

STAT 705 Generalized linear mixed models

Adapted from Timothy Hanson

Department of Statistics, University of South Carolina

Stat 705: Data Analysis II

Generalized Linear Mixed Models

We have considered random block effects in two models: randomized complete block and repeated measures within one factor designs. The idea is that within a block the responses are more similar (i.e. correlated) than across blocks.

Observations often occur in related clusters. Phrases like *repeated measures*, *longitudinal data*, and *panel data*, get at the same thing: there's correlation among observations in a cluster.

One approach, the topic of this lecture, induces correlation through the inclusion of *random effects*, yielding a GLMM.

Another approach, *generalized estimating equations*, accounts for correlation by modifying “independence” likelihood equations. This is not discussed here, but rather in STAT 770 & STAT 771.

Add random effects to linear predictor

Let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})$ be n_i correlated responses in cluster i . Associated with each repeated measure Y_{ij} are fixed (population) effects β and cluster-specific random effects \mathbf{u}_i . Let $\mu_{ij} = E(Y_{ij})$.

In a GLMM the linear predictor is augmented to include random effects:

$$g(\mu_{ij}) = \mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\mathbf{u}_i.$$

For logistic regression, this is

$$\text{logit } P(Y_{ij} = 1) = \mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\mathbf{u}_i.$$

Add random effects to linear predictor

Note that conditional on \mathbf{u}_i ,

$$E(Y_{ij}|\mathbf{u}_i) = \frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i}}{1 + e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i}}.$$

The distinction between the conditional model and a marginal model is subtle, but very important. The interpretation of $\boldsymbol{\beta}$ is slightly different in both cases, and estimation of $\boldsymbol{\beta}$ differs depending on the model. Estimation of marginal models was developed first.

Example

I ask a random sample of *the same* $n = 14$ STAT 705 graduate students “do you like statistics?” once a month for 4 months.

$Y_{ij} = 1$ if “yes” and $Y_{ij} = 0$ if no. Here, $i = 1, \dots, 14$ and $j = 1, \dots, 4$.

Covariates might include m_{ij} , the average mood of the student over the previous month ($m_{ij} = 0$ is bad, $m_{ij} = 1$ is good), the degree being sought ($d_i = 0$ doctoral, $d_i = 1$ masters), the month $t_j = j$, and p_j the number of homework problems assigned in STAT 705 in the previous month.

A GLMM might be

$$\text{logit } P(Y_{ij} = 1) = \beta_0 + \beta_1 m_{ij} + \beta_2 d_i + \beta_3 p_j + \beta_4 j + u_i.$$

This model assumes that log-odds of liking statistics changes linearly in time, holding all else constant. Alternatively, we might fit a quadratic instead or treat time as categorical. Here, u_i represents a student's *a priori* disposition towards statistics.

Let's compare month $j + 1$ to month j for individual i , holding all else (m , d , and p) constant. The difference in log odds is

$$(\beta_0 + \beta_1 m_{ij} + \beta_2 d_i + \beta_3 p_j + \beta_4(j + 1) + u_i) - (\beta_0 + \beta_1 m_{ij} + \beta_2 d_i + \beta_3 p_j + \beta_4 j + u_i) = \beta_4.$$

Not holding everything constant we get

$$\begin{aligned} & (\beta_0 + \beta_1 m_{i,j+1} + \beta_2 d_i + \beta_3 p_{j+1} + \beta_4(j + 1) + u_i) - (\beta_0 + \beta_1 m_{ij} + \beta_2 d_i + \beta_3 p_j + \beta_4 j + u_i) \\ & = \beta_1(m_{i,j+1} - m_{ij}) + \beta_3(p_{j+1} - p_j) + \beta_4. \end{aligned}$$

Either way, we are conditioning on individual i , or *the subpopulation of all individuals with predisposition u_i* ; i.e. everyone “like” individual i in terms of liking statistics to begin with.

How are e^{β_1} , e^{β_2} , e^{β_3} and e^{β_4} interpreted here?

Random effects $\mathbf{u}_1, \dots, \mathbf{u}_n$

The random effects are assumed to come from (in general) a multivariate normal distribution

$$\mathbf{u}_1, \dots, \mathbf{u}_n \stackrel{iid}{\sim} N_q(\mathbf{0}, \mathbf{\Sigma}).$$

The covariance $\text{cov}(\mathbf{u}_i) = \mathbf{\Sigma}$ can have special structure, but is usually unstructured. The free elements of $\mathbf{\Sigma}$ are estimated along with β .

The \mathbf{u}_i can account for heterogeneity caused by omitting explanatory variables and/or are interpreted as blocking effects.

Note that $\text{cov}(\mathbf{Y}_i) \neq \mathbf{\Sigma}$! In general, the covariance of \mathbf{Y}_i does not have closed form unless \mathbf{Y}_i is normal, rather than Bernoulli or Poisson.

Two special cases

Logistic regression with random intercepts:

$$\text{logit } P(Y_{ij} = 1|u_i) = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_i, \quad u_i \stackrel{iid}{\sim} N(0, \sigma^2).$$

Poisson regression with random intercepts:

$$Y_{ij}|u_i \sim \text{Poisson}(\lambda_{ij}t_{ij}), \quad \log(\lambda_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_i, \quad u_i \stackrel{iid}{\sim} N(0, \sigma^2).$$

When $\sigma = 0$ we get the standard logistic or Poisson regression models; when $\sigma > 0$ we account for extra heterogeneity in clustered responses (each i is a cluster with its own random u_i).

GLMMs induce only *positive* correlation between two observations Y_{ij_1} and Y_{ij_2} within the same cluster.

Fitting binary GLMMs

The general model is hierarchical:

$$Y_{ij} | \mathbf{u}_i \stackrel{ind.}{\sim} \text{Bern} \left(\frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i}}{1 + e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i}} \right),$$
$$\mathbf{u}_1, \dots, \mathbf{u}_n \stackrel{iid}{\sim} N_q(\mathbf{0}, \boldsymbol{\Sigma}).$$

Conditional on the random effect \mathbf{u}_i , the elements in $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT_i})$ are independent. So the PDF of $\mathbf{Y}_i | \mathbf{u}_i$ is

$$p(\mathbf{y}_i | \mathbf{u}_i) = \prod_{j=1}^{n_i} \left(\frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i}}{1 + e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i}} \right)^{y_{ij}} \left(\frac{1}{1 + e^{\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i}} \right)^{1-y_{ij}}.$$

However, the $\mathbf{u}_1, \dots, \mathbf{u}_n$ are not model parameters. The model parameters are $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$. We need to maximize the likelihood

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = p(\mathbf{y}_1, \dots, \mathbf{y}_n | \boldsymbol{\beta}, \boldsymbol{\Sigma}).$$

Integrate to get likelihood

The *unconditional* PDF of \mathbf{Y}_i is

$$p(\mathbf{y}_i) = \int_{\mathbb{R}^q} \left[\prod_{j=1}^{n_i} \frac{(e^{\mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\mathbf{u}_i})^{y_{ij}}}{1 + e^{\mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\mathbf{u}_i}} \right] p(\mathbf{u}_i | \boldsymbol{\Sigma}) d\mathbf{u}_i,$$

where $p(\mathbf{u}_i | \boldsymbol{\Sigma})$ is a $N_q(\mathbf{0}, \boldsymbol{\Sigma})$ PDF. The \mathbf{u}_i is integrated out and this is a function of $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ only. The likelihood is the product of these

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \prod_{i=1}^n \int_{\mathbb{R}^q} \left[\prod_{j=1}^{n_i} \frac{(e^{\mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\mathbf{u}_i})^{y_{ij}}}{1 + e^{\mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\mathbf{u}_i}} \right] p(\mathbf{u}_i | \boldsymbol{\Sigma}) d\mathbf{u}_i.$$

This involves n q -dimensional integrals that do not have closed-form.

This is how correlation is induced among the $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})$.

Building a Poisson GLMM likelihood is similar.

- PROC NL MIXED estimates the integrals (for a “current” quasi-Newton value of (β, Σ)) using adaptive Gauss-Hermite quadrature. This approach approximates the integrals above by sums

$$\int_{\mathbb{R}^q} h(\mathbf{u}_i) p(\mathbf{u}_i | \Sigma) d\mathbf{u}_i \approx \sum_{k=1}^Q c_k h(\mathbf{s}_k),$$

for arbitrary $h(\cdot)$ where Q is the number of quadrature points $\mathbf{s}_1, \dots, \mathbf{s}_Q$ and c_1, \dots, c_Q are weights. The (adaptive) quadrature points and weights are chosen from a theory on integral approximations.

- PROC GLIMMIX essentially uses $Q = 1$. GLIMMIX is by far the fastest approach in SAS.
- PROC MCMC uses Markov chain Monte Carlo for updating all parameters, including random effects.

GLIMMIX extends the MIXED procedure to GLM's, and in fact iteratively calls MIXED when fitting GLMM's.

Only normal random effects are allowed.

GLIMMIX uses an approximation when fitting models. The approximation in effect replaces an intractable integral in the likelihood with a simple linear Taylor's expansion. See SAS' GLIMMIX documentation for details on "Pseudo-likelihood Estimation Based on Linearization."

Longitudinal study of mental health

Agresti (2013) considers data from a longitudinal study comparing a new drug with a standard drug for treatment of subjects suffering mental depression. $n = 340$ Patients were either mildly or severely depressed upon admission into the study. At weeks 1, 2, and 4, corresponding to $j = 1, 2, 3$, patient i 's suffering Y_{ij} was classified as normal $Y_{ij} = 1$ or abnormal $Y_{ij} = 0$. Let $s_i = 0, 1$ be the severity of the diagnosis (mild, severe) and $d_i = 0, 1$ denote the drug (standard, new).

We treat time as a categorical predictor and fit a marginal logit model with an exchangeable correlation structure:

```
data depress;
  infile "c:/tim/cat/depress.txt";
  input case diagnose treat time outcome; time=time+1;
run;
proc glimmix; class case time;
  model outcome = diagnose treat time treat*time / dist=bin link=logit solution;
  random case;
  estimate "new vs. std @ week 1" treat 1 treat*time 1 0 0 / exp cl;
  estimate "new vs. std @ week 2" treat 1 treat*time 0 1 0 / exp cl;
  estimate "new vs. std @ week 4" treat 1 treat*time 0 0 1 / exp cl;
run;
```

A conditional, random effects model

Model interpretation is *within the individual*; i.e. blocked on individual.

The model is

$$\begin{aligned} \text{logit } P(Y_{ij} = 1) &= \beta_0 + \beta_1 s_i + \beta_2 d_i + \beta_3 I\{j = 1\} + \beta_4 I\{j = 2\} \\ &\quad + \beta_5 I\{j = 1\}d_i + \beta_6 I\{j = 2\}d_i + u_i \\ \text{where } u_i &\sim N(0, \sigma^2). \end{aligned}$$

The model parameters are β_0, \dots, β_6 along with σ .

Output from PROC GLIMMIX

Covariance Parameter Estimates

| Cov | | Standard |
|------|----------|----------|
| Parm | Estimate | Error |
| case | 0.01838 | 0.1330 |

Solutions for Fixed Effects

| Effect | time | Estimate | Standard Error | DF | t Value | Pr > t |
|------------|------|----------|----------------|-----|---------|---------|
| Intercept | | 0.9813 | 0.1813 | 337 | 5.41 | <.0001 |
| diagnose | | -1.3119 | 0.1470 | 676 | -8.93 | <.0001 |
| treat | | 2.0435 | 0.3060 | 676 | 6.68 | <.0001 |
| time | 1 | -0.9599 | 0.2290 | 676 | -4.19 | <.0001 |
| time | 2 | -0.6205 | 0.2245 | 676 | -2.76 | 0.0059 |
| time | 3 | 0 | . | . | . | . |
| treat*time | 1 | -2.0989 | 0.3894 | 676 | -5.39 | <.0001 |
| treat*time | 2 | -1.0965 | 0.3838 | 676 | -2.86 | 0.0044 |
| treat*time | 3 | 0 | . | . | . | . |

New is better than old

Type III Tests of Fixed Effects

| Effect | Num DF | Den DF | F Value | Pr > F |
|------------|-----------|-----------|---------|--------|
| diagnose | 1 | 676 | 79.69 | <.0001 |
| treat | 1 | 676 | 41.09 | <.0001 |
| time | 2 | 676 | 9.12 | 0.0001 |
| treat*time | 2 | 676 | 14.73 | <.0001 |

| Label | Lower | Upper | Exponentiated Estimate | Exponentiated Lower | Exponentiated Upper |
|----------------------|---------|--------|---------------------------|------------------------|------------------------|
| new vs. std @ week 1 | -0.5296 | 0.4188 | 0.9461 | 0.5888 | 1.5201 |
| new vs. std @ week 2 | 0.4855 | 1.4085 | 2.5780 | 1.6250 | 4.0899 |
| new vs. std @ week 4 | 1.4426 | 2.6444 | 7.7178 | 4.2318 | 14.0757 |

We see a severe diagnosis ($s = 1$) significantly decreases the odds of a normal classification by a factor of $e^{-1.31} = 0.27$.

The odds (of normal classification) ratio comparing the new drug to the standard drug changes with time because of the interaction. At 1 week it's $e^{2.04-2.10} = 0.95$; at week 2 it's $e^{2.04-1.10} = 2.6$, and at 4 weeks it's $e^{2.04-0} = 7.7$. The new drug is better, but takes time to work.

Ache monkey hunting

Data on the number of capuchin monkeys killed by 47 Ache hunters over several hunting trips were recorded. There were 363 total records. Recall that monkey hunting involves splitting into groups, chasing monkeys through the trees, and shooting arrows straight up.

Let Y_{ij} be the number of monkeys killed by hunter i , $i = 1, \dots, 47$ on trip j of length L_{ij} (the trip length serves as an 'offset' in the model fitting). Let λ_i be the hunter i 's kill rate (per day).

$$Y_{ij} \sim \text{Poisson}(\lambda_i L_{ij}),$$

where

$$\log \lambda_i = \beta_0 + \beta_1 a_i + \beta_2 a_i^2 + u_i,$$

$$u_1, \dots, u_{47} \stackrel{iid}{\sim} N(0, \sigma^2).$$

We include a quadratic effect because we expect a “leveling off” effect or possible decline in ability with age.

Of interest is when hunting ability is greatest. Hunting prowess contributes to a man's status within the group. a_i is hunter i 's age-45 years.

An individual's kill rate is given by $\lambda = e^{\beta_0 + \beta_1 a + \beta_2 a^2} e^u$, where a is the individual's age and u is their latent hunting ability, i.e. their blocking effect.

One can compare the effect of age within the span of, say, 20 to 60 years, to the spread of e^u to see which explains more of the variability in terms of hunting ability: age or innate ability.

Monkey hunting SAS code

```
data ache1;
  input TRIP$ PID$ AGE nkills tripdays; ltripday=log(tripdays); age=age-45;
  datalines;
C082697A    3394    31  1  4
C082697A    3327    38  0  4
C082697A    3313    39  0  4
C082697A    3220    50  0  4
C082697A    3157    56  0  4
C082697A    3146    57  0  4
C082697A    3144    58  1  4
C082697A    7089    59  1  4
C082697A    3126    60  2  4
C082697A    7085    62  1  4
C102197A    3394    31  1  3
C102197A    3327    38  0  3
C102197A    3238    48  3  3
C102197A    3220    50  0  3
C102197A    3144    58  2  3
C102197A    3086    67  0  3

    ...et cetera...

T120997A    3182    53  0  5
T120997A    3094    65  0  5
T121597A    3254    46  0  4
T121597A    3128    60  0  4
;
proc glimmix data=ache1; class pid;
  model nkills = age age*age / dist=pois link=log offset=ltripday solution;
  random intercept / subject=pid;
```

GLIMMIX GLMM output

The GLIMMIX Procedure

Model Information

| | |
|----------------------------|----------|
| Response Variable | nkills |
| Response Distribution | Poisson |
| Link Function | Log |
| Variance Function | Default |
| Offset Variable | ltripday |
| Variance Matrix Blocked By | PID |

Dimensions

| | |
|------------------------|----|
| G-side Cov. Parameters | 1 |
| Subjects (Blocks in V) | 47 |
| Max Obs per Subject | 28 |

Covariance Parameter Estimates

| Cov Parm | Subject | Estimate | Standard Error |
|-----------|---------|----------|----------------|
| Intercept | PID | 1.7965 | 0.6505 |

Solutions for Fixed Effects

| Effect | Estimate | Standard Error | DF | t Value | Pr > t |
|-----------|----------|----------------|-----|---------|---------|
| Intercept | -2.4222 | 0.4113 | 46 | -5.89 | <.0001 |
| AGE | 0.02889 | 0.02307 | 314 | 1.25 | 0.2115 |
| AGE*AGE | -0.00405 | 0.002079 | 314 | -1.95 | 0.0525 |

GLMM example with bivariate random effects vectors

Thall and Vail (1990) presented data from a clinical trial of $n = 59$ epileptic patients who were randomized to take either a new drug $d_i = 1$ or a placebo ($d_i = 0$) in addition to standard chemotherapy. Other baseline data included $a_i = \log(\text{age}_i)$ where age_i is age in years and $b_i = \log(\text{base}_i)$, where base_i is number of seizures in preceding 8-week period. The outcome is Y_{ij} the number of seizures within the following 4 2-week periods, up to 8 weeks. So $j = 1, 2, 3, 4$. The model fit is

$$Y_{ij} \sim \text{Poisson}(\lambda_{ij}),$$

where

$$\begin{aligned} \log \lambda_{ij} &= \beta_0 + \beta_b b_i + \beta_d d_i + \beta_{bd} b_i d_i + \beta_a a_i + \beta_v t_j + u_{i1} + u_{i2} t_j \\ &= \begin{bmatrix} 1 \\ b_i \\ d_i \\ b_i d_i \\ a_i \\ t_j \end{bmatrix}' \begin{bmatrix} \beta_0 \\ \beta_b \\ \beta_d \\ \beta_{bd} \\ \beta_a \\ \beta_v \end{bmatrix} + \begin{bmatrix} 1 \\ t_j \end{bmatrix}' \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix} \\ &= \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{u}_i \end{aligned}$$

We further assume

$$\mathbf{u}_1, \dots, \mathbf{u}_{59} \stackrel{iid}{\sim} N_2(\mathbf{0}, \mathbf{\Sigma}) = N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right).$$

This assumes a log linear trend individual i 's seizure rate over the 8 weeks. Specifically,

$$\log \lambda_{ij} = \theta_{0i} + \theta_{1i} \text{weeks}_j,$$

where $\text{cov}(\theta_{0i}, \theta_{1i}) = \sigma_{12}$. This follows from properties of multivariate normal distributions.

```
data seizure;
  input id$ seiz visit treat age base;
  age=log(age);
  base=log(base);
  datalines;
101  11    1  1  18  76
101  14    2  1  18  76
101   9    3  1  18  76
101   8    4  1  18  76
102   8    1  1  32  38
102   7    2  1  32  38
102   9    3  1  32  38
102   4    4  1  32  38
      ...et cetera...
238  13    1  0  22  47
238  15    2  0  22  47
238  13    3  0  22  47
238  12    4  0  22  47
;
proc glimmix; class id;
  model seiz=base|treat age visit / dist=poi link=log solution;
  random intercept visit / subject=id type=un;
  nloptions maxiter=50; * did not converge with default=20;
run;
```

Covariance Parameter Estimates

| Cov Parm | Subject | Estimate | Standard Error |
|-------------|---------|----------|-------------------|
| UN(1,1) | id | 0.4052 | 0.1251 |
| UN(2,1) | id | -0.05393 | 0.02964 |
| UN(2,2) | id | 0.02174 | 0.009264 |

Solutions for Fixed Effects

| Effect | Estimate | Standard Error | DF | t Value | Pr > t |
|------------|----------|-------------------|-----|---------|---------|
| Intercept | -2.3423 | 1.2616 | 55 | -1.86 | 0.0687 |
| base | 0.8700 | 0.1351 | 117 | 6.44 | <.0001 |
| treat | -1.3682 | 0.6884 | 117 | -1.99 | 0.0492 |
| base*treat | 0.3304 | 0.2090 | 117 | 1.58 | 0.1166 |
| age | 0.4627 | 0.3568 | 117 | 1.30 | 0.1973 |
| visit | -0.05276 | 0.03137 | 58 | -1.68 | 0.0980 |

Consider an individual from the population with covariates (a, b) at time t with random effect (u_1, u_2) . The ratio of seizure rates, within this individual, for drug versus placebo is:

$$\frac{\lambda(a, b, t, d = 1|\mathbf{u})}{\lambda(a, b, t, d = 0|\mathbf{u})} = \frac{e^{\beta_0 + \beta_b b + \beta_d + \beta_{bd} b + \beta_a a + \beta_v t + u_1 + u_2 t}}{e^{\beta_0 + \beta_b b + \beta_a a + \beta_v t + u_1 + u_2 t}} = e^{\beta_d + \beta_{bd} b}.$$

Within a subject, the mean number of seizures over a 2-week period is reduced by $e^{-1.37 + 0.330 \log(\text{base})} = (0.25)\text{base}^{0.33}$.

This function crosses unity around 65 baseline seizures within the previous 8 weeks. It's about 0.5 when $\text{base} = 8$. So the drug significantly reduces seizures at any visit, but the reduction rate critically depends on the baseline seizure rate.