Data-Driven Safety Training Introduction Part II

Carlos Sun, Praveen Edara, Yaw Adu-Gyamfi

University of Missouri

Missouri Center for Transportation Innovation

Outline

- Background/motivation
- Subjective vs. objective safety
- **Complexity of traffic crashes & data**
- Regression to the mean bias
- Review of statistics
- Use and application of data-driven safety methods

Definitions

- crash
  - set of events
  - resulting in injury or property damage
  - due to the collision of at least one vehicle with another vehicle, pedestrian, bicyclist, fixed object
- crash frequency
  - # crashes/year for a particular transportation facility
  - e.g. road segment(s), intersection(s)

Why Use Crash Frequency?

- Measurable, data is collected by police
- Quantitative, easy to use
- Standard unit, can compare across different jurisdictions
- Follows a known statistical distribution, so can be modeled
- Can be used to forecast/predict future crash frequency, alternate conditions, or new facility

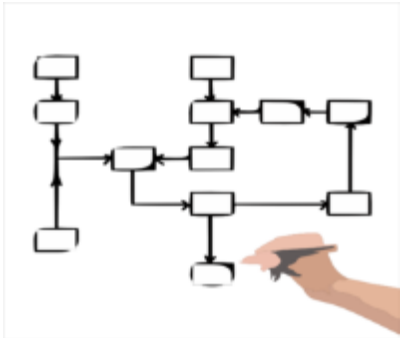
Review of Data-Driven Safety Theory

- Complexity of traffic crashes and data
- Review of statistics fundamentals
- How to overcome challenges in safety analysis

Complexity of Traffic Crashes & Data

How do crashes occur and what determines severity?

- Not a singular cause but convergence of a series of events
  - not before-crash factors only
- Data driven methods examines multiplicity of contributing factors



Complexity of Traffic Crashes & Data

- Examples
  - before – vehicle w/worn brakes/tires, driver inattention
  - during - air-bag deployed properly to cushion impact
  - after - EMS delayed by congestion leading to loss of blood

Complexity of Traffic Crashes & Data

Large number of possible factors, examples:

- behavior of multiple drivers interacting –aggressiveness, inattention, impairment
- vehicle char's – braking, acceleration, center of gravity
- pavement char's – surface friction
- geometric design – horizontal and vertical curves, superelevation, median, type of shoulder
- land use –alcohol establishments, ped generation (e.g. bus stops)
- traffic control – permitted/yield, right turn on red
- weather – snow, ice, rain

Complexity of Traffic Crashes & Data

Large number of possible factors, examples:

- Even when the confluence of these different factors creates a perfect storm, a crash still might not result
  - e.g. driver/ped takes evasive maneuvers to prevent crash

Complexity of Traffic Crashes & Data

Crashes are random & rare events, statistically

- Crashes are random
  - # of crashes fluctuates over time and are influenced by large # of factors
  - does not mean that the underlying factors are not measurable or identifiable
- Crashes are rare
  - only represent a small fraction of total amount of travel
  - 1.18 fatality per 100 million VMT (2016 NHTSA)

Complexity of Traffic Crashes & Data

- Categorize crashes as rare and random events in a statistical sense to understand data-driven safety methods
- Not to trivialize
  - ~35k fatal crashes (~900 in Missouri)
  - ~6M crashes per year

Complexity of Traffic Crashes & Data

Challenges in applying data-driven safety methods

- large # of factors
- some factors – no direct measurement, rely on surrogates, e.g. driver aggressiveness
- measurement errors – estimate vertical grades using topographic map
- subjective assignment of data, e.g. intersection-related crash even though far away?
- conflicting data, e.g. witnesses conflict
- data entry errors

Complexity of Traffic Crashes & Data

- Given crashes are rare random occurrences, they fluctuate naturally over time
- Use of small samples -> likely only random fluctuations are captured and not underlying trend
  - Example
    - A large number of crashes occurred at a particular intersection
    - Was this due to natural fluctuations?
    - Or was the number of crashes unusually high for such a type of intersection?
  - Problem of “regression to the mean” bias