Communications for Statistical Applications and Methods
2022, Vol. 29, No. 4, 413–420

https://doi.org/10.29220/CSAM.2022.29.4.413
Print ISSN 2287-7843 / Online ISSN 2383-4757

Modeling clustered count data with discrete weibull regression model

Hanna Yoo[1,a]

[a]Department of Big Data, Busan University of Foreign Studies, Korea

Abstract

In this study we adapt discrete weibull regression model for clustered count data. Discrete weibull regression model has an attractive feature that it can handle both under and over dispersion data. We analyzed the eighth Korean National Health and Nutrition Examination Survey (KNHANES VIII) from 2019 to assess the factors influencing the 1 month outpatient stay in 17 different regions. We compared the results using clustered discrete Weibull regression model with those of Poisson, negative binomial, generalized Poisson and Conway-maxwell Poisson regression models, which are widely used in count data analyses. The results show that the clustered discrete Weibull regression model using random intercept model gives the best fit. Simulation study is also held to investigate the performance of the clustered discrete weibull model under various dispersion setting and zero inflated probabilities. In this paper it is shown that using a random effect with discrete Weibull regression can flexibly model count data with various dispersion without the risk of making wrong assumptions about the data dispersion.

Keywords: clustered count data, discrete weibull regression, dispersion, random effect

1. Introduction

Clustered count data arise in various areas such as public health, biomedical and behavioral sciences. For example, patients can be clustered within hospitals, multi-center or other communities. Subjects within a same cluster tend to be correlated and the within-cluster dependency must be taken into account. Ignoring the correlation of subjects can lead to incorrect conclusions (Dunlop,1994). Many statistical methods are developed to deal with clustered count data and incorporating random effects into a regression model is widely used. Lawless (1987) described random effects in negative binomial and mixed Poisson regression models and Albert (1992) adopted a Bayesian approach for Poisson random effect model. Also zero-inflated models with random effect models has gained much attention when data include excessive zeros. Hall (2000) proposed a zero-inflated Poisson and binomial regression model with random effects and Yau et al. (2003) developed a zero-inflated negative binomial mixed model. However, all these studies are based on regression models that have strong assumptions of their data. Equal dispersion must be satisfied for Poisson model and negative Binomial model is suitable for over-dispersed data. For under dispersed data generalized Poisson model can be used. Recently Conway-Maxwell-Poisson (COMP) distribution which is an extension of Poisson regression is widely used for count data. It has a merit that it can be used both for under and over dispersed data

This work was supported by the research grant of the Busan University of Foreign Studies in 2021.

[1] Department of Big Data, Busan University of Foreign Studies,korea 65 Geumsaem-ro 485 beon-gil, Geumjeong-gu, Busan 46234 Korea. E-mail: yoohann@hs.ac.kr

Published 31 July 2022 / journal homepage: http://csam.or.kr

however the major difficulty with COMP regression is that the likelihood function contains multiple intractable normalizing constants and is not amenable to standard inference and Markov Chain Monte Carlo (MCMC) techniques (Alan and Nial, 2021).

In this paper we use discrete Weibull (DW) distribution to clustered count data. DW distribution was first introduced by Nakagawa and Osaki (1975). It has an attractive feature that in can handle both over and under dispersed data. Several papers have shown the flexibility of this model (Klakattawi et al. 2018; Kulassekera, 1994). Through DW regression model, one can model count data without the risk of having wrong assumptions about the data dispersion. Also data with many zeros can be modeled through the DW since it has a parameter that can handle zero probabilities. Up to now there are only few papers that deal with clustered count data using DW regression model. Recently Luyts et al (2018) used DW model for longitudinal/clustered data structures and Barbiero (2019) proposed a bivariate count model with discrete Weibull margins where they estimate parameters based on Farlie-Gumbel- Morgenstern copula.

In this paper we extend the discrete DW regression model to clustered count data and compare the performance with other commonly used models for count data. Random intercept is incorporated to account for correlation between subjects within the same cluster. We also conduct a simulation study to investigate the robustness of clustered DW under various dispersion type.

The remainder of the paper is organized as follows. In Section 2, we describe clustered DW regression mode and in Section 3 we apply our statistical method to KNHANES VIII data. The results of the simulation study are summarized in Section 4 and we conclude with a brief discussion in Section 5.

2. Clustered discrete weibull regression model

In this section we introduce the clustered discrete Weibull regression model. Let Y_{ij} denote the j^{th} subject measured for cluster i, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$, with the total number of subjects in the data as $N = \sum_{i=1}^{n} n_i$. We assume that Y_{ij} arises from a clustered DW distribution with parameters q and β. Parameter q is the probability that the outcome random variable Y has value greater than 0 and parameter β controls the range of the values of Y. As β converges to 0, the distribution is highly skewed and as β converges to ∞, the DW distribution approaches to Bernoulli distribution. In the presence of covariates we can denote random variable Y_{ij} following the clustered DW as follows:

$$Y_{ij} \sim \mathrm{DW}\left(q\left(\boldsymbol{X}_{ij}\right), \beta\right)$$

where $\boldsymbol{X}'_{ij} = (1, \boldsymbol{X}_{1,ij}, \ldots, \boldsymbol{X}_{p,ij})$ is a vector of covariates. Covariates can also be linked with parameter β, however in this paper we linked the covariates only through $q(\boldsymbol{X}_{ij})$. The cumulative distribution function with covariates can be denoted as:

$$F_{\mathrm{DW}}\left(Y_{ij}; q\left(\boldsymbol{X}_{ij}\right), \beta\right) = \begin{cases} 1 - q\left(\boldsymbol{X}_{ij}\right)^{(Y_{ij}+1)^{\beta}}, & Y_{ij} = 0, 1, 2, 3, \\ 0, & Y_{ij} < 0 \end{cases} \tag{2.1}$$

and the probability mass function is given by:

$$f_{\mathrm{DW}}\left(Y_{ij}; q\left(\boldsymbol{X}_{ij}\right), \beta\right) = q\left(\boldsymbol{X}_{ij}\right)^{Y_{ij}^{\beta}} - q\left(\boldsymbol{X}_{ij}\right)^{(Y_{ij}+1)^{\beta}} \tag{2.2}$$

for $Y_{ij} = 0, 1, 2, 3, \ldots,\ 0 < q(\boldsymbol{X}_{ij}) < 1,\ \beta > 0.$

For clustered data, subjects in the same cluster tend to have correlation and this association can be incorporated through random effects.Herein we consider a random-intercept model for the clustered DW regression model and it can be denoted as follows:

$$\log\left(-\log\left(q\left(X_{ij}\right)\right)\right) = \boldsymbol{\theta}'\boldsymbol{X}_{ij} + U_{0,i} = \theta_0 + \theta_1 X_{1,ij} + \theta_2 X_{2,ij} + \cdots + \theta_p X_{p,ij} + U_{0,i}, \tag{2.3}$$

where $\boldsymbol{\theta}' = (\theta_0, \theta_1, \theta_2, \ldots, \theta_p)$ the corresponding $(p+1)$ regression coefficients associated with $\boldsymbol{X}'_{ij}$= $(1, \boldsymbol{X}_{1,ij}, \ldots, \boldsymbol{X}_{p,ij})$. $U_{0,i}$ is a random-effect following independently and identically distributed normal distribution with mean 0 and a common dispersion parameter σ_0^2, $U_{0,i} \sim N(0, \sigma_0^2)$.

To estimate the parameters we used h-likelihood which was first introduced by Lee and Nelder (1996). The h-likelihood is an extension of penalized quasi-likelihood and is based on the joint distribution of the random variable Y and the random effect.

Given a random sample $(x_{ij}, y_{ij}), i = 1, \ldots, n, j = 1, \ldots, n_i$, the joint log-likelihood for the clustered DW model is given as:

$$\begin{aligned} l\left(\boldsymbol{\theta}, \beta; \boldsymbol{y}, \boldsymbol{x}, u_0\right) &= l\left(\boldsymbol{\theta}, \beta; \boldsymbol{y}, \boldsymbol{x} | \boldsymbol{u}_0\right) + l\left(\sigma_0^2; u_0\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\log\left(q(x_{ij})^{y_{ij}^{\beta}} - q(x_{ij})^{y_{ij}^{\beta}}\right)\right\} + \sum_{i=1}^{n}\left\{-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\sigma_0^2\right) - \frac{1}{2\sigma_0^2}u_{0,i}^2\right\} \end{aligned} \tag{2.4}$$

and the score function for N observations is given by:

$$S_N\left(\boldsymbol{\theta}, \beta; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{u}_0\right) = \begin{cases} \dfrac{\partial l\left(\boldsymbol{\theta}, \beta; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{u}_0\right)}{\partial \theta_k}, & k = 0, \ldots, p \\ \dfrac{\partial l\left(\boldsymbol{\theta}, \beta; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{u}_0\right)}{\partial \beta}, & \\ \dfrac{\partial l\left(\boldsymbol{\theta}, \beta; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{u}_0\right)}{\partial u_{0,i}}, & i = 1, \ldots, n \end{cases} \tag{2.5}$$

where

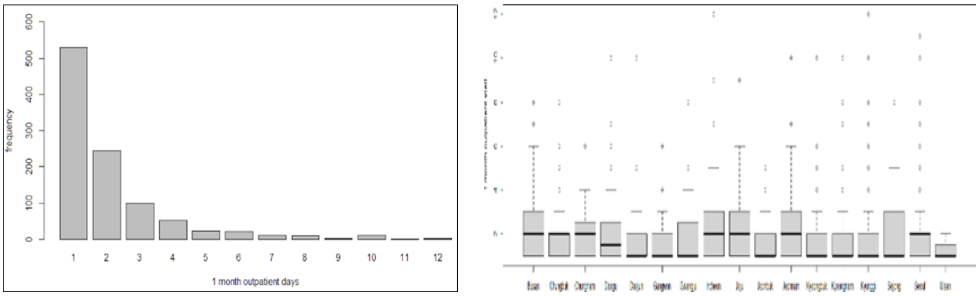
$$\frac{\partial l\left(\boldsymbol{\theta}, \beta; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{u}_0\right)}{\partial \theta_k} = -\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{x_{k,ij}e^{x'_{ij}\boldsymbol{\theta}+u_{0,i}-e^{x'_{ij}\boldsymbol{\theta}+u_{0,i}}}}{w_{ij}\left(\theta_k, \beta\right)}\left[y_{ij}^{\beta}\left(e^{-e^{x'_{ij}\boldsymbol{\theta}+u_{0,i}}}\right)^{y_{ij}^{\beta}-1} - (y_{ij}+1)^{\beta}\left(e^{-e^{x'_{ij}\boldsymbol{\theta}+u_{0,i}}}\right)^{(y_{ij}+1)^{\beta}-1}\right],$$

$$\frac{\partial l\left(\boldsymbol{\theta}, \beta; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{u}_0\right)}{\partial \beta} = -\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{e^{x'_{ij}\boldsymbol{\theta}+u_{0,i}}}{w_{ij}(\boldsymbol{\theta}, \beta)}\left[\left(e^{-e^{x'_{ij}\boldsymbol{\theta}+u_{0,i}}}\right)^{y_{ij}^{\beta}} y_{ij}^{\beta}\log\left(y_{ij}\right) - \left(e^{-e^{x'_{ij}\boldsymbol{\theta}+u_{0,i}}}\right)^{(y_{ij}+1)^{\beta}}(y_{ij}+1)^{\beta}\log(y_{ij}+1)\right],$$

$$\frac{\partial l\left(\boldsymbol{\theta}, \beta; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{u}_0\right)}{\partial u_{0,i}} = \frac{1}{2\sigma_0^2}u_{0,i}^2 - \sum_{j=1}^{n_i}\frac{e^{x'_{ij}\boldsymbol{\theta}+u_{0,i}-e^{x'_{ij}\boldsymbol{\theta}+u_{0,i}}}}{w_{ij}(\theta_k, \beta)}\left[y_{ij}^{\beta}\left(e^{-e^{x'_{ij}\boldsymbol{\theta}+u_{0,i}}}\right)^{y_{ij}^{\beta}-1} - (y_{ij}+1)^{\beta}\left(e^{-e^{x'_{ij}\boldsymbol{\theta}+u_{0,i}}}\right)^{(y_{ij}+1)^{\beta}-1}\right],$$

with $w_{ij}(\boldsymbol{\theta}, \beta)$=$(e^{-e^{x'_{ij}\boldsymbol{\theta}+u_{0,i}}})^{y_{ij}^{\beta}} - (e^{-e^{x'_{ij}\boldsymbol{\theta}+u_{0,i}}})^{(y_{ij}+1)^{\beta}}$. The parameters $\boldsymbol{\theta}, \beta$ and $u_{0,i}$ can be obtained by solving $S_N(\boldsymbol{\theta}, \beta; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{u}_0) = 0$ and the parameter σ_0^2 can be estimated using the procedure of second stage (Lee and Nelder, 1996). The variance-covariance matrix $V(\boldsymbol{\theta}, \beta, u_0)$ of the maximum h-likelihood estimators can be obtained from the expected inverse of the Fisher information matrix as below:

$$V\left(\boldsymbol{\theta}, \beta, u_0\right) = E\left(-\frac{\partial^2 l}{\partial\left(\boldsymbol{\theta}, \beta, u_0\right)}\right). \tag{2.6}$$



(a) (b)

Figure 1: *(a) Barplot of 1month Outpatient day and (b) Boxplot of 1month out patient days in each 17 regions.*

Table 1: Covariate's basic description for the KNHANES VIII data

Variables	Categories	Mean (sd) *n*(%)
Age in years		65.19 (11.1)
Education	Elementary school	410 (40.5)
	Middle school	162 (16.0)
	High school	262 (25.9)
	Graduate school	178 (17.6)
Home income	Low	344 (34.0)
	Middle low	296 (29.2)
	Upper low	191 (18.9)
	High	181 (17.9)
Occupation	Specialist, Service	220 (21.7)
	Agriculture, Simple labor worker.	263 (26.0)
	Unemployed	529 (52.3)
Subjective health status	Good	667 (65.9)
	Bad	345 (34.1)
Sex	Male	371 (36.7)
	Female	641 (63.3)

3. KNHANES VIII data analysis

To demonstrate the approach with real data, we used KNHANES VIII data from with patients who has chronic disease. Chronic diseases have a large socioeconomic loss due to their long duration of illness. In the case of medical users with more than 150 days of outpatient visits per year, more than half of those are patients with complex chronic diseases (Lee, 2004). Especially The average outpatient visits to doctors by Korean patients was 17 visits a year, which is the highest yearly rate among OECD countries in 2018 (Michas, 2021). Accurately predicting the trend of outpatient visits can help policy makers manage hospitals effectively, reasonably organize schedules for human resources.

In this section we analyzed the eighth KNHANES VIII data from 2019 to assess the factors influencing the 1 month outpatient visit of patients with chronic disease in 17 different regions.

The data was collected from 17 different regions and the distribution of 1 month outpatient visit for each region is shown as a boxplot in Figure 1 (a) and (b).

We can see from Figure 1 that the distribution of patients OV is highly skewed and also vary across different regions. Patients in the same region compose a cluster and thus we incorporated a random intercept in the DW regression model to predict the 1month outpatient visits of patients with chronic disease. We considered covariates that affect the 1 month outpatient visit as age in years, education,

Table 2: Parameter estimates (standard error) without random effect for (a) Discrete Weibull model, (b) Poisson model (c) Negative binomial model (d) Generalized Poisson model (e) Conway-Maxwell Poisson model

Variable		DW	Poisson	NB	GP	COMP
Age		−0.00005 (0.0037)	0.0009 (0.0025)	0.0001 (0.0028)	0.0013 (0.0028)	0.0009 (0.0025)
Education	Elementary school	0.2118 (0.1133)	−0.2009* (0.0790)	−0.1950* (0.0876)	−0.1998 (0.0872)	−0.1900* (0.0773)
	Middle school	0.1128 (0.1222)	−0.082 (0.0821)	−0.0808 (0.0913)	−0.0782 (0.0900)	−0.0780 (0.0799)
	High school	0.1574 (0.1043)	−0.1308 (0.0720)	−0.1304 (0.0797)	−0.1311 (0.0788)	−0.1239 (0.0703)
Home income	Low	−0.4480* (0.1096)	0.2876* (0.0783)	0.2841* (0.0861)	0.2563 (0.0863)	0.2720* (0.0771)
	Middle low	−0.1646 (0.1033)	0.1037 (0.0744)	0.1039 (0.0816)	0.0928 (0.0811)	0.0983 (0.0725)
	Upper low	−0.0812 (0.1075)	0.0595 (0.0793)	0.0592 (0.0867)	0.0580 (0.0864)	0.0564 (0.0773)
Occupation	Specialist, Service	0.2362* (0.0950)	−0.1504* (0.0672)	−0.1506* (0.0740)	−0.1352 (0.0736)	−0.1423* (0.0657)
	Agriculture, Simple labor	0.0360 (0.0807)	−0.0152 (0.0548)	−0.0170 (0.0609)	−0.0157 (0.0607)	−0.0144 (0.0533)
Subjective Health Status	Good	−0.2931* (0.0691)	0.2155* (0.0459)	0.2147* (0.0511)	0.2095 (0.0509)	0.2038* (0.0455)
Sex	Male	0.1242 (0.0699)	−0.0782 (0.0493)	−0.07785 (0.0544)	−0.0715 (0.0544)	−0.0739 (0.0480)
Other		β = 1.7087* (0.0377)		α = 0.1086* (0.0210)	ν = 0.1055* (0.0193)	η = 0.9208* (0.0593)
AIC		3507.7	3543.5	3505.7	3508.8	3543.7
BIC		3571.6	3602.6	3569.6	3572.8	3607.7

*are values with p-value < 0.05.

home income, occupation, subjective health status and sex. The basic description of the covariates is shown in Table1.

We compared the results of clustered DW with other commonly used models (Poisson, negative binomial, generalized Poisson and Conway-maxwell Poisson model) for count data. In order to see the effect of a random intercept in the model we also held a model without a random intercept. Table 2 shows the result of ignoring the clustering. According to the AIC and BIC, the negative binomial regression model provides the best fit and DW has values similar to negative binomial. However Poisson and COM-P model produced large values of AIC and BIC.

Table 3 shows the result of considering a random intercept in the model. The random effect in all model was significant. Comparing the results of Table 2 and Table 3, for all models considering a random effect had better fit with smaller AIC/BIC. Among the five models, the clustered DW had the best fit compared to other models and Poisson regression model had the worst fit. Dispersion parameter of negative binomial regression model showed significance implying the data is overdispersed compared Poisson regression model. The dispersion parameter in clustered DW also showed significance. Variables that showed significance in predicting the 1 month outpatient visit are similar. Poisson and negative binomial regression model had same variables (education, home income, occupation and subject health status) that affect the 1 month outpatient visit, however in the clustered DW model education did not effect the 1 month outpatient visit. Based on the clustered discrete Weibull model, the results indicate that respondents of low household income, who were unemployed and those with good subjective health status tend to frequently visit the hospital more than other groups.

Table 3: Parameter estimates (standard error) with random effect for (a) Discrete Weibull model, (b) Poisson model (c) Negative binomial model (d) Generalized Poisson model (e) Conway-Maxwell Poisson model

Variable		DW	Poisson	NB	GP	COMP
Age		−0.0013	0.0008	0.0009	0.0012	0.0008
		(0.0038)	(0.0025)	(0.0028)	(0.0028)	(0.0025)
Education	Elementary school	0.1930	−0.1964*	−0.1914	−0.1984*	−0.1892*
		(0.1157)	(0.0802)	(0.0883)*	(0.0876)	(0.0791)
	Middle school	0.0678	−0.0759	−0.0734	−0.0754	−0.0731
		(0.1255)	(0.0831)	(0.0920)	(0.0904)	(0.0816)
	High school	0.1293	−0.1264	−0.1261	−0.1299	−0.1218
		(0.1058)	(0.0724)	(0.0796)	(0.0788)	(0.0712)
Home income	Low	−0.4453	0.2850*	0.2806*	0.2557*	0.2741*
		(0.1119)*	(0.0789)	(0.0862)	(0.0864)	(0.0783)
	Middle low	−0.1734	0.1011	0.1008	0.0898	0.0972
		(0.1046)	(0.0748)	(0.0817)	(0.0812)	(0.0735)
	Upper low	−0.0816	0.0557	0.0560	0.0555	0.0537
		(0.1090)	(0.0796)	(0.0866)	(0.0863)	(0.0781)
Occupation	Specialist, Service	0.2315*	−0.1538*	−0.1530	−0.1391	−0.1479*
		(0.0955)	(0.0677)	(0.0740)	(0.0737)	(0.0667)
	Agriculture, Simple labor	0.0561	−0.0227	−0.0225	−0.0181	−0.0214
		(0.0827)	(0.0556)	(0.0613)	(0.0608)	(0.0545)
Subjective Health Status	Good	−0.2903*	0.2107*	0.2111*	0.2084*	0.2028*
		(0.0696)	(0.0461)	(0.0511)	(0.0508)	(0.0460)
Sex	Male	0.1151	−0.0822	−0.0798	−0.0738	−0.0788
		(0.0707)	(0.0496)	(0.0544)	(0.0545)	(0.0488)
Other		β = 1.7403*		α = 0.1039*	ν = 0.1014*	η = 0.9432*
		(0.0395)		(0.0209)	(0.0197)	(0.0608)
σ		0.1991*	0.1032*	0.0834*	0.0618	0.0964*
		(0.0654)	(0.0391)	(0.0421)	(0.0409)	(0.0387)
AIC		3500.0	3539.4	3505.1	3509.5	3540.5
BIC		3511.6	3550.2	3516.7	3512.2	3552.1

*are values with p-value < 0.05.

Table 3 shows the result of considering a random intercept in the model. The random effect in all model was significant. Comparing the results of Table 2 and Table 3, for all models considering a random effect had better fit with smaller AIC/BIC. Among the five models, the clustered DW had the best fit compared to other models and Poisson regression model had the worst fit. Dispersion parameter of negative binomial regression model showed significance implying the data is overdispersed compared Poisson regression model. The dispersion parameter in clustered DW also showed significance. Variables that showed significance in predicting the 1 month outpatient visit are similar. Poisson and negative binomial regression model had same variables (education, home income, occupation and subject health status) that affect the 1 month outpatient visit, however in the clustered DW model education did not effect the 1 month outpatient visit. Based on the clustered discrete Weibull model, the results indicate that respondents of low household income, who were unemployed and those with good subjective health status tend to frequently visit the hospital more than other groups.

4. Simulation studies

Simulation study is conducted to investigate the performance of clustered DW regression model with random effect in various distribution settings. To check if the clustered DW predicts well in various dispersion type we simulated data based on four different distributions Poisson, negative binomial, generalized Poisson and Conway-Mawxell Poisson.

Table 4: Comparison of model fit based on AIC/BIC of four distributions and clustered DW

Fitted model		Original model			
		Poisson	NB	GP	COMP
DW	AIC	298.9	412.6	344.7	319.3
	BIC	297.3	411.0	343.1	317.8
Poisson	AIC	297.8	512.6	439.0	324.0
	BIC	296.6	511.5	437.9	322.8
NB	AIC	330.9	414.6	363.6	325.6
	BIC	329.8	413.5	362.4	324.4
GP	AIC	298.9	417.2	342.2	320.1
	BIC	297.3	415.7	340.6	318.5
COMP	AIC	299.0	414.0	349.1	319.4
	BIC	297.5	412.5	347.5	317.9

We considered a single covariate $X \sim \text{Uniform}(0,1)$ with 5 clusters and 20 subjects within each cluster. For all the four distributions we set the mean parameter $\mu = \exp(0.2 + 0.4X)$ for Poisson and extra parameters $\theta = 2$ for negative binomial, $\lambda = 0.50$ for generalized Poisson and $\eta = 0.7$ for Conway-Mawxell Poisson. To introduce clustering of subjects, we generated a latent random effect from a normal distribution with mean0 and standard deviance $\sigma = 0.5$. We compared the AIC and BIC of clustered DW and other four original models. Table 4 shows the AIC and BIC of each model. We can see that the AIC and BIC value calculated from the clustered DW is similar (or even smaller) to the original model and has smaller value compared to other fitted models. Thus it shows that clustered DW performs well on various dispersion type of data.

5. Discussion

In this paper we adapt clustered DW regression model to account for count data which are correlated. Random intercept is incorporated to accommodate the clustered structure. We estimate the regression coefficient through h-likelihood. We compared the performance of clustered DW regression model with other commonly used model for clustered count data. Through our case study using KNHANES VIII data it is shown that incorporating random intercept with DW regression model has the best fit compared to other models (Poisson, negative binomial, generalized Poisson and Conway-Maxwell Poisson).

Also in order to examine the flexibility of clustered DW model we simulated data under various setting. Based on the AIC/BIC the clustered DW model has small values indicating that clustered DW regression model can fit various dispersion type of data. In this paper it is shown that clustered DW regression model can be a suitable choice for when the data is clustered. It can handle both over and under dispersion data without any other additional assumptions. This flexible modeling aspect is the most attractive feature of DW regression.

In this paper we only considered the random intercept effect however random slope effect can be also incorporated in the clustered DW regression model. Even though DW regression model can handle data with many zeros. Modeling the process of zeros and non-zeros in the model inside the clustered DW regression model can be a better choice. Thus clustered zero inflated DW regression model will be our future study.

Acknowledgments

This work was supported by the research grant of the Busan University of Foreign Studies in 2021.

References

Alan B and Nial F (2021). Bayesian inference, model selection and likelihood estimation using fast rejection sampling: The Conway-Maxwell-Poisson distribution, *Bayesian Analysis*, **16**, 905–931.

Albert J (1992). A Bayesian analysis of a Poisson random effects model for home run hitters, *The American Statistician*, **46**, 246–253.

Barbiero A (2019). A bivariate count model with discrete Weibull margins, *Mathematics and Computers in Simulation*, **156**, 91–109.

Dunlop D (1994). Regression for longitudinal data: A bridge from least squares regression, *The American Statistician*, **48**, 299–303.

Hall DB (2000). Zero-inflated Poisson and binomial regression with random effects: A case study, *Biometrics*, **56**, 1030–1039.

Klakattawi HS, Vinciotti V, and Yu K (2018). A simple and adaptive dispersion regression model for count data, *Entropy*, **20**, 142.

Kulasekera K (1994). Approximate MLE's of the parameters of a discrete Weibull distribution with type 1 censored data.Microelectron, *Reliab*, **34**, 1185–1188.

Lawless J (1987). Negative binomial and mixed Poisson regression, *Canadian Journal of Statistics*, **15**, 209–225.

Lee S (2004). Development of a Sustainable Health Management System for Management of Chronic Diseases. Health and Welfare Policy Forum 1, 72–81.

Lee Y and Nelder AJ (1996).Hierarchical generalized linear models, *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**, 619–678.

Nakagawa T and Osaki S (1975). The discrete Weibull distribution.*IEEE Transactions on Reliability* , **24**, 300–301.

Yau KKW, Wang K, and Lee AH (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros, Biometrics, **45**, 437–452.

Luyts M, Geert M, Geert V, Koen M, Eduardo R, Clarice D, and John H. (2019). A Weibull-count approach for handling under- and overdispersed longitudinal/clustered data structures, *Statistical Modelling*, **19**, 569–589.

Michas F (2021). Number of doctor visits per capita in selected countries 2019. Statista, Health Professionals Hospitals. Available from: https://www.statista.com/statistics/236589/number-of-doctor-visits-per-capita-by-country/♯ statisticContainer

Received November 22, 2021; Revised January 7, 2022; Accepted January 13, 2022