Calculating intraclass correlation coefficients in multilevel models for count responses

George Leckie

School of Education

University of Bristol

g.leckie@bristol.ac.uk

bristol.ac.uk/cmm/team/leckie.html

Introduction

- A standard first step when fitting a multilevel model is to justify the need for a multilevel approach by reporting the degree of clustering in the response using
  - Variance partition coefficients (VPCs)
  - Intraclass correlation coefficients (ICCs)

- When fitting continuous response models, simple well-known formula exist and these can also be applied, with minor modification, when fitting binary, ordinal, or nominal response models (via their latent response formulation)

- But what should we do when fitting count response models, especially as here there is no latent response formulation?

Poisson regression
Two-level random-intercept model

- Consider the two-level random-intercept Poisson regression

$$y_{ij}|\mu_{ij} \sim \text{Poisson}(\mu_{ij})$$

$$\ln(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_j$$

$$u_j \sim N(0, \sigma_u^2)$$

- We can't apply the usual VPC/ICC formula

$$\text{VPC} \equiv \text{ICC} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

as here there is no level-1 variance σ_e^2 in the usual sense

Poisson regression
Analytic expression for VPC/ICC

- In Austin et al. (2018) we showed that

$$\text{VPC} \equiv \text{ICC} = \frac{\overbrace{\left(\mu_{ij}^{M}\right)^{2}\{\exp(\sigma_u^2) - 1\}}^{\text{"level−2 variance"}}}{\underbrace{\left(\mu_{ij}^{M}\right)^{2}\{\exp(\sigma_u^2) - 1\}}_{\text{"level−2 variance"}} + \underbrace{\mu_{ij}^{M}}_{\text{"level−1 variance"}}}$$

where $\mu_{ij}^{M} = \exp\left(\mathbf{x}_{ij}'\boldsymbol{\beta} + \sigma_u^2/2\right)$ is the marginal mean of y_{ij}

- Clearly the VPC/ICC now depends on the covariates and so one might want to evaluate the VPC/ICC at different values of the covariates

Negative Binomial regression
Two-level random-intercept model

- In new work, we focus on the two-level random-intercept negative binomial regression (mean dispersion; NB2)

$$y_{ij}|\mu_{ij} \sim \text{Poisson}(\mu_{ij})$$

$$\ln(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_j + e_{ij}$$

$$u_j \sim N(0, \sigma_u^2)$$

$$\exp(e_{ij}) \sim \text{Gamma}(\alpha^{-1}, \alpha)$$

where e_{ij} represents the omitted unit-level variables that are envisaged to induce the overdispersion in the counts

Negative Binomial regression
Analytic expression for VPC/ICC

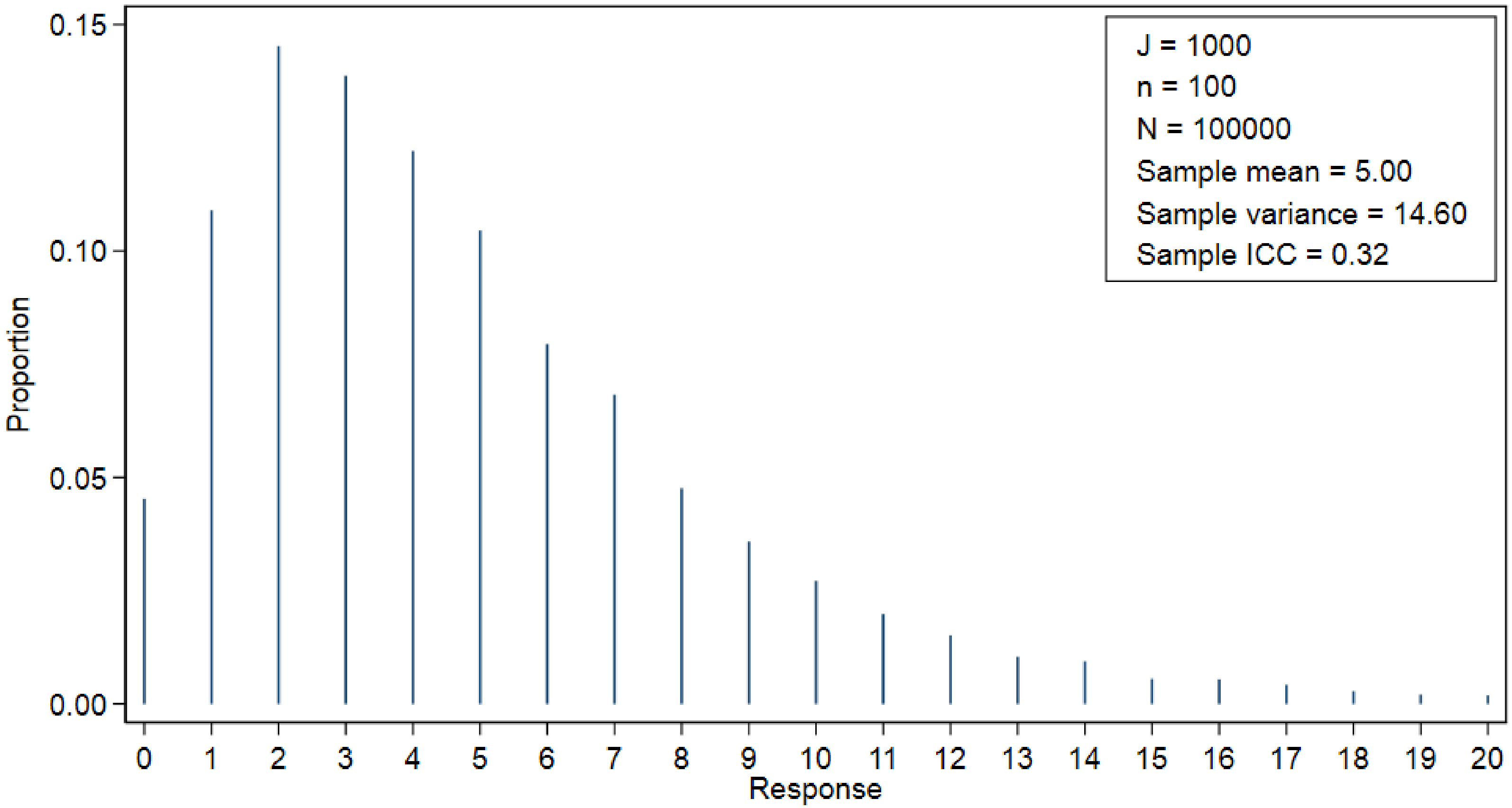
- We show that

$$\text{VPC} \equiv \text{ICC} = \frac{\overbrace{\left(\mu_{ij}^{M}\right)^2\{\exp(\sigma_u^2) - 1\}}^{\text{"level}-2\text{ variance"}}}{\underbrace{\left(\mu_{ij}^{M}\right)^2\{\exp(\sigma_u^2) - 1\}}_{\text{"level}-2\text{ variance"}} + \underbrace{\mu_{ij}^{M} + \left(\mu_{ij}^{M}\right)^2 \exp(\sigma_u^2)\,\alpha}_{\text{"level}-1\text{ variance"}}}$$

where $\mu_{ij}^{M} = \exp\left(\mathbf{x}_{ij}'\boldsymbol{\beta} + \sigma_u^2/2\right)$ is the marginal mean of y_{ij}

- The level-1 variance now has a second term suggesting that the previous Poisson VPC/ICC will be biased upwards in the presence of any overdispersion as it constrains $\alpha = 0$

Illustration
Simulated count response



- Data simulated according to a negative binomial model with no covariates, but with overdispersion

Illustration
Model results

	Single-level (SL)		Multilevel (ML)	
	Poisson	NB	Poisson	NB
β_0	1.61	1.61	1.52	1.52
σ_u^2	0	0	0.18	0.18
α	0	0.35		0.16
Marginal mean	5.00	5.00	5.01	5.01
Marginal variance	5.00	13.69	9.92	14.76
“Level-2 variance”	0	0	4.91	4.87
“Level-1 variance”	5.00	13.69	5.01	9.89
VPC ≡ ICC	0	0	0.50	0.33
AIC	587687	516138	504468	481996

- We fit “empty” single-level and multilevel (two-level random intercept) Poisson and negative binomial models to these data

Illustration (cont'd)
Model results

	Single-level (SL)		Multilevel (ML)	
	Poisson	NB	Poisson	NB
β_0	1.61	1.61	1.52	1.52
σ_u^2	0	0	0.18	0.18
α	0	0.35	0	0.16
Marginal mean	5.00	5.00	5.01	5.01
Marginal variance	5.00	13.69	9.92	14.76
"Level-2 variance"	0	0	4.91	4.87
"Level-1 variance"	5.00	13.69	5.01	9.89
VPC ≡ ICC	0	0	0.50	0.33
AIC	587687	516138	504468	481996

- Models are better fitting as we move from left to right
- As expected, multilevel NB model is the best fitting

Illustration (cont'd)
Model results

	Single-level (SL)		Multilevel (ML)	
	Poisson	NB	Poisson	NB
β_0	1.61	1.61	1.52	1.52
σ_u^2	0	0	0.18	0.18
α	0	0.35	0	0.16
Marginal mean	5.00	5.00	5.01	5.01
Marginal variance	5.00	13.69	9.92	14.76
"Level-2 variance"	0	0	4.91	4.87
"Level-1 variance"	5.00	13.69	5.01	9.89
VPC ≡ ICC	0	0	0.50	0.33
AIC	587687	516138	504468	481996

- All four models recover the sample mean of 5.00

Illustration (cont'd)
Model results

	Single-level (SL)		Multilevel (ML)	
	Poisson	NB	Poisson	NB
β_0	1.61	1.61	1.52	1.52
σ_u^2	0	0	0.18	0.18
α	0	0.35	0	0.16
Marginal mean	5.00	5.00	5.01	5.01
Marginal variance	5.00	13.69	9.92	14.76
"Level-2 variance"	0	0	4.91	4.87
"Level-1 variance"	5.00	13.69	5.01	9.89
VPC ≡ ICC	0	0	0.50	0.33
AIC	587687	516138	504468	481996

- Models differ wildly in their ability to recover the sample variance of 14.60

Illustration (cont'd)
Model results

	Single-level (SL)		Multilevel (ML)	
	Poisson	NB	Poisson	NB
β_0	1.61	1.61	1.52	1.52
σ_u^2	0	0	0.18	0.18
α	0	0.35	0	0.16
Marginal mean	5.00	5.00	5.01	5.01
Marginal variance	5.00	13.69	9.92	14.76
"Level-2 variance"	0	0	4.91	4.87
"Level-1 variance"	5.00	13.69	5.01	9.89
VPC ≡ ICC	0	0	0.50	0.33
AIC	587687	516138	504468	481996

- SL Poisson constrains marginal variance to equal the mean

Illustration (cont'd)
Model results

	Single-level (SL)		Multilevel (ML)	
	Poisson	NB	Poisson	NB
β_0	1.61	1.61	1.52	1.52
σ_u^2	0	0	0.18	0.18
α	0	0.35	0	0.16
Marginal mean	5.00	5.00	5.01	5.01
Marginal variance	5.00	13.69	9.92	14.76
"Level-2 variance"	0	0	4.91	4.87
"Level-1 variance"	5.00	13.69	5.01	9.89
VPC ≡ ICC	0	0	0.50	0.33
AIC	587687	516138	504468	481996

- SL NB allows for overdispersion (but ignores cluster variability)

Illustration (cont'd)
Model results

	Single-level (SL)		Multilevel (ML)	
	Poisson	NB	Poisson	NB
β_0	1.61	1.61	1.52	1.52
σ_u^2	0	0	0.18	0.18
α	0	0.35	0	0.16
Marginal mean	5.00	5.00	5.01	5.01
Marginal variance	5.00	13.69	9.92	14.76
"Level-2 variance"	0	0	4.91	4.87
"Level-1 variance"	5.00	13.69	5.01	9.89
VPC ≡ ICC	0	0	0.50	0.33
AIC	587687	516138	504468	481996

- ML Poisson allows for clustering (but ignores overdispersion)
- ML NB model allows for both clustering and overdispersion

Illustration (cont'd)
Model results

	Single-level (SL)		Multilevel (ML)	
	Poisson	NB	Poisson	NB
β_0	1.61	1.61	1.52	1.52
σ_u^2	0	0	0.18	0.18
α	0	0.35	0	0.16
Marginal mean	5.00	5.00	5.01	5.01
Marginal variance	5.00	13.69	9.92	14.76
"Level-2 variance"	0	0	4.91	4.87
"Level-1 variance"	5.00	13.69	5.01	9.89
VPC ≡ ICC	0	0	0.50	0.33
AIC	587687	516138	504468	481996

- ML NB model allows for both clustering and overdispersion

Illustration (cont'd)
Model results

	Single-level (SL)		Multilevel (ML)	
	Poisson	NB	Poisson	NB
β_0	1.61	1.61	1.52	1.52
σ_u^2	0	0	0.18	0.18
α	0	0.35	0	0.16
Marginal mean	5.00	5.00	5.01	5.01
Marginal variance	5.00	13.69	9.92	14.76
"Level-2 variance"	0	0	4.91	4.87
"Level-1 variance"	5.00	13.69	5.01	9.89
VPC ≡ ICC	0	0	0.50	0.33
AIC	587687	516138	504468	481996

- SL models implicitly assume an ICC of 0 as they ignore clustering
- ML Poisson ICC biased upwards, ML NB recovers the correct ICC

Illustration (cont'd)
Model results

	Single-level (SL)		Multilevel (ML)	
	Poisson	NB	Poisson	NB
β_0	1.61	1.61	1.52	1.52
σ_u^2	0	0	0.18	0.18
α	0	0.35	0	0.16
Marginal mean	5.00	5.00	5.01	5.01
Marginal variance	5.00	13.69	9.92	14.76
"Level-2 variance"	0	0	4.91	4.87
"Level-1 variance"	5.00	13.69	5.01	9.89
VPC ≡ ICC	0	0	0.50	0.33
AIC	587687	516138	504468	481996

- ML Poisson ICC biased up as it ignores overdispersion, and therefore underestimates the level-1 variance

Further extension 1
Other count models

- We have also derived the VPC/ICC for other count models that allow for overdispersion

  - Poisson models with a normally distributed unit-level overdispersion random effect

  - Negative Binomial model: constant dispersion (NB1)

- In each case the expression for the "level-2 variance" stays the same, but the expression for the "level-1 variance" changes to reflect the different way that overdispersion is accommodated in each model

Further extension 2
Higher-level models

- Consider the three-level random-intercept Poisson regression

$$y_{ijk}|\mu_{ijk} \sim \text{Poisson}(\mu_{ijk})$$

$$\ln(\mu_{ijk}) = \mathbf{x}'_{ijk}\boldsymbol{\beta} + v_k + u_{jk}$$

$$v_k \sim N(0, \sigma_v^2), \qquad u_{jk} \sim N(0, \sigma_u^2)$$

- The level-3 VPC/ICC can be derived as follows

$$\frac{\overbrace{\left(\mu_{ijk}^M\right)^2\{\exp(\sigma_v^2) - 1\}}^{\text{level-3 variance}}}{\underbrace{\left(\mu_{ijk}^M\right)^2\{\exp(\sigma_v^2) - 1\}}_{\text{level-3 variance}} + \underbrace{\left(\mu_{ijk}^M\right)^2\exp(\sigma_v^2)\{\exp(\sigma_u^2) - 1\}}_{\text{level-2 variance}} + \underbrace{\mu_{ijk}^M}_{\text{level-1 variance}}}$$

Further extension 3
Random-coefficient models

- Consider the two-level random-coefficient Poisson regression

$$y_{ij}|\mu_{ij} \sim \text{Poisson}(\mu_{ij})$$

$$\ln(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_j$$

$$\mathbf{u}_j \sim N(0, \boldsymbol{\Omega}_{\mathbf{u}})$$

- In the previous two-level random-intercept VPC/ICC expressions, simply replace σ_u^2 with $\mathbf{z}'_{ij}\boldsymbol{\Omega}_{\mathbf{u}}\boldsymbol{z_{ij}}$ (i.e., the cluster-level variance function)

- Analogous substitutions can be made when working with three- and higher-level models

Conclusion

- In Austin et al. (2018) we drew attention to an analytic expression for calculating the VPC/ICC in two-level random-intercept Poisson regression

- In our current work, we have derived analogous expressions for three more flexible count response models
  - Poisson models with an overdispersion random effect
  - Negative Binomial model: mean dispersion (NB2)
  - Negative Binomial model: constant dispersion (NB1)

  as well as extending all these expressions to allow for
  - Additional levels
  - Random coefficients

End of talk – Thank you

g.leckie@bristol.ac.uk

bristol.ac.uk/cmm/team/leckie.html

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, 572-589.

- Stryhn, H., J. Sanchez, P. Morley, C. Booker, and I. R. Dohoo. (2006). Interpretation of variance parameters in multilevel Poisson regression models. In Proceedings of the 11th Symposium of the International Society for Veterinary Epidemiology an Economics, 702–704. Cairns, Australia.