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
- Purpose of the Study
- Research questions
- Background
- Data Analysis
- Summary
- Recommendations
- Acknowledgements
Introduction

- North Carolina Strategic Highway Safety Plan
- What is it?
- How will it be implemented?
- Relation to this 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
- \( \text{logit} = \log \text{ of odds} = \ln \left( \frac{\text{Prob}(\text{success})}{\text{Prob}(\text{failure})} \right) \)
- 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

\[ Y_{\phi}^* = \sum_{h=1}^{H} \beta_h X_{h\phi} + \varepsilon_{\phi} = Z_{\phi} + \varepsilon_{\phi} \]  

(1)

\[ Z_{\phi} = \sum_{h=1}^{H} \beta_h X_{h\phi} = E(Y_{\phi}^*) \]  

(2)

\[
P(Y = 1) = \frac{1}{1 + \exp(Z_{\phi} - \Gamma_1)}
\]

\[
P(Y = 2) = \frac{1}{1 + \exp(Z_{\phi} - \Gamma_2)} - \frac{1}{1 + \exp(Z_{\phi} - \Gamma_1)}
\]

\[
P(Y = 3) = 1 - \frac{1}{1 + \exp(Z_{\phi} - \Gamma_2)}
\]
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

<table>
<thead>
<tr>
<th>Driver Speed (compared to 0-20 mph)</th>
<th>$y &gt; 1$</th>
<th>$y &gt; 2$</th>
<th>Brant test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-35 mph</td>
<td>0.47</td>
<td>0.808</td>
<td>0.292</td>
</tr>
<tr>
<td></td>
<td>3.24</td>
<td>2.53</td>
<td></td>
</tr>
<tr>
<td>Over 35 mph</td>
<td>0.807</td>
<td>1.83</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>2.78</td>
<td>4.16</td>
<td></td>
</tr>
</tbody>
</table>

Table: Binary logit coefficients
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

\( \chi^2 \) 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
Data Analysis - Gologit Model

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

(3)

where

\( \alpha_j = \) threshold or intercept parameters
\( X_i = \) vector of explanatory variables
\( \beta_j = \) vector of coeff. for explanatory variables
\( j = 1, 2, ..., M - 1 \)
Data Analysis - Gologit Model Results

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

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

Verification of the Model

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

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

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

Number of obs = 1,273
LR $\chi^2$ (41) = 173.13
Prob > $\chi^2$ = 0.0000
Log likelihood(model) = -1035.9246
Log likelihood(null) = -1122.488
Pseudo $R^2 = 0.0771$
<table>
<thead>
<tr>
<th>Variables</th>
<th>Coef +/-</th>
<th>Minor/Major/Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicyclist age: 55+</td>
<td>positive</td>
<td>-0.118 / 0.088 / 0.030</td>
</tr>
<tr>
<td>Driver speed: 21-35</td>
<td>positive</td>
<td>-0.117 / 0.094 / 0.023</td>
</tr>
<tr>
<td>(m1)Driver speed: over 35 mph</td>
<td>+0.712</td>
<td>-0.165 / 0.001 / 0.165</td>
</tr>
<tr>
<td>(m2)Driver speed: over 35 mph</td>
<td>+1.980</td>
<td></td>
</tr>
<tr>
<td>Road feature: 4-way-int.</td>
<td>positive</td>
<td>-0.105 / 0.085 / 0.020</td>
</tr>
<tr>
<td>Road feature: T-intersection</td>
<td>positive</td>
<td>-0.145 / 0.116 / 0.030</td>
</tr>
<tr>
<td>(*)Light condition: Dk-no lights.</td>
<td>negative</td>
<td>0.156 / -0.129 / -0.027</td>
</tr>
<tr>
<td>Day of week: Weekend</td>
<td>positive</td>
<td>-0.067 / 0.051 / 0.016</td>
</tr>
<tr>
<td>Season: Spring</td>
<td>positive</td>
<td>-0.119 / 0.088 / 0.031</td>
</tr>
</tbody>
</table>
Summary

▶ Conclusions
▶ Recommendations
▶ Future Work
Acknowledgements

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