

Sensitivity Analysis Of Independent Variables On Traffic Crash Prediction Models By Using Stata

Deo Chimba

Graduate Research Assistant

Department of Civil and Environmental Engineering

Florida State University



CONTENTS

- Purpose of the Study
- What are roadway geometrics and traffic characteristics
- Properties of Crash (Accident)
- Choice of Appropriate Distribution (Model)
- Selection between Poisson and NB
- Modeling
- Model Fitness Parameters
- Finding from the Model
- Sensitivity Analysis
- Conclusion

Purpose of the study

- To evaluate roadway and operational factors influencing crash frequency on six-lane divided roadways in Florida
- Factors considered Include
 - Roadway segment length
 - Number of Vehicles (AADT)
 - Access Density
 - Median Width
 - Shoulder Width
 - Surface Width
 - Percentage of Trucks

Roadway geometrics



Roadway geometrics



Properties of Crash (Accident)

- **Random in nature**
- **Are count data**
- **Are always non-negative**



Appropriate Distribution

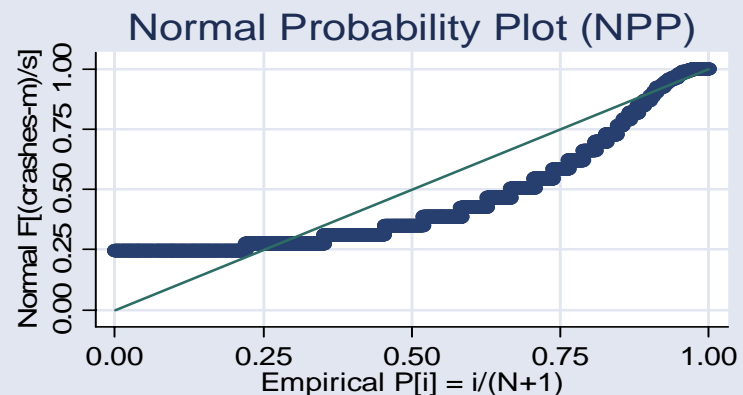
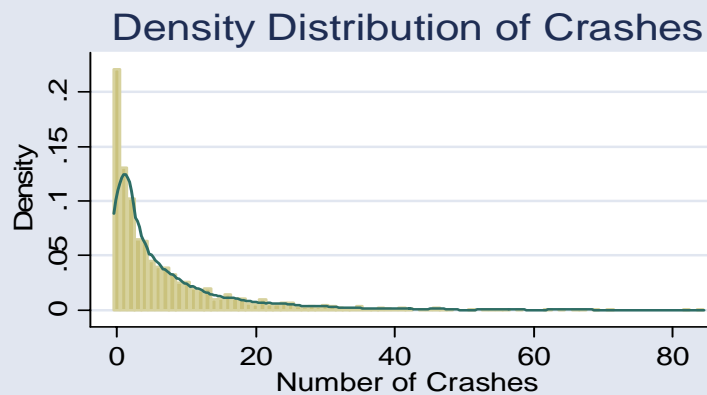
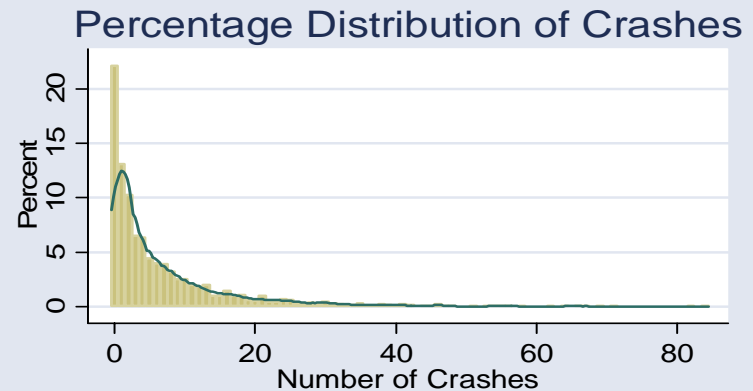
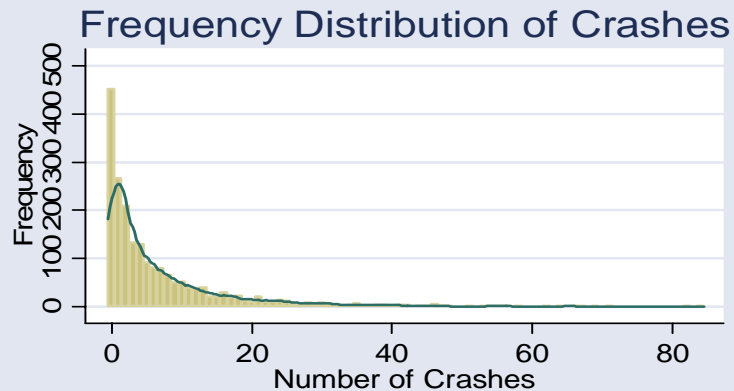
- **Poisson Distribution**
 - **Assume Mean is equal to Variance**
- **Negative Binomial**
 - **Takes care of overdispersed data**
 - **Assume Mean is not equal to variance**



Crash Distribution Plots

- Commands “*histogram crashes*”, “*pnorm crashes*” and “*graph combine*”

Distribution of Crashes



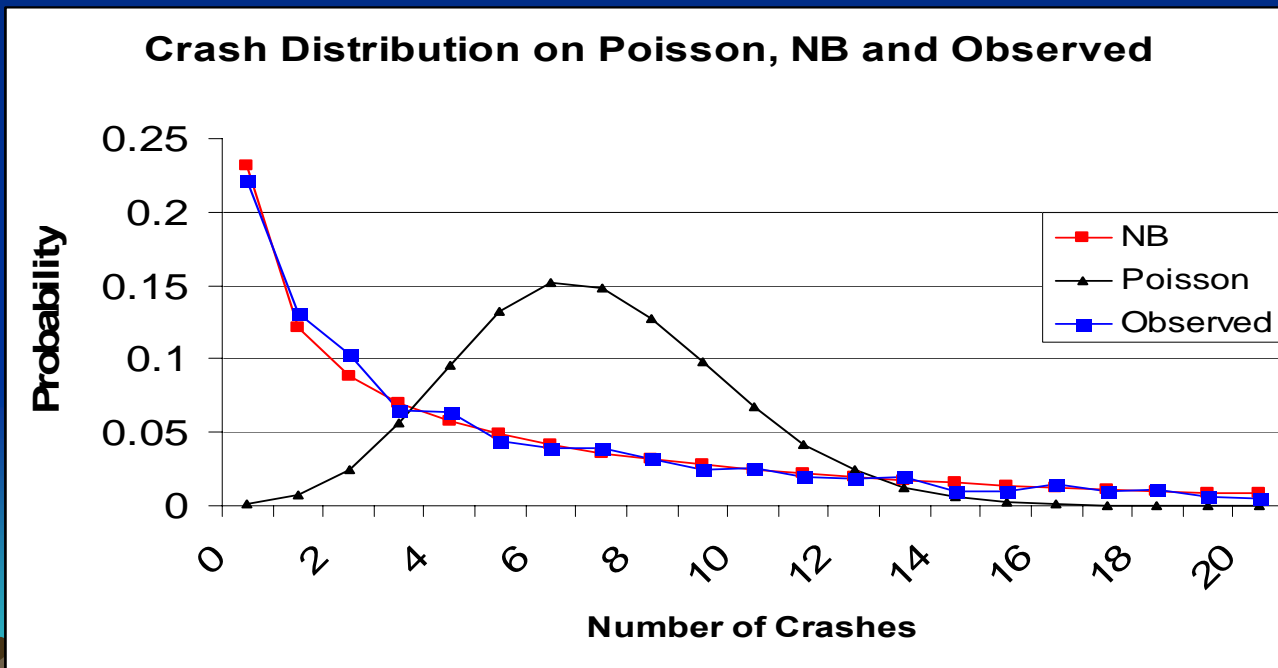
Choice Between Poisson and negative Binomial

- Equality of mean and variance
 - Command *“summarize crashes, detail”*

	Percentiles	Crashes		
25%	1	0	Observations	2038
50%	3		Mean	7
75%	9	69	Std. Dev	10
90%	18	71	Variance	100
95%	27	82	Skewness	2.86
99%	47	84	Kurtosis	14.19

Choice Between Poisson and negative Binomial Cont'

- Overdispersion test
 - Command “*nbvargr crashes*”
 - Overdispersion Factor=1.760794 , NB Favored



Test of alpha

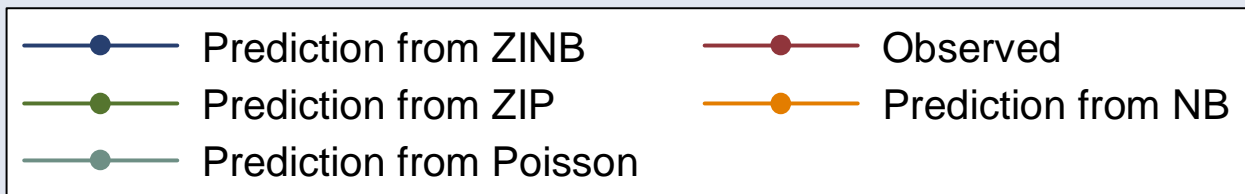
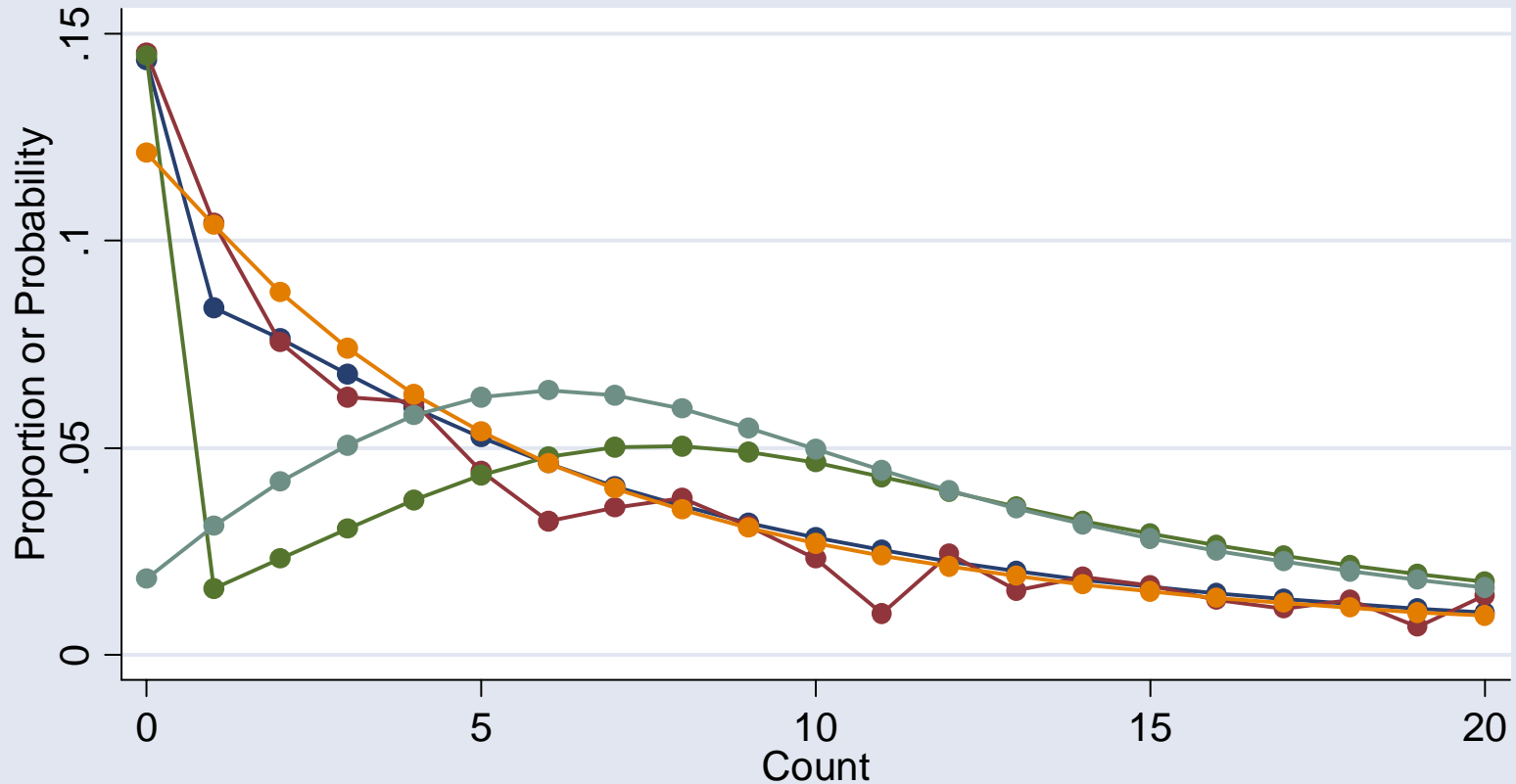
- Negative Binomial takes $\text{Var}(y_i) = \mu_i + \alpha \mu_i^2$.
If $\alpha = 0$, then Poisson is appropriate,
otherwise NB
- Command “*nbreg*”
- Alpha = 1.061775
- Likelihood-ratio test of alpha = 0 gave $\chi^2 = 7441.14$ with a p-value = 0.000.
- Hence α is significantly greater than 0,
NB Favored

Vuong's Test

- Takes care of excess zeroes in the data
- Zero Inflated Negative Binomial (ZINB) and Zero Inflated Poisson (ZIP)
- Command *“zinb crashes length .. surwidth, inflate(length ... surwidth) vuong ”*
- Vuong test of ZINB vs Negative Binomial(NB), $Z = 1.54$, $P\text{-value} = 0.06$, favoring *Negative Binomial*

Comparison of Distributions

Distributions of Crashes, Poisson, NB, ZIP and ZINB



Model

- Command “*nbreg crashes length accdens avgtfact medwidth sectadt sldwidth surwidth*”

Negative binomial regression					Number of Obs =2038		
Log likelihood=-5571.5611					LR chi2(7) =804.57		
					Prob >chi2 =0		
					PseudoR2 =0.0673		
crashes	Coef.	IRR	Std. Err.	z	P> z	[95% Conf.Interval]	
length	2.39931	11.0156	0.110069	21.8	0.000	2.183577	2.615038
accdens	0.00868	1.00872	0.002407	3.61	0.000	0.003966	0.0134
medwidth	-0.0074	0.992584	0.003058	-2.43	0.015	-0.01344	-0.00145
sldwidth	-0.0601	0.94165	0.013189	-4.56	0.000	-0.08598	-0.03428
surwidth	-0.067	0.93516	0.014891	-4.5	0.000	-0.09622	-0.03785
sectadt	2.4E-05	1.00002	1.76E-06	13.46	0.000	2.03E-05	2.72E-05
avgtfact	0.03838	1.03913	0.010137	3.79	0.000	0.018517	0.058252
_cons	2.24622		0.525809	4.27	0.000	1.215654	3.276787
Ln(alpha)	0.05994		0.040936			-0.02029	0.140175
alpha	1.06178		0.043464			0.979915	1.150475
Likelihood-ratio test of Alpha=0: chibar2(01)=7441.14					Prob>=chibar2 = 0.000		

Model Fitness Parameters

- Command *“fitstat”*

Measures of Fit for nbreg of crashes	
Log-Lik Intercept Only=-5969.901	Log-Lik Full Model=-5566.98
Deviance=11133.95	LR(7):805.852
	Prob > LR:0
McFadden's R2=0.067	McFadden's Adj R2=0.066
Maximum Likelihood R2:0.327	Cragg & Uhler's R2=0.328
AIC:5.475	AIC*n=11151.95
BIC=-4317.855	BIC'=-752.517

Finding from the model

- The longer the section length, the higher the crash rate
- The more the number of vehicle, the higher the likelihood of crash
- The higher the access density, the higher the crash rate
- The higher the percentage trucks the higher the probability of crash
- The wider the lane the lower the crash rate
- The wider the median, the lower the crash rate
- The wider the shoulder, the lower the crash rate

Sensitivity Analysis

- **Unit and standard deviation change in independent variables**
- **Marginal Effect**
- **Discrete change of the variable**



Unit and standard deviation change in independent variables

- *Command “listcoef, percent help”*

nbreg (N=2038): Percentage Change in Expected Count

Observed SD: 9.9653801

crashes	b	z	P> z	IRR &%		Std Deviation		SDofX
				e^b (irr)	%	e^bStdX	%StdX	
length	2.39931	21.798	0	11.0155	1001.6	1.8659	86.6	0.26
accdens	0.00868	3.608	0	1.0087	0.9	1.1181	11.8	12.8585
avgtfact	0.03838	3.787	0	1.0391	3.9	1.1077	10.8	2.6656
medwidth	-0.0074	-2.435	0.015	0.9926	-0.7	0.9351	-6.5	9.0195
sectadt	0.00002	13.463	0	1	0	1.4489	44.9 1	5619.36
sldwidth	-0.0601	-4.559	0	0.9416	-5.8	0.8818	-11.8	2.0927
surwidth	-0.067	-4.502	0	0.9352	-6.5	0.873	-12.7	2.0257
ln alpha	0.05994	1.464						

Marginal Effect

- Command “*mfxf compute*”

Marginal effects after nbreg

$y = \text{predicted number of events (predict)} = 5.071057$

Variable	dy/dx	Std. Err	z	P> z	[95% Conf.Interval]		Mean(X)
length	12.167	0.11007	21.8	0.00	2.18358	2.61504	0.255773
accdens	0.04403	0.00241	3.61	0.00	0.00397	0.0134	13.5819
medwidth	-0.0377	0.00306	-2.43	0.015	-0.0134	-0.0015	20.7655
sldwidth	-0.3049	0.01319	-4.56	0.00	-0.086	-0.0343	2.89426
surwidth	-0.3399	0.01489	-4.5	0.00	-0.0962	-0.0379	34.7964
sectadt	1.20E-04	1.8E-06	13.46	0.00	2.1E-05	2.7E-05	47726
avgtfact	0.19465	0.01014	3.79	0.00	0.01852	0.05825	4.52513


Discrete change of the variable

- Command “*prchange, help*”

nbreg: Changes in Predicted Rate for crashes

	min- >max	0->1	-+1/2	-+sd/2	MargEfct		
length	275.3679	27.4951	15.3028	3.2146	12.167		
accdens	12.5078	0.0393	0.044	0.5665	0.044		
avgtfact	4.5498	0.1668	0.1947	0.5191	0.1947		
medwidth	-2.1829	-0.0439	-0.0377	-0.3405	-0.0377		
sectadt	14.6026	0	0.0001	1.8913	0.0001		
sldwidth	-2.7497	-0.3522	-0.3049	-0.6385	-0.3049		
surwidth	-4.2578	-3.3876	-0.34	-0.6891	-0.3399		
exp(xb):	5.0711						
	length	accdens	avgtfact	medwidth	sectadt	sldwidth	surwidth
x=	0.255773	13.5819	4.52513	20.7655	47726	2.89426	34.7964
sd(x)=	0.259974	12.8585	2.66564	90.01953	15619.4	2.09272	2.02567

Conclusion

- **By using Various tests from Stata, Negative Binomial has been found to be appropriate distribution for our crash data**
 - **Effect of the independent variables has been found and their significances**
 - **Sensitivity analysis of how the change in the measure of the roadway geometrics can change the crash frequency**
- 

END

Thank you all

?

