

Index

Note: 'Page numbers followed by "f" indicate figures and "t" indicate tables.'

A

Adjacency-based measurement, 306
AICc. *See* Corrected Akaike information criterion (AICc)
Akaike information criterion (AIC), 46–47, 320
Alternate hypothesis (H_1), 152, 160, 171
Analysis of variance (ANOVA), 160
Annual average daily traffic (AADT), 27–28, 69b–71b, 335–336
Area under ROC curve, 294
Arithmetic mean, 136–137
Artificial intelligence (AI), 399
Artificial neural network (ANN), 416, 418
Association
 measures of, 144–148
 rules, 400–402
Asymmetrical distribution, 141
Autocorrelation, 180–181, 280, 319, 442–444
Automatic machine learning algorithm (AutoML), 422
Autonomous vehicles (AVs), 10, 395
Autoregressive integrated moving average model (ARIMA model), 181–182
Autoregressive model (AR model), 181
Average annual daily traffic (AADT), 173, 184
Average daily rail traffic (ADRT), 73b–74b
Average nearest neighbor (ANN), 279–280

B

Backward elimination rule, 202, 202b–204b
Balanced panel dataset, 185
Bar graphs, 165–166
Basic safety messages (BSMs), 396
Basic spatial unit (BSU), 312b–314b
Bayes factors, 48
Bayes information criterion (BIC), 47, 201–202

Bayes method, 40, 44, 232–233, 237–239, 276–277
Bayesian methods, 42–44, 232–239, 273–277, 347–348. *See also* Geostatistical methods
 Bayes method, 237–239, 276–277
 EB method, 233–237, 273–276
Bayesian networks (BN), 400, 412–413
Bayesian neural networks (BNN), 93–94, 422–423
Before–after studies, 219
 adjusting for site selection bias, 239–242
 Bayesian methods, 232–239
 with comparison groups, 229–232
 critical issues with, 220–223
 RTM, 220–221
 site selection bias, 222–223
 example, 231b–232b
 PSM method, 242–244
 sample size calculations, 246–255
 simple before–after study, 223–225
 using survival analysis, 244–246
Bernoulli trial, 20–21, 63
Bidirectional elimination, 202–204
Binary logistic regression model, 349–351
Binary logit models, 354
Binomial distribution, 21, 64, 86–87, 406
Bins, 145–146, 162–164
Blackspots. *See* Hazardous sites
Block maxima or minima (BM), 388–389, 391–394
Blood alcohol concentration (BAC), 163
Box plot. *See* Box-and-whisker plot
Box-and-whisker plot, 161–162
Bubble chart, 171–172

C

Calibration factor, 209–211
Capacity, 335
 drop, 355–357, 359
Car-following model, 337–338, 337f
100-car naturalistic study, 34
Case-control studies, 212–213
Causal relationships, 62–63
Cell transmission model (CTM), 355–357

- Cell transmission model (CTM) (*Continued*)
 simulation algorithm, 359–360
- Central limit theorem (CLT), 387–388
- Central tendency measures, 136–137, 141
- Centroid update, 342–346
- Chi-square test for independence, 145–146
- CIs. *See* Confidence intervals (CIs)
- Classification and regression trees
 methodology (CART methodology), 405–409
 GBT, 410–411
 RF, 409–410
- Cluster(s), 403
 analysis, 341, 404–405
 assignment step, 342–346
- Clustering analysis (CA), 172–173, 261, 278–281, 399, 403–405
 KC, 403
 LCC, 403–405
- Coefficient of determination R^2 , 50
- Coefficient of variation (CV), 140, 194, 282–283
- Cohort studies, 211–212
- Collision course, 374–376
- Complete spatial randomness (CSR), 278, 309–310
- Composite safety score, 267–268
- Conditional autoregressive function (CAR function), 324, 433–434
- Conditional expectation, 220–221
- Conditional logistic regression models, 351–354
- Conditional probability tables (CPTs), 412
- Confidence coefficient, 148–149
- Confidence intervals (CIs), 50–51, 148–151, 205–207
- Confounders, 190–191
- Confounding variables, 190–191, 349
- Contiguity weights, 306
- Contiguity-based measurement. *See* Adjacency-based measurement
- Continuous distribution. *See* Gamma model
- Continuous risk profile method (CRP method), 283–285
- Contour plot, 173
- Control units, 242–243
- Convolutional neural networks (CNNs), 419–420
- Conway–Maxwell–Poisson model (COM-Poisson model), 72–74
 computing codes, 472
- Corrected Akaike information criterion (AICc), 330–332
- Courant–Friedrichs–Lewy condition (CFL condition), 359
- Crash modification factors (CMFs), 195, 244
- Crash prediction models (CPMs), 37, 268, 335–336
- Crash-frequency models/modeling, 7, 268, 442–444, 446t
 applications of, 60–63
 causal relationships, 62–63
 prediction, 62
 screening variables, 61
 sensitivity of variables, 61–62
 understanding relationships, 61
- confidence intervals, 205–207
 count models, 65–71
 crash variance, 205–207
 finite mixture models, 75–76
 functional form, 194–201
 generalized count models, 71–75
 model selection, 94–96
 model transferability, 209–211
 multi-distribution models, 77–82
 multivariate models, 76–77
 nomenclature for, 60
 outlier analysis, 208–209
 sample size determination, 207
 semi-and nonparametric models, 85–94
 sources of dispersion, 63–65
 unobserved heterogeneity, 82–85
 variable selection, 201–204
- Crash-injury severities, 451t–452t
- Crash-severity models/modeling, 7, 103, 444–447
 characteristics of crash injury severity data, 104–105
 confidence intervals, 205–207
 crash variance, 205–207
 functional form, 194–201
 model interpretation, 130–131
 model transferability, 209–211
 as ordered discrete outcome, 119–130
 outlier analysis, 208–209
 random utility model, 105–107
 sample size determination, 207
 as unordered discrete outcome, 107–119
 variable selection, 201–204
- Crash(s), 1–2, 17
 concentration location methods, 282–285

- CRP method, 283–285
 - peak searching method, 282–283
 - sliding window method, 282
- counts, 7, 41, 59, 76, 85, 103, 182, 444
- data, 6, 23–24, 25t–26t, 319
- frequency method, 261–262
- hazard, 244–245
- hot spots, 312b–314b
- injury severity data, 104–105
- modeling, 360–361, 441–442, 449
- occurrence, 268, 285
- predicting imminent crash likelihood, 346–348
- prediction, 361–363
- process, 18–22
- rate method, 262–263
- variance, 205–207
- Critical value, 153–155, 163
- Cross-classified random effects modeling (CCREM), 82–83
- Cross-K function, 317–318
- Cross-sectional data, 183–184
- Cross-sectional study
 - crash-frequency and crash-severity models, 194–211
 - data
 - aggregation, 193–194
 - and modeling issues, 188–193
 - types, 180–187
- Cross-traffic conflict, 373
- Cross-validation, 312
- Crowdsourcing, 35
 - data, 35
- Cumulative logit model. *See* Ordinal logistic model
- Cumulative residuals (CURE), 50–51, 196–197
- D**
- Data
 - 4-stage modeling framework, 37–44
 - aggregation, 193–194
 - assembling data, 36–37
 - assembly, 36–37
 - collection, 17
 - crash process, 18–22
 - cross-sectional, 183–184
 - evaluating model performance, 44–51
 - error-based methods, 48–51
 - likelihood-based methods, 45–48
 - heuristic methods for model selection, 51–54
 - mining, 9, 399–400
 - CA, 403–405
 - decision tree model, 405–411
 - and modeling issues, 188–193
 - endogenous variables, 191
 - omitted variables bias, 190–191
 - overdispersion, 188–189
 - sample mean and size, 189
 - underdispersion, 188–189
 - underreporting, 189–190
 - unobserved heterogeneity, 192–193
- models, 300
- panel, 184–187
- sources of data and procedures, 22–36
 - data issues, 36
 - disruptive technological and crowdsourcing data, 35
 - naturalistic driving data, 31–35
 - traditional data, 23–31
 - time-series, 180–183
 - types, 180–187
- Data integration. *See* Data assembly
- Deceleration rate to avoid collision (DRAC), 381–382
- Decision errors, 152–153
- Decision tree model, 405–411
 - CART methodology, 405–409
- Degrees of freedom (df), 145–146
- Dependent samples, 155
- Deterministic integration, 36–37
- Deviance, 48
- Deviance information criterion (DIC), 47, 327–328
- Directed acyclic graph (DAG), 412
- Directional test, 153–154
- Dirichlet process models (DP models), 88–93
- Dispersion
 - parameter, 66
 - sources of, 63–65
- Disruptive technology, 35
- Distance decay models, 306–308
- Distance-band weights, 306
- Distribution-free, 85
- Double Poisson model, 74
- Driving while intoxicated (DWI), 20
- E**
- Elasticity, 131
- Empirical Bayes method (EB method), 40, 200–201, 220, 233–237, 273–276
- Endogenous variables, 191

Equi-dispersion, 63
 Equivalent property damage only method (EPDO method), 264–265
 Error bars, 167
 Error-based methods, 48–51
 Euclidean distance, 279–280, 315, 403
 Evolutionary Monte Carlo training
 algorithm (EMC training algorithm), 423
 Expectation-maximization algorithm (EM algorithm), 403
 Exploratory data analyses, 135
 graphical techniques, 161–176
 quantitative techniques, 136–161
 External factors, 229
 Extreme value distributions, 107–108
 Extreme value models (EVM), 383–387
 Extreme value theory (EVT), 383–387

F

False identification test, 292–294
 False negative errors. *See* Type II error
 False negative rate (FNR), 293
 False positive errors. *See* Type I error
 False positive rate (FPR), 293
 Fatal injury, 23–24
 Fatality Analysis Reporting System (FARS), 163
 Federal Highway Administration (FHWA), 4, 371–372
 Feed-forward neural network (FNN), 416
 Field survey of traffic conflicts, 373–374
 Finite Mixture Negative Binomial (FMNB), 75
 First-order process, 309–310
 Flow-only models, 198–199
 with CMFs, 199–200
 Focus crash types, 287
 Focus facility types, 287
 Full Bayes model (FB model), 40, 220, 273–276
 Full information maximum likelihood (FIML), 113–115
 Fundamental diagram (FD), 338
 calibration, 357–359

G

Gamma model, 74
 Gamma-count model, 74
 Gaussian function, 330

Gaussian quadrature, 88
 General flow-only models. *See* Flow-only models
 General Motors model (GM model), 336–337
 General Motors Research Laboratory (GMR), 372
 Generalized additive models (GAMs), 86
 Generalized count models, 71–75
 COM-Poisson model, 72–74
 Generalized estimating equations (GEE), 42
 Generalized event count, 75
 Generalized exponential distribution (GE distribution), 81
 Generalized extreme value model (GEV model), 111–112, 387–388
 block maxima using, 388–389
 Generalized linear autoregressive and moving average (GLARMA), 182
 Generalized linear latent and mixed models (GLLAMs), 85
 Generalized linear mixed model (GLMM), 323–324
 Generalized linear model (GLM), 72–73, 323–328
 Generalized ordered logistic model (gologit), 124
 Generalized Pareto distribution (GP distribution), 387–388
 POT using, 389–391
 Geographic information system (GIS), 36–37, 299
 Geographically weighted negative binomial regression model (GWNBR model), 329
 Geographically weighted Poisson regression (GWPR), 329
 Geographically weighted regression (GWR), 307
 Geostatistical methods, 278–281
 clustering methods, 278–281
 KDE, 281
 Getis–Ord General G* (G* (d)), 301–302
 Getis–Ord G* statistics (G*(d) statistics), 281
 Gibbs sampling method, 436–437
 Gini impurity, 405–406
 Goodness-of-fit (GOF), 44, 201–202
 Goodness-of-fit (GOL), 44, 94–95, 197
 Gradient boosting trees (GBT), 409–411

Gramian Difference Angular Field (GDAP), 420

Graphical techniques, 161–176. *See also* Quantitative techniques

- bar graphs, 165–166
- box-and-whisker plot, 161–162
- bubble chart, 171–172
- contour plot, 173
- error bars, 167
- heatmap, 172–173
- histogram, 162–164
- pie charts, 168
- population pyramid, 174–176
- radar/web plot, 172
- scatterplots, 168–169

Gross domestic product (GDP), 2–3

H

Haddon Matrix, 19, 20t

Hausman test, 192

Hazardous sites, 259

- Bayesian methods, 273–277
- combined criteria, 277–278
- crash concentration location methods, 282–285
- evaluating site selection methods, 288–295
- geostatistical methods, 278–281
- observed crash methods, 261–268
- predicted crash methods, 268–272
- proactive methods, 285–288

Heatmap, 172–173

Heuristic methods for model selection, 51–54

Hierarchical Bayesian model (HBM), 323–328

Hierarchical NB model (HNB model), 43

Highway capacity manual (HCM), 335

Highway safety, 4–5

- analyses, 7–9
- initiatives, 4
- research, 213

Highway Safety Manual (HSM), 4–5, 37, 195, 260, 369

Hill climbing (HC), 412

Histogram, 162–164, 310

Hotspots. *See* Hazardous sites

Hyper-Poisson (hP), 75

Hypothesis testing, 152–161

I

Integer-valued autoregressive (INAR), 182

Interpolation algorithm, 360b

Interquartile range (IQR), 138

Irrelevant and independent alternatives (IIA), 103–104, 119

K

K-means

- algorithm, 343b
- cluster analysis, 341–342

K-means clustering (KC), 278–279, 403

k-nearest neighbors approach (k-NN approach), 361–362, 362b–363b

Kernel density, 164, 311–312

Kernel density estimation (KDE), 281, 307, 310–314

Kurtosis, 143–144

L

Lag, 180–181

Lagged response model. *See* Spatial autoregressive model (SAR model)

Latent class clustering (LCC), 403–405

Leptokurtic distribution, 143–144

Level of service (LOS), 335

Level of service of safety method (LOSS method), 271–272, 339

Likelihood function, 245–246, 429

Likelihood-based methods, 40–42, 45–48

- AIC, 46–47
- Bayes factors, 48
- BIC, 47
- deviance, 48–51
- DIC, 47
- likelihood ratio index, 46
- likelihood ratio test, 45–46
- MLE method, 45
- WAIC, 47

Line of Best Fit, 169

Local G* (d), 303–304

Local indicators of spatial association (LISA), 303

Local Moran's I_i , 304–305

Log-likelihood (LL), 41, 432

- statistic, 412

Log-likelihood ratio (LR), 201–202

Long short-term memory (LSTM), 420–422

Longitudinal data. *See* Panel data

M

Machine learning, 9, 399–400

Manual on Uniform Traffic Control Device (MUTCD), 222–223

- Margin of error, 149
 - Marginal effect, 130
 - Markov Chain Monte Carlo methods (MCMC methods), 40, 69b–71b, 233, 434
 - estimation, 435
 - Poisson-gamma model, 435–437
 - Poisson-gamma-CAR model, 438–439
 - Matched pairs, 155
 - Matched samples. *See* Dependent samples
 - Mathematical expectation. *See* Mean
 - Maximum available deceleration rate (MADR), 381–382
 - Maximum likelihood estimation method (MLE method), 39, 45, 69b–71b, 434–435
 - Maximum-likelihood (ML), 403
 - Maximum-posterior (MAP), 403
 - McFadden R^2 index, 46
 - Mean. *See* Arithmetic mean
 - Mean absolute deviance (MAD), 49
 - Mean prediction bias (MPB), 49
 - Mean response, 206–207
 - Mean squared error (MSE), 49
 - Mean squared prediction error (MSPE), 49
 - Median, 137
 - Mesokurtic distribution, 143–144
 - Method consistency test (MCT), 291
 - Metropolis-Hastings algorithm (MH algorithm), 437
 - Miss Rate. *See* False negative rate (FNR)
 - Mixed effect model, 325–326
 - Mixed function, 116
 - Mixed logit model (ML model), 107–108, 116–119
 - Mixing distribution, 116
 - Mobility, 335
 - Mode, 137
 - Model transferability, 209–211
 - Modeling space between vehicles, 336–338
 - “Moderated” causal relationship, 195
 - Modified time to collision (MTTC), 377–378
 - Monte Carlo simulation (MC simulation), 85, 316
 - Moran’s I index, 280–281, 302–303
 - Mosaic plot, 166
 - Moving average model (MA model), 181
 - Multi-distribution models, 77–82
 - NB-L model, 78–81
 - Multidimensional data. *See* Panel data
 - Multilayer perceptron (MLP), 93–94, 416
 - neural network, 416–419
 - Multilevel model. *See* Random-effects models (RE models)
 - Multinomial logit model (MNL model), 103, 108–111
 - computing codes, 473
 - Multiple regression model, 200–201
 - Multivariate models, 76–77
 - Multivariate normal, 433
- ## N
- Naïve method, 252–253
 - National Academy of Sciences (NAS), 34
 - National Highway Transportation Safety Agency (NHTSA), 4, 18
 - Naturalistic driving data, 31–35
 - Naturalistic driving study (NDS), 396
 - Nearest neighborhood clustering, 279–280
 - Negative binomial (NB)
 - computing codes, 467–468
 - random effects, 469–470
 - random parameters, 470–471
 - with varying dispersion parameter, 468–469
 - distribution, 233
 - model, 9–10, 65–66, 193–194
 - Poisson-gamma model, 429–433
 - estimation methods, 434–439
 - with spatial interaction, 433–434
 - probability density and likelihood functions, 429
 - regression models, 429
 - Negative Binomial Integer-valued Generalized Autoregressive Conditional Heteroscedastic model (NBINGARCH model), 182
 - Negative binomial-Crack model (NB-CR model), 81–82
 - Negative binomial-generalized estimate/exponential model (NB-GE model), 81, 192–193
 - Negative Binomial–Lindley model (NB-L model), 53–54, 78–81, 192–193
 - computing codes, 472
 - Nested logit model (NL model), 107–108, 111–115
 - computing codes, 473–474
 - Network cross-K function, 317
 - Network screening, 259–260
 - Neural network, 416–423

- BNN, 422–423
- CNNs, 419–420
- LSTM, 420–422
- MLP neural network, 416–419
- RNN, 420–422
- Nondirectional test. *See* Two-tailed test
- Nonparametric models, 93–94, 145
- Null hypothesis (H_0), 152
- Null value, 153–154
- O**
- Observed crash methods, 261–268. *See also*
 - Predicted crash methods
 - composite safety score, 267–268
 - crash frequency method, 261–262
 - crash rate method, 262–263
 - EPDO method, 264–265
 - RQC method, 263–264
 - SI, 266–267
 - strengths and limitations of, 269t–270t
- Odds ratio, 108, 147–148, 350–351
- Omitted variables bias, 190–191
- One-tailed hypothesis test, 153–154
- One-tailed test, 153–154
- Opposing left-turn conflict, 373
- Ordered probit model (OP model), 426
- Ordinal logistic model, 121
- Ordinal logit/probit model, 120–124
- Outlier analysis, 208–209
- Over-dispersion parameter, 66
- Overdispersion, 63–64, 188–189. *See also* dispersion
- P**
- Paired samples. *See* Dependent samples
- Paired *t*-test, 156
- Panel data, 184–187
- Panel study
 - crash-frequency and crash-severity models, 194–211
 - data
 - aggregation, 193–194
 - and modeling issues, 188–193
 - types, 180–187
- Partial proportional odds model (PPO), 124–125
- Peak over threshold (POT), 388–394
- Peak searching method, 282–283
- Pearson *Chi*-square, 50
- Pearson's correlation coefficient, 144–145
- Pedestrian conflict, 373
- Percentile, 138
- Pie charts, 168
- Planar K-function, 316
- Point data analysis, 309–318
 - crash hot spots, 312b–314b
 - cross-K function, 317–318
 - first-and second-order process, 309–310
 - KDE, 310–314
 - Ripley's K-function, 314–316
- Point patterns, 309–310
- Poisson mean, 206–207
- Poisson mean differences (PMD), 294–295
- Poisson model, 65
- Poisson trials, 21–22, 63
- Poisson-gamma model, 429–433
 - NB-1 model, 432–433
 - NB-2 model, 430–432
 - with spatial interaction, 433–434
- Poisson-Inverse Gamma, 71
- Poisson-Inverse Gaussian (PIG), 71
- Poisson-lognormal model (PLN model), 53–54, 66–71
 - computing codes, 471–472
- Poisson-mixture models, 71
- Poisson–Dirichlet Process (P-DP), 91
- Poisson–Tweedie distribution models, 71
- Poisson–Weibull distribution model, 71
- Pooling, 419–420
- Population pyramid, 174–176
- Post encroachment time (PET), 371
- Potential for improvement (PI), 268–271
- Predicted crash methods, 268–272. *See also*
 - Observed crash methods
 - LOSS method, 271–272
 - PI, 268–271
 - strengths and limitations of, 273t
- Prediction, 62
- Prediction intervals (PIs), 206–207
- Presence of commercial area (PCA), 73b–74b
- Presence of guide (PG), 73b–74b
- Presence of speed hump (PSH), 73b–74b
- Presence of track circuit controller (PTCC), 73b–74b
- Proactive methods, 285–288
 - focus crash types, 287
 - focus facility types, 287
 - risk factors development, 287
 - screen and prioritize candidate locations, 288
- Probability density function (PDF), 86–87, 310, 348, 429

Probability mass function (PMF), 66, 430–431
 Probability of false alarm. *See* False positive rate (FPR)
 Propensity score matching method (PSM method), 242–244
 Propensity score method (PS method), 220
 Property damage only (PDO), 108
 Proportion of stopping distance (PSD), 381
 Prospective cohort study, 212

Q

Quantile regression (QR), 442–444
 Quantitative techniques, 136–161. *See also*
 Graphical techniques
 confidence intervals, 148–151
 hypothesis testing, 152–161
 measures of
 association, 144–148
 central tendency, 136–137
 variability, 137–144
 Quartiles range, 138

R

R 3.5.0 package, 422
 Radar/web plot, 172
 Radial basis function (RBF), 425
 Random forest (RF), 409–410
 Random parameters logit model. *See* Mixed logit model (ML model)
 Random utility model, 105–107
 Random-effects models (RE models), 82–83
 Random-parameters models (RP models), 83–85, 192
 Random-parameters NB-L model (RPNB-L model), 80
 Randomized control trial (RCT), 213
 Range, 137–138
 Rate quality control method (RQC method), 263–264
 Real-time crash prediction models (RTCPM), 8–9, 336
 Real-time predictive analysis of crashes, 348–354
 binary logistic regression model, 349–351
 binary logit models, 354
 conditional logistic regression model, 351–353
 conditional logistic regression models, 354

Real-time traffic, characterizing crashes by, 340–346
 Receiver operating characteristic curve (ROC curve), 294
 Rectified linear unit (ReLU), 416
 Recurrent neural networks (RNN), 420–422
 Regression, 400
 tree, 405
 Regression-to-the-mean (RTM), 200–201, 220–221
 Rejection regions, 153
 Relative risk, 147–148
 Relative Severity Index (RSI), 267–268
 Relative standard deviation. *See* Coefficient of variation (CV)
 Resampling, 418–419
 Retrospective cohort study, 212
 Right-turn-on-red conflict (RTOR conflict), 373
 Ripley's K-function, 279, 314–316
 Risk, 294
 of collision at signalized intersections, 390f
 factors development, 287
 Risk ratio. *See* Relative risk
 Road Safety Manual (RSM), 4–5, 260
 Roadway data, 26, 26t
 Run-off-the-road events (ROR events), 167

S

Safety, 335
 as function of traffic flow, 338–340
 pyramid, 383
 safety-related data, 30–31
 Safety performance functions (SPFs), 37, 183, 195, 268, 339
 Safety Pilot Model Deployment program (SPMD program), 396
 Same-direction conflict, 373
 Sample mean, 64, 136–137, 189
 Sample size, 189
 calculations, 246–255
 factor influencing, 247–249
 using known crash counts, 249–252
 on variance and ratio, 252–255
 determination, 207
 Sampling distribution, 41, 139–140, 154
 Scatterplots, 168–169
 Schwarz Information Criterion. *See* Bayes information criterion (BIC)
 Screening variables, 61

- Seasonal ARIMA model (SARIMA model), 181–182
- Seasonality, 181
- Second-order process, 309–310
- Secondary conflict, 373
- Selectivity. *See* Specificity (SPEC)
- Seminonparametric Poisson model (SNP Poisson model), 86–88
- Semiparametric models, 86–88
- Sensitivity (SENS), 293, 336–337
analysis, 426
of variables, 61–62
- Sequential logistic/probit regression model, 128–130
- Severity index (SI), 266–267
- Severity of conflict, 389
- Sichel model (SI model), 71
- Significance level, 248
- Simultaneous Autoregressive function (SAR function), 433–434
- Site consistency test (SCT), 290–291
- Site selection
adjusting for, 239–242
bias, 222–223
methods, 288–295
- Skewed distributions. *See* Asymmetrical distribution
- Skewness, 141
- Sliding window method, 282
- Spatial Analysis on NETwork (SANET), 316
- Spatial association
global statistics for, 301–303
local indicators of, 303–305
measurement of, 300–305
spatial patterns of weather-related crashes, 305b
- Spatial autoregressive model (SAR model), 319–320
- Spatial correlation, 6
generalized linear model with, 323–328
positive, 278
- Spatial data
and data models, 300
measurement of spatial association, 300–305
point data analysis, 309–318
spatial regression analysis, 318–332
spatial weights and distance decay models, 306–308
- Spatial econometrics methods, 319–323
- Spatial error model (SEM), 319–323
- Spatial interaction, 433–434
Poisson-gamma model with, 433–434
- Spatial lag model. *See* Spatial autoregressive model (SAR model)
- Spatial random effect, 433
- Spatial regression analysis, 318–332
generalized linear model, 323–328
modeling local relationships, 328–332
spatial econometrics methods, 319–323
- Spatial weights, 306–308
- Spearman rank-order correlation coefficient, 145
- Specificity (SPEC), 294
- Stacked bar graphs, 166
- Standard deviation (SD), 138–140, 247
- Standard error (SE), 138–140, 149, 247
- 4-stage modeling framework, 37–44
computational techniques and tools, 40–44
developing models, 39
inferential goals determination, 39
modeling objective matrix, 37–38
- Stationarity, 181
- Stationary process, 310
- Statistical modeling, 130
- Subcritical zone, 339–340
- Sum of squared error (SSE), 405
- Supercritical zone, 339–340
- Supplemental data, 28–30
- Support vector machine (SVM), 93–94, 423–426
- Support vectors, 423, 425
- Surrogate Safety Assessment Model (SSAM), 394–395
- Surrogate safety measures
comparison of indicators, 384t–386t
field survey of traffic conflicts, 373–374
PET, 371
proximal, 374–383
safety risk pyramid, 387f
TCT, 371–372
theoretical development of, 383–394
traffic conflicts, 370–371
from traffic microsimulation models, 394–395
from video and emerging data sources, 395–396
- Survival analysis, 244–246
- Swedish Conflicts Technique, 383
- Symmetrical distribution, 141

Synthetic Minority Over-sampling
Technique algorithm (SMOTE
algorithm), 418–419

T

Time advantage (TAdv), 381
Time exposed time to collision (TET),
376–377
Time integrated time to collision (TIT),
376–377
Time measured to collision (TMTC), 376
Time to accident (TA), 376
Time to collision (TTC), 371, 376–380
Time-and distance-based proximal
surrogate safety measures, 376–383
encroachment time family, 380–381
other indicators, 381–383
PSD, 381
time to collision family, 376–380
Time-series data, 180–183
Total rank difference test (TRDT), 291–292
Total score test (TST), 292
Traffic analysis zones (TAZ), 309
Traffic conflicts, 370, 373, 396
Traffic conflicts technique (TCT), 370
Traffic crashes, 22, 137, 369, 418
Traffic density, 337, 339–340
Traffic flow data, 26–28
Traffic microsimulation models, 394–395
Traffic simulation to predicting crashes,
354–363
crash modeling, 360–361
crash prediction, 361–363
CTM, 355–357
CTM simulation algorithm, 359–360
FD calibration, 357–359
interpolation algorithm, 360b
Train detector distance (mile) (TDD),
73b–74b
Transitional zone, 339–340
Trend Line, 169

True negative rate. *See* Specificity (SPEC)
True Poisson means (TPMs), 294–295
True positive rate (TPR). *See* Sensitivity
(SENS)
Two-tailed hypothesis test, 153
Two-tailed test, 153
Type I error, 248
Type II error, 248

U

Unbalanced panel dataset, 185
Underdispersion, 63–65, 188–189
generalized count models for, 71–75
observed, 188–189
Underreporting, 189–190
Unobserved heterogeneity, 82–85, 105,
188, 192–193
RE models, 82–83
RP models, 83–85
Utility function, 105–106
Utility maximization, 106

V

Variability, measures of, 137–144
Variable selection, 201–204
Variance, 138–140
reduction, 406
variance-inflation-factors, 62
Vector data, 300
Vehicle-miles traveled (VMT), 169, 287

W

Weight factor, 205, 234–235
Widely applicable information criterion
(WAIC), 47

Z

z statistic, 154
Zero-centered models, 77–78
Zero-inflated models (ZI models), 76,
94–95