Crash Data Analysis Tools

John Ash[a] , Hongyuan Liu[a], and Wenbo Zhu[a]

[a] Department of Civil and Environmental Engineering, University of Washington

BACKGROUND AND MOTIVATION

- Transportation safety is both a "hot" research area and a necessary task for government transportation agencies (Departments of Transportation etc.) to receive funding/prioritize investments
- 2014 National Statistics (NHTSA, 2016):
  - 29,989 fatal crashes
  - 32,675 fatalities
    - 1.07 per 100 million vehicle miles traveled (VMT)
    - 10.3 per 100,000 population
  - ~2.3 million injuries
  - ~6.1 million police-reported crashes
- Practitioners and researchers follow several typical steps in a crash data analysis:
  - Obtain and merge/aggregate data from multiple sources
  - Visualize data
  - Develop regression models and perform associated predictive analyses
  - Prioritize sites for safety improvements under limited funding
- Currently, these steps are performed haphazardly using many tools including Excel, MS Access, SQL Server, R, FHWA Safety Analyst

RESEARCH OBJECTIVE

Develop a comprehensive set of tools to simplify a variety of steps in the crash data analysis for practitioners and researchers

DATA PREPARATION

DATA SOURCES

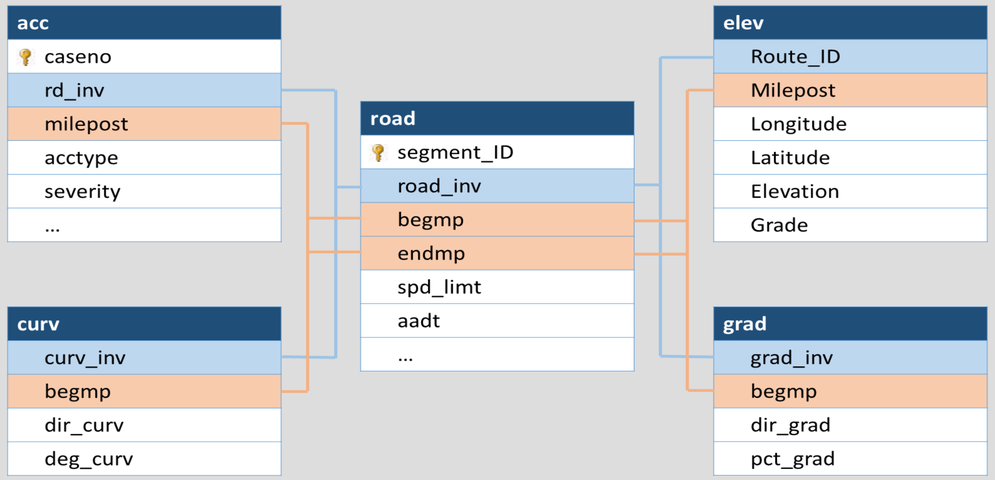
- **Accident (acc)** – Traffic Accident records contained in the HSIS (Highway Safety Information System) database
- **Curvature (curv)** – Roadway curvature information contained in the HSIS database
- **Grade (grad)** – Roadway grade information contained in the HSIS database
- **Road Segment (road)** – Roadway segment information (e.g., roadway inventory, milepost numbers, roadway geometry, etc.) contained in the HSIS database
- **Elevation (elev)** – Freeway elevation database created by research group in a previous research project

DATA MERGING

- Road segment is used as the base table for data merging, while accident, elevation, grade, and curvature data have been aggregated (e.g., count, average, maximum, and minimum) into the final dataset based on roadway inventory and milepost information
- Two data sources (i.e., elev and grad) can provide roadway grade information. The elevation dataset has high resolution roadway elevation information (per 10ft) which is useful in producing grade statistics such as average, maximum and minimum, but only covers the Interstate freeway network. The HSIS grade dataset, although covering more roadway mileage, contains less-detailed grade information. Thus, the HSIS grade data is only applied when the elevation data for the road segment is not available
- HSIS has different data tables for each analysis year. The data merging process first aggregates the annual data table for each year and then creates a final dataset combining all of the annual data tables

DATABASE AND TOOL DESIGN

DATABASE DESIGN



SELECT PYTHON PACKAGES USED

Package	Purpose
ipywidgets	Interactive plotting
matplotlib (Basemap)	Plotting, visualization, and mapping
pandas	Data manipulation
sqlite3	Database construction and querying
statsmodels	Regression modeling

KEY COMPONENTS OF TOOL

Data Merging

- Merge data from multiple sources (e.g., road inventory data, crash data etc.)

Data Summary

- Get summary statistics for continuous and categorical variables

Predictive Crash Modeling

- Estimate negative binomial regression model with offset term for crash prediction (aka safety performance function)

Estimation of Expected Safety of Road Segments

- Implement Empirical Bayes (EB) method to combine predicted crash mean and observed crash count in a weighted sum

Ranking of Sites (i.e., prioritize for safety treatment)

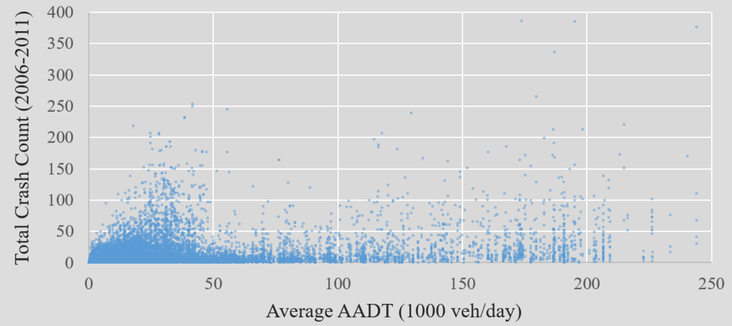
- Calculate accident reduction potential (ARP) based on EB results
- Generate map of crash rates to show "hot spots"

Confidence/Prediction Intervals for Regression Model

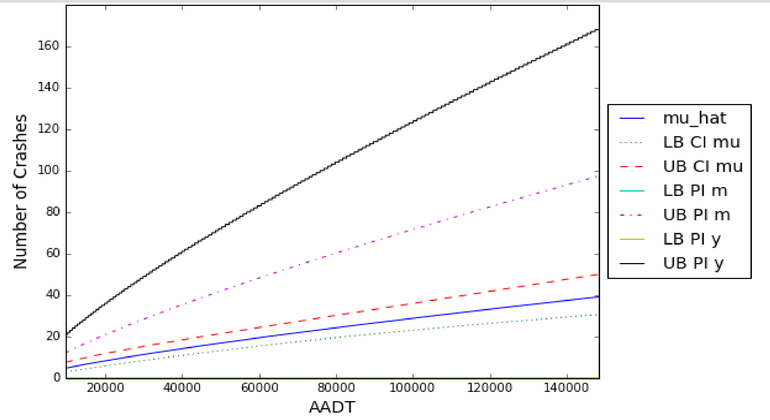
- Calculate and plot confidence and prediction intervals for parameters at different levels of hierarchy in mixed-Poisson model (Poisson mean (μ), Poisson parameter (m), predicted response (y))

SAMPLE RESULTS/OUTPUT

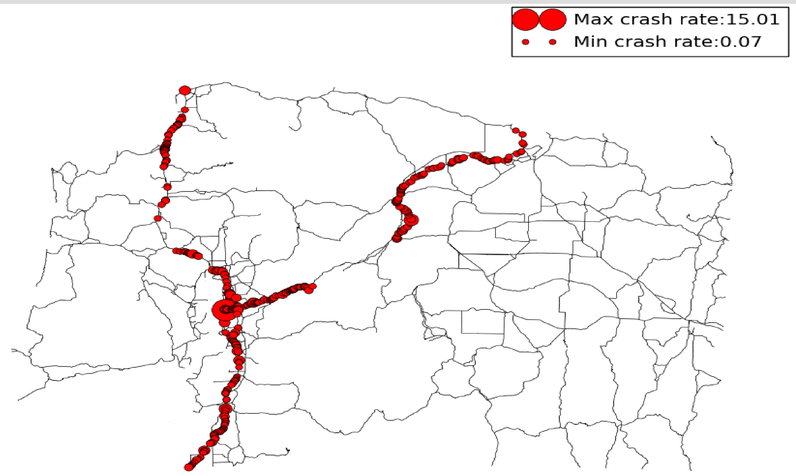
ROAD SEGMENT CRASH SUMMARY



CONFIDENCE AND PREDICTION INTERVALS FOR NB MODEL



WASHINGTON STATE HOT SPOT CRASH RATE MAP



REFERENCE

National Highway Traffic Safety Administration (NHTSA). (2016). "Quick Facts 2014." *FARS Encyclopedia*, <http://www-nrd.nhtsa.dot/gov/Pubs/812234.pdf> (Feb. 23, 2016).