MWERA conference, October 2022: Cincinnati, Ohio

The Ohio State University

On extra-dispersion in clustered count models: A brief tutorial

Ann A. O'Connell, Krisann Stephany, Winifred Wilberforce & Abena Anyidoho
The Ohio State University
Oconnell.87@osu.edu

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

1

Today's presentation

- Review background, applications for Multilevel Count data
- Examine assumption of equidispersion and contributions to over-dispersion
- Provide illustration: Multilevel Poisson, Negative Binomial and Generalized Poisson
- Recommendations and next steps

More details on multilevel counts: O'Connell, Bhaktha & Zhang (2022). Chapter 8: Single and multilevel models for count outcomes. In *Multilevel Modeling Methods with Introductory and Advanced Applications* (O'Connell, McCoach & Bell (Eds.)) Information Age Publishing.

2

2

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, across one or more levels of the data?

3

3

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 (2022). Impacts of the Let's Know curriculum on the language and comprehension related skills of PK and K children. *Journal of Educational Psychology*.

4

4

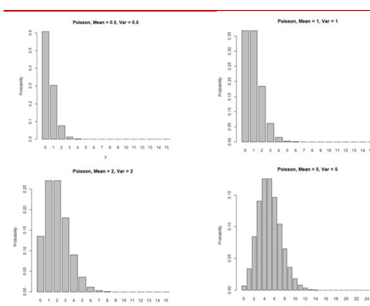
Models for Counts

- *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 rate 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*)

5

5

What does equi – vs over-dispersion look like?



Poisson Density plots for 4 different values of the mean (.5, 1, 2, 5)

As the mean increases, so does the variance

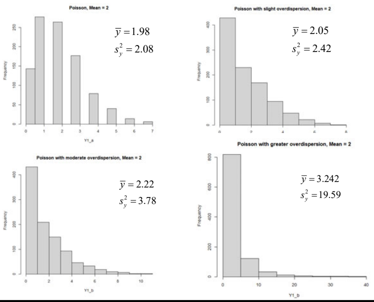
One parameter distribution
- Mean = Variance

However, real count data may not strictly follow a Poisson distribution

6

6

Add some noise to sampled data (Poisson, μ = 2)…



Sampled n = 1000 observations

1st plot are data sampled from a Poisson, with μ = 2

2nd, 3rd & 4th plots are data sampled from a Poisson with μ = 2, but with random error added

- N(0, .25)
- N(0, .50)
- N(0, 1.0)

Ratio of Var to Mean for each sample?

- 1st : 1.05
- 2nd: 1.18
- 3rd: 1.70
- 4th: 6.04

7

7

Contributions to Overdispersion

- Many factors can contribute to overdispersion
  - Unaccounted for clustering in the data
  - Missing covariates or interaction terms
  - Unreliable data collection
- We often don't know the distribution of the extra dispersion
  - Need to make assumptions about how it's distributed
  - Adjust our model to better represent the data
- Overdispersion is a property of the data *relative to* the model we choose
- The data aren't wrong, but our models or methods might be!

8

8

Alternatives when Overdispersion is Present

- Over-dispersed Poisson (quasi-Poisson)
- Negative Binomial (NB)
- Generalized Poisson (GP)
- Zero-inflation models (ZIP or ZINB)
- Hurdle models

We'll focus on NB and GP, and in comparison to the Poisson

- GLMM = generalized linear *mixed* models – for data that are nested or clustered

9

9

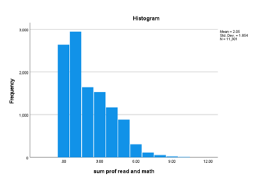
Demonstration Example

- 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).

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

$$\frac{var}{mean} = 1.67$$

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 <u>or</u> Negative Binomial <u>or</u> Generalized Poisson?

- 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
- 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 known as "*Equidispersion*"
- The Negative Binomial (NB) model adjusts for and includes an additional parameter that captures the extra variation present in the data
  - Mixture of Poisson and Gamma distributions
- The Generalized Poisson (GP) estimates a scale parameter that adjusts both the variance <u>and</u> the mean
  - Mixture of Poisson and Poisson distributions

14

14

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:

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

  - 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"

15

15

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

16

16

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

17

17

Negative Binomial Regression Model (single-level)

- Includes a <u>distinct source of dispersion</u> 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
- <u>Interpretation</u> of fixed model parameters and RR <u>same as in the P</u> regression
- To *identify* the NB regression model, we make an assumption, E(δ_i) = 1.0
- <u>Expected</u> counts then are the same as they were in the P regression:

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

18

18

NB Dispersion Parameter

- 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 governing 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 thus the variance of the NB distribution is a function of both the Poisson and Gamma variances

19

19

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

- A distribution that is Gamma(λ, θ) has variance λ^2/θ
- 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

20

20

Generalized Poisson

NOTE: dispersion parameter is defined for *each family* of distribution; the same symbols do not necessarily refer to same quantity or parameter

- Alternative to the NB: Mixture of Poisson with Poisson = Generalized Poisson (Consul, 1998; Joe & Zhu, 2005; Harris, Yang & Hardin, 2012; SAS, 2022)

$$f(y_i, \lambda_i, \theta) = \lambda_i(\lambda_i + \theta y_i)^{y_i - 1} e^{-\lambda_i - \theta y_i} / y!$$
$$\text{for } \lambda_i > 0, \text{ and } [\max(-1, -\lambda_i/4)] \le \theta < 1$$

(NOTE: $\theta < 1$)

- The mean and variance of Y under GP distribution are:

$$E(Y_i) = \mu_i = \frac{\lambda_i}{1-\theta}$$
$$Var(Y_i) = \frac{\lambda_i}{(1-\theta)^3} = \frac{1}{(1-\theta)^2} E(Y_i) = \phi^2 E(Y_i)$$

When $\theta = 0$ there is equidispersion, and GP → Poisson
When $\theta > 0$ we have over-dispersion
When $\theta < 0$ we have under-dispersion

- GP Regression Model: $\log\left(\frac{\lambda_i}{1-\theta}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots \beta_p x_{ip}$

21

21

Clustering and Multilevel P, NB and GP Models

- Presence of clustering in the data can contribute to overdispersion
  - Random intercepts and slopes can be added, similar to standard HLM models
  - NB and GP include random effects at the lower-level even after adjusting for clustering
- Some notes between *R* (glmmTMB) and *SAS (*Proc Glimmix)
- NB (slide 22)
  - *R* estimates what we've called θ
  - SAS estimates what we've called α, which is $1/\theta$
- GP (slide 23)
  - R estimates what we've called ϕ^2 (and calls it "sigma"), where $\phi^2 = 1/(1-\theta)^2$
  - SAS estimates θ, and labels as the "scale" parameter

22

22

Assessing the ICC for Example Data

- For our example data (Y = *profcounts* of K children assessed within schools), we have:
- Normal approximation (for crude comparison)
  - ICC = .24 (24%)
- Multilevel Poisson
  - ICC = .37 (37%)
- Multilevel NB
  - ICC = .28 (28%)

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.

23

23

Comparison of Competing Models

- Several ways to choose between competing models
- Most common is through the *Deviance* or -2LL
  - 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
- Information criteria
- Others: observed to predicted counts, graphical comparison, etc.

- Goal for today is to illustrate how these models work, general comparison between them

24

24

Parameter estimates (standard errors) for proficiency count data using two-level random-intercept Poisson, NB2, and GP Models: Results in SAS PROC GLIMMIX (AQ estimation).

Parameter/Statistic	Random Intercepts Models		
	Random-Intercepts **Poisson**	Random-Intercepts **Negative Binomial**	Random-Intercepts **Generalized Poisson**
γ_{00}= *Intercept*	1.12 (.04) ᵃ	1.12 (.04) ᵃ	1.14 (.04) ᵃ
γ_{10} = *male*	−0.09 (.01) ᵃ	−0.09 (.02) ᵃ	−0.10 (.02) ᵃ
γ_{20} = *c_momed*	0.12 (.004) ᵃ	0.12 (.01) ᵃ	0.12 (.005) ᵃ
γ_{30} = *foodinsec*	−0.20 (.03) ᵃ	−0.21 (.04) ᵃ	−0.21 (.04) ᵃ
γ_{01} = *public*	−0.29 (.04) ᵃ	−0.28 (.04) ᵃ	−0.29 (.04) ᵃ
γ_{02} = *nbhoodprobs*	−0.04 (.005) ᵃ	−0.04 (.005) ᵃ	−0.04 (.005) ᵃ
τ_{00} = intercept.var	0.13 (.01) ᵃ	0.12 (.01) ᵃ	0.12 (.01) ᵃ
NB Disp.Param.		α = 0.12 (.009) ᵃ	
GP Scale			θ = 0.13 (.008) ᵃ
-2LL	40182.89	39955.28	39801.72
Pearson χ^2/df	1.19	0.98	0.92
AIC	40196.89	39971.28	39817.72

ᵃ p < .0001 [Notes: NB θ in R = 8.7; QP phi² ="sigma" in R = 1.30917]

25

25

Rate Ratios (RR) = exp(fixed effects)

Table 2.

Rate Ratios for the GP Model

Fixed Effect	GP Rate Ratios (se)
γ_{00}= *Intercept*	3.11 (.13)
γ_{10} = *male*	0.91 (.01)
γ_{20} = *c_momed*	1.13 (.006)
γ_{30} = *foodinsec*	0.81 (.03)
γ_{01} = *public*	0.75 (.03)
γ_{02} = *nbhoodprobs*	0.96 (.01)

- Intercept: γ_{00} = 1.14 (*sig*)
  - Exp(1.14) = 3.11
  - Expected proficiency count when all predictors are 0
    - Given random effects of 0
    - Dispersion estimate doesn't affect parameter interpretation, just tells us about unaccounted for variation around our estimates

Slope: *c_momed,* γ_{20} = 0.12 (sig): Exp(0.12) = IRR = 1.13

As momed increases by 1 unit (i.e., from 5 = some college to 6 = Bach. Degree), the expected proficiency counts increase 1.13 times

- Controlling for other predictors in the model, and for a "typical" participant (random effects of 0)
- Represents a percentage change of 100(1.13 – 1) = 13% increase

26

26

Summary

- Multilevel GP is a reasonable alternative to the NB
  - Easy to fit and interpret
  - Also appropriate for under-dispersion (while the NB is not)
- In our example, the multilevel GP had better "fit" criteria, although the difference wasn't too dramatic
- Software parameterizations are not consistent across platforms, so some caution is needed when interpreting
- Next steps
  - Are there model conditions for which GP may always outperform the NB? Or vice versa?
    - Sample size
    - Degree of overdispersion
    - Degree of clustering
    - Model complexity
- ICC estimates from the multilevel GP

27

27

Thank you!

- Please visit my website for code in SAS and R

https://u.osu.edu/oconnell.87/res-other-page/

Ann A. O'Connell

Oconnell.87@osu.edu

References follow …

28

28

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.
- Consul, PC & Famoye, F. (1992). Generalized Poisson regression model. *Communications in Statistics, 21(1)*, 89-109
- Fox, J. (2017). *Applied regression analysis and generalized linear models* (3rd ed.). Sage.
- Harris, T, Yang, Z, & Hardin, JW. Modeling underdispersed count data with generalized Poisson regression. *The Stata Journal, 12*(4), 736-747
- Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.
- Hilbe, J. M. (2014). *Modeling count data*. Cambridge University Press.
- Joe, H & Zhu, R. (2005). Generalized Poisson distribution: the property of mixture of Poisson and comparison with Negative Binomial Distribution. *Biometrical Journal, 2*, 219-229.
- 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

29

29