Advocating Safety for Bicyclists at Intersections: Investigating Factors that Influence Bicyclist Injury Severity in Bicycle-Motor Vehicle Crashes at Unsignalized Intersections in North Carolina

Shatoya Covert

Stata Conference 2020

July 30, 2020

Table of Contents

Introduction

- North Carolina Strategic Highway Safety Plan
- What is it?
- How will it be implemented?
- Relation to this study?

Purpose of the Study

The purpose of this study was to answer the following research questions:

- What are the potential factors associated with bicyclist injury severity in bicycle-motor vehicle crashes at unsignalized intersections?
- Do these factors impact bicyclist safety?

Background Definitions

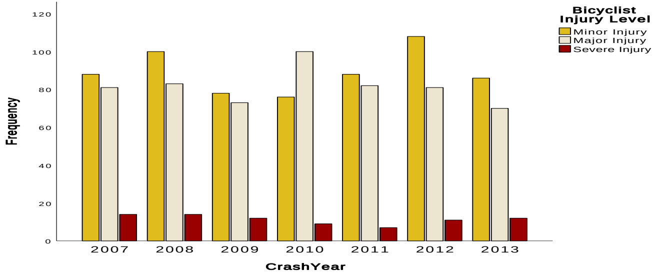
- Bicyclist Injury Severity - 5 types
- Unsignalized Intersections - 3 types

Background Data

- The UNC Highway Safety Research Center - 8,418 bicycle-motor vehicle (2007 to 2015)
- Sample size - 1,273 BMVC's at unsignalized intersections

Background Data

Frequency distribution of Bicyclist Injury Level of BMVC's at unsignalized intersections in North Carolina by year



Background - Variables Selected

- Bicyclist - age, gender
- Driver - age, gender, vehicle, vehicle speed
- Roadway - class, feature, speed limit, traffic control
- Crash - crash type, light condition, day of week
- Environmental - rural/urban land, crash time, season
- ALL VARIABLES ARE CATEGORICAL

Data Analysis - Ordinal Regression

Research question:

What are the potential factors associated with bicyclist injury severity in bicycle-motor vehicle crashes at unsignalized intersections?

- Ordinal Logistic regression - predict outcome of ordinal dependent variable
- Ordinal variable - categorical and has ordered relationship between outcomes

Data Analysis - Ordinal Regression

Ordinal Logistic Regression

- Performs binomial logistic regressions on cumulative logits
- logit = log of odds = $\ln\left[\frac{Prob(success)}{Prob(failure)}\right]$
- A logit can be modelled as a linear expression of a set of independent variables
- Cumulative logit - the odds of an event where that event results in the combination of 1 or more categories of an ordinal dependent variable

Data Analysis - Ordinal Regression Model

$$\mathsf{Y}^*_\phi = \sum_{h=1}^{H} \beta_h X_{h\phi} + \varepsilon_\phi = \mathsf{Z}_\phi + \varepsilon_\phi \tag{1}$$

$$\mathsf{Z}_\phi = \sum_{h=1}^{H} \beta_h X_{h\phi} = E(\mathsf{Y}^*_\phi) \tag{2}$$

$$P(\mathsf{Y}=1) = \frac{1}{1+\exp(\mathsf{Z}_\phi - \Gamma_1)}$$

$$P(\mathsf{Y}=2) = \frac{1}{1+\exp(\mathsf{Z}_\phi - \Gamma_2)} - \frac{1}{1+\exp(\mathsf{Z}_\phi - \Gamma_1)}$$

$$P(\mathsf{Y}=3) = 1 - \frac{1}{1+\exp(\mathsf{Z}_\phi - \Gamma_2)}$$

Data Analysis

Assumptions

- Dependent variable must be measured on an ordered level
- There is at least one independent variable that can be categorical or continuous
- There should be no multi-collinearity
- There are proportional odds

Data Analysis - Ordinal Regression

Proportional Odds (Parallel Regression) Assumption

- The slope on a continuous variable doesn't change across the different levels of your ordinal dependent variable.
- This assumption is tested by running separate binomial logistic regressions on cumulative binary dependent variables

Data Analysis - Ordinal Regression

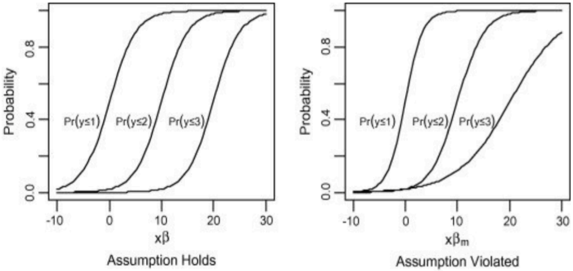


Figure: Proportional Odds Assumption

Data Analysis - Ordinal Regression

Proportional Odds Assumption Example

Driver Speed (compared to 0-20 mph)	y>1	y> 2	Brant test results Sig.
21-35 mph	0.47	0.808	0.292
	3.24	2.53	
Over 35 mph	0.807	1.83	0.024
	2.78	4.16	

Table: Binary logit coefficients

Data Analysis - Ordinal Regression - PO Results

The following variables did not meet the assumption

- Driver speed - Over 35 mph
- Driver vehicle - SUV
- Crash type - Bicyclist induced
- Light condition - Dawn and Dusk
- Crash time - Night
- Season - Fall
- χ^2 statistic for all analyzed variables was significant; Proportional Odds Assumption violated
- An alternative model needed

Data Analysis - Alternative Model for Analysis

Generalized Ordered Logit Model (Gologit)

- Partial proportional odds-relaxed the parallel regression assumption (i.e. relaxed assumption of same intercept shifts in our model with all categorical variables)
- Allowed some coefficients to be the same/different.
- Created a series of binary logistic regressions...dependent categories were combined
- Variables that violated the ordinal regression model also violated the gologit model
- Reference - Williams, R. (2006). Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Response Variables. The STATA Journal, 6, pp. 58-82.

Data Analysis - Gologit Model

$$P(Y_i > j) = g(X\beta_j) = \frac{exp(\alpha_j + X_i\beta_j)}{1 + [exp(\alpha_j + X_i\beta_j)]} \quad (3)$$

where

$$\alpha_j = \text{threshold or intercept parameters}$$
$$X_i = \text{vector of explanatory variables}$$
$$\beta_j = \text{vector of coeff. for explanatory variables}$$
$$j = 1, 2, ..., M - 1$$

Data Analysis - Gologit Model Results

- Wald test of parallel lines assumption: χ^2 is not significant; final model does not violate the proportional odds/parallel lines assumption

$$
\begin{aligned}
= \quad & -3.888 - 0.189 + 0.158X_2 + 0.514X_2 + 0.019X_4 + 0.003X_6 + 0.221X_7 \\
& - 0.088X_8 + 0.496X_{10} + 0.712X_{11a} + 1.980X_{11b} + 0.154X_{13} - 0.196X_{14} \\
& - 0.141X_{15} + 0.221X_{17} + 0.132X_{18} - 0.441X_{19} + 0.451X_{21} + 0.625X_{22} \\
& + 0.278X_{23a} + 1.188X_{23b} - 0.504X_{24} - 0.445X_{25} - 0.176X_{27} + 0.026X_{29} \\
& + 0.276X_{31a} + 1.221X_{31b} - 0.167X_{32} - 0.073X_{33} - 0.684X_{34} - 0.226X_{36a} \\
& + 1.448X_{36b} + 0.288X_{37} + 0.266X_{38} - 0.166X_{39} - 0.167X_{40} \\
& + 0.160X_{42a} + 2.031X_{42b} - 0.313X_{43} + 0.510X_{44} + 0.065X_{45} + 0.090X_{46a} \\
& - 0.634X_{46b}
\end{aligned}
$$

Data Analysis - Gologit estimates

Verification of the Model

$$\chi^2 = -2[\ln(L_0) - \ln(L_f)]$$

$$R^2 = 1 - \frac{\ln(L_f)}{\ln(L_0)}$$

$$\text{AIC} = -2 * \ln(\text{likelihood}) + 2 * k$$

Number of obs = 1,273
LR χ^2 (41) = 173.13
Prob > χ^2 = 0.0000
Log likelihood(model) = -1035.9246
Log likelihood(null) = -1122.488
Pseudo R^2 = 0.0771

Summary - Gologit Significant Variables - Marginal effects

Variables	Coef +/-	Minor/Major/Severe
Bicyclist age: 55+	positive	-0.118 / 0.088 / 0.030
Driver speed: 21-35	positive	-0.117 / 0.094 / 0.023
(m1)Driver speed: over 35 mph	+0.712	-0.165 / 0.001 / 0.165
(m2)Driver speed: over 35 mph	+1.980	
Road feature: 4-way-int.	positive	-0.105 / 0.085 / 0.020
Road feature:T-intersection	positive	-0.145 / 0.116 / 0.030
(*)Light condition: Dk-no lights.	negative	0.156 / -0.129 / -0.027
Day of week: Weekend	positive	-0.067 / 0.051 / 0.016
Season: Spring	positive	-0.119 / 0.088 / 0.031

Summary

- Conclusions
- Recommendations
- Future Work

Acknowledgements

- North Carolina Department of Transportation
- UNC Highway Safety Research Center
- Richard Williams and Hugh Briggs III

The End