Data-Driven Safety Training Introduction Part IV

Carlos Sun, Praveen Edara, Yaw Adu-Gyamfi
University of Missouri
Missouri Center for Transportation Innovation

Outline

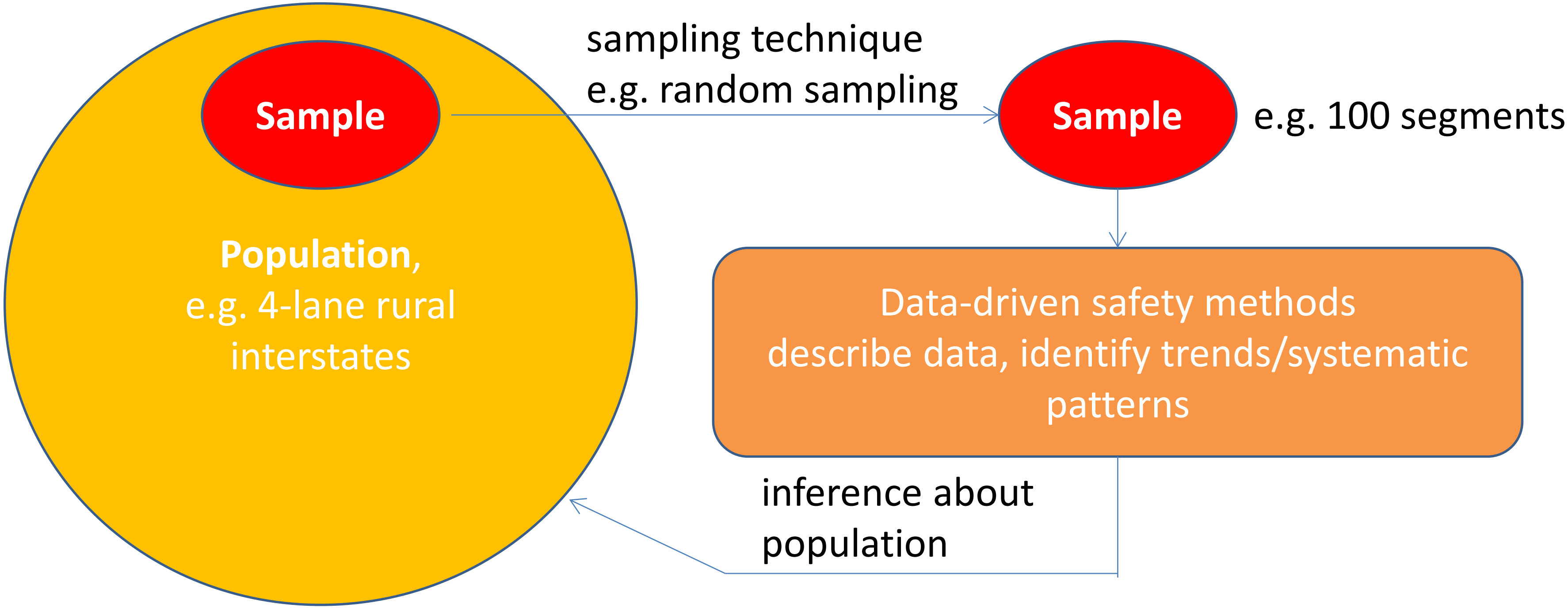
- Background/motivation
- Subjective vs. objective safety
- Complexity of traffic crashes & data
- Regression to the mean bias
- **Review of statistics**
- **Use and application of data-driven safety methods**

Review of Statistics

- De-mystifying safety analysis
- Important two words: sampling distributions
- Lay/simplified explanation
  - sampling – obtaining some data to help us get at overall trends
  - distribution – if enough data, then obtain a picture of how data behaves to get at overall trends/systematic behavior

Review of Statistics

- Sampling distributions



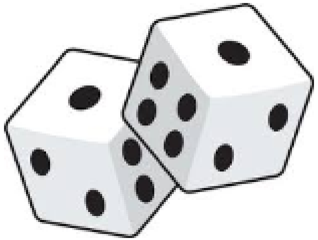
Review of Statistics

- Sampling distributions
  - Data-driven safety methods require good safety data
  - Many challenges in obtaining adequate safety data
    - money & time
    - aforementioned problems: unavailability measurement error, subjectivity, data entry
    - data bias, not being representative of systematic trends
      - e.g. one off, outlier

Review of Statistics

- Statistics is all about distributions
- Distribution – lay explanation
  - an adequate picture of how specific data behaves for different ranges of values
- Distribution – statistical explanation
  - possible values of data and how frequently each type of value occurs
  - central tendency (e.g. average), variability
- Distribution leads to trends/systematic behavior

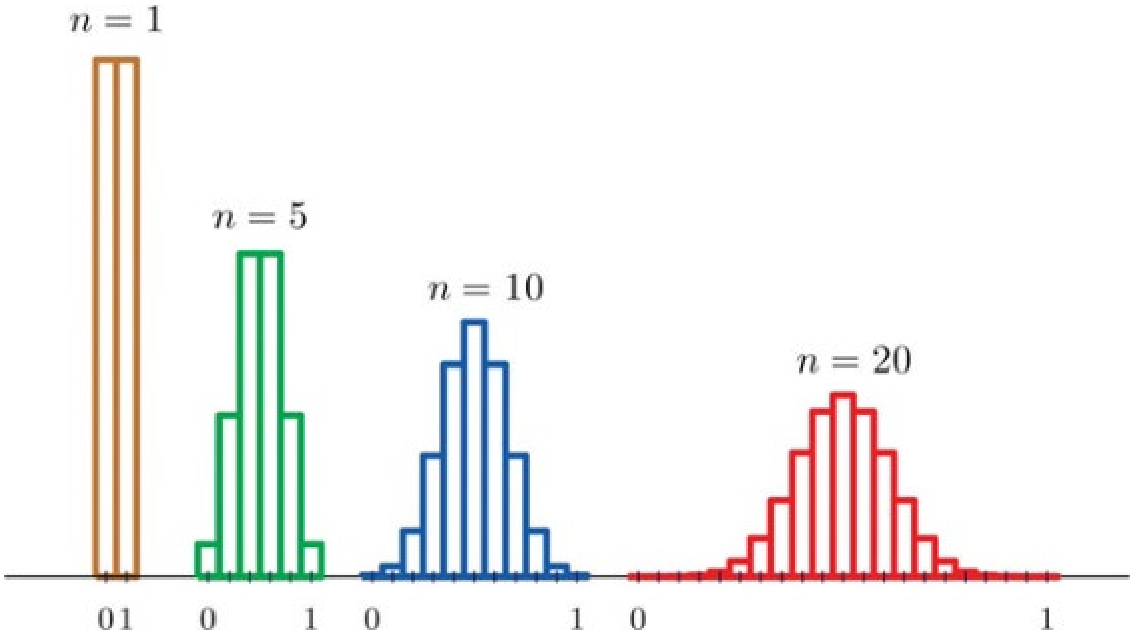
Review of Statistics



- An exercise in sampling distributions
- Roll 2 dies 10 times
  - Use: real dice, web e.g. https://www.random.org/dice, Excel randbetween(1,6) function, etc.
  - Compute average of the 2 dice
  - Plot number of occurrences (y-axis) for each dice roll value (x-axis)
- Do the same for 4 dies

Review of Statistics

- What do we see when the number of dice increases from 2 to 4 to even more?



Simply Psychology 2020

Review of Statistics

- Point of experiment?
- When N increases, i.e. we have a large sample
  - we obtain a nice smooth distribution
  - the distribution shape can be modeled
  - we can use this distribution to find systematic trends in data
  - when sample size is limited we run into problems
    - resulting decisions are less ideal

Data Sources for Safety Analysis

- Crash - police completed crash reports using Missouri Uniform Crash Report
- Traffic/exposure data - annual average daily traffic (AADT), MoDOT Transportation Management Systems
- Facility Data – roadway pavement, geometrics, median, shoulder, etc.
    - measures from aerial photographs using CAD
    - Automated Road Analyzer (ARAN) data

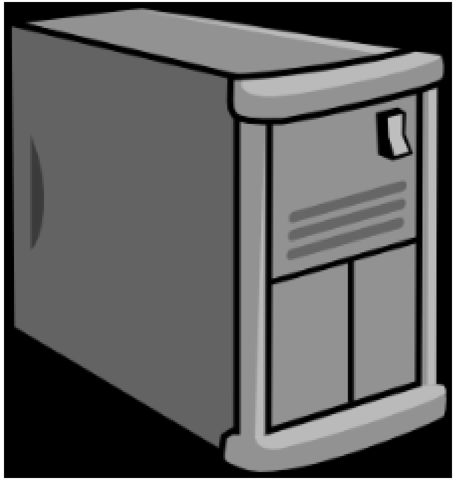
Data Sources for Safety Analysis

- Land-use – e.g. alcohol venues, bus stops
- Signal – signal timing charts, street view photo of signal & signage (e.g. permitted/protected)
- Weather

Code Nast Traveler 2020

Hierarchy of Data Sources

- Anecdotal information
  - e.g. comments from police, staff, citizens
- Use one database, e.g. crash
- Fusion of easily available databases
  - e.g. MoDOT TMS data + aerial photos



Hierarchy of Data Sources

- Fusion of all available databases
  - crash, road information, driver, vehicle, weather (NOAA), detector/probe, district signals, aerial photos, social media, TMC alerts, mass media
- Field measurement of current site conditions
  - e.g. accident reconstruction, total stations, GPS surveying + all available sources of info from databases
- Use what is practicably available to you considering resource and time

Hierarchy of Methodology

- Intuition/experience (not quantitative/data-driven)
- Naïve/simple before/after study
- Data-driven using HSM defaults
- Data-driven using HSM calibrated
- Data-driven using HSM calibrated + MO severity distributions
- MO-specific models

Applications of Data-Driven Safety

- Design exceptions
  - estimate the actual safety of designs vs. just the nominal safety
- Prioritizing safety improvements
  - how to choose sites or systematic improvements from a list of sites or alternatives

missouir-map.org 2020

Applications of Data-Driven Safety

- Traffic impact studies
  - consider not just traffic generation but safety impacts of development
- Safety design-build
  - utilize safety concepts in contracting, maximize return on investment

Typical Safety Prediction Model

- N = SPF x CMFs x C
- N – predicted crash frequency
- SPF – safety performance function
  - relates fundamental exposure variables to crashes
  - AADT and length (for segments)
- CMFs – crash modification factors
  - relate other variables to crashes
- C – calibration factor
  - calibrate safety model to local conditions

Final Thoughts
Practitioner Encouragement

- Not binary safety process
  - not let unavailability of some data or method stop you
    - missing aspects might not be very impactful overall, e.g. CMF examples (0.01% impact)
  - partial info better than no info, limited accuracy better than no data
  - no such thing as perfect safety modeling
    - limitations exist throughout safety modeling process
    - e.g. time, data quality and availability, engineering judgment/discretion

Final Thoughts
Practitioner Encouragement

- Justify your reasonable approach and document your approach
- Data-driven safety method is littered with engineering discretion
    - e.g. section length, def. of intersection crash, which CMFs to apply, calibration sample size
- Incorporate workshop material as time & opportunity allows

Final Thoughts
Practitioner Encouragement

- Reach out to MoDOT safety experts if you have questions
  - Others might have encountered the same issue before