A COMPARISON OF COUNT DATA MODELS WITH AN APPLICATION TO INTENSIVE CARE UNIT STAY

Miss Bhavatharini S[1] , Mrs. Reka K [2], Department of Biostatistics, Christian Medical College, Vellore

Introduction

Count data refers to observations made about events or items that are enumerated
For example the number of products that a consumer buys on online, number of asthma attacks in an occupational cohort
Nature : Non-negative ,discrete, skewed
Ordinary least square based linear regression models can fail miserably on counts based data as it might generate negative and fractional predictions

Objective

To identify the appropriate count regression model for ICU data
To find the risk factors association on the hospital and ICU stay using the count data regression models

Methods

Poisson Regression

These are generalised linear models with the logarithm as the canonical link function
response variable Y has a poisson distribution and the logarithm of its expected value can be modelled by a linear combination of unknown parameters

$$\ln(\mu)=\beta_0+\beta_1 X_1+.....+\beta_k X_k$$

Negative Binomial regression

Poisson over dispersion occurs in data where the variability of the data is greater than the mean which can be modelled by negative binomial regression as It has additional parameter α to accomdate the extra variability in the data

$$\ln(\mu)=\beta_0+\beta_1 X_1+.....+\beta_k X_k+ \alpha$$

Zero- Truncated models

The Poisson, negative binomial distributions each assume the possibility of zero counts even if there may not in fact be any.
If zero counts are not possible,then underlying Pdf has to be adjusted to exclude zeroes
Zero-truncated (ZT) models are constructed for exactly that purpose.

Zero-truncated Poisson

With respect to the Poisson distribution, the probability of a zero count is exp(−μ). This value needs to be subtracted from 1 and then the remaining probabilities rescaled on this difference. That is, the Poisson PDF is rescaled to exclude zero counts by dividing the PDF by 1 − P(y = 0)

$$f(y,\mu)=\frac{e^{-\mu}\mu^{y}}{(1-\exp(-\mu))y!}$$

Methods

Zero Truncated Negative Binomial
Let y be the i observation of the response variable. Considering that the event yi = 0 is not observed, we can obtain the zero-truncated distribution conditioning the probability functions in the point zero. In the case of the negative binomial, we have

$$f(y_i|x_i)=\frac{\Gamma(yi+\alpha^{-1})}{\Gamma(yi+1)\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}}\left(\frac{\mu}{\alpha^{-1}+\mu}\right)^{y_i}$$

$$1-(1+\alpha\mu)^{-\alpha^{-1}}$$

Material

The data was obtained from the Surgical Intensive Care Unit.
The study has data from 489 patients.
The study research the relationships between various factors and the length of ICU stay and hospital stay.
The risk factors considered for association on both the count outcomes, ICU length of stay and hospital length of stay are Age, BMI, Modified Nutric score, SOFA score, APACHE II, Diabetes, Ventilated, Vasopressor, Surgery Done, Polytrauma, Nutrition Risk, Blood culture, Insulin Required, Type of nutrition.

Results

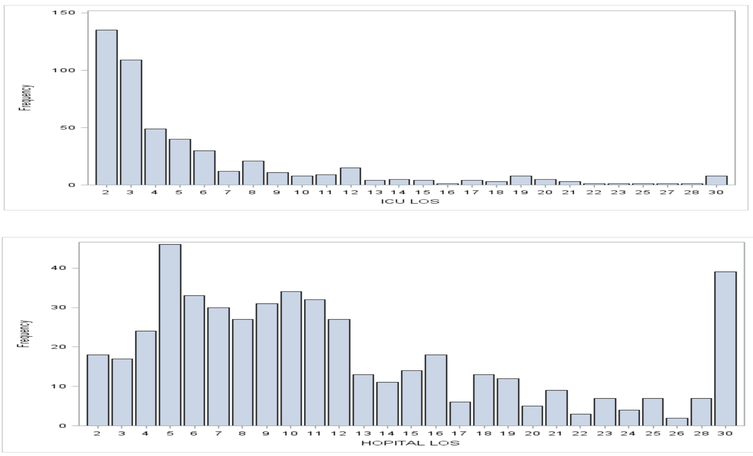

Figure 1 shows the histogram of ICU length of stay and Figure 2 shows the hospital length of stay

Table 1:

Parameter	Estimate	SE	95% CI	t-Value	p-value
Intercept	0.972	0.326	(0.330, 1.613)	2.98	0.0031
Age	-0.004	0.003	(-0.010, 0.001)	-1.41	0.1581
Modified Nutric score	-0.040	0.052	(-0.143, 0.061)	-0.78	0.4357
SOFA baseline	0.034	0.015	(0.004, 0.064)	2.29	0.0225
APACHEII Score	0.032	0.009	(0.013, 0.051)	3.44	0.0006
Ventilated(Yes)	-0.25	0.105	(-0.460,- 0.045)	-2.40	0.0169
Nutrition risk(High)	0.053	0.162	(-0.266, 0.372)	-0.33	0.7434
Type of nutrition(Parenteral)	-0.22	0.087	(- 0.393, -0.0504)	2.54	0.0113
Blood culture positive(Yes)	0.306	0.111	(0.088, 0.525)	2.76	0.0060

Table 2

Parameter	Estimate	SE	95% CI	t value	p-value
Intercept	1.939	0.191	(1.56, 2.315	10.14	<.0001
Age	-0.002	0.001	(-0.006, 0.001)	-1.30	0.1928
BMI	0.015	0.005	(0.004, 0.026)	2.82	0.0051
SOFA baseline	0.011	0.007	(-0.003, 0.027)	1.54	0.1232
DM(Present)	-0.159	0.068	(-0.293, -0.025)	-2.34	0.0198
Blood culture positive(Yes)	0.218	0.079	(0.062, 0.374)	2.76	0.0060

Table 1 is the Multiple Zero Truncated Negative Binomial model for ICU length of stay and table 2 is the Multiple Zero Truncated Negative binomial model for hospital length of stay

Conclusion

Since both our count variables were over-dispersed we wouldn't model Poisson regression in either cases. AIC values were used for comparison of all other models and chi-square by d.f to check for over dispersion. Zero truncated negative binomial fit the best for both the data. But in case of Hospital length of stay, the negative binomial regression model gave similar results to the zero truncated model because the mean for hospital length of stay is 12 so the expected probability of 0 is very less.
It is concluded that zero truncated models need not be necessary if the mean value is high as both the models will give almost same results. It is shown that if the blood culture is positive then the risk of staying in hospital and ICU increase by 24% and 35% respectively. It is odd that the ventilation is protective (RR< 1) for the variable ICU length of stay It can be influenced by the fact that people with ventilation have increased risk of mortality. Using survival analysis might have been useful but it was beyond our scope and we did not have any data on the event of death.

Contact

Bhavatharini S,
M.Sc Biostatistics
[Christian Medical College,Vellore
bhavatharini96@gmail.com

References

. Hilbe JM. Modeling Count Data. Cambridge: Cambridge University Press; 2014 [cited 2020 Aug 22]
Zaninotto P, Falaschetti E. Comparison of methods for modelling a count outcome with excess zeros: Application to Activities of Daily Living (ADL-s) [Internet]. [cited 2020 Aug 22]
. Cameron AC. Regression Analysis of Count Data by A. Colin Cameron