Generalized Linear Mixed Models

Bruce Craig - Purdue University
## Linear Models

<table>
<thead>
<tr>
<th>Response variable</th>
<th>Explanatory</th>
<th>Explanatory / Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fixed - Categorical</td>
<td>Fixed - Continuous</td>
</tr>
<tr>
<td>Gaussian (normal)</td>
<td>“General” Linear Model</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$y = X\beta + e$</td>
<td></td>
</tr>
<tr>
<td>Non-Gaussian (non-normal)</td>
<td>Generalized Linear Model</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GLM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>???</td>
<td></td>
</tr>
</tbody>
</table>
Modeling Non-Normal Responses

\[ y \sim X\beta + Zu + \epsilon \]

Not defined!
Modeling Non-Normal Responses

\[ y = X\beta + Zu \]

We will need a bridge…

**Link function!**
A exponential family distribution has the probability mass/distribution function in the form

\[
f(y; \theta, \phi) = \exp \left( \frac{t(y)\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right)
\]

- \( \theta \) canonical form of parameter representing location (also called natural parameter)
- \( \phi \) dispersion parameter representing the scale
- \( a(\cdot), b(\cdot), c(\cdot), \) and \( t(\cdot) \) are known functions that depend on the data, parameters, or both

Normal, Binomial, Poisson, Beta, Gamma,...
Examples

- $y_i \sim \text{Poisson}(\lambda_i) \rightarrow f(y_i) = \exp\left(y_i \log(\lambda_i) - \lambda_i - \log(y_i !)\right)$
  \[
  \theta_i = \log(\lambda_i), \quad \phi = 1, \quad \text{and } t(y_i) = y_i
  \]
  \[
  a(\phi) = 1, \quad b(\theta_i) = \exp(\theta_i), \quad \text{and } c(y_i, \phi) = -\log(y_i !)
  \]

- $y_i \sim \text{Binomial}(n_i, p_i) \rightarrow f(y_i) = \exp\left( y_i \log\left(\frac{p_i}{1 - p_i}\right) + n_i \log(1 - p_i) + \log\left(\frac{n_i}{y_i}\right)\right)$
  \[
  \theta_i = \log\left(\frac{p_i}{1 - p_i}\right), \quad \phi = 1, \quad t(y_i) = y_i
  \]
  \[
  a(\phi) = 1, \quad b(\theta_i) = n_i \log\left(1 + \exp(\theta_i)\right), \quad c(y_i, \phi) = \log\left(\frac{n_i}{y_i}\right)
  \]
Properties of Exponential Family

- When \( \phi \) is known, the dist belongs to the one parameter exponential family (binomial, Poisson).
- When \( \phi \) is unknown, the dist belongs to the two parameter exponential family (normal, gamma).
- The canonical parameter \( \theta \) plays an important role in generalized linear models (canonical link function).
- The mean and variance can be shown to be a function of \( \theta \)

\[
E(Y) = b'(\theta) \\
\text{Var}(Y) = b''(\theta)a(\phi)
\]
Returning to Examples

- Poisson:
  \[ b'(\theta_i) = \exp(\theta_i) = \lambda_i \]
  \[ b''(\theta_i) = \exp(\theta_i) = \lambda_i \rightarrow \text{Var} = \lambda_i \times 1 = \lambda_i \]

- Binomial
  \[ b'(\theta_i) = n_i \frac{\exp(\theta_i)}{1 + \exp(\theta_i)} = n_i \frac{p_i}{1 - p_i} (1 - p_i) = n_i p_i \]
  \[ b''(\theta_i) = n_i \frac{\exp(\theta_i)}{(1 + \exp(\theta_i))^2} = n_i \frac{p_i}{1 - p_i} (1 - p_i)^2 = n_i p_i (1 - p_i) \]
  \[ \text{Var} = n_i p_i (1 - p_i) \times 1 = n_i p_i (1 - p_i) \]
Generalized Linear Model

- Let Y be response variable with a probability distribution from the exponential family.
- We can associate the E(Y) to a linear combination of predictors through a link function.
- We transform the mean of the response Y through this link function, we do not transform Y itself...thus log-transforming counts and fitting with a traditional LM is not the same.
Modeling Non-Normal Responses

Distribution of the response

\[ y \sim \text{Distribution} \]
\[ E(y) = \mu \]

Data Scale or Inverse-Link Scale

Linear predictor

\[ \eta = X\beta \]
\[ \eta = X\beta + Zu \]

Model Scale or Link Scale

Generalized Linear (Mixed) Model

Link function

\[ \eta = g(\mu) \]
## Link Functions for Common Responses

<table>
<thead>
<tr>
<th>Data Distribution</th>
<th>Type of Response</th>
<th>Range of variable</th>
<th>Mean</th>
<th>Common link function</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Continuous</td>
<td>$-\infty &lt; y &lt; \infty$</td>
<td>$\mu$</td>
<td>Identity $\eta = \mu$</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>Binomial</td>
<td>Discrete proportion</td>
<td>$\frac{0}{N}, \frac{1}{N}, ... \frac{N}{N}$</td>
<td>$\pi = \frac{\mu}{N}$</td>
<td>Logit $\eta = \log \left( \frac{\pi}{1-\pi} \right)$</td>
<td>$N\pi(1-\pi)$</td>
</tr>
<tr>
<td>Poisson</td>
<td>Discrete count</td>
<td>$y = 0,1,2...$</td>
<td>$\lambda$</td>
<td>$\eta = \log(\lambda)$</td>
<td>1</td>
</tr>
<tr>
<td>Negative Binomial</td>
<td>Discrete count</td>
<td>$y = 0,1,2...$</td>
<td>$\lambda$</td>
<td>$\eta = \log(\lambda)$</td>
<td>$\lambda + \phi\lambda^2$ GLIMMIX parameterization</td>
</tr>
</tbody>
</table>
Inference in GLM

- Parameters estimated via maximum likelihood
- Nonlinear optimization problem so common to transform into a series of weighted least squares fits
- Hypothesis tests are constructed using Wald-type statistics based on MLEs
- Statistics assumed asymptotically Normal
- Inference performed on transformed scale
Inference in GLMM

- Parameters also estimated via maximum likelihood

- Two common computational approaches
  1. Pseudo-likelihood
  2. Integral approximation of the log-likelihood
     - Laplace Approximation
     - Adaptive Quadrature

- Inference performed on transformed scale...back-transform and approximate standard errors based on delta method
Pseudo-Likelihood (PL)

- Applicable to a broader set of models
- There is a Kenward-Roger adaptation
  - Ad hoc but appears accurate when the linear approximation used by PL is reasonable
- There are two drawbacks
  1. Biased covariance estimates when sample size is small. Problem enhanced in 2 param EFD
  2. Does not use likelihood so model selection criterion have no clear meaning
Integral Approximation (Laplace, Quad)

- Use the actual likelihood so LR tests and information criteria make sense
- Estimates have better asymptotic behavior and less small sample bias
- Cannot be used in models with \( R \) side effects
  - Often can rewrite model to have \( G \) side only
- As random effects increase, procedures quickly become more computationally intensive
Motivating Example

- Researcher interested in how ecological processes shape the evolutionary trajectories of populations
- Focus on plant communities that inhabit stressful habitats
- Consider an experiment looking at a noxious weed
  - 17 randomly chosen locations in state where weed present
  - At each location, 200 seeds collected (3400 total seeds)
  - In growth chamber at Purdue, half of the seeds planted in dry, hard clay soil. The other half planted in very wet clay soil
  - # of germinating seeds per loc/soil is recorded
## Data for Motivating Example

```plaintext
data weed;
   input Loc $ Soil $ Y N @@;
cards;
1  Dry  68 100 2  Dry  76 100 3  Dry  93 100
4  Dry  73 100 5  Dry  89 100 6  Dry  97 100
7  Dry  82 100 8  Dry  51 100 9  Dry  63 100
10 Dry  62 100 11 Dry  96 100 12 Dry  89 100
13 Dry  97 100 14 Dry  92 100 15 Dry  98 100
16 Dry  97 100 17 Dry  93 100 1 Wet  77 100
2  Wet  43 100 3  Wet  53 100 4  Wet  83 100
5  Wet  85 100 6  Wet  76 100 7  Wet  50 100
8  Wet  43 100 9  Wet  76 100 10 Wet  81 100
11 Wet  98 100 12 Wet  84 100 13 Wet  90 100
14 Wet  96 100 15 Wet  61 100 16 Wet  90 100
17 Wet  87 100
; run;
```

**Response:** Binomial
- # germinating out of 100 seeds

**Explanatory Variables**
- Soil Type → Fixed Effect
- Location → Random blocking factor
- RCBD with random blocks?
Approach 1: Normal Approximation

Sample proportion $\Rightarrow \quad p_{ij} = \frac{Y_{ij}}{N_{ij}}$

Model $\Rightarrow \quad p_{ij} = \mu + \tau_i + b_j + e_{ij} \quad b_j \sim N(0, \sigma^2_b) \quad e_{ij} \sim N(0, \sigma^2)$

```
proc glimmix data=weed nobound;
   class soil loc;
   prop = Y/N;
   model prop = soil;
   random intercept /subject=loc;
   lsmeans soil;
run;
```

Two issues:
1. $0 \leq \pi_{ij} \leq 1$ but nothing guarantees that the estimated $\mu_{ij}$ will be constrained between 0 and 1
2. Constant variance assumption for $p_{ij}$
# Approach 1: Normal Approximation

## Selected Output

### Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>loc</td>
<td>0.01090</td>
<td>0.007379</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>0.01653</td>
<td>0.005845</td>
</tr>
</tbody>
</table>

### Type III Tests of Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>soil</td>
<td>1</td>
<td>16</td>
<td>3.64</td>
<td>0.0746</td>
</tr>
</tbody>
</table>

### soil Least Squares Means

| soil  | Estimate | Standard Error | DF | t Value | Pr > |t| |
|-------|----------|----------------|----|---------|-------|
| Dry   | 0.8329   | 0.04017        | 16 | 20.74   | <.0001|
| Wet   | 0.7488   | 0.04017        | 16 | 18.64   | <.0001|

Not strong enough evidence to reject

SE should depend on estimate
Residual Diagnostics
Approach 2: Variance Stabilizing Transformation

For binomial data expressed as proportions, arc sine square root commonly used

\[ \sin^{-1}\left(\sqrt{\frac{y_{ij}}{N_{ij}}}\right) \]

`proc glimmix` data=example1 nobound plots=studentpanel;
  class soil loc;
  `tprop = arsin(sqrt(Y/N));`
  model tprop = soil;
  random intercept /subject=loc;
  lsmeans soil;
run;
Approach 2: Variance Stabilizing Transformation

**Selected Output**

<table>
<thead>
<tr>
<th>Covariance Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cov Parm</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>Residual</td>
</tr>
</tbody>
</table>

**Type III Tests of Fixed Effects**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>soil</td>
<td>1</td>
<td>16</td>
<td>4.28</td>
<td>0.0551</td>
</tr>
</tbody>
</table>

**soil Least Squares Means**

| soil | Estimate | Standard Error | DF | t Value | Pr > |t| |
|------|----------|----------------|----|---------|-------|
| Dry  | 1.1852   | 0.05104        | 16 | 23.22   | <.0001|
| Wet  | 1.0713   | 0.05104        | 16 | 20.99   | <.0001|

Dry: 85.85% (3.84%) vs Wet 77.06% (4.89%)

Not strong enough evidence to reject
Residual Diagnostics
Approach 3: GLMM

Distribution of the response:

\[ y_{ij} | b_j \sim \text{Binomial} \left( N_{ij}, \pi_{ij} \right) \]

\[ b_j \sim \text{NIID} \left( 0, \sigma_B^2 \right) \]

Linear predictor:

\[ \eta_{ij} = \eta + \tau_i + b_j \]

Link Function: LOGIT

\[ \eta_{ij} = \log \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) \]

Default link function for binomial distribution in GLIMMIX.
Optimization methods

- Most GLMM fitted with GLIMMIX have one or more non-linear parameters
- Estimation requires non-linear optimization methods

- Default → Quasi-Newton Optimization
  \[ \text{nloptions tech=quanew;} \]

- Newton-Raphson Ridge Optimization
  \[ \text{nloptions tech=nrridg;} \]
    - Recommended: Seems to be more stable
Approach 3: GLMM

proc glimmix data=weed;
  class Soil Loc;
  model Y/N = soil / dist=binomial;
  random intercept /subject=loc;
  lsmeans soil/ diff ilink cl;
  nloptions tech=nrridg;
run;
### Approach 3: GLMM

#### Selected Output

<table>
<thead>
<tr>
<th>Model Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Set</strong></td>
</tr>
<tr>
<td><strong>Response Variable (Events)</strong></td>
</tr>
<tr>
<td><strong>Response Variable (Trials)</strong></td>
</tr>
<tr>
<td><strong>Response Distribution</strong></td>
</tr>
<tr>
<td><strong>Link Function</strong></td>
</tr>
<tr>
<td><strong>Variance Function</strong></td>
</tr>
<tr>
<td><strong>Variance Matrix Blocked By</strong></td>
</tr>
<tr>
<td><strong>Estimation Technique</strong></td>
</tr>
<tr>
<td><strong>Degrees of Freedom Method</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class Level Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class</strong></td>
</tr>
<tr>
<td>loc</td>
</tr>
<tr>
<td>soil</td>
</tr>
</tbody>
</table>

- Number of Observations Read: 34
- Number of Observations Used: 34
- Number of Events: 2689
- Number of Trials: 3400
Approach 3: GLMM with PseudoLk

Selected Output

### Dimensions

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>G-side Cov. Parameters</td>
<td>1</td>
</tr>
<tr>
<td>Columns in X</td>
<td>3</td>
</tr>
<tr>
<td>Columns in Z per Subject</td>
<td>1</td>
</tr>
<tr>
<td>Subjects (Blocks in V)</td>
<td>17</td>
</tr>
<tr>
<td>Max Obs per Subject</td>
<td>2</td>
</tr>
</tbody>
</table>

### Fit Statistics

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 Res Log Pseudo-Likelihood</td>
<td>194.10</td>
</tr>
<tr>
<td>Generalized Chi-Square</td>
<td>160.20</td>
</tr>
<tr>
<td>Gener. Chi-Square / DF</td>
<td>5.01</td>
</tr>
</tbody>
</table>

Expect value near 1 if decent fit.

Under PseudoLk, not really well defined as measure of model fit…

Cannot assess overdispersion!
Overdispersion

- Important issue of statistical modeling & very common
  - Type I errors can be wildly inflated
- Unique to non-Gaussian models
  - In particular, one-parameter exponential family distributions
- Frequently associated with an under-specified model
  - Sources of random variation not accounted for
- Diagnose using Pearson $\chi^2/df$ statistic or formal test
  - Requires true Likelihood for computation
Approach 3: GLMM

```sas
proc glimmix data=weed method=quad;
  class Soil Loc;
  model Y/N = soil / dist=binomial;
  random intercept /subject=loc;
  lsmeans soil/ diff ilink cl;
  nloptions tech=nrridg;
run;
```

**Fit Statistics**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 log L(Y</td>
<td>r. effects)</td>
</tr>
<tr>
<td>Generalized Chi-Square</td>
<td>145.12</td>
</tr>
<tr>
<td>Gener. Chi-Square / DF</td>
<td>4.27</td>
</tr>
</tbody>
</table>

**Type III Tests of Fixed Effects**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>soil</td>
<td>1</td>
<td>16</td>
<td>40.11</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Must specify this or “laplace” to look at Pearson $\chi^2$/df

Appears to be overdispersion
Causes for Overdispersion

- Using the wrong probability distribution
- Missing an important predictor or source of variation

In this case,

- We excluded soil*loc interaction as we would in linear mixed model
- But one-parameter EFDs do not have a residual variance so we can include it here
Approach 3: GLMM with soil*loc random effects

```
proc glimmix data=weed method=quad;
  class Soil Loc;
  model Y/N = soil  / dist=binomial;
  random intercept soil /subject=loc;
  lsmeans soil/ diff ilink cl;
  covtest 'soil*loc' . 0;
  nloptions tech=nrridg;
run;
```

Include the soil*loc random effects

Provides test is soil*loc variance is greater than zero.
Approach 3: GLMM

Selected Output

<table>
<thead>
<tr>
<th>Dimensions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>G-side Cov. Parameters</td>
<td>2</td>
</tr>
<tr>
<td>Columns in X</td>
<td>3</td>
</tr>
<tr>
<td>Columns in Z per Subject</td>
<td>3</td>
</tr>
<tr>
<td>Subjects (Blocks in V)</td>
<td>17</td>
</tr>
<tr>
<td>Max Obs per Subject</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fit Statistics for Conditional Distribution</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 log L(Y</td>
<td>r. effects)</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>3.74</td>
</tr>
<tr>
<td>Pearson Chi-Square / DF</td>
<td>0.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tests of Covariance Parameters Based on the Likelihood</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>DF</td>
</tr>
<tr>
<td>loc*soil</td>
<td>1</td>
</tr>
</tbody>
</table>

Have we moved to underdispersion?

 MI: P-value based on a mixture of chi-squares.
## Approach 3: GLMM

### Selected Output

#### Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>loc</td>
<td>0.5479</td>
<td>0.3442</td>
</tr>
<tr>
<td>soil</td>
<td>loc</td>
<td>0.6471</td>
<td>0.2629</td>
</tr>
</tbody>
</table>

#### Type III Tests of Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>soil</td>
<td>1</td>
<td>16</td>
<td>4.85</td>
<td>0.0426</td>
</tr>
</tbody>
</table>

Without soil*loc F=40.11

#### Model Scale

#### Data Scale

#### soil Least Squares Means

<table>
<thead>
<tr>
<th>soil</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th>Alpha</th>
<th>Lower</th>
<th>Upper</th>
<th>Mean</th>
<th>Std Error Mean</th>
<th>Lower Mean</th>
<th>Upper Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry</td>
<td>1.9986</td>
<td>0.2803</td>
<td>16</td>
<td>7.13</td>
<td>&lt;.0001</td>
<td>0.05</td>
<td>1.4044</td>
<td>2.5928</td>
<td>0.8807</td>
<td>0.02946</td>
<td>0.8029</td>
<td>0.9304</td>
<td></td>
</tr>
<tr>
<td>Wet</td>
<td>1.3419</td>
<td>0.2746</td>
<td>16</td>
<td>4.89</td>
<td>0.0002</td>
<td>0.05</td>
<td>0.7599</td>
<td>1.9240</td>
<td>0.7928</td>
<td>0.04510</td>
<td>0.6813</td>
<td>0.8726</td>
<td></td>
</tr>
</tbody>
</table>

Describing variation among location average logits

Describing variation of log-odds ratios among locations

**Model Scale**

**Data Scale**

---

**Describing variation among location average logits**

**Describing variation of log-odds ratios among locations**

---

Without soil*loc F=40.11

---

**Model Scale**

**Data Scale**

---

**Describing variation among location average logits**

**Describing variation of log-odds ratios among locations**

---

Without soil*loc F=40.11
Approach 4: 
Conditional Model II

```plaintext
proc glimmix data=weed method=quad;
  class soil loc;
  model Y/N = soil / dist=binomial;
  random soil / type=cs subject=loc;
  lsmeans soil/ diff ilink cl;
  nloptions tech=nrridg;
run;
```

Conditioning only on soil*loc random effects...results in compound symmetry structure

Results in similar parameter estimates as Approach 3
Approach 5: 
Marginal (GEE) Model

```
proc glimmix data=weed;
   class soil loc;
   model Y/N = soil / dist=binomial;
   random soil / type=cs subject=loc residual;
   lsmeans soil/ diff ilink cl;
   nloptions tech=nrridg;
run;
```

No random effects...Responses have a quasi-likelihood that derives from the binomial....Use the CS working correlation structure

\[
\text{Var}(Y_i) = \begin{bmatrix}
\sqrt{\pi_{i1}(1-\pi_{i1})} & 0 \\
0 & \sqrt{\pi_{i2}(1-\pi_{i2})}
\end{bmatrix} \phi \begin{bmatrix}
1 & \rho \\
\rho & 1
\end{bmatrix} \begin{bmatrix}
\sqrt{\pi_{i1}(1-\pi_{i1})} & 0 \\
0 & \sqrt{\pi_{i2}(1-\pi_{i2})}
\end{bmatrix}
\]
Approach 5: GEE
Selected Output

### Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>loc</td>
<td>6.7360</td>
<td>4.4952</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>9.9352</td>
<td>3.5126</td>
</tr>
</tbody>
</table>

### Type III Tests of Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>soil</td>
<td>1</td>
<td>16</td>
<td>3.59</td>
<td>0.0763</td>
</tr>
</tbody>
</table>

### soil Least Squares Means

<table>
<thead>
<tr>
<th>soil</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th>Alpha</th>
<th>Lower</th>
<th>Upper</th>
<th>Mean</th>
<th>Std Error Mean</th>
<th>Lower Mean</th>
<th>Upper Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry</td>
<td>1.6066</td>
<td>0.2655</td>
<td>16</td>
<td>6.05</td>
<td>&lt;.001</td>
<td>0.05</td>
<td>1.0438</td>
<td>2.1694</td>
<td>0.8329</td>
<td>0.03694</td>
<td>0.7396</td>
<td>0.8975</td>
<td></td>
</tr>
<tr>
<td>Wet</td>
<td>1.0923</td>
<td>0.2283</td>
<td>16</td>
<td>4.78</td>
<td>0.002</td>
<td>0.05</td>
<td>0.6083</td>
<td>1.5764</td>
<td>0.7488</td>
<td>0.04295</td>
<td>0.6476</td>
<td>0.8287</td>
<td></td>
</tr>
</tbody>
</table>
Summary

- Marginal Model (GEE)
  - Results similar to the Normal approximation
  - P-value roughly 0.075
  - Dry 83.3% and Wet 74.9%

- Conditional Model (with soil*loc effects)
  - P-value roughly 0.043
  - Dry 88.1% and Wet 79.3%

- So how do we decide??
What are we modeling…?

For Gaussian data

Marginal Model

\[ y \sim N(X\beta, ZBZ' + R) \]

Conditional Model

\[ y | b \sim N(X\beta + Zb, R) \]

\[ b \sim N(0, B) \]

No ambiguity… equivalent models and results!
What are we modeling…?
For NON-Gaussian data

\[ f(y) = \int f(y | b) f(b) db \]

- Probabilistic behavior of the data, as observed
- Probabilistic processes assumed to give rise to the data

Marginal Model

\( y \sim ?? \)

- Not binomial any more…after integration over random effects
- With some exceptions, lack of closed form expression

Conditional Model

\( y | b \sim Binomial (N, \pi) \)
\( b \sim N(0, B) \)

Models no longer equivalent!
Models have different targets

**Normal Approximation**
- Response $\rightarrow$ Sample proportion of the marginal distribution $f(y)$
  
  
  $p_{ij} \sim N\left(\mu_{p,ij}, \sigma^2\right)$

  $b_j \sim N\left(0, \sigma^2_B\right)$

  Estimates the **mean** of the marginal distribution $f(y)$ for $i^{th}$ Trt

  $\hat{\mu}_{p,i.} = \hat{\eta} + \hat{\tau}_i$

**Conditional GLMM**
- Response $\rightarrow$ Number of successes observed conditional on a location
  
  
  $y_{ij} \mid b_j \sim Binomial\left(N_{ij}, \pi_{ij}\right)$

  $b_j \sim N\left(0, \sigma^2_B\right)$

  Estimates the binomial proportion of $f(y \mid b) \rightarrow$ **Median** of the marginal distribution $f(y)$ for $i^{th}$ Trt

  $\hat{\pi}_i = \frac{1}{1 + e^{-(\hat{\eta} + \hat{\tau}_i)}}$
Why the discrepancy?

Marginal Distribution of Y

The marginal distribution \( f(y) \) is skewed whenever \( \pi \neq 0.5 \)

Mean → Estimation target for normal approximation and GEE
Median → Estimation target for conditional GLMM
# Differences in Approaches

<table>
<thead>
<tr>
<th>Normal Approximation</th>
<th>GLMM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Embedded in the marginal modeling approach ~ GEE</strong></td>
<td><strong>Reflects the conditional modeling approach</strong></td>
</tr>
<tr>
<td><strong>Targets the mean of the marginal distribution $f(y)$</strong></td>
<td>**Targets the true binomial probability $\pi_i$ of $y</td>
</tr>
<tr>
<td>– Overestimates when $\pi_i&lt;0.50$</td>
<td><strong>No problem controlling Type I error</strong></td>
</tr>
<tr>
<td>– Underestimates when $\pi_i&gt;0.50$</td>
<td><strong>Greater power!</strong></td>
</tr>
<tr>
<td>– Inaccurate estimates of odds ratio</td>
<td></td>
</tr>
<tr>
<td><strong>Erroneous precision estimates</strong></td>
<td></td>
</tr>
<tr>
<td><strong>No problem controlling Type I error</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Less power!</strong></td>
<td></td>
</tr>
</tbody>
</table>
Conditional versus Marginal Models

Take Home Message:

Marginal vs. Conditional GLMM models
Not equivalent!

What are we estimating?
What is the target for inference?
Interested in average over population of blocks or interested in the average block?
Specify the model accordingly

Stroup 2011
Conditional versus Marginal Models

Take Home Message II:

Arc sine square root transformed analysis results in percentage estimates that are in between the marginal and conditional models. Not clear what they estimate.

Provides some justification for the use of GLMM.
More on some common issues

- Overdispersion
- Bias correction for standard errors
- Complete or quasi-complete separation of values
  - Extreme category problem
- Variance components and/or their standard errors not estimated
Overdispersion “Fix”

- Add a scale parameter

\[ \text{In GLIMMIX } \rightarrow \text{ random } \_\text{residual}_\; \]

- Creates a Quasi-Likelihood \( \rightarrow \) Cannot use integral approximation for estimation
- Not always effective in controlling inflation of Type I error
- Inference limited to marginal model framework
“Common” Fix:
Scale Parameter

```plaintext
proc glimmix data=data;
  class soil loc;
  model Y/N = soil;
    random intercept /subject=loc;
    random _residual_;  
run;
```

<table>
<thead>
<tr>
<th>Fit Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 Res Log Pseudo-Likelihood</td>
<td>95.69</td>
</tr>
<tr>
<td>Generalized Chi-Square</td>
<td>292.36</td>
</tr>
<tr>
<td>Gener. Chi-Square / DF</td>
<td>9.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Covariance Parameter Estimates</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cov Parm</td>
<td>Subject</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
</tr>
<tr>
<td>Intercept</td>
<td>loc</td>
</tr>
<tr>
<td>Residual (VC)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type III Tests of Fixed Effects</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect</td>
<td>Num DF</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>soil</td>
<td>1</td>
</tr>
</tbody>
</table>

Note that we are creating a quasi-Lk! Cannot use integral approximation… limited to Pseudo-Lk.
Better to consider...

- **GLMM: What are we estimating?**
  - Conditional vs. Marginal Models
  - Specify the model accordingly

- **Overdispersion:**
  - Is the distribution of the response correctly specified?
    - e.g., Instead of Poisson consider Negative Binomial
  - Are we accounting for all sources of random perturbations on the response?
More on some common issues

- Overdispersion
- Bias correction for standard errors
- Complete or quasi-complete separation of values
  - Extreme category problem
- Variance components and/or their standard errors not estimated
Inference

Bias Correction for Std Errors

- Satterthwaite’s approximation of degrees of freedom
- Kenward-Roger’s bias correction for standard errors
- Applicable to LMM on Gaussian $y$ … and to GLMM on non-Gaussian $y$ fitted with PseudoLk
  - “ad-hoc” → Based on $y^*$, not $y$
  - Performs well provided PseudoLk linearization is accurate
Inference

Bias Correction for Std Errors

- Analogous issues in GLMM regardless of estimation method!
- So what to do under integral approximation?
- “Sandwich” (aka “Empirical” or “Robust”) estimator
  - Liang and Zeger (1986) in GEE context
  - In “small” datasets, requires bias correction (Morel et al. 2003)

```plaintext
proc glimmix data=... method=laplace empirical=mbn;
  random intercept / subject=block;
  random block;
```
More on some common issues

- Overdispersion
- Bias correction for standard errors
- Complete or quasi-complete separation of values
  - Extreme category problem
- Variance components and/or their standard errors not estimated
Extreme Category Problem in GLMM

- Complete or quasi-complete separation of values
- Relatively common in binary/binomial data
  - Every value of the response variable in a given level of the Trt factor is 0 \( \Rightarrow \) \( \text{Prob}(\text{Success}|\text{Trt}) = 0\% \)
  - Every value of the response variable in a given level of the Trt factor is 1 \( \Rightarrow \) \( \text{Prob}(\text{Success}|\text{Trt}) = 100\% \)
- Analysis may fail to converge to a solution
- Maximum likelihood estimate for Trt effects goes to infinity!
- Even if convergence, the estimates of regression coefficients or Trt effects, and standard errors, are extremely large
- Validity of inference is questionable
Extreme Category Problem

Alternative Strategies

Complementary log-log link function

- Does not require symmetry over the parameter space
- Useful for binomial data under “boundary conditions” → Most of the probabilities lie close to zero or close to one

\[
\eta = \log(-\log(1-\pi))
\]

```latex
proc glimmix ...
   model y/N = Trt / link=cll;
   random ... ;
   lsmeans Trt/ ilink;
```
Penalized Likelihood Method (Firth 1993; Heinze and Schmepner 2002)

- Limited to fixed effects models → Cannot handle random effects!
- Yields finite estimates of parameters under complete or quasi-complete separation
- Estimates have relatively small bias, even under extreme conditions
- Computationally feasible even for large samples
- Available beginning SAS Version 9.2

```
proc logistic ...;
model y/N = ... / firth;
```
More on some common issues

- Overdispersion
- Bias correction for standard errors
- Complete or quasi-complete separation of values
  - Extreme category problem
- Variance components and/or their standard errors not estimated
Variance Components Not Estimated

- Symptom of model estimation process gone bad
- Careful! Ripple effects for inference on fixed effects of interest
- Not an issue to leave unattended, regardless of direct interest on variances
Variance Components Not Estimated
Consequences

- Inflation of F-test statistics for fixed effects
- Misspecification of degrees of freedom for tests on fixed effects
- Inference??? P-values ???

Suggestions that might help:
- Integral approximation
  
  ```
  Proc glimmix ... method=laplace;
  or method=quad;
  ```

- Non linear optimization methods
  
  ```
  nloptions tech=nrridg;
  ```

- Consider re-specifying the model → Pool random effects?
- Consider aggregating binary responses into binomials
Acknowledgements

Walter Stroup
Department of Statistics
University of Nebraska, Lincoln

Upcoming book (Spring 2012?):
Generalized Linear Mixed Models
A 21st Century Introduction to Linear Models
Concepts, Application and Methodology"
References

J. Dairy Sci. 92:1–15
doi:10.3168/jds.2008-1404

Invited review: Assessing experimental designs for research conducted on commercial dairies

R. J. Tempelman
Department of Animal Science, Michigan State University, East Lansing 48824-1225
Additional Resources

2006

2009