Car Crash Analysis

Michael Vaughn
David Lowe

Introduction

- The effect of an automobile crash can be classified by the severity of the injuries/casualties caused by the crash.
- Public Policy is enacted to decrease the number of fatal and severe injury inducing crashes.
- Public officials seek to take precautions against severe crashes.
    - Legislation
    - Traffic planning

Goal of Analysis

- Understand the relationship between factors that are generally associated with severe and non-severe crashes so that policies can be enacted to help protect drivers from having severe crashes.

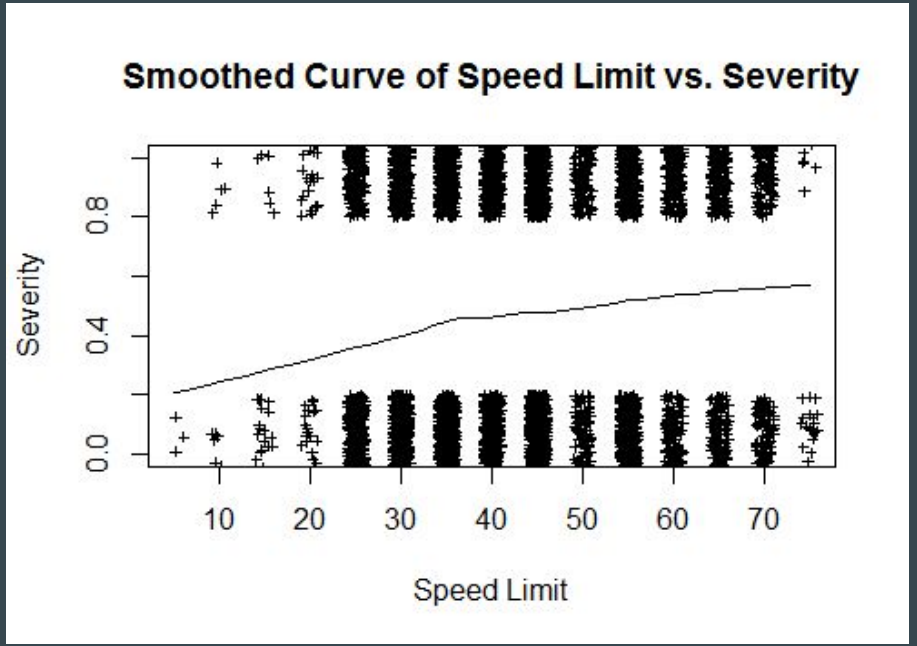
Data Description

- Data from 8,603 automobile crashes were collected
- Each crash is classified as either severe or not severe
  - Severe: (1 or more individuals sustained serious, or fatal, injuries)
  - Not Severe: No one sustained serious, or fatal, injuries
- The other variables in the data set attempt to describe the environment in which the crash took place, i.e. weather, speed limit, light conditions, etc.

Variable Type	Variable
Continuous	Speed Limit
Categorical	Hour (Military Time) Light Condition Weather Alcohol Use Type of Intersection Restraint Use Airbag Type of Trafficway Number of Lanes Road Alignment Surface Condition

Data Exploration

- Though speed limits are commonly marked in increments of 5 mph, we decided to treat speed limit as a continuous variable for this analysis.
- This jittered scatterplot seems to indicate that there may be a linear relationship between speed limit and crash severity.



Data Exploration

- Because some variables had many possibilities for classification, we looked at tables showing the number of crashes in our data set that fall into each classification.

Roadway Surface Conditions

Dry	Wet	Snow	Ice/Frost	Sand	Water (Flooding)	Other	Slush	Mud, Dirt, Gravel
6969	1310	117	149	2	16	4	28	8

- After reviewing these tables, we became concerned about some of the classifications that were rare occurrences in our data, i.e. a crash on a sandy roadway surface.

Data Adjustments

Light Conditions: Combined ‘Dusk’ and ‘Dawn’

Weather: Combined ‘Rain’ and ‘Freezing Rain’; ‘Sleet’, ‘Snow’, and ‘Blowing Snow’; ‘Severe Crosswinds’ and ‘Blowing Sand and Dirt’

Type of Intersection: Combined ‘L’ and ‘T’ intersections; ‘Traffic Circle’ and ‘Roundabout’

Data Adjustments

Restraint Use: Combined ‘Shoulder Belt Only’ and ‘Lap Belt Only’

Airbag: Combined ‘Curtain (roof)’ and ‘Other (knee, air belt, etc.)’

Number of Lanes: Combined ‘5 Lanes’, ‘6 Lanes’, and ‘7 or more Lanes’

Surface Condition: Combined ‘Snow’, ‘Slush’, and ‘Ice’; ‘Water’, ‘Sand’, ‘Mud, Dirt, Gravel’, and ‘Other'

Model

$$Y_i \overset{ind}{\sim} Bern(p_i)$$

Y_i represents the severity of the i^{th} crash. $Y_i \in \{0, 1\}$

p_i represents the probability of severity for the i^{th} crash. $p_i \in [0, 1]$

Model

$$log\left(\frac{p_i}{1-p_i}\right) = x_i'\beta$$

x_i' represents the crash characteristics for the i^{th} crash.

β represents the effect of each crash characteristic (x_i') on crash severity.

Model Selection

- Expand each variable level as individual factor
- Combine similar factors (e.g. traffic circle & roundabout)
- Select variables using forward variable selection
- Only used the significant variables from the forward selection model (this had minimal effect on AIC and Pseudo-R^2)

Selected Variables

Airbags Deployed - Combination
Airbags Deployed - Unknown
Airbags Deployed - Front
Airbag Not Deployed
Speed Limit
3 Lane Road
Dark, but Lighted
Curve - Left
Curve - Unknown Direction
Cloudy Weather
No Restraints Used
Restraint Used, but Unknown
DOT-Compliant Motorcycle Helmet
During 1am Hour
During 3am Hour
During 4am Hour
During 7am Hour
4-Way Intersection
T or L Intersection
Alcohol Involved

Assumptions

- Each crash is considered independent of all other crashes.
- Log-odds probability is monotone in each predictor variable.
- No significant collinearity is shown between predictors.

Variance Inflation Factors

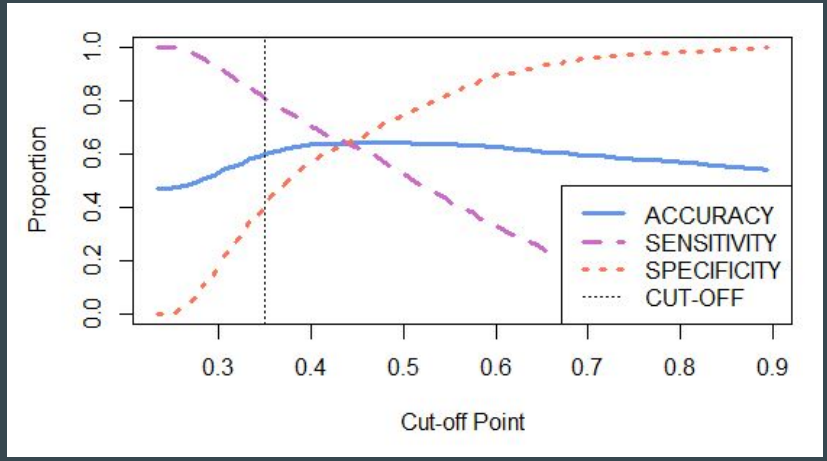
Airbag_{NoDep}	Restraint_{None}	Alcohol	Speed Limit	$\text{Dark}_{Lighted}$
4.128	1.039	1.100	1.175	1.105
Helmet	Curve_{Left}	$\text{Restraint}_{Used/Unknown}$	3am	Lanes-3
1.400	1.051	1.056	1.020	1.034
1am	$\text{Curve}_{Unknown}$	Airbag_{Front}	Airbag_{Combo}	$\text{Airbag}_{Unknown}$
1.035	1.008	3.422	1.551	2.110
4am	7am	$\text{Intersection}_{4-way}$	$\text{Intersection}_{TorL}$	Cloudy
1.024	1.014	1.227	1.125	1.007

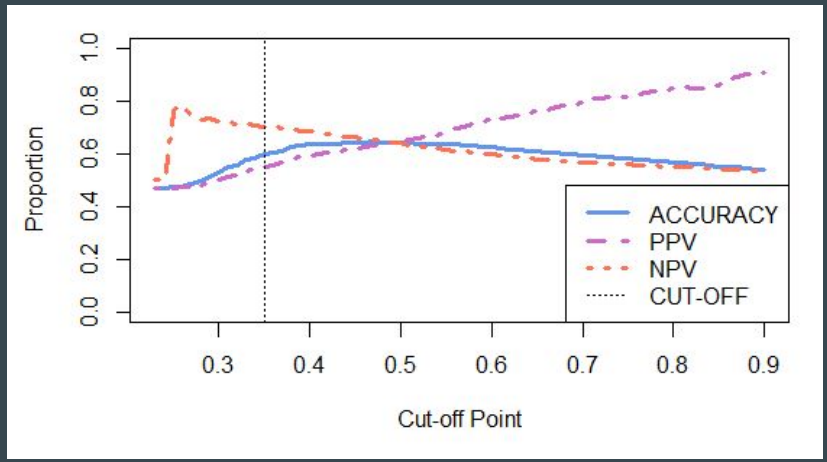
Assessing the Model

- Pseudo-R^2 of 8.47%
- Accuracy is above 50% for the majority of possible cut-off points.
- Sensitivity is the main concern of the model

Cut-off = .35

Accuracy	59.83%
Sensitivity	80.14%
Specificity	41.72%
Positive Predictive Value	55.06%
Negative Predictive Value	70.22%





Model Estimates

Coefficient	$e^{\hat{\beta}}$	2.5%	97.5%	Coefficient	$e^{\hat{\beta}}$	2.5%	97.5%
$\beta_{intercept}$	0.439	0.339	0.568	β_{1am}	1.634	1.153	2.315
$\beta_{Airbag_{NoDep}}$	0.709	0.590	0.851	$\beta_{Curve_{Unknown}}$	1.683	1.102	2.570
$\beta_{Restraint_{None}}$	3.751	3.078	4.572	$\beta_{Airbag_{Front}}$	1.612	1.326	1.960
$\beta_{Alcohol}$	1.620	1.393	1.884	$\beta_{Airbag_{Combo}}$	1.896	1.433	2.509
$\beta_{SpeedLimit}$	1.010	1.006	1.014	$\beta_{Airbag_{Unknown}}$	1.495	1.185	1.886
$\beta_{Dark_{Lighted}}$	1.377	1.216	1.559	β_{4am}	1.504*	0.989	2.288
β_{Helmet}	0.756*	0.549	1.042	β_{7am}	1.213*	0.982	1.497
$\beta_{Curve_{Left}}$	1.446	1.187	1.760	$\beta_{Intersection_{4-way}}$	0.884	0.793	0.986
$\beta_{Restraint_{Used/Unknown}}$	1.607	1.263	2.044	$\beta_{Intersection_{TorL}}$	0.864*	0.742	1.006
β_{3am}	2.025	1.311	3.128	β_{Cloudy}	1.111*	0.983	1.256
$\beta_{Lanes-3}$	1.228	1.082	1.395				

* Indicates a variable that is close to significant at an α=0.05 level.

Conclusion

- **Model Advantages**
  - Easy interpretability of the effect of predictor variables
  - Limited number of factors explaining severe car crashes
  - At a cut-off of 0.35, the model predicts 80% of severe car crashes
- **Model Shortcomings**
  - The model only explains a small portion of what causes severe car crashes
  - Time of day is considered a factor, rather than a continuous variable

Conclusion

- **Policy Recommendations**
  - Ensure proper seat belt use
- **Future Research**
  - Cell phone use and other in-car distractions
  - Assess cause of severe crash during early hours of the day (1 a.m. - 7 a.m.)