Using Random-Effects Zero-Inflated Poisson Model to Analyze Longitudinal Count Data with Extra Zeros

Haiyi Xie[1], Kwan Hur[2], and Gregory McHugo[1]

[1]*Dartmouth Medical School, PRC*
2 Whipple Place, Suite 202
Lebanon, NH 03766, USA
Haiyi.Xie@Dartmouth.Edu; Gregory.J.McHugo@Dartmouth.Edu

[2]*Hines VA Hospital, CSPCC*
5th Ave and Roosevelt
Hines, IL 60141, USA
hur@research.hines.med.va.gov

1. Introduction

In healthcare research, count variables with many zeros are quite common in which case a classical Poisson regression model may not be appropriate in that it underestimates zero counts. A zero-inflated Poisson regression (ZIP) model was developed (Lambert, 1992) to deal with counts with extra zeros, but this model has limitations for longitudinal and/or clustered count data.

Recently, random-effects zero-inflated Poisson regression (RE-ZIP) models that accommodate both correlated and extra-zero count data have been developed (Hur, 1998; Hall, 2000). In this paper, we employed a RE-ZIP model to analyze hospital admissions collected over 3 years from the New Hampshire Dual Diagnosis Study (Drake et al., 1998).

2. Random Effects Zero-Inflated Poisson Regression Model

A ZIP model is derived from a mixture distribution of Poisson and logistic components. A RE-ZIP is an extension of ZIP by taking the effect of clustering into account. The normally distributed RE- ZIP model can be written as:

$$\Pr(y_{ij}) = (1 - \pi_{ij})f(y_{ij}) + I(y_{ij})\pi_{ij} \qquad (1)$$

where
$$f(y_{ij}) = \exp(-\lambda_{ij}) \cdot \lambda_{ij}^{\,y_{ij}} / y_{ij}!, \qquad (2)$$
$$\text{Logit}(\pi_{ij}) = \gamma' \boldsymbol{w}_{ij} + \upsilon_i = \gamma' \boldsymbol{w}_{ij} + \sigma_1\theta_{1i}, \qquad (3)$$
and
$$\text{Log}(\lambda_{ij}) = \beta' \boldsymbol{x}_{ij} + \nu_i = \beta' \boldsymbol{x}_{ij} + \sigma_2\theta_{2i} \qquad (4)$$

y_{ij} is the response variable, and $\boldsymbol{x}_{ij} = (1, \boldsymbol{x}_{ij1},\ldots, \boldsymbol{x}_{ijp})$ and $\boldsymbol{w}_{ij} = (1, \boldsymbol{w}_{ij1},\ldots, \boldsymbol{w}_{ijp})$ are the $(p+1) \times 1$ and $(r+1) \times 1$ vectors of known covariates for the Poisson and logistic parts, respectively, from the *j*th individual in the *i*th cluster. $\beta_{ij} = (\beta_0, \beta_1, \ldots, \beta_p)'$ and $\gamma_{ij} = (\gamma_0, \gamma_1, \ldots, \gamma_p)'$ are Poisson and logistic regression parameter vectors associated with covariates $\boldsymbol{x}_{ij}$ and $\boldsymbol{w}_{ij}$. An indicator function, $I(y_{ij})$, takes a value of 1 if the observed response is zero ($y_{ij} = 0$) and a value of 0 if the observed response is positive ($y_{ij} > 0$). Here, π_{ij} is considered to be a mixing parameter for the mixture of a binary and a Poisson process. υ_i and ν_i are random effects and are assumed to be normally distributed, $\upsilon_i \sim (0, \sigma_1^2)$ and $\nu_i \sim (0, \sigma_2^2)$. The maximum marginal likelihood estimation method is used to estimate parameters of the RE-ZIP model.

3. Application

The outcome variable for this analysis is the number of hospital admissions, collected every 6 months over 3 years of the New Hampshire Dual Diagnosis Study. Eligible persons with co-occurring severe mental illness and substance use disorder from New Hampshire were randomly assigned to two treatment groups: assertive community treatment (ACT) and standard case management (SCM). Twohundred-threeparticipants completedthestudy. Treatment group (ACT vs. SCM), diagnosis (schizophrenia vs. bipolar), time, time squared and the baseline outcome are used as covariates. The number of admissions ranged from zero to eight, and the number of zero outcomes ranged from 60 to 72 percent over time.

Table 1 shows the parameter estimates and standard errors for ZIP and RE-ZIP models. The likelihood ratio test indicates that the RE-ZIP model provides a better fit of the data ($\chi 2$ = 108.2 with 2 df).

Table 1. Parameter Estimates (and S.E.) for ZIP and Random-Effects ZIP models

Variables	ZIP				Random-Effects ZIP			
	Poisson		Logistic		Poisson		Logistic	
	Parameter (S.E.)		Parameter (S.E.)		Parameter (S.E.)		Parameter (S.E.)	
Intercept	-0.599 (0.224)	**	-0.420 (0.492)		-1.171 (0.225)	**	-1.4440 (0.797)	
Baseline	0.304 (0.048)	**	-0.214 (0.090)	*	0.249 (0.042)	**	-0.6180 (0.194)	**
Diagno. (Schiz=1)	-0.090 (0.119)		-0.812 (0.214)	**	0.287 (0.098)	**	-0.4290 (0.349)	
Group (SCM=1)	0.142 (0.100)		0.500 (0.196)	*	0.087 (0.077)		0.6980 (0.292)	*
Time	0.437 (0.135)	**	0.638 (0.288)	*	0.284 (0.123)	*	1.0260 (0.481)	*
Time2	-0.067 (0.020)	**	-0.074 (0.040)		-0.062 (0.018)	**	-0.1790 (0.079)	*
σ					0.823 (0.048)	**	0.0570 (0.176)	
log likelihood	-1192.9				-1138.8			

** p < .01, * p < .05

Note that the logistic part models whether an individual is admitted or not admitted to the hospital, and the Poisson part models how often an individual is admitted (number of admissions), conditional on the usage. The results show that there is a quadratic time effect; the number of hospital admissions toward the end of the 3-year study period declined. Clients in the ACT group are more likely to be admitted, but they are less frequently admitted.

REFERENCES

Drake,R.,McHugo,G.,Clark,R.,Teague,G.,XieH.,Miles,K.,&Ackerson,T. (1998). Assertive community treatment for patients with co-occurring severe mental illness and substance abuse disorder: a clinical trial. <u>American Journal of Orthopsychiatry</u>, 68(2):201-215.

Hall, D.B. (2000). Zero-inflated Poisson and binomial regression with random effects: a case study, <u>Biometrics</u>, 56, 1030-1039.

Hur, K. (1998). A random-effects zero-inflated Poisson regression model for clustered extra-zero counts, <u>Ph.D. dissertation</u>.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to detects in manufacturing, <u>Technometrics</u>, 34, 1-14.

RÉSUMÉ

Un modèle de régression de Poisson à effets randomisés et à extension d'ordre zéro qui satisfait aux compte de données corrélées ainsi qu'au compte de données « extra-zéro » sera discuté et démontré.