996). Generalized linear mixed models. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds.) *Markov Chain Monte Carlo in Practice*, pp 275–301. London: Chapman and Hall.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Goldstein, H. (2003). *Multilevel Statistical Models*, 3rd edn. London: Arnold.

Goldstein, H. and Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A* **159**, 505–513.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.

McCulloch, C. E., Searle, S. R., and Neuhaus, J. M. (2008). *Generalized, Linear and Mixed models*, 2nd edn. New York: Wiley.

Rabe-Hesketh, S. and Skrondal, A. (2008). Generalized linear mixed effects models. In Fitzmaurice, G. M., Davidian, M., Verbeke, G., and Molenberghs, G. (eds.) *Longitudinal Data Analysis*, pp 79–106. Boca Raton, FL: Chapman and Hall/CRC.

Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata*, 2nd edn. College Station, TX: Stata Press.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika* **69**, 167–190.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable

models with nested random effects. *Journal of Econometrics* **128**, 301–323.
Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational and Behavioral Statistics* **18**, 321–349.
Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models*, 2nd edn. Thousand Oaks, CA: Sage.
Raudenbush, S. W., Yang, M.-L., and Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics* **9**, 141–157.
Ritz, J. and Spiegelman, D. (2004). A note about the equivalence of conditional and marginal regression models. *Statistical Methods in Medical Research* **13**, 309–323.
Singer, J. D. and Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics* **18**, 155–195.
Skrondal, A. and Rabe-Hesketh, S. (2003). Multilevel logistic regression for polytomous data and rankings. *Psychometrika* **68**, 267–287.
Skrondal, A. and Rabe-Hesketh, S. (2007). Redundant overdispersion parameters in multilevel models. *Journal of Behavioral and Educational Statistics* **32**, 419–430.
Skrondal, A. and Rabe-Hesketh, S. (in press). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society, Series A*.
Snijders, T. A. B. and Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.

Relevant Websites

http://www.gllamm.org – Gllamm, State Programs for estimating, predicting, stimulating Generalized Linear Latent and Mixed Models.
http://www.ssicentral.com/htm – SSI, Scientific Software International, Hierarchical Linear and Nonlinear Models.
http://www.cmm.bristol.ac.uk – University of Bristol, Centre for Multilevel Modelling.

EDUC: 01332

Non-Print Items

Abstract:
Generalized linear mixed models are used to model response variables of different types, including continuous and binary responses and counts, when the data are clustered. Clustered data are common in education, canonical examples being students nested in schools and longitudinal responses nested in students. This article introduces the models through an example with a binary response variable and discusses conditional and marginal effects, estimation and software, and assignment of values to random effects.

Generalized linear mixed models extend linear mixed models, or hierarchical linear models, to accommodate noncontinuous responses, such as binary responses or counts. Such models are useful when the data are clustered in some way, a canonical example in education being students nested in schools. Another important example is longitudinal data where repeated observations of the same students are nested within students. Another term for generalized linear mixed models is hierarchical or multilevel generalized linear models. The terms random coefficient models or random effects models are also often used for either linear or generalized linear mixed models.

We start by fitting a particular kind of generalized linear mixed model, a logistic random intercept model, to data on reading proficiency of students nested in schools. We then discuss random coefficient models and models for different response types, conditional and marginal relationships, estimation and software, and assignment of values to random effects. We close by briefly considering some useful extensions to the models discussed here.

Keywords: Adaptive quadrature; Conditional effect; Empirical Bayes; Generalized linear mixed model; Gllamm; Hierarchical generalized linear model; Logistic regression; Logit link; Marginal effect; Multilevel generalized linear model; Poisson regression; Probit link; Random coefficients; Random effects

Author and Co-author Contact Information:

S Rabe-Hesketh
3659 Tolman Hall #1670
Graduate School of Education
University of California
Berkeley
CA 94720
USA
sophiarh@berkeley.edu

A Skrondal
Division of Epidemiology
Norwegian Institute of Public Health
PO Box 4404 Nydalen
Olso N-0403
Norway
anders.skrondal@fhi.no

Biographical Sketch for Online Version

Sophia Rabe-Hesketh is a professor at the Graduate School of Education and the Graduate Group in Biostatistics at the University of California, Berkeley. She is also chair of social statistics at the Institute of Education, University of London. She is an applied statistician interested in multilevel, longitudinal, and latent variable modeling. Her recent research has included the development of the generalized linear latent and mixed (GLLAMM) modeling framework that unifies and extends multilevel and latent variable models. She has written the gllamm software to estimate all the models within the GLLAMM framework. The software has been used in over 250 peer-reviewed papers indexed in the Web of Science between 2002 and 2008. Rabe-Hesketh is co-editor of *Statistical Methods in Medical Research*, associate editor for several journals including *Psychometrika*, and a member of several editorial boards. She has co-authored 82 papers in journals including *Psychometrika, Biometrics, Journal of Econometrics*, and *Scandinavian Journal of Statistics*. Her five books include the best-sellers *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models* and *Multilevel and Longitudinal Modeling Using Stata*.

Anders Skrondal is senior biostatistician in the Division of Epidemiology, Norwegian Institute of Public Health. He has previously held positions as professor of statistics and director of the Methodology Institute at the London School of Economics. His research interests include topics in biostatistics, social statistics, econometrics, and psychometrics. Recently, Skrondal has concentrated on the development of the generalized linear latent and mixed model (GLLAMM) framework. Outcomes include papers published in (bio)statistical, psychometric, and econometric journals, and two bestselling books: *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models* published by Chapman and Hall/CRC in 2004 and *Multilevel and Longitudinal Modeling using Stata* published by Stata Press in 2005 and 2008 (second edition). Skrondal is involved in numerous collaborative projects within medical, social, and behavioural research, and is currently co-editor of *Statistical Methods in Medical Research*. Anders Skrondal was awarded the 1997 Psychometric Society Dissertation Prize and is an elected member of the Research Section Committee of the Royal Statistical Society and the International Statistical Institute.