

Symbol

! operator 77
 != operator 77
 # symbol 8
 %a symbol 81
 %A symbol 81
 %B symbol 82
 %b symbol 82
 %d symbol 81
 %m symbol 81
 %Y symbol 82
 %y symbol 82
 * operator 75, 178
 ** operator 75
 ... option 58, 61
 . symbol 178
 / operator 75
 : symbol 178
 ? function 11
 ?? function 11
 ^ operator 75, 178, 181
 ~ symbol 178
 + operator 75, 178
 < operator 77
 <<- operator 29
 <= operator 77
 == operator 77
 > operator 77
 >= operator 77
 -1 symbol 178
 brackets 29
 3D pie charts 127
 3D scatter plots 274–278

A

abline() function 60, 265
 abs() function 93
 absolute widths 67
 acos() function 93
 acosh() function 93
 AER package 421

aggr() function, VIM package 357
 aggregate() function 113, 240
 aggregating data 112–113
 AIC() function 179, 208
 all subsets regression 210, 213
 alpha option 390
 alternative= option 255
 Amelia package 365, 369, 421
 analyses, excluding missing values from 80–81
 analysis of covariance (ANCOVA)
 one-way 230–233
 assessing test assumptions 232
 visualizing results 232–233
 overview 222
 analysis of variance (ANOVA)
 219–245, 252–253
 fitting models 222–225
 aov() function 222–223
 order of formula terms 223–225
 MANOVA 239–243
 assessing test assumptions 241–242
 robust 242–243
 one-way 225–230
 assessing test assumptions 229–230
 multiple comparisons 227–229
 one-way ANCOVA 230–233
 assessing test assumptions 232
 visualizing results 232–233
 as regression 243–245
 repeated measures 237–239
 terminology of 220–222
 two-way factorial 234–236
 analytic packages, for large datasets 431

ANCOVA. *See* analysis of covariance
 ancova() function, HH package 232
 AND operator 77
 annotating datasets 42
 annotations 62–64
 ANOVA. *See* analysis of variance
 anova() function 179, 208
 Anova() function, car package 225, 239
 aov() function 222–223
 append option 13
 apply() function 102–103
 apropos() function 11
 aq.plot() function, mvoutlier package 242
 arithmetic operators 75
 arrayImpute package 370 421
 arrayMissPattern package 370, 421
 arrays 26–27
 Arthritis dataset 19
 as.character() function 83
 ASCII file 35
 as.datatype() function 84
 as.Date() function 81, 88
 asin() function 93
 asinh() function 93
 aspect option 378
 assumptions
 linear model, global validation of 199
 of MANOVA tests, assessing 241–242
 of OLS regression, assessing 188–199
 of one-way ANCOVA tests, assessing 232
 of one-way ANOVA tests, assessing 229–230
 asypow package 261

at option 58
 atan() function 93
 atanh() function 93
 attach() function 28, 30, 88
 auto.key option 384
 avPlots() function 193, 203
 axes 57, 60
 axes option 57
 axis() function 57

B

background color (bg)
 option 61
 backslash character 13, 102
 bar plots 120–125
 fitting labels in 124
 for mean values 122–123
 simple 120–121
 spinograms 124–125
 stacked and grouped
 121–122
 tweaking 123–124
 barplot() function 120,
 122–123
 base package 274
 batch processing 17
 Beta distribution 97
 bg option. *See* background color
 option
 bg parameter 52
 biganalytics package 431
 biglars package 431
 bigmemory package 430–431
 bigtabulate package 431
 Binomial distribution 97
 bitro.diameter variable 339
 bivariate relationships 184
 block comments 33
 bmp() function 47
 boot package 422
 bootstrap package 214
 bootstrapping 89, 303–309
 box plots 133–138
 parallel, comparing groups
 with 134–137
 violin variation of 137–138
 box type (bty) option 61
 boxplot() function 47, 238
 boxplots option 267
 boxplot.stats() function 133
 boxTidwell() function 206
 Box-Tidwell transformations
 206
 breaks option 128

Brodingnag package 431
 bty option. *See* box type option
 bubble plots 278–279
 by function 146
 by option 113
 byrow option 25
 bzfile() function 36

C

c() function 9, 24, 43
 ca package 422
 car package 225, 230, 239, 266,
 268
 case identifiers 23, 30
 case-wise deletion 364–365
 cast() function 114–115
 casting 114, 116
 cat() function 101, 111
 cat package 370, 422
 categorical variables 23
 Cauchy distribution 97
 cbind() function 43, 85,
 105–106, 240
 ceiling() function 93
 cex() option 61
 cex parameter 51, 53
 cex.axis parameter 53
 cex.lab parameter 53
 cex.main parameter 53
 cex.names option 123
 cex.sub parameter 53
 CFA. *See* confirmatory factor
 analysis
 character functions 99–101
 character variables, converting
 date values to 83
 Chi-square tests 255–256
 Chi-squared (noncentral)
 distribution 97
 class() function 43
 cld() function 228
 CLI. *See* command-line interface
 close() function 40
 cm.colors() function 53
 cmdscale() function 350
 code editors, list of 403–404
 coefficients() function 179
 coin package 422
 col option 52, 58, 122, 134,
 136, 359, 378
 col.axis parameter 52
 colClasses option 36, 430
 col.corrgram() function 287
 colfill vector 132
 col.lab parameter 52
 col.main parameter 52
 color option 139, 390
 colorRampPalett()
 function 287
 colors, graphical parameters
 52–53
 colors() function 53
 col.sub parameter 52
 columns
 adding 85
 data frames 27
 combine objects. *See* c() *function*
 combining graphs. *See* page
 arrangement of graphs
 command prompt 7
 command-line interface
 (CLI) 403
 command-line options 407
 command-line prompt 403
 comments, # symbol 8, 33
 common factors 342
 comparisons, multiple
 227–229
 complete() function 369
 complete-case analysis 364–365
 complete.cases()
 function 356–357, 364
 components, principal
 extracting 339
 rotating 339–341
 scores 341–342
 selecting number to
 extract 335
 comprehensive GUIs, for
 R 405
 Comprehensive R Archive
 Network (CRAN) 7,
 406
 conditional execution 107–109
 if-else construct 108–109
 ifelse construct 109
 repetition and looping
 107–108
 switch construct 109
 conditioning variables 376,
 379–380
 confint() function 179, 188
 confirmatory factor analysis
 (CFA) 349
 constant residual variance 191
 contrasts() function 244
 contr.helmert function 244
 control flow 107–109

contr.poly function 244
 contr.SAS function 244
 contr.sum function 244
 contr.treatment function 244
 conversions, type 83–84
 Cook's distance 18, 189–191,
 202–204, 317
 cooks.distance() function 18
 cor() function 184
 corrective measures 205–207
 deleting observations 205
 variables
 adding or deleting 207
 transforming 205–207
 correlations
 tests of significance
 162–164, 253
 types 159–162
 using to assess missing data
 patterns 360–361
 correlograms 283, 287
 corrgram() function, corrgram
 package 284
 corrgram package 284, 422
 corrperm package 422
 cos() function 93
 cosh() function 93
 cov() function 240
 cov2cor() function 343
 Cox proportional hazards
 regression 175
 cpairs() function 269–270
 CRAN. *See* Comprehensive R
 Archive Network
 crimedat dataframe 40
 cross-tabulations 151–155
 crossval() function 214
 cross-validation 213, 215
 crPlots() function 193, 196
 curly braces 107
 cut() function 78, 101, 379

D

D plots, Cook. *See* Cook's
 distance
 D values, Cook. *See* Cook's
 distance
 data
 exporting of 408–409
 delimited text file 408
 Excel spreadsheet 409
 missing. *See* missing data
 for statistical
 applications 409
 long format 114

time-stamping 82
 data() function 11
 data frames 22–23, 27–30
 applying functions to
 102–103
 attach(), detach(), and with()
 functions 28–30
 case identifiers 30
 using SQL statements to
 manipulate 89–90
 data management
 aggregating 112–113
 control flow 107–109
 conditional execution
 108–109
 repetition and
 looping 107–108
 datasets
 merging 85–86
 subsetting 86–89
 date values 81–83
 example 73–75
 functions 93–103
 applying to matrices and
 data frames 102–103
 character 99–101
 mathematical 93–94
 probability 96–99
 statistical 94–96
 missing values 79–81
 excluding from
 analyses 80–81
 recoding values to
 missing 80
 restructuring
 reshape package 113–116
 transpose 112
 sorting 84–85
 type conversions 83–84
 user-written functions
 109–111
 using SQL statements
 to manipulate data
 frames 89–90
 variables
 creating new 75–76
 recoding 76–78
 renaming 78–79
 data objects
 applying functions to 102
 functions for working
 with 42–44
 data option 299, 300, 306, 308,
 319, 323, 325, 328, 365,
 368, 375
 data storage, outside of
 RAM 430–431
 data structures 23–33
 arrays 26–27
 data frames 27–30
 attach(), detach(), and
 with() functions 28–30
 case identifiers 30
 factors 30–31
 lists 32–33
 matrices 24–26
 vectors 24
 data type, converting from one
 to another 84
 database interface (DBI) related
 packages 41
 database management systems
 (DBMSs), accessing
 39–41
 DBI-related packages 41
 ODBC interface 39–40
 data.frame() function 27
 datasets
 annotating 42
 data structures 23–33
 arrays 26–27
 factors 30–31
 frames 27–30
 lists 32–33
 matrices 24–26
 vectors 24
 description of 22–23
 functions for working with
 data objects 42–44
 input 33–42
 accessing DBMSs 39–41
 entering data from
 keyboard 34–35
 importing data 35–39,
 41–42
 webscraping 37
 large 18, 429–431
 analytic packages for 431
 efficient
 programming 429–430
 storing data outside of
 RAM 430–431
 merging 85–86
 adding columns 85
 adding rows 86
 subsetting 86–89
 excluding variables 86–87
 random samples 89
 selecting observations
 87–88

selecting variables 86
 subset() function 88–89
 transposing 112
 date() function 82
 date values 81–83
 DBI related packages. *See*
 database interface
 related packages
 DBMSs, accessing. *See* database
 management systems,
 accessing
 deleting old versions of R 433
 deletion, pairwise 370–371
 delimited text files
 exporting data to 408
 importing data from 35–36
 demo() function 9–10
 density() function 130
 densityplot() function 386
 dependent variable 220
 detach() function 28, 30
 dev.new() function 47
 dev.next() function 47
 dev.off() function 13, 47
 dev.prev() function 47
 dev.set() function 47
 diagnostics, regression
 188–200
 enhanced approach
 192–198
 global validation of linear
 model assumption 199
 multicollinearity 199–200
 typical approach 189–192
 diag.panel option 285
 diff() function 95
 difftime() function 83
 dim() function 43
 dimensions
 of an array 26
 of graphs and margins
 54–56
 dimnames 26
 dir.create() function 13
 directory initialization file 406
 distribution functions,
 normal 97–98
 dmat.color() function, gclus
 package 270
 doBy package 422
 dollar sign character 33
 dot plots 138, 140
 dotchart() function 138
 Durbin-Watson test 196
 durbinWatsonTest()
 function 193, 196

E

echo option 412
 edit() function 34–35
 EFA. *See* exploratory factor
 analysis
 effect() function 187, 231
 defined for ANOVA 252
 defined for chi-square
 tests 255
 defined for correlation 253
 defined for linear
 models 253, 254
 defined for test of
 proportions 254
 defined for t-test 250
 effect size 248–260
 effect size benchmarks
 257–258
 effects library 231
 effects package 187, 231, 422
 environment, customizing
 startup 406–407
 environment variables 407
 errors, independence of 196
 escape character 13, 102
 ES.w2() function 255
 eval option 412
 example() function 11
 example.Rnw file 411
 Excel, Microsoft
 accessing files with
 RODBC 36
 exporting data to
 spreadsheet 409
 importing data from 36–37
 excluding
 missing values from
 analyses 80–81
 observations 87–89
 variables 86–87
 exp() function 94
 exploratory factor analysis
 (EFA) 331–334,
 342–349
 deciding number of common
 factors to extract
 343–344
 FactoMineR package 349
 factors
 extracting common
 344–345
 rotating 345–348
 scores 349
 FAiR package 349
 GPArotation package 349

 nFactors package 349
 other latent variable
 models 349–351
 exponential distribution
 97, 315
 exponentiation operator 75
 exporting data 408–409
 delimited text file 408
 Excel spreadsheet 409
 for statistical
 applications 409
 expression() function 386
 expression statement 107
 extracting
 common factors 344–345
 principal components 339

F

F distribution 97
 fa() function 333, 344, 349
 facets, ggplot2 package
 390–394
 facets option 390
 factanal() function 333
 FactoMineR package 349, 422
 factor() function 30, 42
 factor intercorrelation
 matrix 346
 factor pattern matrix 346
 factor structure matrix 346
 factorial ANOVA design 221
 factor.plot() function 333, 347
 factors
 as dimensions in principal
 components or factor
 analysis deciding
 number of common to
 extract 343–344
 extracting common
 344–345
 rotating 345–348
 scores 349
 as R data structures 23–24,
 30–31
 fa.diagram() function
 333, 347
 FAiR package 349, 422
 family parameter 54
 fan plots 127–128
 fan.plot() function 127
 fa.parallel() function 333,
 335, 343
 fCalendar package 83, 422
 ff package 430–431
 fg parameter 52

- fgui package 405
 - fig graphical parameter 69–71
 - fig option, in Sweave 412
 - figures, creating with fine control 69–72
 - file() function 36
 - filehash package 430
 - fill option 390
 - fine control, creating figure arrangements with 69–72
 - First() function 406–407
 - fit lines 5
 - fitted() function 179
 - fitting ANOVA models 222–225
 - aov() function 222–223
 - order of formula terms 223–225
 - fitting regression models, with lm() function 178–179
 - fix() function 35, 43, 78
 - FlexMix package 350
 - floor() function 93
 - fmi, fraction of missing information 367–368
 - font families
 - changing 54
 - examples on Windows platform 64
 - font parameter 54
 - font.axis parameter 54
 - font.lab parameter 54
 - font.main parameter 54
 - font.sub parameter 54
 - for loop 108
 - foreign package 38, 409, 422
 - format() function 82
 - formulas, in R 178, 223–225
 - frame.plot option 57
 - freq option 128
 - frequency tables 149–155
 - Friedman test 168
 - functions
 - applying to data objects 102
 - character 99–101
 - date 81–83
 - for debugging 111
 - mathematical 93–94
 - numeric 93–99
 - other useful 101
 - probability 96–99
 - for saving graphic output 14
 - statistical 94–96
 - type conversion 84
 - user-written 109–111
- G**
-
- Gamma distribution 97
 - gap package 261
 - gclus package 16, 269–270, 422
 - gcolor option 138
 - generalizability 174
 - genome-wide association studies (GWAS) 261
 - geom option 390
 - geometric distribution 97
 - geostatistical data 14
 - getwd() function 12, 406
 - GGObi program 399
 - ggplot2 package 374–375, 390–394
 - Gibbs sampling 366
 - glht() function, multcomp package 227
 - glm() function 431
 - glmPerm package 423
 - global validation, of linear model assumption 199
 - gls() function, nlme package 239
 - gmodels package 423
 - GPArotation package 349
 - gplots package 123, 226, 235, 423
 - graph dimensions 54, 56
 - graphic output 13–14
 - graphic user interfaces (GUIs) 5, 403–405
 - IDEs for 403–404
 - for R 405
 - graphical parameters 49–56
 - colors 52–53
 - graph and margin dimensions 54–56
 - reference lines 60
 - symbols and lines 50–51
 - text characteristics 53–54
 - graphics 373–399
 - four systems of 374–375
 - ggplot2 package 390–394
 - interactive graphs 394–399
 - identifying points 394
 - iplots package 397–398
 - latticist package 396–397
 - playwith package 394–395
 - rggobi package 399
 - lattice package 375–389
 - graphic parameters 387–388
 - page arrangement 388–389
 - panel functions 381–383
 - variables 379–380, 383–387
 - parameters 387–388
 - graphs
 - axis and text options 56–64
 - annotations 62–64
 - axes 57–60
 - legend 60–62
 - reference lines 60
 - titles 57
 - bar plots 120–125
 - for mean values 122–123
 - simple 120–121
 - spinograms 124–125
 - stacked and grouped 121–122
 - tweaking 123–124
 - box plots 133–138
 - parallel 134–137
 - violin variation of 137–138
 - combining 65–72
 - creating 46
 - dot plots 138–140
 - example 48
 - graphical parameters 49–56
 - colors 52–53
 - graph and margin dimensions 54–56
 - symbols and lines 50–51
 - text characteristics 53–54
 - histograms 128–130
 - interactive 394–399. *See also* intermediate graphs
 - identifying points 394
 - iplots package 397–398
 - latticist package 396–397
 - playwith package 394–395
 - rggobi package 399
 - kernel density plots 130–132
 - pie charts 125–128
 - single enhanced 69
 - gray() function 53
 - grep() function 37, 100
 - grid function 374
 - grid package 374, 423
 - grouped bar plots 121–122
 - grouping variables 383–387
 - groups option
 - dot plots 138
 - lattice package 378, 384
 - gsub() function 37
 - GUIs. *See* graphic user interfaces
 - gvlma() function 199
 - gvlma package 193, 199, 423
 - GWAS. *See* genome-wide association studies
 - gzfile() function 36

H

hat statistic 201
HDF5 files. *See* Hierarchical Data Format files
hdf5 package 39, 423
head() function 43
header value 35
heat.colors() function 53
height variable 339
height vector 120
heights option 67
help() or ? function 11
help facilities 11, 16
help.search() or ??
 function 11
help.start() function 11
hexbin() function 272
hexbin package 272, 423
HH package 232, 235–236, 423
Hierarchical Data Format (HDF5) files 39
high-density scatter plots 271–274
high-leverage
 observations 201–202
hist() command 47
hist() function 66
histograms 128–130
 of bootstrapped statistics 306–308
 in ggplot2 plots 390
 in iplots 397
 in lattice plots 375, 377
 in scatterplot matrices 269
 of studentized residuals 195
history() function 12
Hmisc package 38, 59, 370, 423
homoscedasticity 191, 197–198
 regression 190
 statistical assumption 177
horiz option 120
hsv() function 53
hypergeometric
 distribution 97
hypothesis testing 247–249

I

I() operator 178, 181
ibar() function 397
ibox() function 397
identify() function 394–395
IDEs. *See* integrated development environments

id.method option 193, 266
IDPmisc package 273–274
if-else construct 108–109
ifelse construct 109
if-else control structure 108
ihist() function 397
imap() function 397
imosaic() function 397
importing data
 from database management systems 39–41
 from delimited text file 35–36
 from HDF5 files 39
 from the keyboard 34–35
 from Microsoft Excel 36–37
 from netCDF files 39
 from SAS datasets 38
 from SPSS datasets 38
 from Stata datasets 38–39
 via Stat/Transfer application 41–42
 from web pages 39
 from XML files 37
imputation
 multiple 365–369
 simple 371–372
incomplete data. *See* missing data
independence, of errors 177, 190, 196
index.cond option 378
indices in R, 33
infile 17
influencePlot() function 193, 204
influential observations 190, 202, 204
input 13–14, 18
installations, updating 432–433
installed.packages()
 function 16, 432, 433
installing
 packages 116
 R application 7
 setting default CRAN site 407
install.packages() function 16, 407, 432
integrated development environments (IDEs) 403–404
interaction2wt() function, HH package 235–236
interaction.plot()
 function 235, 238

interactions, multiple linear regression with 186–188
interactive graphs 394–399
identifying points 394
iplots package 397–398
lattice package 396–397
playwith package 394–395
rggobi package 399
intermediate graphs 263
 bubble plots 278–279
 correlograms 283–287
 line charts 280–283
 mosaic plots 288
 scatter plots 264–279
 3D, 274–277
 high-density 271–274
 matrices 267–271
ipairs() function, IDPmisc package 274
ipcp() function 397
iplot() function 273, 397
iplots package 394, 397–398
is.datatype() function 84
is.infinite() function 355
is.na() function 79, 355
is.nan() function 355
isoMDS() function 350
isTRUE() operator 77

J

JGR/Deducer GUI 405
jpeg() function 47

K

kernel density estimation 6
kernel density plots 130, 132
key (or auto.key) option 378
keyboards, entering data from 34–35
k-fold cross-validation 214
kmi package 370, 424
Kruskal-Wallis test 168

L

labels, fitting in bar plots 124
labels option 58, 193, 266
lapply() function 103
las option 58
Last() function 406–407
latent variable models 349, 351
LaTeX documents, R code + (Sweave package) 410–415

- lattice package 48, 374–375, 378–381, 424
 - graphic parameters 387–388
 - graphs types 377
 - page arrangement 388–389
 - panel functions 381–383
 - variables
 - conditioning 379–380
 - grouping 383–387
 - lattice package 396–397, 424
 - lavaan package 350, 424
 - layout() function 65–69
 - layout option 378
 - lcda package 350, 424
 - lcm() function 67
 - lmm package 350
 - leadership data frame 86
 - leaps package 211, 424
 - legend() function 60, 132
 - legend option 60
 - legend.plot option 266
 - legends 60–62
 - in bar plots 122
 - in kernel density plots 131–132
 - in lattice plots 384–386
 - in line plots 282–283
 - in mosaic plots 289
 - in scatter plots 264, 273
 - legend.text parameter 122
 - length() function 43, 101, 143, 148
 - level option, cld() function 228
 - leverage value, of observations 190
 - .libPaths() function 15, 407
 - library() function 15–16
 - line() function 59
 - line charts 280–283
 - linear models 253–254, 257
 - assumption, global validation of 199
 - versus nonlinear model 183
 - linear regression
 - multiple 184–186
 - simple 179–181
 - linearity 196–197
 - regression 190
 - statistical assumption 177
 - lines
 - graphical parameters 50–51
 - reference 60
 - lines() function 123, 129–130, 282
 - link function 315
 - list() function 32
 - lists 32–33
 - list-wise deletion 81
 - listwise deletion 364–365
 - lm() function 178–179, 184–188
 - lme4 package 239
 - lmer() function, lme4 package 239
 - lmfit list object 18
 - lmPerm package 424
 - load() function 12, 13
 - loadhistory() function 12
 - location option 60, 62
 - locator() function 132
 - loess() function 266
 - log() function 93
 - log10() function 93
 - logical operators 76
 - logistic regression 175, 314, 315, 317–323
 - extensions 323
 - fitting 317–320
 - interpreting parameters 320–322
 - overdispersion 322–323
 - lognormal distribution 97
 - logregperm package 424
 - long data format 116
 - longitudinalData package 370, 424
 - looping, repetition and 107–108
 - lower.panel option 285
 - lowess() function 265–266
 - ls() function 12, 43
 - lsa package 350, 424
 - .ls.objects() function 430
 - ltm package 424
 - lty option 58, 378
 - lty parameter 51
 - lty.smooth option 183–184, 268–269
 - lubridate package 83, 424
 - lwd option 378
 - lwd parameter 51
- M**
-
- mad() function 94
 - mai parameter 55
 - main option 378, 390
 - Mallows Cp statistic 211
 - MANCOVA. *See* multivariate analysis of covariance
 - Mann-Whitney U test 166–167
 - MANOVA. *See* multivariate analysis of variance
 - manova() function 241
 - MAR. *See* missing at random
 - mar parameter 55
 - margin dimensions 54–56
 - marginplot() function, VIM package 359
 - MASS package 98, 209, 350, 424
 - math annotations 64
 - mathematical functions 93–94
 - matrices 24–26
 - applying functions to 102–103
 - matrix algebra with R 419–420
 - of scatter plots 267–271
 - matrix function 24–25
 - matrixplot() function, VIM package 358
 - max() function 95
 - MCAR. *See* missing completely at random
 - md.pattern() function, mice package 357
 - MDS. *See* multidimensional scaling
 - mean() function 94, 102, 105, 356
 - mean substitution 371
 - mean values, bar plots for 122–123
 - median() function 94
 - melt() function 114–115
 - melting 114
 - merging datasets 85–86
 - adding columns 85
 - adding rows 86
 - metafile format, Windows 47
 - method option 390
 - mfrow parameter 65
 - MI. *See* multiple imputation
 - mi package 365, 369
 - mice() function 366
 - mice package 353, 357, 365, 369
 - Microsoft Excel, importing data from 36–37
 - min() function 95
 - minor.tick() function 59
 - minus sign 87
 - missing at random (MAR) 354
 - missing completely at random

(MCAR) 354
 missing data 352–372
 approaches for dealing with
 incomplete data
 363–364
 complete-case analysis
 364–365
 exploring patterns 356–361
 exploring missing data
 visually 357–359
 missing values 357,
 360–361
 identifying 355–356
 multiple imputation
 365–369
 pairwise deletion 370–371
 rational approaches for
 correcting 363–364
 simple imputation 371
 steps in dealing with
 353–355
 understanding sources and
 impact of 362–363
 missing values 79–81
 excluding from analyses
 80–81
 recoding values to missing 80
 mix package 370
 mixed-model ANOVA
 design 221
 mlogit package 425
 mode() function 43
 MODULUS operator 75
 mosaic() function, vcd
 library 288
 mosaic plots 288
 mosaicplot() function 288
 mtcars data frame 29, 46, 377
 mtext() function 59, 62
 multcomp package 227, 231,
 425
 multicollinearity 199–200
 multidimensional scaling
 (MDS) 350
 multiline comments 33
 multiple comparisons
 nonparametric 169
 parametric
 one-way ANCOVA 231
 one-way ANOVA 227–229
 multiple graphs per page. *See*
 page arrangement of
 graphs
 multiple imputation (MI)
 365, 369

multiple linear regression 175,
 179, 184–188
 multiple regression 184
 multivariate analysis
 of covariance
 (MANCOVA) 222
 multivariate analysis of variance
 (MANOVA) 222,
 239–243
 assessing test
 assumptions 241–242
 robust 242–243
 multivariate normal data,
 generating 98–99
 multivariate regression 175
 mvnml package 370, 425
 mvoutlier package 242, 425
 mvnorm() function 98

N

NA. *See* not available
 names() function 43, 79
 names.arg argument 123
 NaN. *See* not a number
 na.omit() function 81, 364
 na.rm option 80
 ncdf package 39, 425, 431
 ncdf4 package 39, 425, 431
 nchar() function 100
 ncvTest() function 193, 197
 netCDF files. *See* network
 Common Data Form files
 netCDF library, Unidata 39
 network Common Data Form
 (netCDF) files 39
 new option 70
 nFactors package 349, 425
 NHST. *See* null hypothesis
 significance testing
 nlme package 239
 NMAR. *See* not missing at
 random
 nonlinear model, versus linear
 model 183
 nonlinear regression 175
 nonparametric regression 175
 nonparametric tests 166–170
 nonstochastic imputation
 371–372
 no.readonly option 49
 normal data, generating
 multivariate 98–99
 normal distribution
 functions 97–98
 normal Q-Q plot 190
 normality 193, 196
 regression 190
 statistical assumption 177
 not a number (NaN) 79, 355
 not available (NA) 79, 355
 not missing at random
 (NMAR) 355
 notched box plots 134–135
 noweb file 411, 414
 npmc package 425
 nrows option 430
 null hypothesis significance
 testing (NHST) 246

O

obcConnectExcel()
 function 37
 objects 23
 oblique rotation 340, 345
 observations
 deleting 205
 deleting with na.omit()
 function 81
 selecting 87–88
 unusual 200–204
 high leverage 201–202
 influential 202–204
 outliers 200–201
 ODBC interface. *See* Open
 Database Connectivity
 interface
 odbConnect() function 40
 ODF. *See* Open Documents
 Format
 odfTable() function 415
 odfWeave package 410–415
 OLS regression. *See* ordinary
 least squares regression
 one-way analysis of covariance
 (ANCOVA) 230–233
 assessing test
 assumptions 232
 visualizing results 232–233
 one-way analysis of variance
 (ANOVA) 225–230
 assessing test
 assumptions 229–230
 multiple comparisons
 227–229
 power and effect size
 252, 257
 terminology 220–221
 one-way between-groups

- ANOVA design 220
 - one-way within-groups ANOVA design 220–221
 - Open Database Connectivity (ODBC) interface 39–40
 - Open Documents Format (ODF) 410
 - openMx package 350, 425
 - OpenOffice documents, creating reports with odfWeave package 415–417
 - OpenOffice ODT file 415
 - options() function 12, 245
 - OR operator 77
 - order() function 84, 107
 - order of formula terms 223, 225
 - order.clusters() function 17
 - ordered factor 31
 - order.single() function, gclus package 270
 - ordinary least squares (OLS) regression 175–218
 - cross-validation 213–215
 - fitting regression models with lm() function 178–179
 - multiple linear regression 184–186
 - multiple linear regression with interactions 186–188
 - polynomial regression 181–184
 - regression diagnostics 188–207
 - simple linear regression 179–181
 - variable (model) selection 207–213
 - variable importance 215–217
 - orthogonal rotation 340, 345
 - outlier observation 190
 - outliers 200–201
 - outlierTest() function 193, 200, 230
 - output
 - graphic 13–14
 - text 13
 - using as input 18
- P**
-
- packages 14–17
 - analytic, for large datasets 431
 - description of 15–16
 - function of 16–17
 - installing 16
 - list of 421–428
 - loading 16
 - updating 432
 - page arrangement of graphs
 - with base graphics 65–71
 - with lattice graphs 388–389
 - pairs() function 267
 - pairwise deletion 370–371
 - pamm package 261
 - pan package 370
 - panel functions 381, 383
 - panel option 285, 378
 - panel.abline() function 383
 - panel.ellipse option 285
 - panel.grid() function 381
 - panel.lmline() function 381
 - panel.loess() function 383
 - panel.minmax option 285
 - panel.pie option 285
 - panel.pts option 285
 - panel.rug() function 381
 - panel.shade option 285
 - panel.txt option 285
 - panel.xyplot() function 381, 383
 - par() function 49–50, 53–56, 59–61, 64–67, 69–70, 124
 - parallel analysis approach 335
 - parallel box plots, comparing groups with 134–137
 - parameters, graphic 387–388
 - paste() function 86, 100
 - pasteecs package 425
 - PCA. *See* principal components analysis
 - pch parameter 49–51
 - pdf() function 47
 - pedantics package 261
 - period character 33
 - PHYS variable 334
 - pie charts 125–128
 - pie3D() function 127
 - piface package 260, 425
 - pin parameter 55
 - playwith package 394–396, 425
 - plot() function 47–48, 59, 282, 388–389
 - plot3d() function, rgl package 277
 - plotmath() function 64
 - plotmeans() function, gplots package 226, 235
 - plot.new() function 71
 - plotrix package 127
 - plots, power analysis 258–259
 - plotting symbols, with pch parameter 51
 - pmm. *See* predictive mean matching
 - png() function 47
 - points, identifying 394
 - poisson regression 175, 324–330
 - fitting 325
 - interpreting parameters 326
 - overdispersion 327
 - robust 330
 - with varying time periods 329
 - zero-inflated 329
 - poLC package 350
 - poLCA package 425
 - polygon() function 130
 - polynomial regression 175, 179, 181, 184
 - pool() function 366
 - pos option 58, 62
 - position option 379
 - postscript() function 47
 - power analysis 246–262, 313
 - hypothesis testing 247–249
 - implementing with pwr package 249–258
 - ANOVA 252–253
 - Chi-square tests 255–256
 - correlations 253
 - effect size benchmarks 257–258
 - linear models 253–254
 - tests of proportions 254–255
 - t-tests 250–252
 - plots 258–259
 - specialized packages 260–262
 - powerGWASinteraction package 261
 - powerpkg package 261
 - powerSurvEpi package 261
 - predict() function 179
 - predictive mean matching (pmm) 368
 - pretty() function 101
 - primary variables, in lattice functions 376
 - principal() function 333, 336, 341

- principal components analysis (PCA) 331–351
 - other latent variable models 349–351
 - principal components extracting 339
 - obtaining scores 341–342
 - rotating 339–341
 - selecting number of components to extract 335
 - princomp() function 333
 - print() statement 412
 - probability functions 96–99
 - generating multivariate normal data 98–99
 - setting seed for random number generation 98
 - profr package 430
 - prooftools package 430
 - proportions, tests of 254–255
 - ps parameter 54
 - pseudo-random numbers 96–99
 - multivariate normal 98–99
 - probability distributions 96–97
 - setting the seed 98
 - psych package 333, 426
 - publication-quality reports
 - OpenOffice documents with odfWeave package 415–417
 - typesetting with Sweave package (R code + LaTeX documents) 410–415
 - pundat dataframe 40
 - pwr package, implementing
 - power analysis with ANOVA 252–253
 - Chi-square tests 255–256
 - correlations 253
 - effect size benchmarks 257–258
 - linear models 253–254
 - tests of proportions 254–255
 - t-tests 250–252
 - pwr.2p2n.test function 250
 - pwr.2p.test() function 250, 254
 - pwr.anova.test() function 250, 252, 257
 - pwr.chisq.test() function 250, 255
 - pwr.f2.test() function 250, 253
 - PwrGSD package 261
 - pwr.p.test function 250
 - pwr.r.test() function 250, 253, 258
 - pwr.t2n.test function 250, 252
 - pwr.t.test() function 250
- Q**
-
- q() function 9, 12
 - qcc package 426
 - qqplot() function 390–393
 - qqPlot() function 193
 - quadratic regression 184
 - quantile() function 95, 105
 - quartzFonts() function 54
- R**
-
- R AnalyticFlow GUI 405
 - R CMD BATCH statement 17
 - R code + LaTeX documents (Sweave package), typesetting with 410–415
 - R Commander GUI 405
 - R Data Import/Export manual 34
 - .R extension 17
 - R GUI Generator (RGG) 405
 - R GUI Projects page 405
 - R icon 8
 - R language
 - batch processing 17
 - example 18
 - getting help 11
 - getting started 8–10
 - input 13
 - large datasets and 18
 - obtaining and installing 7
 - output
 - graphic 13–14
 - text 13
 - using as input 18
 - packages 14–17
 - description of 15–16
 - function of 16–17
 - installing 16
 - loading 16
 - reasons to use 5–7
 - workspace 11–13
 - R programming, common mistakes in 17
 - R statements 107
 - R_HOME value 406
 - R_HOME/etc directory 406
 - R2HTML package 411
 - rainbow() function 53, 127
 - Random Access Memory (RAM), storing data outside of 430–431
 - random number generation, setting seed for 98
 - random samples 89
 - randomized block design 223
 - randomLCA package 350, 426
 - range() function 95
 - Rattle (for data mining) GUI 405
 - rbind() function 43, 86, 232
 - Rcmdr package 277, 426
 - Rcpp package 430
 - RCurl package 37
 - .RData file 13
 - readLines() function 37
 - read.spss() function 38
 - read.ssd() function 38
 - read.table() function 35–36, 430
 - read.xlsx() function 37
 - recode() function 78
 - recodevar() function 78
 - recoding, values to missing 80
 - Red R GUI 405
 - reference lines 60
 - regression 173–218
 - ANOVA as 243–245
 - corrective measures 205–207
 - deleting observations 205
 - variables 205–207
 - diagnostics 188–200
 - enhanced approach 192–198
 - global validation of linear model assumption 199
 - multicollinearity 199–200
 - typical approach 189–192
 - measuring
 - performance 213–218
 - cross-validation 213–215
 - relative importance 215–218
 - OLS 177–188
 - fitting regression models with lm() function 178–179
 - multiple linear regression 184–186
 - polynomial regression 181–184

simple linear regression 179–181
 selecting model 207–213
 comparing models 208–209
 variable selection 209–213
 unusual observations 200–204
 high leverage 201–202
 influential 202–204
 outliers 200–201
 varieties of 174–177
 regression diagnostics 188–207
 regsubsets() function 210
 relaimpo package 215
 relative importance, of variables 215–218
 relative weights 216
 relative widths 67
 relweights() function 216
 rename() function 78
 rep() function 101
 repeated measures, ANOVA 221, 237–239
 repetition, and looping 107–108
 reports, publication-quality
 OpenOffice documents with
 odfWeave package 415–417
 typesetting with
 Sweave package (R code + LaTeX
 documents) 410–415
 resampling statistics 89, 291–303
 reshape package 78–79, 113–116, 426
 casting 114–116
 melting 114
 residplot() function 195
 residual versus leverage plot 190
 residuals() function 179
 restructuring data
 reshape package 113–116
 casting 114–116
 melting 114
 transpose 112
 results option, Sweave 412
 RGG. *See* R GUI Generator
 rggobi package 394, 399, 426
 rgl package 277, 426
 .Rhistory file 13
 RJDBC package 426

Rkward GUI 405
 rm() function 12, 43, 430
 rms package 426
 RMySQL package 41, 431
 robust MANOVA 242–243
 robust package 427
 Robust regression 175
 RODBC package 36–40, 427–431
 ROracle package 41, 427–431
 rotating
 3D scatter plots 277
 factors 345–348
 principal components 339–341
 round() function 93
 .Rout extension 17
 rownames 23
 row.names option 30, 36
 rows, adding 86
 RPostgreSQL package 41, 431
 Rprof() function 430
 .Rprofile file 406
 Rprofile.site file 406–407, 433
 rrcov package 242, 427
 RSiteSearch() function 11
 SQLite package 41, 431
 R-square statistic 214
 rug plots 6, 129
 runif() function 98

S

sample() function 89
 sample size quantity 248
 samples, random 89
 sampling package 427
 sapply() function 103, 106
 sas.get() function 38
 save() function 12
 savehistory() function 12
 save.image() function 12
 scalar values 33
 scale() function 96, 104, 215
 scales option 379
 scan() function 430
 scatter plots 264–279
 3D 274–277
 high-density 271–274
 matrices 267–271
 scatter3d() function, Rcmdr
 package 277
 scatterplot() function 183, 193, 266
 scatterplot3d() function,
 scatterplot3d
 package 274
 scatterplot3d package 274, 427
 scatterplotMatrix()
 function 6, 184, 193, 268
 scores
 factors 349
 principal components 342
 scores option 341
 scree() function 333
 sd() function 94
 search() command 16
 seed, setting for random
 number generation 98
 SELECT statements 40, 89
 selecting
 observations 87–88
 variables 86
 SEM. *See* structural equation
 modeling
 sem package 350, 427
 seq() function 101
 SeqKnn package 370, 427
 sequence, in for
 statements 107–108
 set.seed() function 98
 setwd() function 12–13
 Sexpr statement 412
 shadow matrix 360
 shape option 390
 show.settings() function 387
 shrinkage() function 214
 side option 58, 62
 signif() function 93
 significance levels 248
 simple imputation 371–372
 simple linear regression 175, 179, 181
 simulate option 196
 sin() function 93
 singer dataset 392
 sinh() function 93
 sink() function 13, 15
 site initialization file 406
 size option, in ggplot2 390
 sm packages 130–131, 427
 sm.density.compare()
 function 130, 132
 smoothScatter() function, base
 package 272, 274
 software, updating 432–433
 sorting, data management 84–85

- source() function 13–15
 - speedglm package 431
 - sphericity 239
 - spine() function 124
 - spinning 3D scatter plots 277
 - spinograms 124–125
 - split() function, in bigtabulate package 431
 - split option, in lattice plots 379, 388
 - splitting output 13
 - spread option 183, 269
 - spreadLevelPlot() function 193, 197, 206
 - SPSS datasets. *See* Statistical Package for the Social Sciences datasets, importing data from spss.get() function 38
 - SQL. *See* Structured Query Language
 - sqldf() function 89
 - sqldf package 90
 - sqlDrop() function 40
 - sqlFetch() function 40
 - sqlQuery() function 40
 - sqlSave() function 40
 - sqrt() function 93
 - ssize.fdr package 261
 - stacked bar plots 121–122
 - standardizing data 96
 - startup environment customizing 406–407
 - Stata datasets exporting data to 409 importing data from 38–39
 - statements, in R 107
 - state.x77 dataset 184
 - Statistical Analysis System (SAS) datasets, importing data from 38
 - statistical applications, exporting data for 409
 - statistical functions 94–96
 - Statistical Package for the Social Sciences (SPSS) datasets, importing data from 38
 - Stat/Transfer application, importing data via 41–42
 - stepAIC() function 209
 - stepwise regression 209–210
 - stop() function 111
 - storage, outside of RAM 430–431
 - str() function 31, 43
 - stringsAsFactors option 36
 - strip option 379
 - strsplit() function 100, 106
 - structural equation modeling (SEM) 349
 - Structured Query Language (SQL) 89–90
 - sub() function 100
 - sub option 390
 - subset() function 88–89
 - subsets() function 211
 - subsetting datasets 86–89 selecting observations 87–88 selecting variables 86 subset() function 88–89
 - substr() function 100
 - sum() function 81, 95, 356
 - sum of squares Type I (sequential) 224 Type II (hierarchical) 224 Type III (marginal) 224
 - summary() function 31, 179, 188
 - summary.aov() function 241
 - summaryRprof() function 430
 - Sweave() function 411, 414
 - Sweave package (R code + LaTeX documents), typesetting with 410–415
 - switch construct 109–110
 - symbols graphical parameters 50–51 plotting with pch parameter 51 in R formulas 178
 - symbols() function 278
 - Sys.Date() function 82
 - Sys.getenv() function 406
 - system.time() function 430
- T**
-
- t() function 112
 - tab-delimited files 36
 - table() function 120–121, 431
 - tabulating missing values 357
 - tail() function 43
 - tan() function 93
 - tanh() function 93
 - tapply() function 431
 - tck option 58
 - terrain.colors() function 53
 - Test of proportions 257
 - TeX file 414
 - text() function 62–64
 - text characteristics, graphical parameters 53–54
 - text files, delimited exporting data to 408 importing data from 35–36
 - text options annotations 62–64 legend 60–62 titles 57
 - text output 13
 - text.col() option 61
 - text.panel option 285
 - tick marks 59–60
 - tick.ratio 59
 - tiff() function 47
 - time-series regression 175
 - time-stamping data 82
 - title() function 57, 123, 132
 - title option 60
 - titles 57
 - tolower() function 100
 - topo.colors() function 53
 - toupper() function 100
 - transform() function 76
 - transformations, of variables 205–207
 - transmission disequilibrium test (TDT) 261
 - transpose 112
 - trellis.par.get() function 387
 - trellis.par.set() function 387
 - trt2times variable 245
 - trunc() function 93
 - t-tests dependent groups 165–166 independent groups 164–165 power 250–252, 257
 - twiddler package 405
 - two-way ANOVA design 221
 - two-way factorial ANOVA 223, 234–236
 - type conversions 83–84
 - Type I (sequential) approach 224
 - Type II (hierarchical) approach 224
 - Type III (marginal) approach 224
 - type option 379
- U**
-
- unbalanced design 220
 - Unidata netCDF library 39

uniquenesses 337
 unz() function 36
 update() function 379
 update.packages()
 function 16, 432
 updating installations 432–433
 upper.panel option 285
 url() function 36
 user-written functions
 109, 111

V

validation, of linear model
 assumption 199
 value labels 42
 var() function 94
 variable <- expression
 statement 75
 variable labels 42
 variable[condition]
 <- expression
 statement 77
 variables
 adding or deleting 207
 conditioning 379–380
 creating new 75–76
 data frames 27
 excluding 86–87
 grouping 383–387
 recoding 76–78
 renaming 78–79

 selecting 86, 209–213
 in stepwise regression
 209–210
 in subsets regression
 210–213
 transforming 205–207
 variance inflation factor
 (VIF) 200
 varimax rotation 340
 varwidth option 134
 vcd package 18–19, 120, 124,
 288, 427
 vcov() function 179
 vectors 24
 vegan package 427
 VIF. *See* variance inflation factor
 vif() function 193, 200
 vignette() function 11
 VIM package 353, 359, 428
 violin plots 137–138
 vioplot() function 137
 vioplot package 137

W

warning() function 111
 webscraping 37
 which() function 88
 while loop 108
 widths option 67
 Wilcoxon rank sum test
 166–167

Wilks.test() function, rrcov
 package 242
 Windows metafile format 47
 windowsFont() function 54
 win.metafile() function 47
 with() function 28, 30, 77, 366
 within() function 77
 workspace 11, 13
 write.foreign() function 409
 write.table() function 408
 write.xlsx() function 409

X

xaxt option 57
 xfig() function 47
 xlab option 120, 379, 390
 xlim option 379, 390
 xlsx package 37, 409, 428
 XML files, importing data
 from 37
 XML package 37, 428
 xtable() function 412, 415
 xtable package 412
 xyplot() function 381
 xzfile() function 36

Y

yaxt option 57
 ylab option 120, 379, 390
 ylim option 379, 390

