10. Hierarchical Generalized Linear Models (HGLM) EPSY 8268

The models reviewed so far are appropriate for outcome data that are normally distributed, typically when the outcome variable is continuous. When continuous data are not normal, or skewed, transformations might be available to transform data so that they function in a more normal manner. In some cases, the nature of the variable yields non-normal distributions and no transformation is available to improve the normality of the distribution.

Most forms of categorical data do not present distributions that are normal or continuous. This includes data that are nominal, dichotomous, ordinal, or counts.

A model can be designed that links the nature of the outcome data to the general linear model. This requires the use of a sampling model, a link function, and a structural model.

**2-Level HLM**

We can conceive of a simple 2-level HLM as a normal sampling model with an identity link function and a linear structural model.

*Level-1 Sampling Model*

 *Yij* | μ*ij* ~ NID(μ*ij*, σ2)

 The level-1 outcome, *Yij*, given the predicted value μ*ij*, is normally and independently distributed, with an expected value of μ*ij* and a constant variance σ2. This can also be written as:

 E(*Yij* | μ*ij* ) = μ*ij* where Var(*Yij* | μ*ij* ) = σ2

*Level-1 Link Function*

 It is possible to transform the level-1 predicted value μ*ij* to ensure that predictions are constrained to be within a specified interval. The transformed predicted value is η*ij*. When the data are normally distributed, no transformation is needed. We can represent the decision to not transform as η*ij* = μ*ij* . This is the “identity link function,” where the transformed predicted value is the predicted value (the case of no transformation).

*Level-1 Structural Model*

 The transformed predicted value η*ij* can now be related to the predictors of the model through the linear structural model

 η*ij* =β0*j* + β1*jX*1*ij* + β2*jX*2*ij* + … + β*pjXpij*

 Note that this is a structural model explicitly declaring the model of the predicted values, where there is no error term or residual.

**Two- or Three-Level Models for Binary Outcomes**

The binary outcome model relies on the binomial sampling model and a logit link function.

*Level-1 Sampling Model*

 Consider *Yij* as the number of successes in *mij* trials, where φ*ij* is the probability of success on each trial. This is the standard binomial model, where

 *Yij* | φ*ij* ~ B(*mij*, φ*ij*)

 *Yij* has a binomial distribution with *mij* trials and a probability of success per trial of φ*ij*. This can also be written as:

 E(*Yij* | φ*ij* ) = *mij* φ*ij* where Var(*Yij* | φ*ij*) = *mij* φ*ij*(1- φ*ij*)

 When *mij* = 1, *Yij* is a binary variable taking on the value of 0 or 1. This is a special case, the Bernoulli distribution.

*Level-1 Link Function*

 The most common link function for binomial sampling models is the logit link:

 $η\_{ij}=log\left(\frac{φ\_{ij}}{1-φ\_{ij}}\right)$

 Here, η*ij* is the log odds of success (the log of the odds, probability of success divided by the probability of failure). Consider the following examples to understand the nature of log-odds.

* Note that when the probability of success φ*ij* is .5, the odds of success $\frac{φ\_{ij}}{1-φ\_{ij}}$ is .5/(1-.5) = 1.0 and the log-odds (or logit) is log(1) = 0.
* When the probability of success is less than .5, the odds are less than 1.0 and the logit is negative.
* When the probability of success is greater than .5, the odds are greater than 1.0 and the logit is positive.
* Although the probability of success φ*ij* is constrained to the interval (0, 1), η*ij* can take on any real value.

*Level-1 Structural Model*

 The structural model is the same as in the linear case above. This results in a predicted log-odds η*ij*.

 η*ij* =β0*j* + β1*jX*1*ij* + β2*jX*2*ij* + … + β*pjXpij*

 Note that a predicted log-odds can be converted to a predicted probability:

 $φ\_{ij}=\frac{1}{1+exp⁡\{-η\_{ij}\}}$

 So, whatever the value of η*ij*, φ*ij* will be constrained within 0 to 1.

*Level-2 and Level-3 Models*

 The higher-level models will follow the same notation and form that have been presented previously.

*In Practice*

A national survey was conducted in Thailand in 1988 to examine the probability that a child in primary education would repeat a grade, including 7516 students in 6th grade across 356 primary schools. In total, 14% of students repeated at least one grade. An unconditional model is

η*ij* =β0*j*

with the level-2 model

β0*j* = γ00 + *u*0*j* where *u*0*j* ~ N(0, τ00)

 γ00 is the average log-odds of repetition across schools, τ00 is the variance between schools of the school-average log-odds of repetition. Recall these can be transformed to odds through exp{γ00} which corresponds to a probability of 1/1 + exp{γ00}.

Explanatory variables can be added to explain that variation between students within schools and variation between schools. All of the associated coefficients will be in log-odds metric.

*A note on Unit-Specific versus Population-Average results*

In part because of the normalization of the distribution due to the use of the link function, the location of the mean relative to the median shifts (they are equal in a normal distribution, but not in a skewed distribution).

Unit-specific questions are those that inquire about the processes occurring in each level-2 unit, which is captured by the level-1 coefficients. The level-2 model describes how differences in explanatory variables at level-2 relate to the differences in level-1 processes in each level-1 unit. These are unit-specific questions, for which unit-specific results can be used.

Population-average models address population-average questions. If our interest is to use the regression model to simulate how differences in processes function across schools, not holding constant school attended, we a need a population-average estimate. Population-average models specify the marginal distributions and are based on fewer assumptions, thus less rich in interpretation flexibility.

**Count Data**

A typical approach to modeling count data is through a Poisson sampling model and a log link function. Count data can include, for example, the number of crimes a person commits, or the number of times a student is sent to the office for disciplinary action or because of bullying behavior.

*Level-1 Sampling Model*

 The level-1 outcome, *Yij*, is the number of events that occur during an interval of time having length *mij* (referred to as exposure).

*Yij* | λ*ij* ~ P(*mij*, λ*ij*)

 An example is where *Yij* is the number of crimes committed by person *i* in neighborhood *j* during five years (*mij* = 5). *Yij* has a Poisson distribution with exposure *mij* and event rate per time period of λ*ij*. This can be expressed as

 E(*Yij* | λ*ij* ) = *mij* λ*ij* where Var(*Yij* | λ*ij*) = *mij* λ*ij*.

 We can interpret this as meaning that the expected number of events, *Yij*, for person *i* in group *j* is its event rate λ*ij* times its exposure *mij*. In this distribution, the variance equals the mean. The exposure *mij* does not necessarily need to be measured in time. It could be the area of a neighborhood where the events are counted

*Level-1 Link Function*

 The standard link function for the Poisson sampling model is the log link.

 η*ij* = log(λ*ij*)

 Note that when the event rate λ*ij* is 1, the log is 0. When the event rate is less than 1, the log is negative; when the event rate is greater than 1, the log is positive.

 Although λ*ij* is constrained to be nonnegative, its log can take on any real value.

**Ordinal Data**

Ordered categories are common in education and social sciences, particularly in the context of survey research. But we can also conceive of achievement outcomes as ordinal, as in the case of state accountability testing programs that yield levels of performance, such as (1) does not meet standards, (2) partially meets standards, (3) meets standards, and (4) exceeds standards). We also regularly see ordinal data in the response to rating scale items, such as strongly disagree to strongly agree. There is a tradition of modeling such outcome data in standard single-level regression models.

*The Cumulative Probability Model for Single-Level Data*

We may have *M* possible ordered categories. Considering the response variable *R*, it can take on the value of *m* (one of the ordered categories) with probability

φ*m* = Prob(*R* = *m*)

If we have a 3-point rating scale with the categories (1) not proficient, (2) proficient, and (3) exceeds proficient (with respect to academic achievement), we can represent these outcomes as

φ1 = Prob(*R* = 1) = Prob(“not proficient”)

φ2 = Prob(*R* = 2) = Prob(“proficient”)

φ3 = Prob(*R* = 3) = Prob(“exceeds proficient”)

In these models, we work with cumulative probabilities, since we want to capture the ordinal nature of the response options (outcome variable). We consider cumulative probabilities as

$φ\_{m}^{\*}$ = Prob(*R* ≤ *m*) = φ1 + φ2 + … + φ*m*

In our example,

$φ\_{1}^{\*}$ = φ1

$φ\_{2}^{\*}$ = φ1 + φ2

$φ\_{3}^{\*}$ = φ1 + φ2 + φ3 = 1

Also note that the third equation is redundant, because if you know the first two, the third is known. Generally, we only need *M* -1 cumulative probabilities are estimated.

The cumulative logit function can be used to represent the idea of cumulative probabilities

 $η\_{m}=log\left(\frac{φ\_{m}^{\*}}{1-φ\_{m}^{\*}}\right)$ = $log\left(\frac{Prob(R\leq m)}{Prob(R>m)}\right)$

A logistic regression can be formulated

 η*mi* = θ*m* + β*X*

In this model, there is an intercept θ*m* for each category *m* (called thresholds) and a common slope β. There are a couple of assumptions in the use of this model, including the assumption that *X* affects the odds ratios for each category *m* in the same way.

*Level-1 Model*

To continue the example above with the 3-level performance scale, the level-1 model would be

 η1*ij* = θ1*j* + β1*j* *X*1*ij*

 η2*ij* = θ2*j* + β2*j* *X*1*ij*

To avoid the complexity introduced by potentially allowing both thresholds to randomly vary (θ), implying that the underlying characteristic being measured translates into response categories differently in different schools (an inconsistent rating-scale structure), we impose a common intercept and work with differences in the thresholds, where θ1 – θ2 = δ.

 η1*ij* = β0*j* + β1*j* *X*1*ij*

 η2*ij* = β0*j* + β2*j* *X*1*ij* + δ

This allows overall level of proficiency to vary randomly over schools *j*; the slope for *X*, the β1 can vary randomly, but typically the threshold δ would be held constant.

*Level-1 Structural Model*

$$η\_{mij}=β\_{0j}+\sum\_{q=1}^{Q}β\_{qj}X\_{qij}+\sum\_{m=2}^{M-1}D\_{mij}δ\_{m}$$

Here *Dmij* is an indicator variable for category *m*, as a design matrix.

*Level-1 Sampling Model*

The sampling model involves layers of the cumulative probabilities, employing *M* – 1 dummy variables *Y*1*ij*, …, *Ym-*1*ij* for case *i* in unit *j*

 *Ymij* = 1 if *Rij* ≤ *m*, 0 otherwise.

In our example of *M* = 3 proficiency levels,

 *Y*1*ij* = 1 if *Rij* = 1

 *Y*2*ij* = 1 if *Rij* ≤ 2

The probabilities are then cumulative, $φ\_{mij}^{\*}$ = Prob(*Ymij* = 1)

 E(*Ymij* | $φ\_{mij}^{\*}$ ) = $φ\_{mij}^{\*}$ where Var(*Ymij* | $φ\_{mij}^{\*}$) = $φ\_{mij}^{\*}$(1-$φ\_{mij}^{\*}$)

and COV(*Ymij*, *Ym’ij* | $φ\_{mij}^{\*}, φ\_{m'ij}^{\*}$ = $φ\_{mij}^{\*}$(1-$φ\_{m'ij}^{\*}$)

*In Practice*

A survey of 650 teachers in 16 schools was completed in California and Michigan in 1990. The outcome variable is a three-category measure of teacher commitment. The unconditional model is

$$η\_{mj}=β\_{0j}+D\_{2ij}δ\_{2j}$$

Where *D*2*ij* is a dummy variable indicating whether *m* = 2. So it will = 1 if *m* = 2 or 0 if *m* = 1. We can summarize this over two equations

$$η\_{1j}=β\_{0j}$$

$$η\_{2j}=β\_{0j}+δ\_{2j}$$

The level-2 model is

 Β0*j* = γ00 + *u*0*j*

 δ2*j* = δ2

**Multinomial Data**

Sometimes we are interested in nominal or categorical outcomes, where the categories are not explicitly ordered (ordinal). An example is post-high school educational plans, including (a) obtaining a career certificate, (b) attending a 2-year college, (c) attending a 4-year college, (d) work, (e) military, or (f) uncertain.

We need a regression model that allows level-1 predictors to have different associations with different categories of outcomes (different probabilities of being associated with each outcome). Thus we have *M* response categories and a response, *R*, with the value of *m* with

Prob(*R* = *m*) = φ*m*

If we consider three postsecondary plan options (a) no higher education, (b) 2-year college, or (c) 4-year college, we have *M* = 3

Prob(*Rij* = 1) = φ1*ij*

Prob(*Rij* = 2) = φ2*ij*

Prob(*Rij* = 3) = φ3*ij* = 1 – φ1*ij* – φ2*ij*

*Level-1 Sampling Model*

Again, we employ dummy variables to indicate the specific category

 *Ymij* = 1 if *Rij* = *m*, 0 otherwise.

Then we have

 E(*Ymij* | $φ\_{mij}$ ) = $φ\_{mij}$ where Var(*Ymij* | $φ\_{mij}$) = $φ\_{mij}$(1 – $φ\_{mij}$) and covariances

*Level-1 Link Function*

A common link function for multinomial regression models is the multinomial logit link, for each category *m* (for *M* – 1 categories)

 $η\_{mij}=log\left(\frac{φ\_{mij}}{φ\_{Mij}}\right)$ = $log\left(\frac{Prob(R\_{ij}=m)}{Prob(R\_{ij}=M)}\right)$

Here, the level 1 outcome is the log-odds of selecting category *m* relative to category *M*, the reference category.

*Level-1 Structural Model*

$$η\_{mij}=β\_{0j(m)}+\sum\_{q=1}^{Qm}β\_{qj(m)}X\_{qij}$$

In our case of *M* = 3, we have two level-1 equations

$$η\_{1ij}=β\_{0j(1)}+\sum\_{q=1}^{Q1}β\_{qj(1)}X\_{qij}$$

$$η\_{2ij}=β\_{0j(2)}+\sum\_{q=1}^{Q2}β\_{qj(2)}X\_{qij}$$

*Level-2 Model*

$$β\_{qj(m)}=γ\_{q0(m)}+\sum\_{s=1}^{S\_{q}}γ\_{qsm}W\_{sj}+u\_{qj(m)}$$

In our example of *M* = 3, there would be two level-2 equations.