**APPENDIX A**

*Table A1.* Descriptive Statistics by Intersection

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Description** | **Average** | **St. Dev.** | **Minimum** | **Maximum** |
| **Geometry** | | | | |
| Number of Left Turn Lanes | 4.53 | 1.83 | 1.00 | 10.00 |
| Number of Through Lanes | 8.23 | 1.91 | 4.00 | 14.00 |
| Number of Right Turn Lanes | 2.50 | 1.61 | 0.00 | 5.00 |
| **Traffic** | | | | |
| Major Road AADT (vpd) | 28,700 | 8,169 | 777 | 48,917 |
| Minor Road AADT (vpd) | 9,456 | 7,972 | 777 | 48,917 |
| **Signal Control** | | | | |
| Legs with Permissive Signal (P) | 0.99 | 1.25 | 0.00 | 4.00 |
| Legs with Protective/Permissive Signal (PP) | 1.60 | 1.41 | 0.00 | 4.00 |
| Legs with Protected Only Signal (PO) | 1.42 | 1.68 | 0.00 | 4.00 |
| **Speed** | | | | |
| Posted Speed Limit Major Road in kph (mph) | 66.35(41.23) | 7.30(4.54) | 40.23(25.00) | 88.52(55.00) |
| Posted Speed Limit Minor Road in kph (mph) | 49.47(30.74) | 10.09(6.27) | 24.14(15.00) | 72.42(45.00) |
| **Weather** | | | | |
| Snowfall in centimeter/year (inches/year) | 120.95(47.62) | 14.35(5.65) | 108.15(42.58) | 161.87(63.73) |
| Days a Year with Snowfall ≥ 2.54 cm (1 inch) | 14.98 | 1.18 | 13.42 | 17.08 |
| Intersection-Weather Station in kilometers (mi) | 6.24(3.88) | 2.83(1.76) | 0.90(0.56) | 14.02(8.71) |
| Number of Weather Stations | 17 | | | |
| **Crashes** | | | | |
| Fatal and Injury (FI) (crashes/year) | 4.22 | 2.78 | 0.33 | 15.17 |
| Property Damage Only (PDO) (crashes/year) | 7.40 | 4.79 | 0.50 | 25.08 |
| Total (TOT) (crashes/year) | 11.62 | 7.30 | 0.92 | 39.75 |
| **Other** | | | | |
| Years of Data per Intersection | 8.38 | 3.55 | 1.00 | 12.00 |
| Number of Intersections | 182 (model development) and 138 (model assessment) | | | |
| Overall Period of Analysis | 2005-2016 | | | |

*Table A2.* Wisconsin Southeast Jurisdiction-Specific Crash Prediction Model Coefficients

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | 0 | 1 | 2 | 3 | 4 | 5 |
| Variable Introduced | None |  |  |  |  |  |
| AIC | -51302.20 | -51389.62 | -51412.13 | -51418.56 | -51422.42 | -51427.10 |
| Log-lik | 25659.10 | 25704.81 | 25718.06 | 25723.28 | 25726.21 | 25729.55 |
|  | 3.620 | 6.093 | 7.181 | 7.693 | 7.946 | 8.154 |
|  | 0.857 | 0.494 | 0.489 | 0.411 | 0.409 | 0.410 |
|  | 0.670 | 0.386 | 0.382 | 0.321 | 0.320 | 0.321 |
|  | 0.712 | 0.410 | 0.406 | 0.341 | 0.340 | 0.341 |
|  | 0.735 | 0.423 | 0.419 | 0.352 | 0.351 | 0.352 |
|  | 1.224 | 0.575 | 0.563 | 0.381 | 0.378 | 0.388 |
|  | 2.270 | 1.065 | 1.043 | 0.707 | 0.702 | 0.720 |
|  | 7.545 | 7.545 | 7.234 | 2.409 | 2.383 | 2.619 |
|  |  | 19.449 | 19.862 | 17.226 | 17.978 | 14.676 |
|  |  | -0.815 | -0.941 | -0.913 | -0.933 | -0.931 |
|  |  |  | 0.096 | 0.098 | 0.102 | 0.091 |
|  |  |  | -0.286 | -0.342 | -0.367 | -0.324 |
|  |  |  |  | 0.596 | 0.612 | 0.546 |
|  |  |  |  | -0.010 | -0.010 | -0.008 |
|  |  |  |  |  | 0.270 | 0.248 |
|  |  |  |  |  |  | -0.213 |
|  |  |  |  |  |  | 2.508 |
| AIC Improvement | | 87.42 | 22.51 | 6.43 | 3.86 | 4.68 |
| Log-lik. Improvement | | 45.71 | 13.25 | 5.21 | 2.93 | 3.34 |
| Improvement | | 2.47 | 1.09 | 0.51 | 0.26 | 0.21 |

Note: The fully loaded model coefficients are for Model 5 (coefficients in las column).

**Measures of Goodness of Fit**

The measures of goodness of fit for model development consisted of the log-likelihood, inverse Overdispersion, CURE plots, and AIC.

The model parameters that maximized the Multivariate Negative Multinomial likelihood function (Eq. 1) are those that maximize the sum of resulting in the log-likelihood. An increase in log-likelihood is desired as predictor variables with optimal functional forms are introduced in the model. The overdispersion parameter indicates the variability of the model in comparison to a Poisson distribution with the same mean. The reliability of the model is likely to be higher with a larger value of the inverse overdispersion coefficient (𝒷 = 1/k).

CURE plots track model performance throughout the range of predictor variables. A satisfactory CURE plot is one that follows a random walk around the horizontal axis and within the confidence intervals (±2σ). In contrast, sudden vertical changes represent outliers, and long increasing or decreasing walks of residuals represent regions of consistent under- or over-prediction (Hauer 2015).

The AIC estimates the quality of the model in relation to other models based on information theory. Models with the minimum AIC values are preferred since the AIC rewards goodness of fit but penalizes models based on the number of estimated parameters, discouraging overfitting. Throughout the process of adding more variables, trying different functional forms, or changing the order in which the variables were introduced, all measures of goodness of fit were evaluated.

Models at different stages of the development can be used based on data availability. For instance, if the AADT and posted speed limit were the only data available, Model 2 coefficients could be used. The order of introduction of variables in the model showed a consistent magnitude in the model coefficients. All predictor variables contributed to the model in terms of measures of goodness of fit. The AADT had the most significant contribution to the models with an improvement of 45.71 in log-likelihood. Similarly, the improvement in inverse overdispersion was of 2.47. The posted speed limit, number of lanes, and weather variables also had significant contributions to the model in both log-likelihood and inverse overdispersion. The fully loaded model CURE plots as a function of AADT are presented in Figure A1. The walk of cumulative residuals showed optimal oscillations within the confidence intervals and around the value of zero at all ranges of AADT.

(a)

(b)

*Figure A1*. Cumulative Residual Plots as a Function of AADT for (a) FI and (b) PDO Crashes.

**Model Assessment**

Model predictions were compared with observed crash rates. Percent variation from the annual estimate was used for model assessment. Data from 138 intersections which contained an uninterrupted period of analysis for 12 years between 2005 and 2016 were used. Observed seasonal crash rates were calculated with seasonal traffic counts adjusted using traffic count sensor data (geometry and weather is not accounted for in crash rates). Thus, the comparison intends to verify that the model predictions follow observed seasonal fluctuations compared to the annual average. Figure A2 illustrates the results of the seasonal analysis. From visual inspection, the percent variations of model estimates and observed crash rates track each other well through the seasons. The sequence of sample odds ratio (Hauer 1997) was used to quantitatively assess the results. The results of the sample odds ratio showed that the sample mean was 0.96 (0.80 to 1.13) for total crashes (95% confidence interval in parenthesis). The sample mean odds ratio was sufficiently close to 1.0 and confidence intervals also include 1.0; thus, the seasonal model prediction and observed crash rate variations as percentage of annual estimates were similar.

The seasonal model showed that crash estimates were the highest during the winter season and the lowest during the spring. Model crash estimates remained below or near the annual average for the rest of the seasons.

*Figure A2.* Seasonal Variations as Percentage of Annual

Estimates of Total Crashes (FI+PDO).