Cross-sectional Studies
Part 2

January 19, 2021

Instructor: Srinivas Geedipally
Texas A&M Transportation Institute
Srinivas-g@tti.tamu.edu

Recap of Part 1

- Study types:
  - Observational
  - Experimental
- Data Types:
  - Time-series
  - Cross-sectional
  - Panel

Recap of Part 1

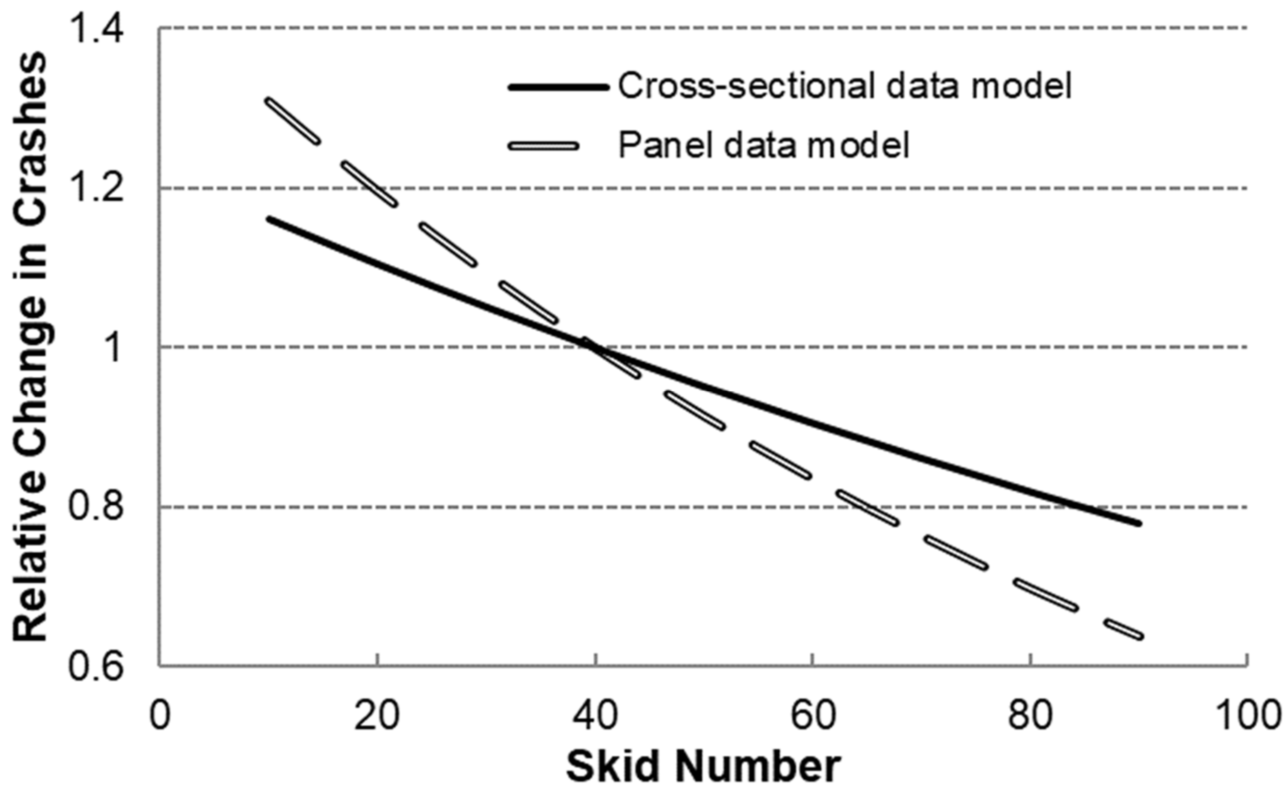
Time series models

TABLE 6.1 Appropriate regression model for time-series crash count data (Quddus, 2018).

Aggregation level	Sample mean	Recommended model
Highly aggregated	>50	ARIMA
Disaggregated	10–20	Poisson INAR(1), NBINGARCH, or GLARMA
Highly disaggregated	<10	NBINGARCH, or GLARMA

Recap of Part 1

Cross-sectional and Panel data models



Recap of Part 1

- Data and modeling issues
- Data aggregation

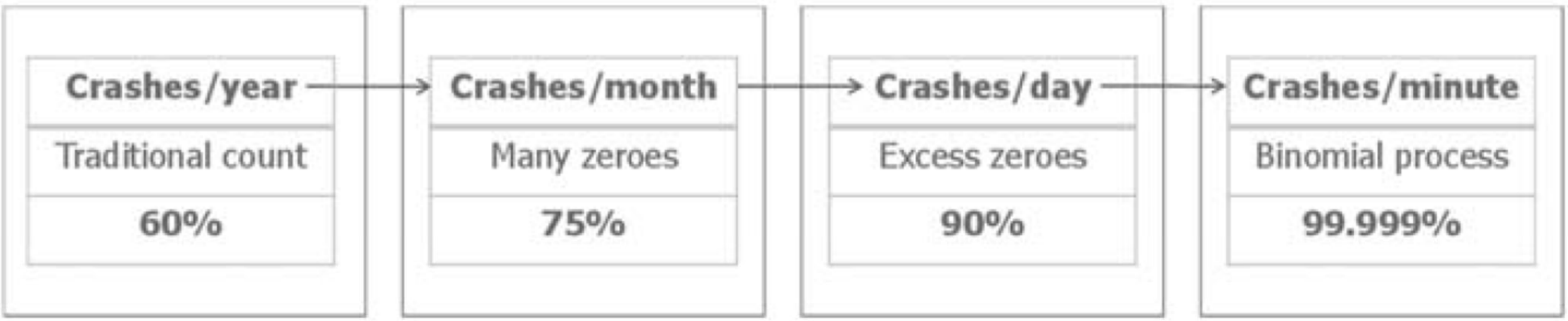


FIGURE 6.1 Percentage of zero responses when changing the time scale (Lord and Geedipally, 2018).

Functional Forms

- Most common assumption – linear relationship.

$$\mu_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}).$$

  - No logical reason, except for simplicity.
  - May not truly represent the complexity of the process.

- Nonlinear functions better characterize the relationships.

$$\mu_i = \exp(\beta_0 + \beta_1 LN(F)) \times CMF_R \times \cdots$$

$$CMF_R = 1 + \beta_2 (0.147V)^4 \frac{(1.47V)^2}{32.2R^2}$$

Functional Forms

TABLE 6.2 Functional form for different variables (Hauer, 2015).

Exposure variables	Influential variables
1 Power: X^{β_1}	**5** Exponential: $e^{\beta_1 X}$
2 Polynomial: $\beta_1 X + \beta_2 X^2 + \beta_3 X^3 \ldots$	**6** Linear: $1 + \beta_1 X$
3 Logistic: $1/(1+\beta_1 e^{\beta_2 X}) - 1/(1+\beta_1)$	**7** Quadratic: $1 + \beta_1 X + \beta_2 X^2$
4 Weibull: $1 - e^{-(X/\beta_1)^{\beta_2}}$	

$$\mu_i = e^{\beta_0} \times \mathrm{L} \times F^{\beta_1} \times CMF_R \times CMF_{LW} \times CMF_{SW} \times CMF_{SN}$$

$$CMF_R = 1 + \beta_2 (0.147V)^4 \frac{(1.47V)^2}{32.2R^2}$$

$$CMF_{LW} = e^{\beta_3 (LW-12)}$$

Functional Forms

The functional form can be known by developing a scatterplot.

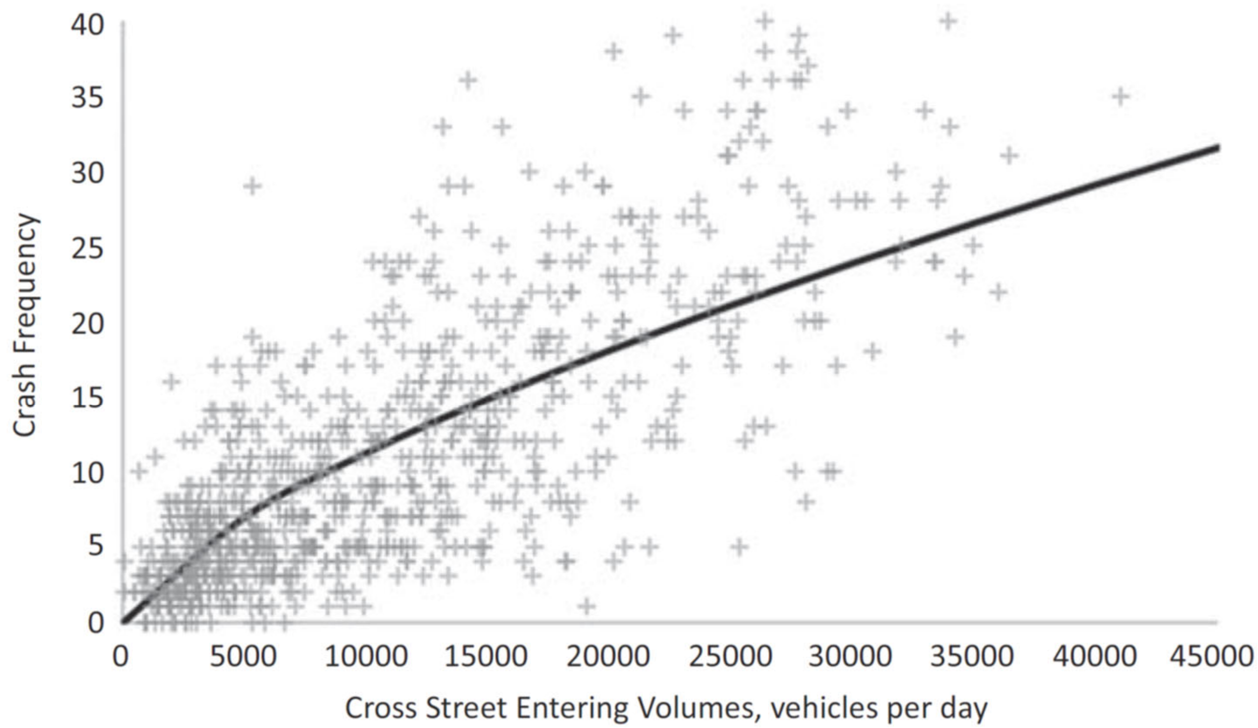


FIGURE 6.2 Relationship between cross street entering volumes and crash frequency.

Flow-only models

- Traffic flow is the only variable.
- Applicable for average conditions for each transportation element.
- Used when limited information about the geometric design features is available.
- Typically used in the network screening process to identify hazardous sites.

Functional Forms

Road Segments

$$\mu_{rs} = \beta_0 \times L \times AADT^{\beta_1}$$

$$\mu_{rs} = \beta_0 \times L^{\beta_1} \times AADT^{\beta_2}$$

Intersections

$$\mu_{int} = \beta_0 \times AADT_{maj}^{\beta_1} \times AADT_{min}^{\beta_2}$$

$$\mu_{int} = \beta_0 \times (AADT_{maj} + AADT_{min})^{\beta_1} \times \left(\frac{AADT_{min}}{AADT_{maj}}\right)^{\beta_2}$$

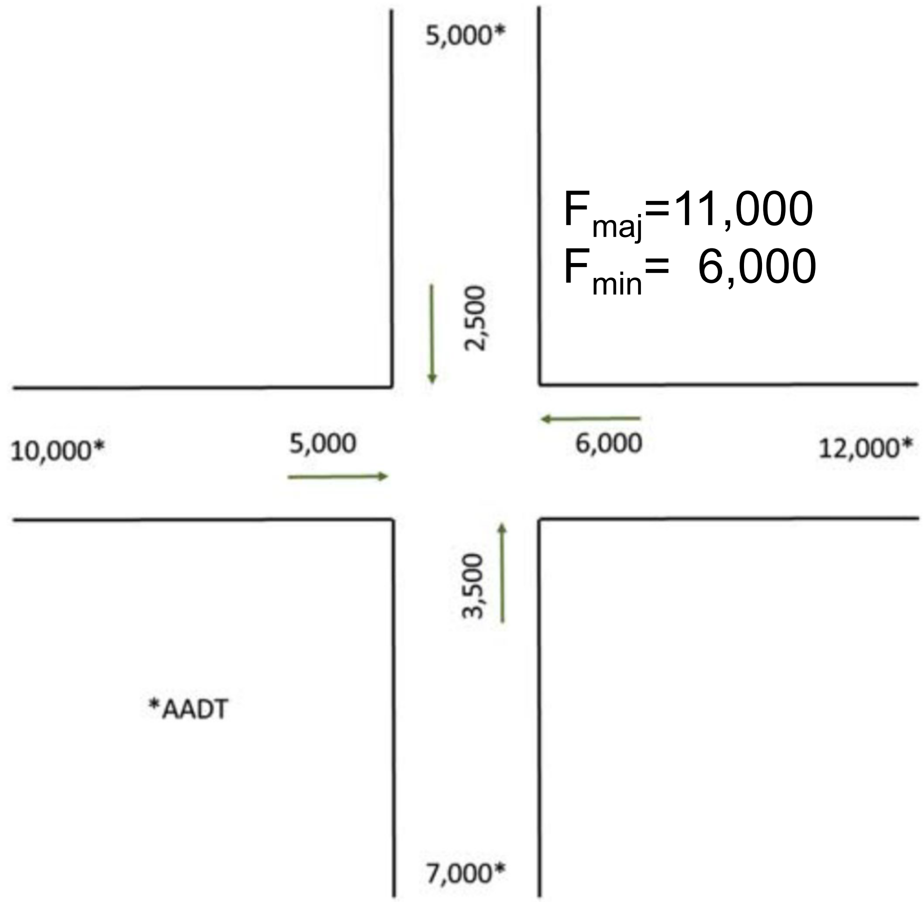
$$\mu_{int} = \beta_0 \times (AADT_{maj} + AADT_{min})^{\beta_1}$$

$$\mu_{int} = \beta_0 \times (AADT_{maj} \times AADT_{min})^{\beta_1}$$

$$\mu_{int} = \beta_0 \times AADT_{maj}^{\beta_1} \times AADT_{min}^{\beta_2} \times e^{\beta_3 \times AADT_{min}}$$

$$\mu_{int} = \left(AADT_{maj} \times e^{\beta_0 + \beta_1 \times AADT_{min}}\right) + \left(AADT_{min} \times e^{\beta_2 + \beta_3 \times AADT_{maj}}\right)$$

Functional Forms



F_{maj}=11,000
F_{min}= 6,000

FIGURE 2.2 Entering flows in vehicles per day (AADT).

Functional Forms

Flow-only models

SITEID	Crashes	LENGTH	AADT	LaneWID	ShoulderWID	Speed limit	Intersections	Horizontal curves
1	0	0.913	12660	12	8	55	6	2
2	2	0.613	9640	12	8	40	2	0
3	7	2.918	4020	13	2	55	4	0
4	5	0.814	4580	12	2	50	0	1
5	12	1.729	4580	11	2	55	2	2
6	0	0.2	6220	12	2	55	1	2
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
6362	6	0.43	4360	12	6	35	6	0

$$\mu_i = e^{-8.605} \times \mathrm{L} \times AADT^{1.11}$$

Functional Forms – Flow-only models with CMFs

- Data that represent a given set of baseline conditions.
- Reflect the nominal conditions agencies most often used (e.g., 12-ft lanes and 6-ft shoulders).
- Model is calibrated using a database that include only sites that have characteristics equal to base conditions.
- This functional form is the one that has been adopted by the Highway Safety Manual (AASHTO) and the FHWA among others.

Functional Forms – Flow-only models with CMFs

Example: HSM Chapter 11

The base conditions of the SPF for undivided roadway segments are:

- Lane width (LW): 12 feet
- Shoulder width: 6 feet
- Shoulder type: Paved
- Sideslopes: 1V:7H or flatter
- Lighting: None
- Automated speed enforcement: None

Functional Forms – Flow-only models with CMFs

SITEID	Crashes	LENGTH	AADT	LaneWID	ShoulderWID	Horizontal curves
1	2	0.613	9640	12	8	0
2	0	0.23	3480	12	8	0
3	6	1.352	4800	12	8	0
4	0	0.157	6100	12	8	0
5	1	0.652	4380	12	8	0
.	.	.	.	.	.	.
.	.	.	.	.	.	.
599	0	0.121	5340	12	8	0

Observations reduced from 6362 to 599

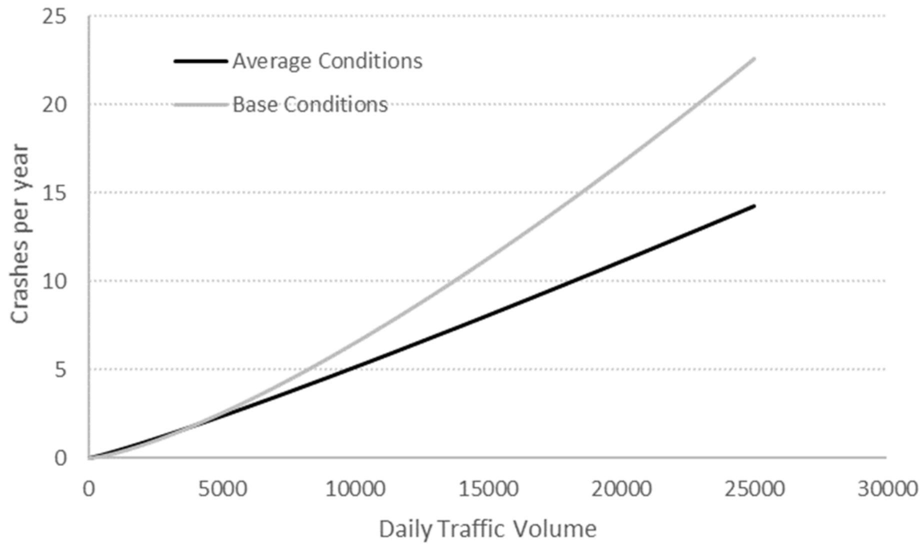
Functional Forms – Flow-only models with CMFs

Exercise 6.1

Variable	Flow-only Model for Average Conditions			Flow-only Model for Baseline Conditions		
	Estimate	Std. Error	*P*-value	Estimate	Std. Error	*P*-value
Intercept (β_0)	−8.605	0.433	<.0001	−10.635	1.332	<.0001
Ln(AADT) (β_1)	1.112	0.050	<.0001	1.358	0.157	<.0001
Dispersion	0.768	0.042	<.0001	0.836	0.150	<.0001
AIC	6362			599		

Functional Forms – Flow-only models with CMFs

Exercise 6.1



Functional Forms – Model with covariates

- Database contains each safety-related variable (e.g., lane width, median width) that has a representative range of values.
- Each variable is included in the model and their coefficients are calibrated using regression analysis.

Functional Forms – Model with covariates

$$\mu_i = e^{\beta_0} \times \mathrm{L} \times F^{\beta_1} \times CMF_R \times CMF_{LW} \times CMF_{SW} \times CMF_{SN}$$

$$CMF_R = 1 + \beta_2 (0.147V)^4 \frac{(1.47V)^2}{32.2R^2}$$

$$CMF_{LW} = e^{\beta_3 (LW-12)}$$

$$CMF_{SW} = e^{\beta_4 (SW-8)}$$

$$CMF_{SN} = e^{\beta_5 (SN-40)}$$

Variable Selection

Selection rules

- Forward
- Backward
- Bidirectional

Variable Selection

Variable	First Iteration		
	Estimate	Std. Error	p-value
Intercept (β_0)	−6.029	0.723	<.0001
Ln(AADT) (β_1)	1.047	0.046	<.0001
Lane Width (β_2)	−0.074	0.044	0.0922
Shoulder Width (β_3)	−0.029	0.007	<.0001
Railroad Crossing Presence (β_4)	−0.045	0.158	0.7763
Posted Speed Limit (β_5)	−0.024	0.003	<.0001
Minor Intersection Density (β_6)	0.037	0.005	<.0001
Horizontal Curve Density (β_7)	0.026	0.012	0.026
Dispersion	0.600	0.036	<.0001
AIC	6187		

Variable Selection

Variable	Second Iteration		
	Estimate	Std. Error	p-value
Intercept (β_0)	−6.036	0.723	<.0001
Ln(AADT) (β_1)	1.047	0.046	<.0001
Lane Width (β_2)	−0.073	0.044	0.0949
Shoulder Width (β_3)	−0.029	0.007	<.0001
Railroad Crossing Presence (β_4)			
Posted Speed Limit (β_5)	−0.024	0.003	<.0001
Minor Intersection Density (β_6)	0.037	0.005	<.0001
Horizontal Curve Density (β_7)	0.026	0.012	0.0262
Dispersion	0.600	0.036	<.0001
AIC	6185		

Variable Selection

Variable	Third Iteration		
	Estimate	Std. Error	p-value
Intercept (β_0)	-6.964	0.464	<.0001
Ln(AADT) (β_1)	1.054	0.046	<.0001
Lane Width (β_2)			
Shoulder Width (β_3)	-0.031	0.007	<.0001
Railroad Crossing Presence (β_4)			
Posted Speed Limit (β_5)	-0.024	0.003	<.0001
Minor Intersection Density (β_6)	0.036	0.005	<.0001
Horizontal Curve Density (β_7)	0.027	0.012	0.0247
Dispersion	0.603	0.036	
AIC	6186		

Crash Variance

$$Var(y) = \mu + \alpha\mu^2$$

$\alpha \rightarrow 0$ means the data are Poisson distributed

Confidence Intervals

$$\left[\frac{\hat{\mu}}{e^{1.96\sqrt{Var(\hat{\eta})}}}, \hat{\mu}e^{1.96\sqrt{Var(\hat{\eta})}}\right]$$

$$\left[0, \left\lfloor \hat{\mu} + \sqrt{19}\sqrt{\hat{\mu}^2 Var(\hat{\eta}) + \frac{\hat{\mu}^2 Var(\hat{\eta}) + \hat{\mu}^2}{\phi} + \hat{\mu}} \right\rfloor\right]$$

Sample Size

TABLE 6.5 Recommended minimum sample size for Bayesian Poisson-lognormal models (Miranda-Moreno et al., 2008).

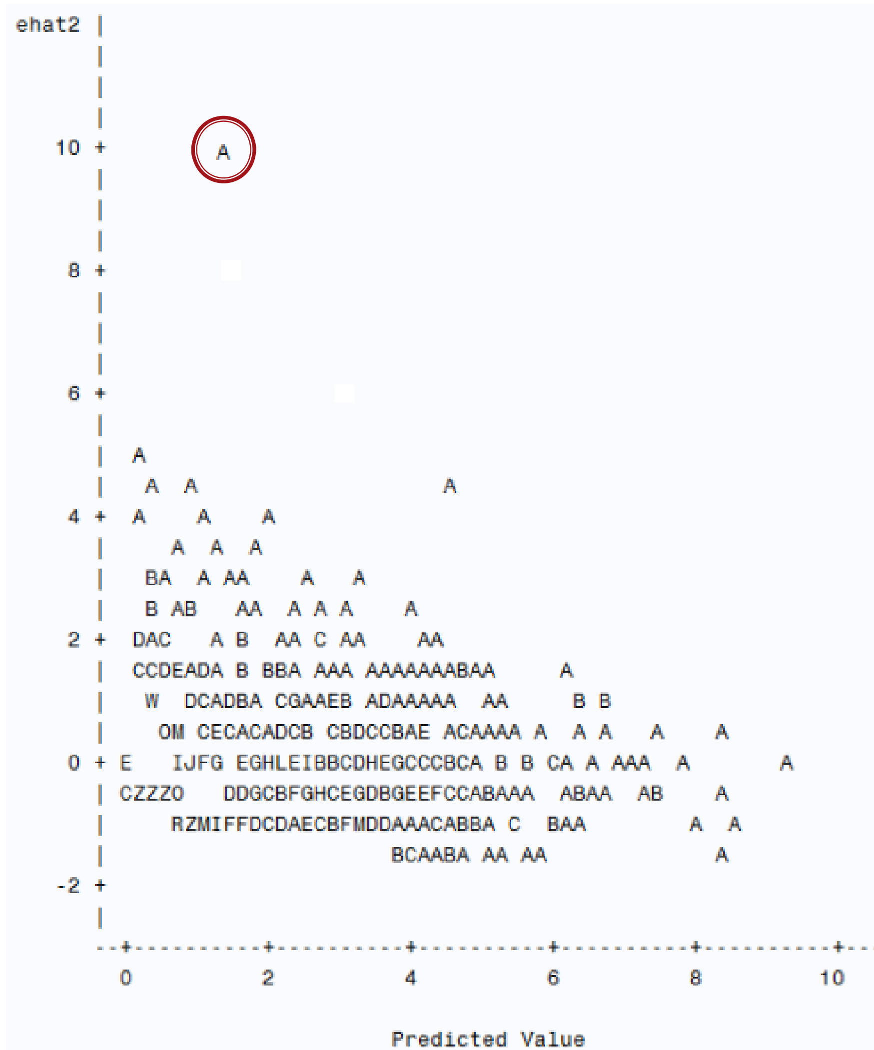
Population sample mean	Minimum sample size
≥2.00	20
1.00	100
0.75	500
0.50	1000
0.25	3000

Outlier Analysis

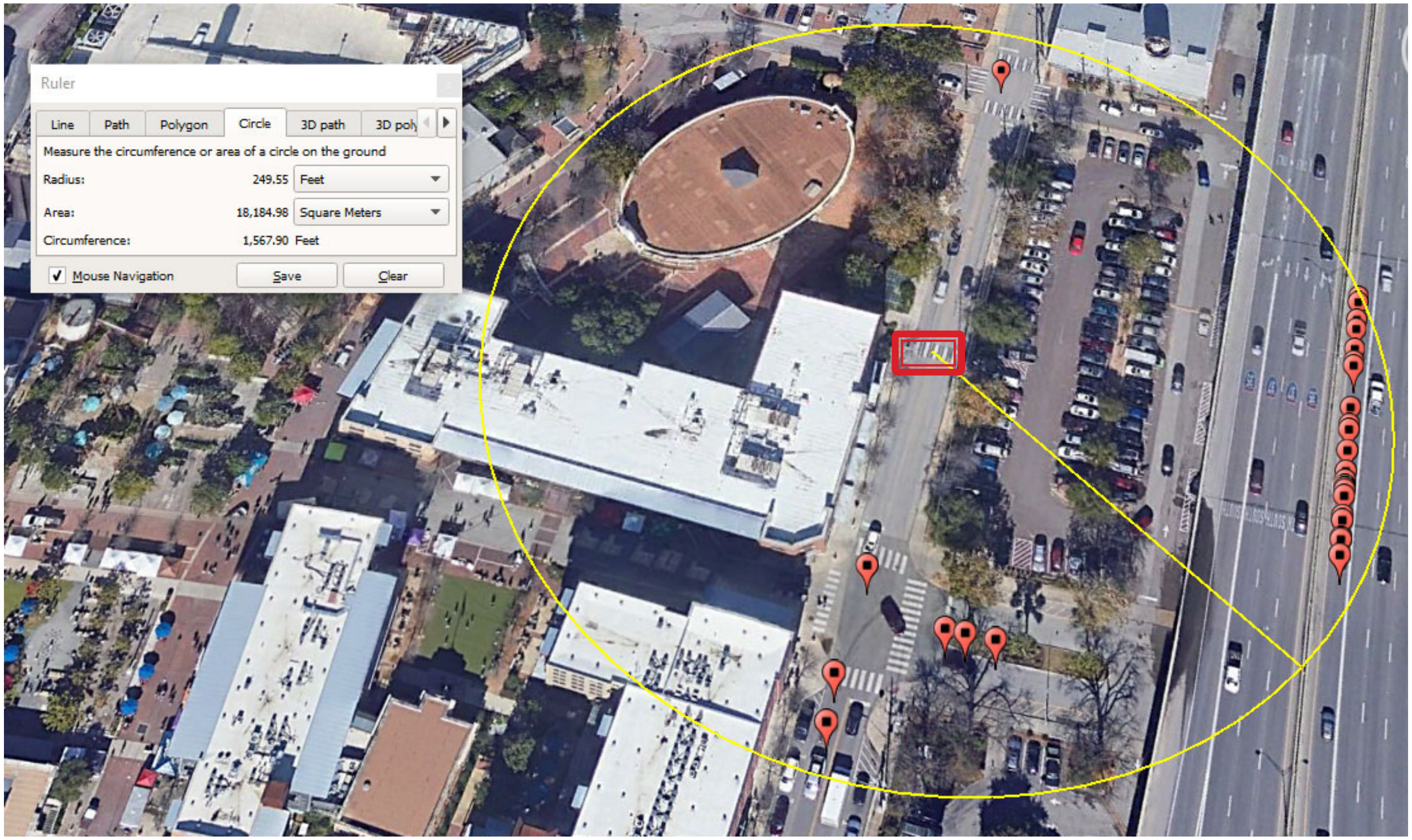
- Cook's distance
- Cumulative Residual (CURE) plots
- Standardized residuals

Outlier Analysis

Standardized residuals



Outlier Analysis



Study Designs

Observational study types

- Cross-sectional study
- Cohort study
- Case-control study

Other Study Types

- Cohort studies
  - Cohorts are first identified
  - Followed at intervals through time
  - Whitlock et al. (2003) investigated the relationship between driver injury and socioeconomic status
  - In 2004, they investigated relationship of injury with marital status

Other Study Types

- Case-control studies
  - Alternatives to before-after studies
  - Controls are often matched to cases
  - One-to-one matching
  - Frequency matching (e.g., four controls per one case)

Other Study Types

- Randomized control trials
  - Performed under controlled conditions
  - Experimental group has the intervention
  - Control group has placebo or no intervention

Thank you