Hierarchical Generalized Linear Models
Hierarchical Generalized Linear Models

- So now we are moving on to the more advanced type topics.

- To begin we will start with a general type of model that deals with generalized linear models.

- This is a general set of models that can deal with observations that are not normally distributed.
Overview

- We will first talk about the generalized linear model.
  - HLM is a special case of HGLM.

- The HGLM handles many types of data:
  - Binary Outcomes.
  - Count Data.
  - Ordinal Data.
  - Multinomial Data.

- We will focus only on the binary case in this lecture.
GLM Versus the Basic Linear Model

- First we start by talking about our basic linear model.
- We then rephrase it as a generalized linear model.
  - This will show us a more general approach to linear models.
- Then we start to talk about how we can use this new model to model non-normal responses.
To begin we think of your basic linear model.

\[ Y_i = \beta_0 + \beta_1 X_i + r_i \]

- Have a part that is being predicted
- Have a part that defines what we would expect to see
- Have a residual part

Note the Sampling distribution of \( Y_i \) given our predicted value:

\[ Y_i | \mu_i \sim N(\mu_i, \sigma^2) \]
Basic Linear Model

- In this model, we assume that:
  - The observed values have a wide range of possible values (neg. inf. to pos. inf. approximately).
  - There is a linear relationship between X and Y.
  - The errors are normal independent and have the same variance.
Basic Linear Model

- This works fine unless we try to model responses that are binary, counts, or categories (ordinal or nominal).

- Notice in this case (with binary data) our basic assumptions seem unreasonable.

- For example, think of a binary response (pass/fail).
Example: Binary Responses

- In this case the first assumption is inaccurate because we can only observe a value of 0 and 1.

- The second assumption does not seem reasonable because it could lead to predicting a value outside of 0 or 1.

- Finally, the third assumption cannot be correct because they are describing the differences between our predicted value and the observed value (which is only a 0 or 1).

- So we can see that we need to generalize our model to account for this.
In expanding the models, we will focus on certain parts of the basic regression model.

Mainly, we will need to change our variable so that it is something that we can predict and change our basic assumptions of the errors.
GLM with the Basic Regression

Now we start by assuming that we have a linear relationship

\[ \eta_i = \beta_0 + \beta_1 X_i + r_i \]

This is a generic notation for a transformation of the expected response

• Conceptually, we will transform our original expected value to \( \eta \) to ensure that we can use our regression.

• This is the same as putting constraints on our expected responses.
Link Function

- In basic regression we have a continuous response variable $Y_i$ with expected outcome $\mu$.

- This value is fine with the basic assumptions so and we will simply have:

$$\eta_i = \mu_i$$

- Finally, we will specify the error mean and variance based on our original variables.
  - Mainly we will need to only worry about our link function
This is the Link function, taking our expected response and changing it to something that we can predict with a regular regression model.

\[ f(u_i) = \eta_i \]

This is the regression part (The Structural Model): \[ \eta_i = \beta_0 + \beta_1 X_i + r_i \]

The Sampling Distribution

This is the conditional distribution of \( Y_i | A \) where \( A \) is some expected response.
Now, one other thing needs an introduction before we get to the actual models:

- The hierarchical part.

The last thing is that we also have nested data, so we not only have a GLM, but we also have to model the coefficients as part of an HLM.
Hierarchical Generalized Linear Models

\[ f(u_i) = \eta_i \]

\[ \eta_i = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij} \]

Notice that now we have added the subscript \( j \) for the nesting of our data.

In modeling the hierarchical part we will make all of the same basic assumption the we have in the past...this part is unchanged.
The Models

- From here we now are going to talk about the different types of models. For each model we will need to talk about
  - The sampling distribution.
  - The link function.
  - The structural model.

- Now we go to binary responses (leaving counts ordinal, and nominal responses to the book).
Binomial

- In the case of Binary data we now have observed a set of binary responses.

- Our goal is to model the probability of a dichotomous response as a function of a set of covariates.

- The problem, using basic regression we could get values outside of 0-1 so we will use a logit link.
Binomial HGLM

This is the Link function, taking our expected response (probability) and changing it to something that we can predict with a regular regression model.

\[
\log \left( \frac{\varphi}{1 - \varphi} \right) = \eta_i
\]

This is the regression part (The Structural Model)

\[
\eta_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}
\]

The Sampling Distribution

This is the conditional distribution of \(Y_i|\varphi\) is Binomial\( (m_{ij}, \varphi_{ij})\)
So what does this mean?

We model the expected log-odds of the response as a linear function of the covariates.

In some cases it is enough to simply interpret the log-odds or the odds in other cases you may want to covert it back to a probability.
Odds

- By using the probability of a 1 divided by the probability of a 0 we have something that tells us the likelihood of either occurring.
  - “How many successes would I expect to see for each failure I see”

- Note that if the odds are “1” this is the same as “1 to 1 odds”, which means each event is equally likely to occur.

- What about the odds of .25?
Odds

- If the odds are greater than 1...then a 1 is more likely:
  - Log-odds are positive.

- If the odds are less than 1...then a 0 is more likely:
  - Log-odds are negative.
Interpretation (using HGLM)

\[ \eta_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \]

**\( \beta_{0j} = \gamma_{00} + u_{0j} \)**

This is the expected log-odds when all other variables are zero plus the random error. Random Error is still normal \((0, \tau)\)

**\( \beta_{1j} = \gamma_{10} + u_{1j} \)**

Again we have random error, but the intercept describes the change in the log-odds for unit change in \(X\). This is related to the increase or decrease in the odds.
Example

- From Hedeker (http://tigger.uic.edu/~hedeker/long.html)

- NIMH Schizophrenia Study (437 patients)
  - 108 given placebo
  - 329 given drug

- 7 week study

- Modeling the probability a patient was rated as worse than moderately ill.
  - 7-point scale.
HLM Setup: Random Intercept Model

LEVEL 1 MODEL (level 1: group-level covariates, level 2: grand-mean centering)
\[ y_{ij} = \beta_0 + \beta_1(x) + e_{ij} \]
\[ \eta = \beta_0 + \eta_{ij} \] (SWEEK)

LEVEL 2 MODEL (level 2: grand-mean centering)
\[ y_{ij} = \gamma_{00} + \gamma_{01}(x) + r_{ij} \]
\[ F_i = \gamma_{10} + \gamma_{11}(x) - r_i \]

Mixed Model
\[ \eta = \gamma_{00} + \gamma_{01}(x) + \gamma_{20} + \gamma_{21}(x) + \text{SWEEK} + \gamma_{30}(x) + \text{SWEEK} + \epsilon_i \]
Model Parameter Estimates: Fixed Effects

- Here we find our fixed effect parameters.
- The main effect of treatment was not significant.
- The interaction between treatment and week was significant.
- As time goes on, treatment improves the condition.

```
<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>T-ratio</th>
<th>Approx. d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, G0</td>
<td>3.171360</td>
<td>0.336054</td>
<td>9.475</td>
<td>435</td>
<td>0.000</td>
</tr>
<tr>
<td>TX, G01</td>
<td>-0.261861</td>
<td>0.430767</td>
<td>-0.608</td>
<td>435</td>
<td>0.543</td>
</tr>
<tr>
<td>INTRCPT2, G10</td>
<td>-1.055894</td>
<td>0.203152</td>
<td>-5.198</td>
<td>1599</td>
<td>0.000</td>
</tr>
<tr>
<td>TX, G11</td>
<td>-0.496923</td>
<td>0.226900</td>
<td>-2.190</td>
<td>1599</td>
<td>0.029</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Odds Ratio</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, G0</td>
<td>3.571360</td>
<td>35.564945</td>
<td>(18.740, 73.557)</td>
</tr>
<tr>
<td>TX, G01</td>
<td>-0.261861</td>
<td>0.769018</td>
<td>(0.551, 1.062)</td>
</tr>
<tr>
<td>INTRCPT2, G10</td>
<td>-1.055894</td>
<td>0.347881</td>
<td>(0.234, 0.518)</td>
</tr>
<tr>
<td>TX, G11</td>
<td>-0.496923</td>
<td>0.608400</td>
<td>(0.390, 0.949)</td>
</tr>
</tbody>
</table>
```
Model Parameter Estimates: 
Variance Components

We can look at the results for our Level-2 variance ($\tau_{00}$).

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Standard Deviation</th>
<th>Variance Component</th>
<th>df</th>
<th>chi-square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, U0</td>
<td>1.25328</td>
<td>1.57071</td>
<td>435</td>
<td>715.16051</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Linear Model: Logits and Square Root of Week

Logit Scale

- Control
- Treatment

Square Root of Week

Logit
Non-linear Model: Logits and Week

![Logit Scale Graph]

- Control
- Treatment
Nonlinear Model: Probabilities and Square Root of Week
Nonlinear Model: Probabilities and Week

Probability Scale

Week [0-6]
Probability [0-0.9]

- Control (solid line)
- Treatment (dashed line)
Summary

- In general, we can see that the only hard thing about the application of these model is the interpretation.

- Otherwise, input using HLM is the same.

- Estimation of such models in HLM can be problematic.
  - Uses an estimator known to be biased in certain situations.
    - Not maximum likelihood.
  - Other mechanisms for estimation are recommended:
    - Mplus.
    - SAS proc nlmixed.
    - Bayesian algorithms (such as glmmgibbs in R).