Index

NUMBERS

16PF (Sixteen Personality Factors Questionnaire), 421
911 calls case study, 194–195. *See also* Geospatial Association Rules
  broken-windows theory of crime, 195
  problem-oriented policing, 195

A

A Practical Guide to Support Vector Classification (Hsu, Chang, and Lin), 462–463
Abbott, Dean, *Applied Predictive Analytics*, 337–338
absolute percent, mean absolute percent, 344–349
Advanced Options panel, 52
AGGREGATE command, 169–170
aggregation, datasets, 410–412
ALSCAL, 253
AMOS, 2
  Baseline Comparisons, TLI, 33–35
  case study and, 4–5
  causation, direction, 37–40
  Chi-Square tests, 28–31
  CMIN, 33–34
  CMIN/DF, 35
  diagrams, 28
  factor analysis, 26–27
  general model, 29–30
    S/N, 31–35
  goodness of fit, 28–29, 31
  Graphics, 26–27
  GUI, 3, 27
  Hoelter results, 34–35
  icons, 29
  model, 27
  Model Fit Summary, 33–34
  null hypothesis, 28–29
  PCLOSE, 35
  RMSEA, 29, 35
    results, 34
  Tools and Plugins, 30–31
  User's Guide, 3–4
AMOS PSAT data.sav file, 4
An Introduction to Statistical Learning with Applications in R (James, Hastie, and Tibshirani), 462
Analysis node, 295–297
Analyze menu, 23, 47
ANCOVA (Analysis of Covariance), 4, 8–9
  multivariate tests, 25–26
  univariate analysis of variance, 13

COPYRIGHTED MATERIAL

ANN (artificial neural network), 272, 286, 325, 326–333
classification, Titanic dataset and, 349–353
Error Backpropagation, 337–339
hidden layers, 332–333
interactions, 331–332
learning, 337–338
mean absolute percent, 344–349
MLP (multilayer perception), 328
output, weights and, 334–335, 337–339
Parameter Estimates, 334, 335
perceptrons, 327
multilayered, 343
regression, 347
bank salary case study, 341–353
output, 342
slope, 335
XOR and, 329–330
topology, 333–334
complex neural net, 348
XOR
Nominal Output Variable, 336
truth table, 328
variables and, 333–341
Anomaly Case Index list, 311
Anomaly Index Summary table, 314
ANOVA (Analysis of Variance), 10–11
data mining results, 281
techniques, 4–5
versus SEM via AMOS, 4
ANOVA table, categorical regression, 94
Applied Predictive Analytics (Abbott), 337–338
arrow case study, GPL and, 163–167
Arrows on a Reference map, 188–191
Asymptomatic Significance, 360
automated linear modeling
regression, 72

B

bar charts
case study, 138–143
Legacy Dialogs, 132–133
mapping and, 174
bars on maps, 184–185
BEGIN GPL ... END GPL command, 151
Berry, Michael J.A., *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management,* 326
bias, KNN and, 386–388
binary logistic regression, 72, 284
Binary Logistic Regression Model, 47
Binning summary table, 322–232
Bivariate Correlations procedure, 422, 432–433
Bohrnstedt and Felson, 37
Bonferroni method, 282
bootstrapping, 43–44, 59–62
Bourdieu, Pierre, 218
Distinction: A Social Critique of the Judgement of Taste, 221
BR (Barrodale-Roberts) estimation method, 454
Breiman, Leo, *Classification and Regression Trees,* 368
broken-windows theory of crime, 195
bubble charts
bands, 159–160
case study, 143–145
Rosling, 137
color mapping, 158–159
GPL case study, 147–160
MBTI case study, 167–172
polygon and, 160

C

Case Processing Summary table, 82
categorical regression, 93–94
Identify Unusual Cases, 311
case sensitivity in datasets, 409–410
case studies
911 calls, 194–195
arrows, GPL and, 163–167
bank salary (ANN), 341–353
bar charts, 138–143
bubble chart/GPL, 147–160
bubble charts, 143–145

double regression lines, GPL and, 160–163
MBTI bubble chart, GPL and, 167–172
media distribution datasets, 404–420
Monte Carlo Simulation, 43
predicting shootings, 207–215
PSAT tests
learning styles and, 6
overview, 4–6
reading comprehension, 20
categorical regression, 86–87
dialogs
Categorical Regression, 87–88
Discretization, 89
Missing Values, 90
Options, 90
Output, 92
Plots, 93
Regularization, 91–92
Save, 92
output
ANOVA table, 94
Case Processing Summary table, 93–94
Coefficients table, 95
Correlations and Tolerance table, 95
Model Summary table, 94
Quantifications table, 96–97
transformation plot, 97–99
Categorical Variable Norms table, 313
causation, direction, 37–40
CHAID algorithm, Decision Tree building, 355–360
Chi-Square, 361–362
Crosstabs, 360–362
results comparison, 372
setting changes, 363–365
CHAID nugget, 296–297
Chang, C., *A Practical Guide to Support Vector Classification*, 462–463
Chart Builder, 132
Element Properties, 153
Gallery tab, 148
Groups/Point ID tab, 150
main menu, 133–134
Basic Elements submenu, 134
scatterplots
colored, 9
grouped, 149
Chart Options dialog, 57
charts
aesthetics, 137
bubble charts, 137
elements, 137
Probability Density, 56–57
edited, 58
regression lines, 10
scatter plots, colored, 9
as static diagrams, 137
Chihara, Laura M., *Mathematical Statistics with Resampling and R*, 63
Chi-Square, 28–31
CHAID algorithm, Decision Trees, 361–362
Correspondence Analysis, 222
Choropleth of Counts map, 175–179
Choropleth of Sums map, 182–184
Choropleth of Values map, 179–181
CISS (Campbell Interest and Skill Survey), 421
classical hypothesis testing, 280–283
Classification and Regression Trees (Breiman, Friedman, Stone, and Olshen), 368
CLASSWEIGHTS keyword, 467
Clementine. *See* Modeler
code, inefficient, 403–404
Coefficients table, categorical regression, 95
coeffiecients, linear mixed models and, 103
Cohen's d, 8
T-Tests and, 468–471
color mapping, bubble plot, 158–159
Command Syntax Reference guide, 151, 400–401
commands. *See also* extension commands
CROSSTABS, 402–403

DATASET ACTIVATE, 397
DATASET DECLARE, 411
DO REPEAT, 414–415
EXECUTE, 400
FREQUENCIES, 397
Syntax Editor, 399
GPL (Graphics Production Language)
AGGREGATE, 169–170
BEGIN GPL ... END GPL, 151
GGRAPH, 151, 169–170
GRAPHDATASET, 151, 169–170
GRAPHSPEC, 151
statement commands, 152
IBM SPSS Statistics Command Syntax Reference, 443
INDEX, 410
LOWER( ), 409–410
MERGE, 415–417
subcommands, ORDER, 398
UPPER( ), 409–410
COMPUTE statement, 415
Coordinates on a Choropleth of Counts map, 187–188
Coordinates on a Reference map, 186–187
Correlations and Tolerance table, 95
Correlations panel, 50, 51
Correlations table, 55, 56, 268–269
Correspondence Analysis, 217
Chi-Square, 222
Cramer's V, 222
Crosstabs menu, 220–223
Define Range submenus, 224
dimension reduction, 225–229, 235
dimension summary, 227, 232
dimensions, 220
covariance, Estimates of Covariance Parameters table, 115–116
Covariates, 11
Cox Regression Model, 47
Cramer's V, 222
CROSSTABS command, 402–403
Crosstabs menu, Correspondence Analysis, 220–223
cross-validation, V-fold cross-validation, 387
CRT algorithm
binary split, 368
Classification and Regression Trees (Breiman, Friedman, Stone, and Olshen), 368
Decision Trees and, 355, 366–367
results comparison, 372
missing values, 369
scale variables, 368
settings, 369–371
curve estimation regression, 72
Custom Dialog Builder, 442

D

Data Format dialog, 253–254
data management, 393–394
data mining
algorithms
Association, 301
Classification, 301–302
Classification (Statistical), 301
Clustering, 302
Decision Trees, 301
Numeric Prediction, 301
ANOVA results, 281
balancing, 294–295
Decision Trees, 272
deployment, 276
ensembles, 297–299
historical data and, 276
interactions, 284–287
K Nearest Neighbors, 272
models, 276
effect size, 284
goodness of fit, 284
omnibus test, 284
significance test, 284
neural nets, 272, 286
Nine Laws of Data Mining (Khabaza), 332
overview, 275–276
partitioning, 288–291
patterns, 276

procedures, 287–288
results comparison, 295–297
SPSS Statistics, 278–280
Tom Khabaza's 9 Laws of Data Mining, 372
validation, 288–291
Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management (Linoff and Berry), 326
data preparation for modeling, 303–304
Data Preparation module, 272, 303
Data menu
Identify Unusual Cases, 303–306
Validation, 303
Transform menu
Data Preparation for Modeling, 303
Optimal Binning, 303–304, 315–316
Data Structure dialog, 117
data validation, 304
data visualization, 129, 131–132
DATASET ACTIVATE command, 397
DATASET DECLARE command, 411
datasets
aggregation, 410–412
case sensitivity, 409–410
Data View, 404–405
Define Variable Properties dialog, 405
parsing strings, 410
restructuring, 410–412
Type and Label subdialog, 408
value labels, 405
Variable View, 404–405
variables, pasting, 412–414
ZIP codes, 407–408
Decision Trees, 272
C5.0 Tree extension command, 355
CHAID algorithm, 355–363
Crosstabs, 360–362
setting changes, 363–365
CRT algorithm, 355, 366–367
Growing Methods
CHAID, 355
CRT, 355
Exhaustive CHAID, 355
QUEST, 355
results comparison, 371–373
impurity, 368
main menu, 356
Post-Hoc tests, 283
purity, 368
Scoring Wizard, 374–378
Test Sample tree, 357–359
Training Sample tree, 357–359
validation options, 373–374
Validation submenu, 357
restricting variables, 356–357
variables, restricting, 356–357
Decomposition of Normalized Raw Stress pivot table, 262
Define Groups subdialog, 7
Define Variable Properties dialog, 405
Density functions panel, 52, 53
Descriptive Statistics of Scale Inputs, 56
Descriptive Statistics of Scale Targets, 56
Descriptive Statistics report, 390
Descriptive Statistics table, 321
DESCRIPTIVES command, 452
dialogs
Categorical Regression, 87–88
Chart Builder, 9
Chart Options, 57
Custom Dialog Builder, 442
Data Format, 253–254
Data Structure, 117
Define Groups subdialog, 7
Define Variable Properties, 405
Discretization, 89
EM Means, 112
Estimation, 110–111
Factor Analysis, 24
Fields & Effects: Fixed Effects, 119
Fields & Effects: Random Effects, 120
Fields & Effects: Target, 118
Fit Details, 50
Fixed Effects, 107–108
Frequencies, 397
General Linear Model, 10

Identify Unusual Cases, 305
Data menu, 305
Missing Values, 309
Options dialog, 310
Output, 307
Save, 308
Variables, 306
Legacy Dialogs, 132–133
Linear Mixed Models, 106–107
Linear Regression, 46–47
Maps, 178
MDS (multidimensional scaling) options, 253
Missing Values, 90
Model, 254–255
Optimal Binning
Missing Values, 319
Options, 320–321
Output, 318–319
Save, 319
Variables, 316–317
Options, 256–257
ordinal regression, 77–81
Output, 92, 258–259
Plots, 93, 257–258
Proximities in Matrices Across Columns, 254–255
Random Effects, 108–110
Regularization, 91–92
Restrictions, 256
Save, 92, 112–113
Simulation Model Source, 45
Specify Subjects and Repeated, 105–106
Statistics, 111
Syntax, 401–402
T-Test, 469
TURF, 446
Univariate, 11
dimension interpretation
statistical approach, 266–269
subjective approach, 264–266
dimension reduction, Correspondence Analysis, 225–229
Discretization dialog, 90
DISCRIMINANT extension command, 464
Discriminant Model, 47
Stepwise, 278
Distinction: A Social Critique of the Judgement of Taste (Bourdieu), 221
Distribution, Monte Carlo Simulation, 49
DO REPEAT command, 414–415
double regression lines, case study, GPL and, 160–163

E

ecological fallacy, 194
effect size, 284
EM Means dialog, 112
ensembles
building, 391–392
data mining and, 297–299
Error Backpropagation, 337–339
Estimates of Covariance Parameters table, 115–116
Estimates of Fixed Effects table, 114
Estimation dialog, 110–111
estimation of effects, linear mixed models and, 103
EXECUTE command, 400
extension commands, 441
C5.0, 355
DESCRIPTIVES, 452
DISCRIMINANT, 464
installed extensions, 443
MATCH FILES, 457
OMS and, 439
Predictive Analytics download, 443
Python and, 394, 396
R and, 394, 396
SPSSINC QQPLOT2, 457
SPSSINC QUANTREG, 458–460
SPSSINC RANFOR, 461
SPSSINC RANPRED, 461
STATS SVM, 461, 467
SUMMARIZE, 452
UNIANOVA, 454

F
Factor Analysis, 293–294
factor analysis, 23–26
  AMOS, 26–27
  Analyze menu, 23–24
  results, 24
Factor Analysis dialog, 24
  Extraction subdialog, 24
Feature Selection (Modeler), 291–294
Felson and Bohrnstedt, 37
Fields & Effects: Fixed Effects dialog, 119
Fields & Effects: Random Effects dialog, 120
Fields & Effects: Target dialog, 118
Final Coordinates pivot table, 263
Fit Details dialog, 50
fixed coefficients, 124
Fixed Effects dialog, 107–108
flowcharts, Modeler and, 276–277
focal points, 382–383
FREQUENCIES command, 397
  Syntax Editor, 399
Frequencies dialog, 397
  Charts subdialog, 402
Frequencies table, 63–65
Friedman, Jerome, *Classification and Regression Trees,* 368

G
General Linear Model, 47
General Linear Model dialog, 10
  MANOVA and, 14
Generalized Linear Model, 47
geographical coordinates, Coordinates on a Reference map, 185–187
Geospatial Association Rules, 193, 195. *See also* 911 calls case study
  Binning and Aggregation, 204–205
  map files, 196–197
    condition variables, 201
    context data, 197–198
    event data, 197–198
    prediction variables, 201
    WKID, 200
  output, 202–206
  overview, 194
  Rule Support, 203
Geospatial Modeling Wizard, 195–196
ggplot2 package, 132
GGRAPH command, 151, 169–170
GGRAPH subcommands, 151
Gini Coefficient, 368
goodness of fit statistics, 284
  AMOS, 28–29, 31
Goodness-of-Fit table, 83
GPL (Graphics Production Language), 394
  arrow case study, 163–167
  bubble chart case study, 156–160
  commands
    AGGREGATE, 169–170
    BEGIN GPL ... END GPL, 151
    DATA, 152
    ELEMENT, 152
    GGRAPH, 151, 169–170
    GRAPHDATASET, 151, 169–170
    GRAPHSPEC, 151
    GUIDE, 152
    SOURCE, 152
    statement commands, 152
  Correspondence Analysis and, 217
  double regression line case study, 160–163
  ELEMENT statement, 157
  functions
    aestheticMaximum, 156
    aestheticMinimum, 156
    null( ), 157
    scale, 152
  GUIDE statements, 157
  MBTI bubble chart case study, 167–172
  MCA (Multiple Correspondence Analysis) and, 217
  Perceptual Map, 242–247

GPL (Graphing Production Language), 129–130, 132, 147
bubble chart case study, 147–155
Chart Builder, 134
statement commands, 152
GPL Reference guide, 155–156
The Grammar of Graphics (Wilkinson), 132, 136–138
Graphboard Editor, 141
regions, 142
Graphboard Template Chooser, 129, 132, 135–136
aesthetics, 137
Basic tab, 138–139
Detailed tab, 140
Bubble chart, 144
elements, 137
fields, specification, 139
The Grammar of Graphics and, 136–137
Graphboard Editor, 141
mapping
Arrows on a Reference map, 188–191
bar chart, 174
bars, 184–185
Choropleth of Counts, 175–179
Choropleth of Sums, 182–184
Choropleth of Values, 179–181
Coordinates on a Choropleth of Counts, 187–188
Coordinates on a Reference, 186–187
maps, 175
Pie of Counts, 181–182
GRAPHDATASET command, 151, 169–170
graphics
editing, 229
Wilkinson, 137
graphing
Chart Builder, 132
Graphboard Template Chooser, 132
Legacy Dialogs, 132
Wilkinson, 137
Graphs menu
Chart Builder, 132
Graphboard Template Chooser, 132, 135–136
Legacy Dialogs, 132–133
GRAPHSPEC command, 151
grouped scatterplot, 149
Grouping Variable, 7
GUI (graphical user interface)
AMOS, 3, 27
Syntax and, 395

H

Hastie, T., *An Introduction to Statistical Learning with Applications,* 462
Help, 398–399
Hesterberg, Tim C., *Mathematical Statistics with Resampling and R,* 63
hidden layers, neural nets and, 332–333
hierarchical linear mixed models. *See* linear mixed models
Hierarchical Linear Models, 2
historical data, data mining and, 276
Hoelter results, 34–35
Hold-Out Validation, 289
Hsu, Chih-Wei, *A Practical Guide to Support Vector Classification,* 462–463
hyperwiggle, 462
hypothesis testing, 280–283

I

IBM SPPS Modeler Cookbook (McCormick), 380
IBM SPSS Statistics Command Syntax Reference, 443
ID3 Decision Tree, 279
Identify Unusual Cases (Data Preparation), 303–304
dialogs, 305
Data menu, 305
Missing Values, 309
Options dialog, 310

Output, 307
Save, 308
Variables, 306
output
Anomaly Case Index list, 311
Anomaly Case Reason List, 312
Anomaly Index Summary table, 314
Case Processing Summary table, 311
Categorical Variable Norms table, 313
Scale Variable Norms table, 313
impurity, decision trees and, 368
INDEX command, 410
INDSCAL (weighted multidimensional scaling), 253
inefficient code, 403–404
Inputs Distributions table, 55
interactions, data mining and, 284–287

J

James, G., *An Introduction to Statistical Learning with Applications,* 462

K

keywords, 398
TO, 413–415
CLASSWEIGHTS, 467
Syntax, 398
Khabaza, Tom, 279
8th Law of Data Mining, 372
Nine Laws of Data Mining, 332
K-Means Cluster Model, 47
KNN (K Nearest Neighbor), 272, 379
bias, 386–388
classification tables, 384, 388
finding neighbors, 380–381
as lazy learner, 379
Nearest Neighbor Analysis main menu, 382
Partitions submenu, 383
Peers chart, 384
predictions, saving, 388–389
Titanic dataset, 380–385
variance, 386–388

L

labels, OMS, 425–426
Latent Class analysis, 12
lazy learners, 379
learning styles, case study and, 6
Legacy Dialogs, 132–133
Bar Charts, 132–133
LeRoux, Brigitte, *Multiple Correspondence Analysis,* 217–218
Levene's Test, 7
Lin, C., *A Practical Guide to Support Vector Classification,* 462–463
linear hierarchical models, 101
linear mixed models, 103
EM Means dialog, 112
Estimation dialog, 110–111
Fixed Effects dialog, 107–108
generalized, 116–117
Data Structure dialog, 117
Fields & Effects: Fixed Effects dialog, 119
Fields & Effects: Random Effects dialog, 120
Fields & Effects: Target dialog, 118
output, 120–126
hierarchical, 101–102
mixed models, 101–102
two-level example, 102–104
Linear Mixed Models dialog, 106–107
output
Estimates of Covariance Parameters table, 115–116
Estimates of Fixed Effects table, 114
Information Criteria table, 113–114
Model Dimension table, 113
Tests of Fixed Effects table, 113–114
Random Effects dialog, 108–110
Save dialog, 112–113
Specify Subjects and Repeated dialog, 105–106
Statistics dialog, 111
structure adjustments, 126–128
Linear Mixed Models dialog, 106–107

linear regression, 71, 72
Monte Carlo Simulation, 48
SPSS Bootstrap, 68–69
Stepwise, 278
Linear Regression dialog, 46–47
Linear Regression models, 47
Linoff, Gordon S., *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management,* 326
Logistic node (Modeler), 278
logistic regression, 71, 73, 464–465
Stepwise, 278
Logistic Regression dialog, Optimal Binning, 323
LOWER( ) command, 409–410
LSD (Least Square Difference), 282

M

MANCOVA (Multivariate Analysis of Covariance), 4
covariants
four, 18–21
single, 16–18
multivariate tests, 18, 25–26
alternate, 22
results, 17
Parameter Estimates, 18, 21
alternate, 22
Pillai's Trace, 17
Tests of Between-Subjects Effects, 20
results, 17
MANOVA (Multivariate Analysis of Variance), 4, 13
Fixed Factors, 14
General Linear Model, 14
General Linear model, 15
Multivariate dialog, 14
multivariate tests, 15
Tests of Between-Subjects Effects, 15
MAPE (Mean Absolute Percent Error), 296, 326
mapping, 173
Graphboard Template Chooser
Arrows on a Reference map, 188–191
bar chart, 174
bars, 184–185
Choropleth of Counts, 175–179
Choropleth of Sums, 182–184
Choropleth of Values, 179–181
Coordinates on a Choropleth of Counts, 187–188
Coordinates on a Reference, 186–187
maps, 175
Pie of Counts, 181–182
Maps dialog, 178
MATCH FILES command, 457
Mathematical Statistics with Resampling and R (Chihara and Hesterberg), 63
MBTI bubble chart case study, 167–172
MCA (Multiple Correspondence Analysis), 217
crosstabulations, 234–242
Discrimination Measures Plot, 239
Discrimination Measures table, 240
sorting, 241
Joint Category Plots, 237–238
main menu, 237
Optimal Scaling, 236
Perceptual Map, 242–247
variable plots, 237–238
McGrayne, Sharon Bertsch, 61
MDS (multidimensional scaling), 249
ALSCAL, 253
Data Format dialog, 253–254
dialog options, 253
dimension coordinates, 266–267
INDSCAL (weighted multidimensional scaling), 253
metric, 251
Model dialog, 254–255
nonmetric, 251
psychology sub-disciplines, 251–253
object points plot, 250
Options dialog, 256–257
output
Correlations table, 268–269
Decomposition of Normalized Raw Stress, 262
Final Coordinates pivot table, 263
Object points plot, 263–264

statisical approach to interpretation, 266–269
Stress and Fit Measures table, 260–262
subjective approach to interpretation, 264–266
Output dialog, 258–259
Plots dialog, 257–258
Proximities in Matrices Across Columns dialog, 254–255
PROXSCAL, 253
Restrictions dialog, 256
Transformation Plot, 264–265
TURF analysis and, 265
mean absolute percent, 344–349
media distribution case study, 404–420
MERGE command, 415–417
metric multidimensional scaling, 251
Miles, Andrew, 218
Minsky, Marvin, *Perceptrons,* 326–327
missing values, CRT algorithm, 369
Missing Values dialog, 90
Optimal Binning, 319
mixed models, linear, 101–104
generalized linear, 116–120
structure adjustments, 126–128
MLP (multilayer perception), 328, 337
XOR patterns and, 328–329
Model dialog, 254–255
Model Dimension table, 113
Model Entropy table, 321–322
model files
creating, 45–59
Linear Regression dialog, 46–47
Model Fitting Information table, 83
Model Summary table, categorical regression, 94
Model tab, 51
Model Type table, 54
Modeler, 271–273, 276–277
balancing, 294–295
Factor Analysis, 293–294
Feature Selection, 291–294
history, 279–280
KNN and, 380
nodes, 277
Analysis, 295–297
Distribution, 294–295
Ensemble, 297–299
Logistic, 278
Type node, 277–278
results comparison, 295–297
scoring new records, 300–302
streams, 276–277, 289–291
models
Binary Logistic Regression Model, 47
Cox Regression Model, 47
Discriminant Model, 47
General Linear Model, 47
Generalized Linear Model, 47
K-Means Cluster Model, 47
Linear Regression models, 47
Multinomial Logistic Regression Model, 47
Neural Net Model, 47
Ordinal Multinomial Regression Model, 47
Tree Model, 47
Two-step Cluster Model, 47
Monte Carlo Simulation, 43
Correlations panel, 50, 56
Correlations table, 55
Density Functions panel, 52, 53
Descriptive Statistics of Scale Inputs, 56
Descriptive Statistics of Scale Targets, 56
Distribution, 49
Fit Details dialog, 50
Frequencies table, 63–65
Inputs Distributions table, 55
linear regression, results, 48
model files, 45
Model Type table, 54
options, 44–45
Output panel, 52, 53
overview, 44
Probability Density charts, edited, 58

Simulated Fields panel, 49
Simulation Summary table, 56
Stopping Criteria table, 55
multinomial logistic regression, 72
Multinomial Logistic Regression Model, 47
Multiple Correspondence Analysis (LeRoux & Rouanet), 217–218
Multiple Imputation, 220
Multivariate dialog, 14
multivariate tests, summary, 25–26

N

Nearest Neighbor Analysis main menu, 382
Neighbors submenu, 386
Partitions submenu, 383
nearest neighbors. *See* KNN (K Nearest Neighbor)
Neural Net Model, 47
neural nets. *See* ANN (artificial neural network)
Neural Networks module, 325
Nine Laws of Data Mining (Khabaza), 332
nodes, Modeler
Analysis, 295–297
Distribution, 294–295
Logistic, 278
Type node, 277–278
nonlinear regression, 73
nonmetric multidimensional scaling, 251
psychology sub-disciplines, 251–253

O

object points, multidimensional scaling and, 250
Object points plot, 263–264
OLS (Ordinary Least Square), 450–451
QR comparison, 455–460
Olshen, R.A., *Classification and Regression Trees,* 368
omnibus test, 284
OMS (Output Management System), 217, 218, 394
Bivariate Correlations procedure, 432–433
manipulated Correlations, 434–435
Control Panel, 229–230, 244, 423–424
completed dialog, 425
Correlations dataset, 436–438
Crosstabs dialog, 431
dataset, 244–245
Descriptives dialog, 437
Identifiers dialog, 439
labels, 425–426
MCA Perceptual Map and, 242–247
Options dialog, 427–428
output
formats, 424–429
suppressing, 426–427, 429–435
overview, 422–423
requests, 428–429
results views, 230–232
running
from menus, 423–424
from Syntax, 438–439
supported formats, 422
Optimal Binning (Data Preparation), 303–304, 315–316
dialogs
Missing Values, 319
Output, 318–319
Save, 319
Variables, 316–317
Options dialog, 320–321
output
Binning summary table, 322–232
Descriptive Statistics table, 321
Logistic Regression dialog, 323
Model Entropy table, 321–322
Variables in the Equation table, 323–324
Optimal Scaling (CATREG) regression, 73
Options dialog, 256–257
Identify Unusual Cases, 310
Optimal Binning, 320–321
ORD (Ordinary Least Squares), quantile regression and, 455–460

ORDER subcommand, 398
ordinal data, four-category variables, 12
Ordinal Multinomial Regression Model, 47
ordinal regression, 72
  dialogs, 77–81
  output, 81
    Case Processing Summary table, 82
    Goodness-of-Fit table, 83
    Model Fitting Information table, 83
    Parameter Estimates table, 84
    Pseudo R-Square table, 83
    Test of Parallel Lines table, 85
ordinal regression theory, 74
  assumptions of models, 77
  cumulative logit models, 74
  distribution, 74–75
  GZLM (generalized linear modeling), 75
  link functions, 74
    Cauchit, 76
    complementary log-log, 76
    Logit, 76
    negative log-log, 76
    probit, 76
  location components, 75–76
  scale components, 76
outlier detection, 304
output. *See also* OMS (Output Management System)
  categorical regression
    ANOVA table, 94
    Case Processing Summary table, 93–94
    Coefficients table, 95
    Correlations and Tolerance table, 95
    Model Summary table, 94
    Quantifications table, 96–97
    transformation plot, 97–99
  Geospatial Association Rules, 202–206
  Identify Unusual Cases, 307
    Anomaly Case Index List, 311
    Anomaly Case Reason List, 312–313
    Anomaly Index Summary table, 314
    Case Processing Summary table, 311
    Categorical Variable Norms table, 313
    Scale Variable Norms table, 313
  linear mixed models
    Estimates of Covariance Parameters table, 115–116
    Estimates of Fixed Effects table, 114
    generalized, 120–126
    Information Criteria table, 114
    Model Dimension table, 113
    Tests of Fixed Effects table, 114
  MDS (multidimensional scaling)
    Decomposition of Normalized Raw Stress, 262
    Final Coordinates pivot table, 263
    Object points plot, 263–264
    scree plot, 260
    Stress and Fit Measures table, 260–262
  Monte Carlo Simulation, 52, 53
  neural nets
    regression, 342
    weights and, 334–335, 337–339
  Optimal Binning
    Binning summary table, 322–232
    Descriptive Statistics table, 321
    Logistic Regression dialog, 323
    Model Entropy table, 321–322
    Variables in the Equation table, 323–324
  ordinal regression, 81
    Case Processing Summary table, 82
    Goodness-of-Fit table, 83
    Model Fitting Information table, 83
    Parameter Estimates table, 84
    Pseudo R-Square table, 83
    Test of Parallel Lines table, 85
Output dialog, 92, 258–259
  Identify Unusual Cases, 307
  Optimal Binning, 318–319
Output Management System. *See* OMS
Output panel, 52, 53

P
p values, 360, 362–363
Pappert, Seymour A., *Perceptrons,* 326–327
Parameter Estimates, 11, 15, 20–21
  MANCOVA, 18
    alternate, 22
  table, 12, 84
parsing strings, 410
partial least square regression, 72
partitioning, 340–341
  data mining and, 288–291
PCLOSE, 35
perceptrons, 327
  multilayered, 333, 343
Perceptrons (Minsky and Pappert), 326–327
Pie of Counts map, 181–182
Pillai's Trace, 17, 18
  results, 19
Plots dialog, 93, 257–258
plug-ins, 442
Post-Hoc tests
  Bonferroni, 282
  Decision tree, 283
  LSD (Least Square Difference), 282
Pratt's measure of relative importance, 96
Predicted by Observed view, 122
predictive analytics, 271–272. *See also* data mining
predictive policing, case study and, 207–215
predictors, 381–384
  chart focal point, 383
Probability Density charts, 56–57
  edited, 58
probit regression, 73
programmability, 393–394
Programming and Data Management, 442
Proximities in Matrices Across Columns dialog, 254–255
proximity matrices, 252, 254
PROXSCAL, 253
PSAT tests case study
  learning styles and, 6
  overview, 4–6
  reading comprehension, 20
Pseudo R-Square table, 83
psychology sub-disciplines nonmetric multidimensional scaling, 251–253
purity, decision trees and, 368
Python, extension commands and, 394, 396

Q
Quantifications table, categorical regression, 96–97
quantile regression, 450–454
  coefficients, 455–456
    by quantile, 458
  estimation, 458
  OLS comparison, 455–460
  ORD (Ordinary Least Squares) and, 455–460
  residuals histogram, 455
QUANTREG coefficients, 456

R
R language
  extension commands and, 394, 396
  graphics, 132
R Squared, 326
Random Effects dialog, 108–110
RBF (Radial Basis Function), 462–463
recalculate correlations, 50
regression
  automated linear modeling, 72
  binary logistic, 72
  categorical, 86–87
    dialogs, 87–93
  curve estimation, 72
  linear, 71, 72
    Monte Carlo Simulation, 48
    SPSS Bootstrap, 68–69
  logistic, 71, 73, 464–465
  multinomial logistic, 72

neural nets, 347
mean absolute percent, 345–346
slope, 335
neural nets bank salary case study, 341–353
nonlinear, 73
Optimal Scaling (CATREG), 73
ordinal, 72, 74
assumptions of models, 77
cumulative logit models, 74
dialogs, 77–81
distribution, 74–75
GZLM, 75
link functions, 74, 76
location components, 75–76
output, 81–86
scale components, 76
partial least square, 72
probit, 73
quantile regression, 450–454
R Squared, 326
regression lines, 10
SVM and, 461–462
two-stage least squares, 73
weight estimation, 73
Regularization dialog, 91–92
relative importance, 96
Restrictions dialog, 256
Restructure Data Wizard, 410–412
RMSEA (Root Mean Square Error of Approximation), 29, 34, 35
Rosenblatt, F., 326
Rosling, Hans, 137, 143
Rouanet, Henry, *Multiple Correspondence Analysis,* 217–218
Run Selection button, 398

S

SAT origins, 21
.sav data file, 4
Save dialog, 92, 112–113
Identify Unusual Cases, 307
Optimal Binning, 319
scale function, 152
Scale Variable Norms table, 313
scaling
INDSCAL (weighted multidimensional scaling), 253
multidimensional (*See* MDS (multidimensional scaling))
scatterplots
colored, 9
grouped, 149
Scoring Wizard, 374–378
SEM (Structural Equation Modeling), 2, 3
Shearer, Colin, 279
significance tests, 284
Simulated Fields panel, 49
Simulation Builder, Monte Carlo Simulation, 45
Simulation Model Source dialog, 45
Simulation Summary table, 56
Spatial Modeling, 195–196
Spatio-Temporal Prediction, 193, 196, 207
Specify Subjects and Repeated dialog, 105–106
Split File menu (Bootstrapping), 65–66, 390
SPPS-X, 395
SPSS Algorithms Guide, 337
SPSS AMOS. *See* AMOS
SPSS Bootstrap
Advanced Statistics module and, 60
frequencies report, 63–64
Frequencies table, 63–65
linear regression, 68–69
mean, 66–67
proportions, 63–66
Regression module and, 60
Split file menu, 65–66
SPSS Statistics GUI, 3
SPSSINC QQPLOT2 extension command, 457
SPSSINC QUANTREG extension command, 458–460
SPSSINC QUANTREG procedure, 442

SPSSINC RANFOR extension command, 461
SPSSINC RANPRED extension command, 461
SPSSINC TURF procedure, 442
stand-in variables, 414–415
Statistics Base, procedures, 60
Statistics dialog, 111
STATS SVM extension command, 442, 461, 467
Stepwise, 278
  Binary Logistic Regression, 284
  cautions, 286–287
Stone, Charles J., *Classification and Regression Trees,* 368
Stopping Criteria table, 55
Stress and Fit Measures table, 260–262
strings, parsing, 410
Sullivan, Alice, 218
SUMMARIZE command, 452
surrogates, missing values and, 369
SVM (support vector machines), 461–468
Syntax, 393–396
  code, inefficient, 403–404
  Command Syntax reference, 400–401
  commands
    COMPUTE, 415
    CROSSTABS, 402–403
    DATASET ACTIVATE, 397
    DATASET DECLARE, 411
    DO REPEAT, 414–415
    FREQUENCIES, 397
    INDEX, 410
    LOWER( ), 409–410
    ORDER subcommand, 398
    UPPER( ), 409–410
  datasets
    aggregation, 410–412
    case sensitivity, 409–410
    Data View, 404–405
    Define Variable Properties dialog, 405
    parsing strings, 410
    restructuring, 410–412
    Type and Label subdialog, 408
    value labels, 405
    Variable View, 404–405
    ZIP codes, 407–408
  dialogs, 401–402
  final file, 417–420
  Help, 398–399
  TO keyword, 413–414
  media distribution case study, 404–420
  MERGE command, 415–417
  OMS, running, 438–439
  Restructure Data Wizard, 410–412
  returns, 417
  Run Selection button, 398
  variables, stand-in, 414–415
Syntax Editor, 397
  FREQUENCIES command, 399
  Frequencies main dialog, 397
    Charts subdialog, 402
  indenting, 417
Syntax Reference Guide, 398
syntax window, 150

T

T-Tests, Cohen's d and, 468–471
tables
  Anomaly Index Summary, 314
  ANOVA, 94
  Binning summary, 322–232
  Case Processing Summary, 82, 93–94
    Identify Unusual Cases, 311
  Categorical Variable Norms table, 313
  Coefficients, 95
  Correlations, 55, 56, 268–269
  Correlations and Tolerance, 95
  Decomposition of Normalized Raw Stress, 262
  Descriptive Statistics, 321
  Descriptive Statistics of Scale Inputs, 56
  Descriptive Statistics of Scale Targets, 56

Discrimination Measures table, 240
Estimates of Fixed Effects, 114
Final Coordinates, 263
Frequencies table, 63–65
Goodness-of-Fit, 83
Information Criteria table, 114
Inputs Distributions, 54
Model Dimension, 113
Model Entropy, 321–322
Model Fitting Information, 83
Model Summary, 94
Model Type, 54
Parameter Estimates, 12, 84
Pseudo R-Square table, 83
Quantifications, 96–97
Scale Variable Norms table, 313
Simulation Summary, 56
Stopping Criteria, 55
Stress and Fit Measures, 260–262
Test of Parallel Lines, 85
Tests of Fixed Effects table, 114
Variables in the Equation, 323–324
Test of Parallel Lines table, 85
Test Variable, 7
Tests of Between-Subjects Effects, 15
MANCOVA, 20
results, 17
The Theory that Would not Die (McGrayne), 61
Tibshirani, R., *An Introduction to Statistical Learning with Applications,* 462
Titanic dataset
Decision Trees
CHAID algorithm, 355–360, 364, 369
Scoring Wizard, 375–378
KNN and, 380–385
neural nets and, 349–353
TLI (Tucker Lewis Index), 33–35
TO keyword, 413–415
topology
bank salary case study, 343
neural nets, 333–334, 348
transformation plot, categorical regression, 97–99
Transformation Plots, 264–265
Tree Model, 47
T-test, 4, 7
Cohen's d, 8
dialog, 16
Levene's Test, 7
T-Test dialog, 469
TURF analysis, 265, 444–450
two-level hierarchical linear model, 102–104
two-stage least squares regression, 73
Two-step Cluster Model, 47
Type and Label subdialog, 408
Type I error, 281–283
Type node (Modeler), 277–278

U

UNIANOVA command, 454
univariate analysis of variance, 13
Univariate dialog, 11
unobserved variables, 23
UPPER( ) command, 409–410
Utilities menu, OMS Control Panel, 423–424

V

validation, 304
alternative options, 373–374
data mining and, 288–291
Hold-Out Validation, 289
V-fold cross-validation, 387
Validation (Data Preparation), 303
value labels, datasets, 405
variables
CRT algorithm, 368
Decision Trees, restricting, 356–357
pasting names, 412–414
stand-in, 414–415
unobserved, 23
Variables dialog, 316–317
Variables in the Equation table, 323–324

variance
  KNN and, 386–388
  linear mixed models and, 103
V-folds, cross-validation, 387
views, Predicted by Observed, 122
visual programming, 276–277

W
weight estimation regression, 73
Wheeler, Andrew, 130
Wilkinson, Leland, *The Grammar of Graphics,* 132, 136–138
windows, syntax, 150
wizards, Geospatial Modeling Wizard, 195–196

X
XOR
  dataset creation, 330–331
  Nominal Output Variable, 336
  truth table
    flat regression line, 329–330
    neural nets and, 328
  variables, neural nets and, 333–341