PREDICTING ROAD ACCIDENTS: A RARE-EVENTS MODELING APPROACH

Dr. Athanasios Theofilatos, National Technical University of Athens (NTUA), Athens, Greece
Prof. George Yannis, National Technical University of Athens (NTUA), Athens, Greece
Dr. Pantelis Kopolias, University of Thessaly (UTH), Volos, Greece
Mr. Fanis Papadimitriou, Attika Tollway Operations Authority – Attikies Diadromes S.A., Paiania, Greece

INTRODUCTION

Modeling road accident occurrence has gained increasing attention over the years. Considerable efforts have been made from researchers and policy makers in order to explain road accidents and improve road safety performance of highways. In reality, road accidents are rare events. In such cases, the binary dependent variable is characterized by dozens to thousands of times fewer events (accidents) than non-events (non-accidents).

OBJECTIVE

The objective of this study is to investigate accident likelihood on freeways by utilizing real-time traffic data and by considering accidents as rare events.

DATA

- The Attica Tollway (Attiki Odos) was chosen.
- Attiki Odos is a modern freeway in Greater Athens Area, Greece.
- Accident and non-accident cases for 2008-2011.
- 8 random locations
- Basic Freeway Segments (BFS) were considered.
- Real-time traffic data
- Closest loop detectors
- 1-hour intervals
- Traffic flow
- Speed
- Occupancy
- Percentage of heavy vehicles in traffic

METHODOLOGY

- Accidents are considered as rare-events.
- Traditional logit coefficients are biased.

- Rare-events logistic regression (King and Zeng, 2001a and 2001b).
- Case-control sampling design based on stratified sampling.
- All events and a random selection of non-events.
- A proportion of 1:10 for the ratio of events (accidents) to non-events (non-accidents) was used in each sample.

- A number of corrections are applied:
  \[ a_i = \alpha - \ln\left( \frac{\hat{p}_i - p_i}{1 - \hat{p}_i - p_i} \right) \]
  where \( a_i \) is the new corrected constant term, \( \alpha \) is the uncorrected constant term, \( \tau \) is the proportion of events in the population and \( \gamma \) is the proportion of events in the sample.

- The corrected logit function:
  \[ \logit \hat{p}_i = \ln \left( \frac{\hat{p}_i}{1 - \hat{p}_i} \right) = a_0 + \sum a_i x_i \]
  where \( a_i \) is the correction factor
  \[ C_i = (0.5 - \hat{p}_i) \times (1 - \hat{p}_i) \times \hat{p}_i \]
  \( \hat{p}_i \) is the probability of an event estimated using the corrected estimated coefficient \( a_i \).
  \( x_i \) is the \( 1^{\text{st}} \) variance-covariance matrix of values for each independent variable, \( V(\beta) \) is the variance-covariance matrix, and
  \[ \hat{p}_i \] is the \( x_i \) transposed.

RESULTS

<table>
<thead>
<tr>
<th>Trial 1</th>
<th>( \beta )</th>
<th>S.E.</th>
<th>z value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>26.4158</td>
<td>11.3706</td>
<td>2.3232</td>
<td>0.0212</td>
</tr>
<tr>
<td>Truck.Prop.</td>
<td>-0.0394</td>
<td>0.1072</td>
<td>-0.3684</td>
<td>0.7129</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial 2</th>
<th>( \beta )</th>
<th>S.E.</th>
<th>z value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>20.0631</td>
<td>12.6321</td>
<td>1.5816</td>
<td>0.1159</td>
</tr>
<tr>
<td>Truck.Prop.</td>
<td>-0.0344</td>
<td>0.0864</td>
<td>-0.4004</td>
<td>0.6840</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial 3</th>
<th>( \beta )</th>
<th>S.E.</th>
<th>z value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>26.2153</td>
<td>12.7865</td>
<td>2.0111</td>
<td>0.0442</td>
</tr>
</tbody>
</table>

CONCLUSIONS

- Stratified sampling.
- 3 trials.
- All accident cases and a random sample of non-accident cases were included in each trial.
- The best models are presented.
- Adequate model fit.
- The risk factors were identified.
- Consistent effect of speed and truck proportion in the different trials.
- The logarithm of speed significantly affects accident occurrence.
- Consistent findings with past literature.

REFERENCES