Original Article
http://mjiri.iums.ac.ir

Medical Journal of the Islamic Republic of Iran (MJIRI)
Iran University of Medical Sciences

Zero inflated Poisson and negative binomial regression models: application in education

Masoud Salehi[1], Masoud Roudbari*[2]

Received: 2 February 2015 Accepted: 7 June 2015 Published: 17 November 2015

Abstract

Background: The number of failed courses and semesters in students are indicators of their performance. These amounts have zero inflated (ZI) distributions. Using ZI Poisson and negative binomial distributions we can model these count data to find the associated factors and estimate the parameters. This study aims at to investigate the important factors related to the educational performance of students.

Methods: This cross-sectional study performed in 2008-2009 at Iran University of Medical Sciences (IUMS) with a population of almost 6000 students, 670 students selected using stratified random sampling. The educational and demographical data were collected using the University records. The study design was approved at IUMS and the students' data kept confidential. The descriptive statistics and ZI Poisson and negative binomial regressions were used to analyze the data. The data were analyzed using STATA.

Results: In the number of failed semesters, Poisson and negative binomial distributions with ZI, students' total average and quota system had the most roles. For the number of failed courses, total average, and being in undergraduate or master levels had the most effect in both models.

Conclusion: In all models the total average have the most effect on the number of failed courses or semesters. The next important factor is quota system in failed semester and undergraduate and master levels in failed courses. Therefore, average has an important inverse effect on the numbers of failed courses and semester.

Keywords: Zero inflated, Course, Semester, Failure, Student.

Cite this article as: Salehi M, Roudbari M. Zero inflated poisson and negative binomial regression models: application in education. *Med J Islam Repub Iran* 2015 (17 November). Vol. 29:297.

Introduction

When the objective of the study is investigation of count data using some variables, statistical modeling is used. In statistics, Poisson distribution is used for investigating the count data if the mean and variance of the distribution are the same. If these two measures are not the same (variance is larger than the mean= over dispersion), then negative binomial distribution is preferred.

Statistical modeling is from suitable methods to investigate the relationship between various phenomena, especially in medicine and health. Poisson regression is of the models which is used in count variables such as the number of blood donations, the number of stopping addiction, the number of failed courses or semester etc. and is applicable in many medical researches. In this method we build a model on the mean of the response variable using statistical methods. The Poisson regression is a subset of a large set of statistical modeling which is called Generalized Linear Model (GLM). Sometimes the count variables which are used to build a statistical modeling, have an inflation in zero which are divided into

[1]. Assistant Professor, Department of Biostatistics, School of Public Health, Iran University of Medical Sciences, Tehran, Iran. salehi74@yahoo.com
[2]. (**Corresponding author**) Professor, Antimicrobial Resistance Research Center, Rasoul-e-Akram Hospital, Department of Biostatistics, School of Public Health, Iran University of Medical Sciences, Tehran, Iran. mroudbari@yahoo.co.uk

Poisson zero inflated or negative binomial zero inflated distributions according to the nature of the variables.

There are various researches that used statistical modeling on count data which applied negative binomial or Poisson regressions; zero inflated Poisson or negative binomial regressions.

Rafiee (1) used negative binomial distribution for modeling of the period of hospitalization of mothers after child birth as the best model. Wong and Lam (2) applied Poisson regression with zero inflated for modeling of DMF for the students' health situation. Barondess et al (3) used Poisson regression with zero inflated to model the estimated number of cigarettes which is used by new smokers of different races in the USA in 2010. Mohammadfam et al (4) applied a model for the number of work accidents in 2009 and showed the best model is a Poisson regression with zero inflation.

One of the applications of this method is in the modeling of educational count data such as the number of failed courses or the number of failed semesters in university students. This application which is considered as a part of the students' academic performance has a great efficient. The academic performance has relationship with some educational and demographic factors of the students as well.

There are many studies regarding the academic performance of the university students. Mlambo (5) showed that gender, age and entry qualifications have a relationship with the academic performance. Foster (6) proved that gender, age, motivation, prior academic performance are variables which have relationship with the academic performance. Kooi (7) showed that age and academic background have strong relationship with the academic performance. Garkaz et al (8) showed that the type of diploma, students' interest, employment status and gender are factors which have relationship with the academic performance. Trockel et al (9) showed that exercise, eating and sleep habit, time management, religious habit, period of work in a week, gender and age are from the variables which affect the academic performance of the students. Since there are many demographic, educational, and economical factors which affect the academic performance of the students, in each research only some of them are presented. Also, in most recent research, descriptive statistics or ordinary or logistic regressions are used to analyze or model the educational data and there is very few work on academic performance of the university students using Poisson of negative binomial regression with zero inflation.

Unfortunately, there is nothing in the literature about the application of zero inflated Poisson or negative binomial for modeling of the failed courses and semesters in university students.

This study intend to apply statistical modeling, especially Poisson and negative binomial regressions with zero inflated to model the number of failed courses and semesters in the students of Iran University of Medical Sciences. In these models some educational and demographic factors which affect the academic performance of the students were used as the predictor variables.

Methods

This cross-sectional study was performed in 2009 and 2010 in Iran University of Medical Sciences. The target population was all current students of the university in 2009 which were almost 6000 in different educational levels. The sample of 670 students was selected using stratified random sampling. To choose the sample from the target population, simple random sampling in each stratum was used and the sample size was proportional to the strata.

The demographic and educational data was collected from the university education file and the information regarding to attendance of the student in university accommodation was collected from the deputy head of the university in students' affairs. The data consist of the first and surnames of the students, the student ID, dis-

cipline, faculty (one for each faculty and zero for others), the educational level (one for each level and zero for others), gender (one for males and zero for females), the year of entrance, marital status according to year of entrance (single= 1 and married= 0), using university accommodation, being native, quota system (no quota= 1, others= 0), university average, the number of failed courses and semesters.

The students' score are from zero to 20 in all Iranian university systems and if a student receives a score of less than 10 in undergraduate or less than 14 in postgraduate courses, he fails the course. Also, the mean of less than 12 in undergraduate and less than 14 in postgraduate semesters are considered as the failed semester. The students can not pursue their study if they have more than four failed semesters.

The students' demographic and educational data were combined to the file of the number of failed courses and semesters to produce the final data set.

For data analysis and modeling, STATA software 9.1 was used. The descriptive statistics and zero inflated Poisson regression and zero inflated negative binomial regression were used to analyze the final data set. P values less than 0.05 were considered statistically significant. The research was approved in research council of the University.

Results

The number of students with no failed courses were 478 (71.3%) and 192 (28.7%) students had one or more failed courses. The students with one, two, three, four or more failed courses were 92 (13.7%), 37 (5.5%), 22 (3.3%) and 41 (6.2%), respectively. The mean±SD of failed courses was 2.18±0.83) and its maximum was 19. The number of students with no failed semesters was 615 (91.8%). The number of students with one or more failed semester were 24 (3.6%) and 31 (4.6%), respectively.

The results of zero inflated (ZI) Poisson regression fit are shown in Table 1.

In Poisson regression for the number of failed courses (Table1), one unit increase in total university average resulted in 0.64 decreases in logarithm of the number of failed courses. Also, one unit increase in the number of failed semesters caused 2.73 decreases in the logarithm of the odd of inflated zero. Furthermore, the change of educational level from other levels to bachelor or master levels resulted in 0.27 and 0.96 decrease of the expected logarithm of the numbers of failed courses. Finally, the change of gender from male to female caused 0.24 decrease of expected logarithm of the numbers of failed courses, and the change of marital status from married to single increased 0.05 of the expected logarithm of the numbers of failed courses.

In the regression of failed semesters, the university average had negative relationship with the numbers of failed semesters and a unit increase in university average resulted in 0.24 decrease in expected logarithm of the numbers of failed semesters. Also, with the change of quota system from other quota system to free quota system, the expected logarithm of the numbers of failed

Table 1. The results of the ZIP regression models in failed course and semester at the Iran University of medical sciences in 2008

Variable	The Regression with the response variable of numbers of failed courses			The Regression with the response variable of numbers of failed semesters		
	Estimation	SE	p	Estimation	SE	p
University average	-0.64	0.034	<0.001	-0.42	0.088	<0.001
Quota system	-	-	-	-0.47	0.209	0.023
The master level	-0.96	0.375	0.011	-	-	-
The bachelor level	-0.27	0.098	0.005	-	-	-
Gender	-0.24	0.095	0.013	-	-	-
Married	0.05	0.019	0.016	-	-	-
The number of failed semesters (zero inflated)	-2.73	1.180	0.021	-1.56	0.283	<0.001

Table 2. The results of the ZINB regression models in failed course and semester at the Iran University of medical sciences in 2008

Variable	The Regression with the response variable of numbers of failed courses*			The Regression with the response variable of numbers of failed semesters		
	Estimation	SE	p	Estimation	SE	p
University average	-0.83	0.050	<0.001	-0.42	0.088	<0.001
Quota system	-	-	-	-0.47	0.209	0.023
School of management	1.24	0.327	<0.001	-	-	-
School of rehabilitation	1.07	0.324	0.001	-	-	-
Bachelor level	-0.81	0.264	0.002	-	-	-
Certificate level	-0.88	0.285	0.002	-	-	-
Master level	-1.45	0.502	0.004	-	-	-
The number of failed semesters	-35.88	1.180	0.021	-1.56	0.283	<0.001

*- Using robust method

semester decreased by 0.47. Furthermore, a unit increase in the number of failed courses caused the logarithm odd of inflated zero to decrease by 1.56. The results of negative binomial regression fit are presented in Table 2.

As one can see, in regression of the number of failed courses, a unit increase in university average resulted in 0.83 decreases in logarithm of the number of failed courses. Also, with a unit increase in the number of failed semesters, the logarithm of odd of the zero inflated decreased by 35.88. The change of faculties from other to faculty of management and rehabilitation caused 1.24 and 1.07 increase in expected logarithm of the number of failed courses, respectively. The change of educational level from other levels to bachelor, certificate and master levels, the expected logarithm of the number of failed courses decreased by 0.81, 0.8847 and 1.45, respectively.

In regression of failed semesters, the university average has negative relationship with the number of failed semesters and with a unit increase in university average, the expected logarithm of the number of failed semesters decreased by 0.42. Also, with the change of quota system from other system to free quota system, the expected logarithm of the numbers of failed semesters decreased by 0.47. Furthermore, with a unit increase of the numbers of failed courses, the logarithm odd of inflated zero decreased by 1.56. All of these results are the same as Table1 on the regression with the response variables of the number of failed semesters.

Discussion

The research results showed that in Poisson and negative binomial regressions with zero inflated using the number of failed semesters as the response variable, the variables of the University average and quota system have inverse relationship with the response variable, so the increase of the University average and the change from other quota system to free quota system caused a decrease in failed semesters. In a research (10) it was shown that in the students with free quota system, the number of failed semesters is less than other students which is the same as this study. Also, we have to consider the fact that failing is less in clever students and in these students the failure is a rare outcome. Furthermore, the students with free system are almost clever and have very less failure. Therefore, quota system is from the factors which decrease the failed semesters. On the other hand, a unit increase in the number of failed courses in both models, decrease the logarithm odd of inflated zero. Thus, the increase in the number of failed courses resulted in the number of failed semesters. This is in the same direction due to the high correlations (0.81) of the failed courses and semesters. Therefore, with increase of the number of failed courses, the chance of having failed semesters increase.

In both regression models with the response variable of the number of failed courses, the University average and the bachelor and master educational levels have negative relationship. In both models, the increase of University average resulted

a decrease of the number of failed courses, and the decrease in negative binomial regression is more than Poisson regression. In a research it was shown that there is a negative relationship between the university average and the number of failed courses which is the same as this research (10). Also, the change from other educational levels to bachelor and master levels caused a decrease of the number of failed courses in these levels. This decrease in negative binomial regression is more than Poison regression. The decrease in the number of failed courses in bachelor and master levels is logical due to the little failure courses in these levels.

In Poisson regression with the failed courses as the response variable, the variables of gender and marital status have inverse and positive relationship to the number of failed courses, respectively, and the failed courses are more in males than females and in singles than married. In different research (5,6,8,10,11), it was shown that gender and marital status have effect to the failed courses or semesters. This result has the same direction with the presented research. In (12) research it is shown that the singles are more successful in their educations than married and have less failed courses which is opposite of this study (12).

In the negative binomial regression, with the number of failed courses, the variables of bachelor level and the faculties of management and rehabilitation caused the increase in the number of failed courses. Since the number of failed courses in bachelor level is more than master level, so the result makes sense. Also, rehabilitation faculty has the most failure in the university which is the same as the model result. Finally, in seems that the entrance of the management faculty to the model, with many failures, is for adjustment of the effects of other variables in the model. On the other hand, a unit increase in the failed semester, in both models, increases the logarithm odd of inflated zero and this reduction in negative binomial model is 13 times of the Poisson model, and this result was expected due to the high correlation between the failed courses and semesters.

Conclusion

It is concluded from all results that in the failed courses regression model, the University average and quota system have the most roles and the increase in University average resulted in the reduction of failed semesters. Also, the change from other quota system to free quota system resulted in the decrease of failed semesters. Furthermore, in these models, the increase in the number of failed courses cause a decrease in the logarithm odd of zero inflated. Therefore, with the increase of the university average and choosing free quota system it is possible to reduce the failed semesters.

In the regressions of the number of failed courses, the University average has an important role and its increase causes the decrease of the number of failed courses. Also, the bachelor and master levels can reduce the number of failed courses. Other affected variables to the number of failed courses are different in both models. On the other hand, the increases in the number of failed semesters, decrease the logarithm of zero inflated in both models and the amounts of these changes are different. Therefore, the increase in University average and choosing more bachelor and master students can reduce the failed courses.

Acknowledgements

This work was supported by Iran University of Medical Sciences under the grant number of 370.

Conflicts of interest

The authors have no conflicts of interest.

References

1. Rafiee M, Ayatollahi M, Behboodian J. Zero-inflated negative binomial modeling, efficiency for analysis of length of maternity hospitalization. Yafteh 2005;6(4):47-58. (Persian)

2. Wong KY, Lam KF. Modeling zero-inflated count data using a covariate-dependent random effect model. Stat Med 2013;32(8):1283-93.

3. Barondess DA, Meyer EM, Boinapally PM, Fairman B, Anthony JC. Epidemiological evidence on count processes in the formation of tobacco dependent. Nicotine Tob res 2010;12(7):734-41.

4. Mohammadfam I, Moghimbeigi A. Evaluation of injuries among a manufacturing industry staff in Iran. J Res Health Sci 2009;9(1):7-12.

5. Mlambo V. An analysis of some factors affecting student academic performance in an introductory biochemistry course at the University of the West Indies. Caribbean Teaching Scholar 2011;1(2):79-92.

6. Foster C. Factors affecting academic performance of in-service students in science education: A case of the University of Zambia. [The master dissertation], The University of Zambia, Zambia, 2011.

7. Kooi LT, Ping TA. Factors Influencing Students Performance in Wawasan Open University: Does Previous Education Level, Age Group and Course Load Matter. 2007, Available from: weko.wou.edu.my/index.php

8. Karkaz M, Banimaha B, Esmaeili H. Factors Affecting Accounting Students' Performance: The Case of Students at the Islamic Azad University. Procedia - Social and Behavioral Sciences 2011;29: 122-128.

9. Trockel MT, Barnes MD, Egget DL. Health-Related Variables and Academic Performance among First-Year College Students: Implications for Sleep and Other Behaviors. J Am College health 2000;49(3):125-131.

10. Roudbari M, Ahmadi A, Roudbari S, Sedghi S. The effective factors on the academic progress of the students of Tehran University of Medical Science. JPMA 2014;64(1):45-48.

11. Shams B, Farshidfard M, Hassanzadeh A. The comparison of demographic and personality characteristics in fail and succeed students of Isfahan University of medical sciences. Res Med Sci 1997;4(2):222-226.

12. Yousefi Mashoof R, Saeedi Jam M. Study in quality of education status of medical students in basic sciences courses in Hamadan university of medical sciences 1989-94. Sci J Hamadan Uni Med Sci 2001;7(4):25-29.