Presentation for Michigan State HDFS Methods Seminar
February 1, 2022

The Ohio State University

Clustering in Count Data: Working with Multilevel Poisson and Negative Binomial Models

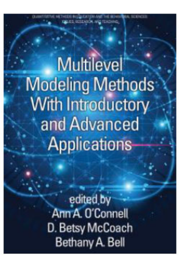
Ann A. O'Connell, Ed.D.
Quantitative Research, Evaluation & Measurement
The Ohio State University
Oconnell.87@osu.edu

MSU presentation, data and R code are at:
https://u.osu.edu/oconnell.87/res-other-page/

1

Goals for today's presentation

- Review some applications and background for multilevel count data
- Demonstration: proficiency counts modeled after the ECLSK
- Provide theoretical background and distinctions between Multilevel Poisson and Multilevel Negative Binomial Models
- Considerations on assessing degree of clustering for multilevel count data
- Example and comparisons ML P and NB



Co-authors for chapter 8:
Nivedita Bhaktha (GESIS Leibniz-Institut für Sozialwissenschäften)
Jing Zhang (Kent State)

O'Connell, A.A., Bhaktha, N. & Zhang, J. (2022). Multilevel models for count outcomes. In A.A. O'Connell, D.B McCoach, & B.A. Bell (Eds.), *Multilevel Modeling Methods with Introductory and Advanced Applications.* Information Age Publishing.

2

2

GLMMs for Counts

- Multilevel Poisson
- Multilevel Negative Binomial
- Approach:
  - Theoretical background of both models
  - Assessment of clustering
  - Present guiding questions to aid researchers in model comparison and choice, given their research questions
  - Model interpretation
  - Software issues and possible extensions or other analysis options

> - Conditional, or "cluster-specific" models
> - Challenging for researchers to model and interpret, or recognize conditions under which one (P or NB) may be preferred

3

3

Counts are common outcomes of interest

- Used to capture frequency counts of events or behaviors (0, 1, 2, …)
- These data are often nested within higher-level units or clusters
- Could result from experimental or non-experimental studies
  - How many times a teacher uses specific instructional practices in a classroom (e.g., outcomes are counts, for teachers in intervention versus control schools)
  - Student scores on a math proficiency assessment (number correct) between public and private schools (outcome are counts, for students within multiple schools)
  - Counts of health events (e.g., number of antenatal care visits for women in low/middle income countries)
  - Change over time for events (e.g., number of patient seizures) between treatment and placebo groups (repeated measures design)
  - Number of words recalled during an auditory exam (correlational study, if no intervention)
  - Others?

4

4

Some recent (multilevel) examples in the literature

- Geremew, et al. (2020). Factors affecting under-five mortality in Ethiopia: A multilevel negative binomial model. *Pediatric Health, Medicine and Therapeutics, 11*, 525-534.
- Barth & Schmitz (2021). Interviewers' and respondents' joint production of response quality in open-ended questions: A multilevel negative-binomial regression approach. *methods, data, analysis, 15*(1), 43-76.
- Chambers & Erausquin (2018). Race, sex, and discrimination in school settings: A multilevel analysis of associations with delinquency. Journal of School Health, 88(2), 159-166.
- LARRC, Lo & Xu (*in press*). Impacts of the Let's Know curriculum on the language and comprehension related skills of PK and K children. *Journal of Educational Psychology*.

5

5

Special nature of count data

- Numbers or rates of events typically vary across individuals or cases in a sample
  - Non-negative integers: 0, 1, 2, …
- Distribution of counts is often skewed/reversed J-shaped
  - Lots of people with no occurrence of the event
  - Some people with one or more occurrences
  - A few people with a very high number of occurrences
- May occur within a specific range, time span, or geographic area (i.e., "exposure" opportunities, which also may vary across cases)

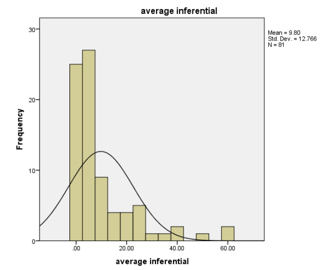
➔ Q: Can variation in counts or rates be explained by one or more predictors, and across one or more levels of the data?

6

6

Historical approaches…

- Depending on the shape of the dependent variable, people have sometimes used OLS or standard HLM, but:
  - Predictions may be negative or non-integer
  - Typically yields heteroscedastic residuals
  - Negatively affects precision of regression coefficients
    - Standard errors are downwardly biased
  - Contributes to flawed inferences
    - Can mask important relationships in the data
  - Larger sample won't help with accuracy of regression estimates



Distribution of number of inferential-type questions asked by teachers during interactive shared book-reading (Binici, 2014)

7

7

Better approach: *Multilevel Count Models*

- Multilevel models for counts can adequately capture the <u>shape</u> and <u>nature</u> of the data – and adjust for clustering
- *Poisson regression* most common *initial* model for count data
  - Poisson process counts the number of times that an event occurs
  - If events occur during a given time span or other index of size, models can estimate the <u>rate</u> at which events occur.
  - Poisson models assume that the (conditional) *mean* of the data is equal to its (conditional) *variance*
    - "*equidispersion*"
- Issue to watch out for: common occurrence with count data is "overdispersion," so need to check and adjust for this.
  - Overdispersion occurs when the variance of observed counts is greater than the mean (*var* > *mean*)

8

8

Demonstration Example

- Data modeled after proficiency counts from the ECLS-K
- https://nces.ed.gov/ecls/kindergarten.asp
- Baseline data from 1998 – 99, followed up through 1st grade, 3rd grade, 5th grade, 8th grade
- Students sampled from both public and private schools
- Modified/simulated for demonstration purposes (not meant to be representative of the ECLS-K original sample)
- Y_{ij} = proficiency counts (*profcount*) were collected for a sample of n = 11,301 Kindergarten children sampled from within J = 720 early-grade schools.
  - Scored based on the number of correct responses to a set of twelve early reading and early math skills items, with the total possible count ranging from 0 to 12 ($\bar{X}$ = 2.05 items, S = 1.85, S^2 = 3.44).

9

9

11301 Students, in 720 Schools

Alphabetic List of Variables and Attributes

	#	Variable	Type	Len	Format	Informat	Label
L1	6	c_momed	Num	8	F8.2		centered mom's highest level of education
ID1	1	childid	Num	8	F8.2		child id
L1	7	foodinsec	Num	8	FOODINS.		0 'none' 1 'food insecure'
L1	4	male	Num	8	MALE.		child sex
L1	5	momed	Num	8	MOMED.		mom's highest level of education
L2	9	nbhoodprobs	Num	8	F8.2		nbhoodprobs scale
L1	3	profcount	Num	8	F8.2		sum prof read and math
L2	8	public	Num	8	PUBLIC.		school type pub or private
ID2	2	schoolid	Char	8	$8.	$8.	school code

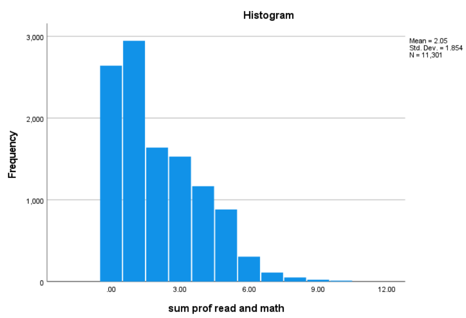
L1 = student level   L2 = school level

- Range of 5 to 24 Kindergarteners per school
- Average of about 16 children per school

10

10

Proficiency Count Distribution – Kindergarten



(Not adjusted for clustering by school)

Statistics

profcount sum prof read and ma

N	Valid	11301
	Missing	0
Mean		2.0526
Median		2.0000
Mode		1.00
Std. Deviation		1.85431
Variance		3.438
Range		10.00
Minimum		.00
Maximum		10.00

var > mean

11

11

Descriptive Statistics – Student Level (n = 11301)

Statistics

		male child sex	momed mom's highest level of education	c_momed centered mom's highest level of education	foodinsec 0 'none' 1 'food insecure'
N	Valid	11301	11301	11301	11301
	Missing	0	0	0	0
Mean		.5017	4.45	.0000	.0760
Median		1.0000	5.00	.5506	.0000
Mode		1.00	3	-1.45	.00
Std. Deviation		.50002	1.714	1.71383	.26503
Range		1.00	8	8.00	1.00
Minimum		.00	1	-3.45	.00
Maximum		1.00	9	4.55	1.00

- 50% male
- 8% food insecure
- Median Mom's Ed is 5 – at least some college
- Mean Mom's Ed is 4.45, so vocational/tech, 2 yr college or at least some college

12

12

Descriptive Statistics – School Level (J = 720)

Statistics

		public	nbhoodprobs
N	Valid	720	720
	Missing	0	0
Mean		.7750	6.2944
Median		1.0000	5.0000
Mode		1.00	4.00
Std. Deviation		.41787	3.38564
Range		1.00	18.00
Minimum		.00	3.00
Maximum		1.00	21.00

- 78% public schools
- Nbhoodprobs represents degree of crime/conflict issues in vicinity of school (crime, trash, vacant buildings, drug-use, etc.), possible range from 0 to 21:
  - Mean = 6.29
  - Std dev = 3.39

13

13

Poisson or Negative Binomial?

- Poisson Regression models the *conditional* distribution of Y, that is, the (count) distribution that occurs when Y is *conditional* on the values of the predictors
  - We are interested in modeling this conditional process
- However, the Poisson (P) model makes a very strong assumption that the variance of the distribution is equal to the mean: E(Y) = Var(Y) = λ
  - "Equidispersion"
- Adaptations are needed when overdispersion is present
- Clustering is also a source of overdispersion, if ignored
- Multilevel extensions for P and NB adjust for the clustering in the data
- The Negative Binomial (NB) model adjusts for and includes an additional parameter that captures the extra variation present in the data

14

14

Choice Between NB and P

- Unfamiliarity with NB can be challenging for applied researchers
- Four Guiding Questions
  1. To what degree is clustering present in the (count) data, and does accounting for clustering improve model fit (over a single-level model)
  2. After adjusting for clustering, is the assumption of equidispersion reasonable, or should this assumption be relaxed?
  3. Does inclusion of covariates improve model fit, and should any level-one predictors be treated as random or fixed?
  4. How are parameter estimates interpreted in our final model (for multilevel P or NB?)
- These questions mirror similar approaches for "standard" (continuous outcome) multilevel models
  - Added complexity given the count nature of the outcome

Some background and terminology

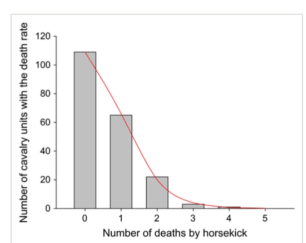
Brief review of Poisson

Model comparisons

Adjustments for overdispersion

Władysław Bortkiewicz: Author of *The Law of Small Numbers* (1898)
image from *mathshistory.st-andrews.ac.uk*

Bredows Charge at Mars-la-Tour
from *theminiaturespage.com*



Bortkiewicz data on deaths in Prussian Cavalry from *associationofanaesthetists*

Poisson distribution

- *Y* is a non-negative integer, and can take on response values *y* = 0, 1, 2, ...
- We are interested in the *probability* of observing a specific count response:

$$P(Y = y) = \frac{\lambda^{y} e^{-\lambda}}{y!} \text{ for y} = 0, 1, 2, \ldots$$

  - where λ = E(Y) = V(Y) (property of equidispersion)
  - λ can be any positive value (not necessarily an integer, since it's an average)
    - it is the mean, or the "rate parameter"
    - Average number of events (counts) that occurs in the identified observation period
    - "exposure" time – if it varies for each case – can be adjusted for as an "offset"

17

17

Poisson Regression Model (single-level)

- In the regression model, the number of events is assumed to follow a Poisson distribution with a conditional mean based on each individual's characteristics:

$$P(y_i \mid \mathbf{x}_i) = \frac{\lambda_i^{\,y_i} e^{-\lambda_i}}{y_i!}$$

$$\lambda_i = E(y_i \mid \mathbf{x}_i) = \exp(\mathbf{x}_i \boldsymbol{\beta})$$

No random error term in the regression model, since we already assume that the variance is equal to the mean

$$\log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots \beta_p x_{ip}$$

A generalized linear model, for which the parameter estimates are found through maximum likelihood estimation

18

18

Coefficient Interpretation (single-level)

- Each estimated coefficient, $\hat{\beta}_j = b_j$, represents the expected change in the natural log of counts given a one-unit change in the predictor, holding other predictors constant
- Coefficients can be exponentiated to represent change in actual counts (rather than change in the log-counts) to yield the *rate ratio* (RR) or *incidence rate ratio* (IRR)
  - exp(b_j) = RR
  - Each RR explicitly compares the *rates* of two values of the predictor as it increases by one unit
  - For each one-unit increase in the predictor, the expected count will multiplicatively increase (or decrease, if the parameter is negative) by the value of the RR, holding other predictors constant

19

19

Negative Binomial Regression Model (single-level) (1)

- Includes a distinct source of dispersion over the assumed equidispersion Poisson variance

$$\log(\tilde{\lambda}_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots \beta_p x_{ip} + \varepsilon_i$$

$$\tilde{\lambda}_i = \exp(\mathbf{x}_i \boldsymbol{\beta}) \cdot \exp(\varepsilon_i) = \lambda_i \cdot \exp(\varepsilon_i) = \lambda_i \cdot \delta_i$$

But: How to characterize the $\delta_i = \exp(\varepsilon_i)$

- Allows for count estimation models where the variance can be different from (greater than) the mean
- Rates may vary across individuals with the same collection of predictors
  - Randomly perturbed or vary by δ_i
- Interpretation of fixed model parameters and RR same as in the P regression
- To *identify* the NB regression model, we make an assumption, $E(\delta_i) = 1.0$
- Expected counts then are the same as they were in the P regression:

$$E(\tilde{\lambda}_i) = E(\lambda_i \delta_i) = \lambda_i E(\delta_i) = \lambda_i$$

20

20

Negative Binomial Regression Model (single-level) (2)

- The distribution of $P(y_i \mid x_i, \delta_i)$ is still Poisson, but we can't solve this equation without specifying the form of pdf for the δ_i. (Cameron & Trivedi, 2013, p. 115-119; Long, 1997,p. 231-232; Long & Freese, 2014, p. 507-508; Stroup, 2013, p. 352-354)
  - (note these refs do not always use the same symbols when presenting their equations)
- The most commonly used distribution for the extra dispersion represented by δ_i is the *Gamma* distribution.
- This form of NB model is a mixed Gamma-Poisson model (NB2 or quadratic form)
- The Gamma distribution in general has two parameters: one for scale or location and one for shape (θ).
  - With $E(\delta_i) = 1.0$ for identification, we are interested in the single shape parameter, θ
- Degree of overdispersion in the conditional means is governed by θ, and the variance of the NB distribution is a function of both the Poisson and Gamma variances

21

21

Variance of NB (NB2) as the Poisson-Gamma Mixture

- A distribution that is Gamma(λ, θ) has variance λ^2/θ
- Thus the variance for the mixture is: $V(Y) = \lambda + \lambda^2/\theta$
- This represents an *indirect* relationship between θ and dispersion
  - As θ tends to infinity, the extra dispersion component tends to 0
- To capture a *direct* association, the NB model is often parameterized so that $\alpha = 1/\theta$ represents the extra dispersion: $V(Y) = \lambda + \alpha\lambda^2$
  - As α increases, so does the extra variance component
  - As α tends to 0, less extra dispersion is added, and the variance of the mixture converges to that of the Poisson
- Under this parameterization, α is referred to as the heterogeneity or NB dispersion parameter.
- There are other ways to express an NB model, but this representation is referred to as the *NB2* or simply the *NB* model

(McCullogh & Nelder, 1989)

Poisson is a reduced form of the P-G mixture, thus allowing for nested comparisons of the models

22

22

Clustering and Multilevel P and NB Models

- Presence of clustering in the data can contribute to overdispersion

Multilevel Poisson

$$y_{ij} \mid \lambda_{ij} \sim Poisson(\lambda_{ij})$$
$$\ln(\lambda_{ij}) = \gamma_{00} + u_{0j}$$
$$u_{0j} \sim N(0, \tau_{00})$$

Multilevel NB

$$y_{ij} \mid \lambda_{ij} \sim Poisson(\lambda_{ij})$$
$$\ln(\lambda_{ij}) = \gamma_{00} + u_{0j} + e_{ij}$$
$$u_{0j} \sim N(0, \tau_{00})$$
$$\exp(e_{ij}) \sim Gamma\left(\frac{1}{\alpha}, \alpha\right)$$

- Random intercepts and slopes can be added, similar to standard HLM models
- NB adds random effect at the student-level to capture contribution of omitted variables thought to be related to overdispersion
- The NB exponentiated dispersion effects, $\exp(\varepsilon_{ij})$, are Gamma distributed.
  - Only one parameter is estimated
    - $1/\alpha$ = scale or location
    - α = shape
- Thus the mean for the $\exp(\varepsilon_{ij})$ is 1.0, and overdispersion (variance) parameter is α

23

23

Degree of Clustering

- For data that are hierarchical, nested or clustered, one of the first steps we pursue in a multilevel analysis is the degree to which systematic differences between clusters is present (ICC or intracluster coefficient).
- There have been challenges until recently to accurately assess degree of clustering for count data
  - Equidispersion assumption of Poisson
  - Level-1 and level-2 components of variance are measured on different scales (count response scale at level 1, continuous scale at level 2)
- Austin et al. (2018) and Leckie et al. (2020) developed and provide code for "exact" approach to estimating the ICC for counts

24

24

Assessing the ICC for Example Data

- For our demonstration data (Y = *profcounts* of K children assessed within schools), we have:
- Normal approximation (for crude comparison)
  - ICC = .24 (24%)

Not recommended as an estimate of correlation among students within the same school, due to skew, J-shape of the data

- Multilevel Poisson
  - ICC = .37 (37%)

We followed Leckie et al. (2020) & Austin et al. (2018), focusing on Variance Partition Coefficients for random intercept models with no covariates (variance component models), using marginal/population-averaged model estimates.

- Multilevel NB
  - ICC = .28 (28%)

Calculations, next page

Summary: Once clustering is accounted for, the NB model indicates somewhat smaller estimated correlation between students from the same school, but 28% of the (marginal) variance in *profcounts* can be attributed to systematic difference between schools.

Based on Leckie et al. (2020)

Marginal expectation of y_{ij} (Poisson): $\lambda_{ij}^{M} \equiv E(y_{ij}) = \exp\left(\gamma_{00} + \frac{\tau_{00}}{2}\right)$

Marginal variance (Poisson): $Var(y_{ij}) = \left(\lambda_{ij}^{M}\right)^2 \bullet [\exp(\tau_{00}) - 1] + \left(\lambda_{ij}^{M}\right)$

$$\text{ICC(Poisson)} = \text{VPC}_{ij}(P) = \frac{\left(\lambda_{ij}^{M}\right)^2 \bullet [\exp(\tau_{00}) - 1]}{\left(\lambda_{ij}^{M}\right)^2 \bullet [\exp(\tau_{00}) - 1] + \left(\lambda_{ij}^{M}\right)}$$

$$\text{ICC(NB)} = \text{VPC}_{ij}(\text{NB}) = \frac{\left(\lambda_{ij}^{M}\right)^2 \bullet [\exp(\tau_{00}) - 1]}{\left(\lambda_{ij}^{M}\right)^2 \bullet [\exp(\tau_{00}) - 1] + \left[\left(\lambda_{ij}^{M}\right) + \left(\lambda_{ij}^{M}\right)^2 \exp(\tau_{00})\alpha\right]}$$

Comparison of Competing Models

- Several ways to choose between competing models
- Most common is through the *Deviance*
  - Sum of the squared deviance residuals
  - Can compare sum of the squared Pearson residuals to df, as additional assessment of presence of overdispersion
- Deviance statistics represent badness of fit or discrepancy between a model and the actual data
- As with other generalized and mixed linear models, the deviance itself may not follow the expected χ^2 distribution, but the *difference* in deviances or differences in the -2LL between two nested models will have *df* equal to the difference in number of parameters between the models being compared
- To address our guiding questions (slide 15), we utilize (in part) the Chi-square difference test for competing nested models.

27

27

Results – Table 1 – Random Intercepts P and NB

Table 1. Selected Analysis Results of Proficiency Counts using Two-level Random-Intercept Poisson and Negative Binomial Models, Parameter estimates (standard errors); Results in SAS PROC GLIMMIX (Laplace estimation).

	Random Intercepts Poisson			Random Intercepts NB		
Parameter/Statistic	Unconditional	Level-1 predictors	Level-1 and 2 predictors	Unconditional	Level-1 predictors	Level-1 and 2 predictors
γ_{00}= *Intercept*	0.59 (.02)[a]	0.65 (.02)[a]	1.12 (.04)[a]	0.60 (.02)[a]	0.65 (.02)[a]	1.12 (.04)[a]
γ_{10} = *male*		−.09 (.01)[a]	−0.09 (.01)[a]		−0.09 (.02)[a]	−0.09 (.02)[a]
γ_{20} = *c_momed*		0.12 (.004)[a]	0.12 (.004)[a]		0.13 (.005)[a]	0.12 (.01)[a]
γ_{30} = *foodinsec*		−0.21 (.03)[a]	−0.20 (.03)[a]		−0.22 (.04)[a]	−0.21 (.04)[a]
γ_{01} = *public*			−0.29 (.04)[a]			−0.28 (.04)[a]
γ_{02} = *nbhoodprobs*			−0.04 (.005)[a]			−0.04 (.005)[a]
τ_{00} = school intercept variance	0.25 (.02)[a]	0.17 (.01)[a]	0.13 (.01)[a]	0.24 (.02)[a]	0.15 (.01)[a]	0.12 (.01)[a]
α = overdispersion				0.15 (.01)[a]	0.12 (.009)[a]	0.12 (.009)[a]
-2LL	41209.50	40321.52	40182.89	40859.01	40091.42	39955.28
Pearson χ^2/df	1.23	1.18	1.19	0.96	0.97	0.98
AIC	41213.50	40331.52	40196.89	40865.01	40103.42	39971.28

[a] $p < .0001$

28

Results – Table 2 – Random Coefficients NB

Table 2. Random Coefficients Results for Negative Binomial Models, Parameter estimates (standard errors); Results in SAS PROC GLIMMIX (Laplace estimation).

	Random Coefficients NB			
Parameter/Statistic	Model A	IRR	Model B	IRR
γ_{00}= *Intercept*	0.67 (.02) [a]	1.95	1.05 (.04) [a]	2.86
γ_{10} = *male*	−0.09 (.02) [a]	.91	−.09 (.02) [a]	.91
γ_{20} = *c_momed*	0.15 (.006) [a]	1.16	0.14 (.006) [a]	1.15
γ_{30} = *foodinsec*	−0.21 (.04) [a]	.81	−0.20 (.04) [a]	.82
γ_{01} = *public*			−0.21 (.03) [a]	.81
γ_{02} = *nbhoodprobs*			−0.04 (.005) [a]	.96
τ_{00} = school intercept variance	0.42 (.05) [a]		0.33 (.04) [a]	
τ_{22} = *c_momed* slope variance	0.0031 (.001) [b]		0.0025 (.001) [b]	
α = overdispersion	0.11 (.009) [a]		0.11 (.009) [a]	
Pearson χ^2/df	0.97		0.97	
-2LL	40001.79		39896.83	
AIC	40017.79		39916.83	

[a] $p < .0001$
[b] $p < .001$

29

29

Guiding Questions (1)

Table 1 contains results for random intercept multilevel P and NB models

- *To what degree is clustering present in the data, and does accounting for clustering improve model fit (over a single-level model)*
- We know that the effect of clustering is salient for these data
  - By design of the study/data (students clustered within schools)
  - By the presence of positive ICC (P (37%) and NB (28%))
- Single-level Poisson is nested within the multilevel Poisson
  - -2LL(single-level P) = 44497.10 (genmod, null)
  - -2LL(multi-level P) = 41209.50 (null)

$$\chi^2_1 = 3287.50 \quad (sig)$$

- Single-level NB is nested within the multilevel NB
  - -2LL(single-level NB) = 42569.07 (genmod, null)
  - -2LL(multi-level NB) = 40859.01 (null)

$$\chi^2_1 = 1710.06 \quad (sig)$$

Summary: systematic differences between schools is present (i.e, clustering effect), and supports multilevel approach to analysis.
For both P and NB, chi-square diff is sig, supporting the multilevel model.

30

30

Guiding Questions (2)

- *After adjusting for clustering, is the assumption of equidispersion reasonable, or should this assumption be relaxed?*
- Because the P is nested within the NB model, we can use a similar likelihood ratio test to compare the two (null) models
  - -2LL(multilevel P) = 41209.50 (null)
  - -2LL(multi-level NB) = 40859.01 (null)

$$\chi_1^2 = 350.49 \quad (sig)$$

- Summary: Even after adjusting for clustering (in P model), comparison supports the need to adjust for overdispersion as in the NB model
- Additionally, for the NB dispersion parameter, α = 0.15 (s.e. = .01).
  - Wald's test indicates that α is significantly greater than zero, .15/.01 = 15 (*sig*)
- Can revisit these as needed once covariates are added to the model

31

31

Guiding Questions (3)

Table 2 contains random coefficients results for the NB models

- *Does inclusion of covariates improve model fit, and should any level-one predictors be treated as random or fixed?*
- First considered adding covariates at level 1 for the NB model as fixed
  - *Male*, *c_momed*, *foodinsec*
- From Table 1 (NB Random Intercepts)
  - -2LL(multilevel NB, null) = 40859.01
  - -2LL(multi-level NB, 3 covs) = 40091.42

$$\chi_3^2 = 767.59 \quad (sig)$$

- From Table 2 (NB Random Coefficients results)
  - Only *c_momed* was included with a significant random effect
  - -2LL(with 3 fixed effects for covariates) = 40091.42
  - -2LL(with *c_momed* as random effect) = 40001.79

$$\chi_2^2 = 89.63 \quad (sig)$$

Summary: included *c_momed* as the only random slope

32

32

Guiding Questions (3) continued

- Next, added the two school-level covariates, *public* and *nbhoodprobs* to the NB model with the random slope for *c_momed*
- This full model fit better than the NB model without level-2 predictors
  - -2LL(without level-2 covariates) = 40001.79
  - -2LL(with level-2 covariates) = 39896.83

$$\chi^2_2 = 104.96 \quad (sig)$$

- Note that the AIC is smallest for this full model (AIC = 39916.83)
  - Overdispersion parameter is also statistically significant in the full NB model
  - .11/.009 = 12.11 (*sig*)

Summary: Model to interpret here is the model with one random slope (c_momed) and the two school-level predictors.

Note that additional comparisons may be relevant, given one's approach to modeling in general.

33

33

Guiding Questions (4)

- *How are parameter estimates interpreted in our final model (for multilevel P or NB?)*
- Predicted outcome for the multilevel NB models, ln(Y_{ij}), is the log of the proficiency counts (number of correct responses by Kindergartners to early literacy and numeracy items)
- The values of the parameter estimates for covariates do not fluctuate too much across models for these data, although the intercept does, as does its variance in the random coefficients models
- All fixed effects are statistically significant (likely given large sample size)
- Adding predictors reduces variability; the NB dispersion does get somewhat smaller as model complexity increases

34

34

Guiding Questions (4) (cont.)

- Manner of interpretation is similar regardless of P or NB model
  - Focus on Table 2, Model B (NB random coefficients with 1 random slope and 2 school-level predictors)
- Intercept: γ_{00} = 1.05 (*sig*)
  - IRR = Exp(1.05) = 2.86
  - Expected count for female child with mom's education at the midpoint (of at least some college), and no presence of food insecurity, in a private school with no indication of community problems in vicinity of the school (crime, vacant buildings, trash, etc.)
    - Given random effects of 0
- Addition of NB dispersion doesn't affect how parameters are interpreted
  - Just tells us about unaccounted variation around our estimates

35

35

Guiding Questions (4) (cont.)

- Within Schools (using IRR = exp(b))
- Effect of *c_momed*: As momed increases by 1 unit (i.e., from 5 = some college to 6 = Bach. Degree), IRR = 1.16
  - Controlling for other predictors in the model, and for a "typical" participant (random effects of 0), the expected proficiency counts increase 1.16 times
  - Represents a percentage change of 100(1.15 – 1) = 15% increase
- Effect of *male*: (*male* = 0 for females, 1 for males) IRR = .91
  - Controlling for other predictors, the number of expected items correct for males (assuming random effects of 0) is estimated to be .91 times the number of items correct for females (i.e., smaller for males)
  - Represents a percentage change (decrease) of 100(.91 – 1) = 9% decrease
- Similarly: For children from families experiencing *foodinsec* (dichotomous), IRR = .82, a 18% decrease

36

36

Guiding Questions (4) (cont.)

- Between Schools (using IRR = exp(b))
- Effect of *public*: IRR = .81. Children from public schools experience an expected 19% decrease in predicted counts, holding all else constant (and random effects of 0).
- Effect of *nbhoodprobs*: IRR = .96. As the number of neighborhood problems increases by 1 unit, expected counts in proficiency decrease by 4%, holding all else constant (including random effects of 0).

37

37

Other considerations

- Software for NB models: α vs θ
  - Software provides either α or θ (i.e., *lme4, glmmTMB* packages provide θ rather than α)
  - Read all documentation carefully to understand the nature of dispersion in NB models
- Graphs of expected versus actual counts, and graphs demonstrating patterns of change for different values of the covariates, can be very helpful in understanding your final model (or comparing across models)
- Other models for counts
  - While P and NB2 models seem to be most commonly used in the applied literature, they are not the only potential approaches to consider
  - Data with excess zeros could be analyzed using zero-inflated or hurdle models, depending on your theory as to how the excess zeros
  - GEE has also been recommended as an alternative method to deal with overdispersion (Stroup, 2013)
- Bolker's github FAQs very helpful for discussion and recommendations on GLMMs for count models in R.

38

38

Thank you!

- You are welcome to contact me for additional information or questions

Prof. Ann A. O'Connell
210B Ramseyer Hall
College of Education and Human Ecology
The Ohio State University
Columbus OH 43210
Oconnell.87@osu.edu

MSU presentation, data and R code are at:
https://u.osu.edu/oconnell.87/res-other-page/

References follow …

39

39

Selected References

- Austin, P. C., Stryhn, H., Leckie, G., & Merlo, J. (2018). Measures of clustering and heterogeneity in multilevel Poisson regression analyses of rates/count data. *Statistics in Medicine*, *37*(4), 572-589.
- Binici, S. (2014). *Preschool teachers inferential questions during shared reading and their relation to low-income children's reading comprehension at K and 1st grade*. [Doctoral Dissertation]. The Ohio Sate University.
- Bolker, B., et al. (2021, June 28). GLMM FAQ. https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (2nd ed.). Cambridge University Press.
- Fox, J. (2017). *Applied regression analysis and generalized linear models* (3rd ed.). Sage.
- Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.
- Hilbe, J. M. (2014). *Modeling count data*. Cambridge University Press.
- Leckie, G., Browne, W. J., Goldstein, H., Merlo, J., & Austin, P. C. (2020). Partitioning variation in multilevel models for count data. *Psychological methods*.
- Long, J.S. (1997). *Regression models for categorical and limited dependent variables*. Sage.
- Long, J. S., & Freese, J. (2014). *Regression models for categorical dependent variables using Stata* (3rd ed). Stata Press.
- McCullagh, P., & Nelder, J. A. (1989). Generalized Linear Models, vol. 37 CRC Press.
- Snijders, T.A.B. & Bosker, R.J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd Ed.). Sage.
- Stroup, W.W. (2013). *Generalized linear mixed models: Modern concepts, methods and applications*. CRC Press

40

40