PO-109

Exploring the Use of Negative Binomial Regression Modeling for Pediatric Peripheral Intravenous Catheterization

Jennifer Mann, The Ohio State University; Jason Brinkley, East Carolina University;
Pamela Larson, East Carolina University Brody School of Medicine

ABSTRACT

A large study conducted at two southeastern US hospitals from October 2007 through October 2008 sought to identify predictive variables for successful intravenous catheter (IV) insertion, a crucial procedure that is potentially difficult and time consuming in young children. The data was collected on a sample of 592 children that received a total of 1,195 attempts to start peripheral IV catheters in the inpatient setting. The outcome here is number of attempts to successful IV placement, for which the underlying data appears to have a negative binomial structure. The goal of this paper is to illustrate the appropriateness of a negative binomial assumption using visuals obtained from PROC SGPLOT and to determine the goodness of fit for a negative binomial model.

Negative binomial regression output from PROC GENMOD will be contrasted with traditional ordinary least squares output. Akaike's Information Criterion (AIC) illustrates that the negative binomial model has a better fit and comparisons are made in the inferences of covariate impact. Many scenarios of negative binomial regression follow from an application to overdispersed Poisson data; however, this project demonstrates a data set that fits well under the traditional ideology and purpose of a negative binomial model.

INTRODUCTION

The purpose of this paper is to illustrate SAS® the code that assesses whether the assumptions of the negative binomial distribution are violated for the data set under consideration. It is necessary to explore the independence of each attempt to insert an IV using confidence intervals. The code to calculate the expected values provides output which illustrates whether the observations match the expectations under different assumptions. Finally, the paper compares a traditional ordinary least square regression analysis with the more specific negative binomial regression modeling in PROC GENMOD.

The data set under analysis includes many different variables, some of which were excluded from the analysis. The variables considered for the regression models here are: shift (whether the procedure was performed during the night or day shift); diff1 (whether the medical professional performing the procedure assessed the patient as difficult before the first stick attempt); dehydrated (a patient was either coded as dehydrated or not dehydrate/unknown); coopch1 (whether the medical professional performing the procedure assessed the patient as cooperative before the first stick attempt); Nurse1Exp (the self-reported level of experience for the medical professional performing the first stick attempt); and osbdm (the mean Obeservational Scale of Behavioral Distress score for the patient. The number of total insertion attempts ranged from 1 to 10.

A link the data set has been included in the reference section, for those who wish to explore it further.

ASSESSING INDEPENDENCE

This assumption of independence is critical to confirming that the data can be considered under a negative binomial assumption. Considering the exact 95% confidence intervals for the binomial proportion of successes per attempt was the main method of examining the independence of the attempts. The code to output these values follows, along with a table summarizing the output. The DATA steps were omitted for conciseness.

```
*Output Exact CI;
ods output BinomialCLs;
proc freq order=data data=Needlesticks;
tables succss1 / binomial(exact) alpha=0.05;
tables succss2 / binomial(exact) alpha=0.05;
tables succss3 / binomial(exact) alpha=0.05;
tables succss4 / binomial(exact) alpha=0.05;
tables succss5 / binomial(exact) alpha=0.05;
tables succss6 / binomial(exact) alpha=0.05;
tables succss7 / binomial(exact) alpha=0.05;
tables succss8 / binomial(exact) alpha=0.05;
tables succss9 / binomial(exact) alpha=0.05;
tables succss10 / binomial(exact) alpha=0.05;
```

```
ods select BinomialCLs;
run;
```

Table 1: Proportions and Bounds for 95% Confidence Intervals, by Attempt

Number of Attempt	***Proportion of Successes***	***Lower Bound***	***Upper Bound***	***Sample Size***
1	0.463	0.422	0.504	592
2	0.438	0.3823	0.495	315
3	0.439	0.363	0.516	171
4	0.330	0.235	0.436	91
5	0.527	0.388	0.664	55
6	0.409	0.207	0.637	22
7	0.364	0.109	0.692	11
8	0.333	0.043	0.777	6
9	1	0.158	1	2

Table 1. Output from PROC FREQ organized into lists of the proportion actual proportion of successes, along with lower and upper bounds for a 95% confidence interval, for each attempt to start an IV. Once a patient experienced a successful IV, they were removed from the sample. Thus, the sample size for each attempt is diminished as the number of the attempt increases.

GRAPH OF CONFIDENCE INTERVALS

To better illustrate these confidence intervals, a graph of the proportions was created in SGPLOT. The code and picture of the output is below.

```
proc sgplot data=sample;
scatter x = Stick y=Prop/yerrorlower=Lower yerrorupper=Upper Markerchar=SampleSize;
run;
```

Table 1: Proportions and Bounds for 95% Confidence Intervals, by Attempt

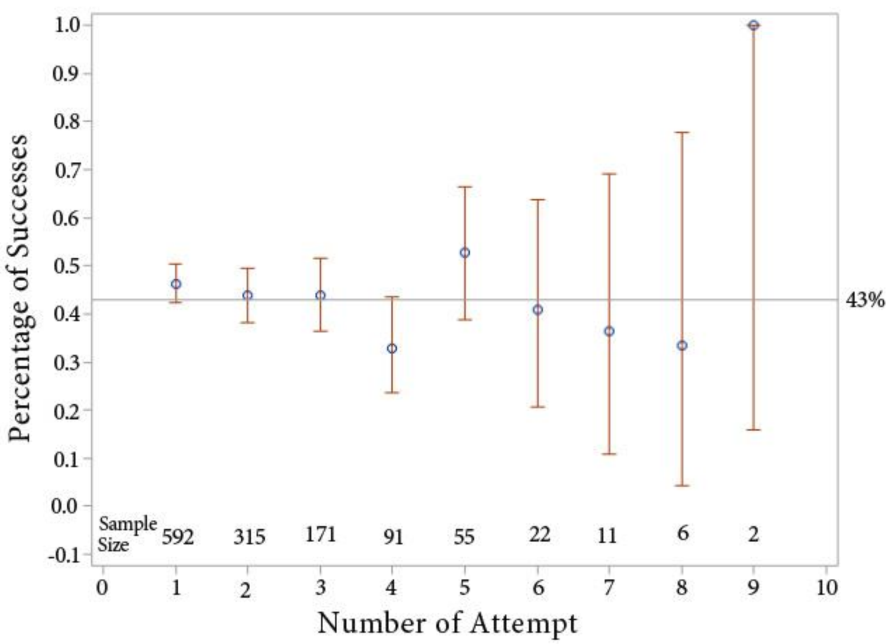

Figure 1. This is a visual representation of the 95% confidence intervals for the proportion of successes per IV attempt. A line has been drawn through the graph at 43%, to illustrate that each interval overlaps, indicating that the assumption of independence required for the negative binomial setting is not violated.

CONFIRMING EXPECTATIONS

The expected values under the negative binomial, zero-inflated negative binomial, and Poisson distributions were calculated and output. The information for the zero-inflated negative binomial distribution was calculated for thoroughness. The Poisson distribution was selected because it is the most common distribution for count data.

Note that at least one attempt to insert an IV is necessary. Since all of the distributions under consideration include the possibility of success on the 0 count, the table lists the values by the number additional attempts before success. This can also be thought of as the number of failed attempts.

The SAS® coding used was to fit intercept only regression models for each distribution, and is detailed below.

```
*Create counts for bar graph, code modified from Morel and Neerchal
(2012) by fitting intercept only regression models;

title "Poisson Model using COUNTREG";
ods output ParameterEstimates=Parms_P;
proc countreg data=Needlesticks;
model y = / dist=Poisson;
output out=exp_p prob=prob_p;
freq freq;
run;

title "Negative-binomial Model using COUNTREG";
ods output ParameterEstimates=Parms_NB;
proc countreg data=Needlesticks;
model y = / dist=negbin(p=1);
output out=exp_nb prob=prob_nb;
freq freq;
run;

title "Zero-inflated Negative-binomial Model using COUNTREG";
ods output ParameterEstimates=Parms_ZINB;
proc countreg data=Needlesticks method=qn;
model y = / dist=zinb;
zeromodel y ~  / link=logistic;
output out=exp_zinb prob=prob_zinb;
freq freq;
run;
```

Table 2: Observed Values and Expected Values under the Poisson, Negative Binomial, and Zero-Inflated Negative Binomial Assumptions

Number of Additional Attempts	***Observed Frequency***	***Expected Frequency Poisson***	***Expected Frequency NB***	***Expected Frequency ZINB***
0	278	188.66	275.97	278
1	142	215.74	147.48	141.55
2	84	123.36	78.69	80.64
3	33	47.02	41.96	43.79
4	31	13.44	22.37	23.18
5	11	3.075	11.92	12.09
6	5	0.59	6.36	6.24
7	5	0.096	3.39	3.20
8	3	0.014	1.80	1.63
9+	0	0.002	2.06	1.67

Table 2. This table compares the actual observed frequency of successes by additional attempts (or by number of failures), along with the expected frequencies under the Poisson, Negative Binomial, and Zero-Inflated Negative Binomial distributions.

GRAPH OF EXPECTED VS. OBSERVED VALUES FOR NEGATIVE BINOMIAL DISTRIBUTION

To better illustrate how closely the observed values match the expected values, a graph was created using PROC SGPLOT. First, the outputs from the intercept only regression had to be combined into a single data set. In the code below it is called “counts”. The DATA step for this manipulation has been omitted for conciseness. The code and figure are below. The Chi-squared value for the data under the assumption of a negative binomial distribution was computed to be 9.80 (p-value =0.2), which verifies that such an assumption is appropriate.

```
*Figure 2*;
proc sgplot data=counts;
yaxis label= "Number of Responses";
vbar stick/ response = Actual;
vbar stick/ response = Predicted barwidth=0.5 transparency= 0.2;
run;
```

Table 12: Frequency of Predicted Values and Observed values, by Additional Stick Attempt

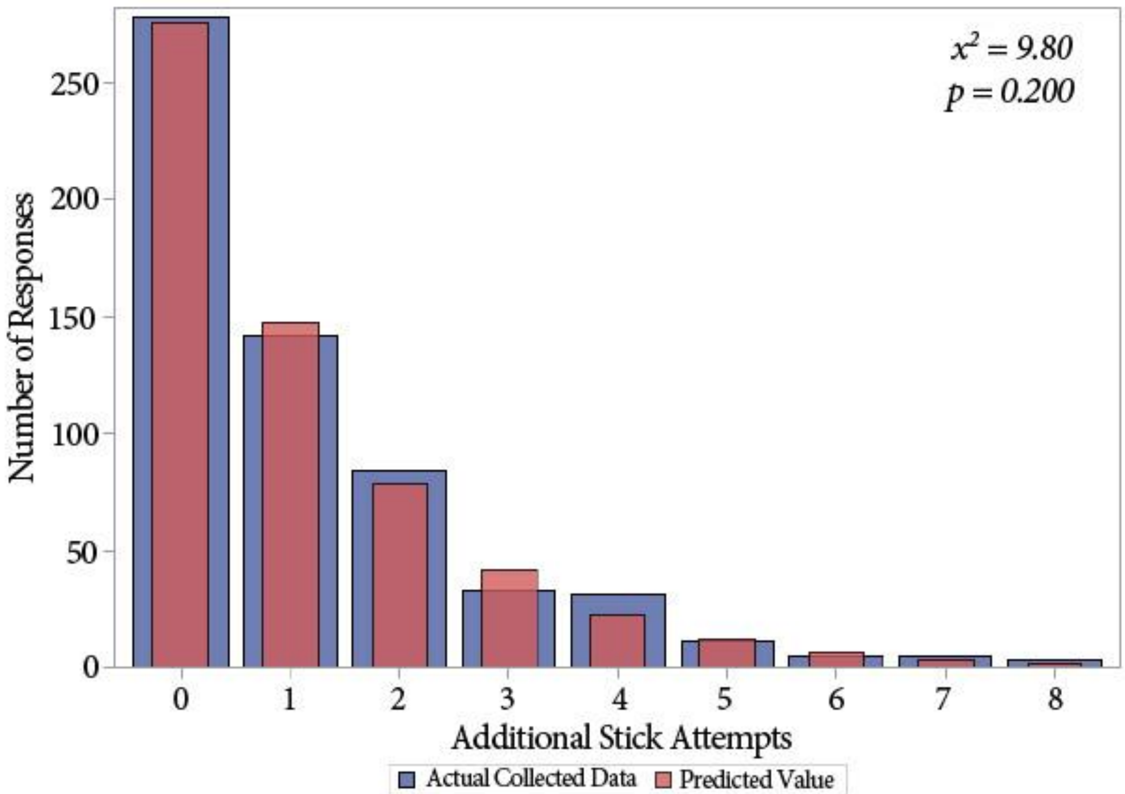

Figure 2. In this illustration, the actual observed frequencies of IV placement success are depicted, along with the expected frequencies under the assumption of negative binomial data. The data have been plotted based on the number of additional IV attempts beyond the first (which can also be thought of as the number of failures).

REGRESSION MODELS

To analyze the data, two different regression models were fit. An ordinary least squares model was created, since such an analysis is the most common choice for many clinicians. Additionally, a negative binomial regression model was also created, since the previous work verified that the data seem to fit well to a negative binomial distribution.

Caution must be used when comparing the two models. Because the NB regression modeling requires the use of a loglink function, direct comparison of the coefficients is not appropriate. Instead, the least squared adjusted means are provided and analyzed.

The code and accompanying tables detailing the output are detailed below.

```
*Reformat Data for Models Used in Paper;
Data Needlesticks;
set Needlesticks;
sticks = totnmbrstks-1;
Nurse1Exp = 0;
if nurexp1 > 1 then Nurse1Exp = 1;
run;

*OLS Regression and NB Regression models;
proc mixed;
class shift diff1 dehydrated coopch1 Nurse1Exp;
model sticks = shift diff1 dehydrated coopch1 Nurse1Exp osbdm/solution;
lsmeans shift diff1 dehydrated coopch1 Nurse1Exp/cl;
run;

proc genmod descending;
class shift diff1 dehydrated coopch1 Nurse1Exp;
model sticks = shift diff1 dehydrated coopch1 Nurse1Exp osbdm/dist=negbin(p=1);
ods output parameterestimates=pe;
lsmeans shift diff1 dehydrated coopch1 Nurse1Exp;
run;
```

Table 3: Results of the Ordinary Least Square regression model and the Negative Binomial regression model.

	OLS Model (AIC = 1919)			Negative Binomial Model (AIC = 1531)					
Effect	**p-value**	**Least Sq. Adj. Mean**	**St. Error**	**p-value**	**Least Sq. Adj. Mean**	**St. Error**	**Par. Estimate**	**St. Error**	**Chi-Square**
Intercept	<.0001			0.0497			0.33	0.17	3.85
Shift (Day)	0.0002	1.72	0.09	0.0001	1.63	0.08	0.44	0.11	15.04
Shift (Night)	Ref	1.28	0.10	Ref	1.05	0.09	Ref	Ref	Ref
Not Difficult	<.0001	1.16	0.09	<.0001	0.98	0.08	-0.58	0.12	24.95
Difficult	Ref	1.84	0.11	Ref	1.75	0.09	Ref	Ref	Ref
Dehydrated (No/Unknown)	<.0001	1.10	0.07	<.0001	0.98	0.07	-0.57	0.12	21.62
Dehydrated (Yes)	Ref	1.90	0.13	Ref	1.74	0.10	Ref	Ref	Ref
Not Cooperative Child	0.0004	1.75	0.10	0.0003	1.63	0.08	0.44	0.12	12.95
Cooperative Child	Ref	1.25	0.11	Ref	1.05	0.09	Ref	Ref	Ref
Nurse Experience (< 1 Year)	0.0013	1.70	0.11	0.0012	1.56	0.09	0.36	0.11	10.43
Nurse Experience (1 Year +)	Ref	1.30	0.09	Ref	1.10	0.08	Ref	Ref	Ref
OSBDM Score	0.0804	*	*	0.2492	*	*	-0.0214	0.02	1.33
Dispersion							0.56	0.11	

Table 3. This table lists the results of the regression models for both the ordinary least squares (OLS) model and the negative binomial (NB) model. Parameter estimates for the OLS model are not provided, as direct comparison of these values to the NB model is not appropriate. Instead, the model based adjusted means for each factor are listed.

CONCLUSION

Based on the confidence intervals for each stick attempt, as well as the observed values being similar to the expected values of negative binomial data, the negative binomial assumption is a good fit for modeling the IV insertion process data. The smaller AIC of the negative binomial regression model indicates that it is a better fit than the OLS regression model. Additionally, the model based adjusted means for the significant factors are generally smaller with smaller standard errors under the negative binomial regression model. Thus, the negative binomial regression model appears to give more precise effects for each significant factor. It is interesting to note that the general clinical inferences implied by both models are the same, that is to say, the variables that had a significant impact on number of IV placement attempts under a negative binomial model were also identified in the OLS model.

Some limitations of the data should be noted. The analysis did not take into consideration the changes in nurses, if any, between stick attempts on the same patient. Furthermore, data on the actual site of each stick attempt was not collected. Thus, the difficulty of the stick site (only the perceived difficulty of the patient by the health provider) was not assessed as a possible factor in either the original or this new analysis.

The current study is important in that it adds an important dataset to the study of negative binomial regression analysis. The complete results and analysis can be found in an article of the same title as this paper, which can be found online and is included in the References section.

REFERENCES

- The data set used in this paper can be located at < http://www.hoajonline.com/medicalstat/2053-7662/2/6>
- J. Mann, P. Larsen, J. Brinkley: **Exploring the Use of Negative Binomial Regression Modeling for Pediatric Peripheral Intravenous Catheterization**. *Journal of Medical Statistics and Informatics* 2014, **2(6)**. < http://www.hoajonline.com/medicalstat/2053-7662/2/6>
- P. Larsen, D. Eldridge, J. Brinkley, D. Newton, D. Goff, T. Hartzog, N.D. Saad, & R. Perkin: **Pediatric Peripheral Intravenous Access: Does Nursing Experience and Competence Really Make a Difference?.** *Journal of Infusion Nursing* 2010, **33(4):**226-235.
- J. G. Morel & N. K. Neerchal: **Overdispersion Models in SAS®**. 2012.

ACKNOWLEDGMENTS

The authors would like to thank the East Carolina Brody School of Medicine for their involvement in data collection.

RECOMMENDED READING

- *Base SAS® Procedures Guide*
- *SAS® For Dummies®*
- J.M. Hilbe: Negative Binomial Regression. 2008.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author(s) at:

Name: Jennifer Mann
Enterprise: The Ohio State University, Department of Statistics
Address: 1958 Neil Avenue, 420 Cockins Hall
City, State ZIP: Columbus, OH 43210
Work Phone: (614) 292-2866
E-mail: mann.426@buckeyemail.osu.edu

Name: Jason Brinkley
Enterprise: East Carolina University, Department of Biostatistics
Address: 2435A Health Sciences Building
City, State ZIP: Greenville, NC 27834
Work Phone: (252) 744-6048
E-mail: brinkleyj@ecu.edu
Web: http://www.ecu.edu/cs-dhs/bios/brinkley.cfm

* For information on the collection of data under consideration only, not data analysis, contact
Name: Pamela Larsen
Enterprise: East Carolina University, Brody School of Medicine, Department of Internal Medicine
Address: Brody Medical Sciences Building 3E-115
City, State ZIP: Greenville, NC 27834
Work Phone: (252) 744-4169
E-mail: larsenp@ecu.edu