

# Subject index

References to the HELP examples are denoted in *italics*.

- 3-D plot, 188
- 95% confidence interval
  - mean, 72
  - proportion, 74
- absolute value, 38
- accelerated failure time model
  - frailty, 147
- access elements
  - R, 276
- access files, 38
- access variables, 15
- add lines to plot, 195
- add normal density, 196
- adding marginal rug plot, 197
- adding straight line, 192
- adding text, 198
- age variable, 90, 287
- agreement, 77
- AIC, 103, 129
- Akaike information criterion (AIC), 103, 129
- alcohol abuse, 288
- alcoholic drinks
  - HELP dataset, 288
- analysis
  - strata, 260
  - analysis of variance
    - one-way, 101
    - two-way, 102, 124
  - analytic power calculations, 223
  - and operator, 61
  - angular plot, 189
  - annotating datasets, 58
  - ANOVA tables, 102
  - Aotearoa (New Zealand), 269
  - Apple R FAQ, 270
  - arbitrary quantiles, 71
  - area under the curve, 190
  - ARIMA model, 146
  - arrays, 54, 60
    - extract elements, 277
  - arrows, 199
  - assignment operators in R, 274
  - association plot, 187
  - attributable risk, 74
  - attributes
    - R, 279
  - AUC (area under the curve), 190
  - Auckland, University of, 269
  - autoregressive integrated moving average

- time series model, 146
- available datasets
  - R, 284
- average number of drinks
  - HELP dataset, 288
- axes
  - labels, 203
  - multiple, 182
  - omit, 205
  - range, 203
  - style, 203
  - values, 203
- barplot, 183
- Base SAS, 251
- baseline interview, 285
- basic concepts
  - SAS, 257
- batch mode, 267, 273
- Bates, Douglas, 269
- Bayesian inference
  - task view, 237
- Bayesian information criterion, 103
- BCA intervals, 73
- best linear unbiased predictors, 143
- beta function, 39
- bias-corrected and accelerated intervals, 73
- BIC, 103
- binomial family, 136
- bivariate loess, 148
- bivariate relationships, 85
- BMDP files, 8, 12
- BMP
  - exporting, 208
- book website, 3
- Boolean
  - operations, 19, 34, 61
  - R, 276
- Boolean operator, 218
- bootstrap sample, 28
- bootstrapping, 73
- Bourque, Susan, 3
- box around plots, 201
- boxplot, 185
  - parallel, 172
  - side-by-side, 185
- Breslow estimator, 147
- “broken stick” models, 144
- bug reports, 284
- c statistic, 137
- calculate derivatives, 41
- calculus, 41
- calling functions
  - R, 279
- case sensitivity, 2, 271
- case-sensitivity, 267
- categorical covariate parameterization, 98
- categorical data
  - plot, 187
- categorical from continuous, 21
- categorical predictor, 98
- categorical tables, 86
- categorical variable, 65
- Cauchy link function, 136
- censored data, 147, 191
- Center for Epidemiologic Studies Depression (CESD) scale, 287
- centering, 72
- CESD, 59, 60
- cesd variable, 58, 287
- cesdtv variable, 166
- Chambers, John, 269
- change working directory, 37
- character translations, 20
- character variable, *see* string variable
- characteristics
  - test, 75
- characters
  - plot, 193
- chi-square statistic, 78
- Cholesky decomposition, 143
- choose function, 39
- circadian plot, 189
- circular plot, 189
- class methods
  - R, 279
- class variable, 65
- class variables, 99
  - ordering of levels, 99
- classification, 244
- clear graphics settings, 122
- clinical trial, 285
- closing a graphic device, 209
- cluster analysis
  - task view, 239
- clustering
  - task view, 248
- cocaine, 288
- Cochran-Mantel-Haenszel test, 78
- code examples

- downloading, 3
- coefficient of variation, 73
- coercing, o numeric34
  - character variable from numeric, 19
  - dataframes into matrices, 278
  - date from character, 7
  - factor variable from numeric, 22
  - matrices into dataframes, 278
  - numeric from character, 7
  - numeric from integer, 18
- color
  - palettes, 204
  - selection, 204
- column width, 56
- comma separated files (CSV), 8
- command file
  - SAS, 259
- command history
  - R, 270
- comments
  - R, 277
- comparison
  - floating point variables, 40
  - operators, 276
- complementary log-log link function, 136
- complex fixed files, 230
- complex fixed format files, 7
- complex survey design, 248
- Comprehensive R archive network, 269
- concatenate, 227
- concatenating datasets, 30
- concatenating strings, 19
- conditional execution, 51
- conditional logistic regression model, 141
- conditioning plot, 187, 212
- confidence intervals
  - for predicted observations, 113
  - for the mean, 112
- confidence level
  - default, 280
- confidence limits
  - for individual (new) observations, 110
  - for the mean, 111
- contingency table, 77
  - plot, 187
- contingency tables, 86
- contour plot, 188
- contrasts, 133
- control flow, 51
- control structures, 51
- controlling graph size, 201
- controlling Type-I error rate, 105
- convergence diagnosis for MCMC, 237, 239
- converting datasets
  - long (tall) to wide format, 30
  - wide to long (tall) format, 29
- Cook's Distance, 108
- coordinate systems (maps), 234
- correlated binary variables
  - generating, 229
- correlated data, 169
  - regression models, 141
- correlated residuals, 143
- correlation
  - Kendall, 76
  - matrix, 85, 219
  - Pearson, 76
  - Spearman, 76
- cosine function, 39
- count models
  - goodness of fit, 138
  - negative binomial regression, 139, 158
  - Poisson regression, 138, 155
  - zero-inflated negative binomial, 139
  - zero-inflated Poisson regression, 138, 157
- covariance matrix, 114, 169
- Cox proportional hazards model, 147, 179
  - frailty, 147
- CPU time, 36
- CRAN, 269
  - task views, 282
- create
  - ASCII datasets, 13
  - categorical variable from continuous, 21
  - categorical variable using logic, 22
  - datasets for other packages, 12
  - date variable, 34
  - files for other packages, 12
  - lagged variable, 26
  - numeric variable from string, 18
  - observation number, 25
  - recode categorical variable, 21
  - string variable from numeric, 17
  - time variables, 35
- create matrix, 42
- creating factors, 99
- Cronbach's  $\alpha$ , 239
- cross-classification table, 63, 77
- cross-validation
  - generalized, 148

- crosstabs, 77, 86
- CSV (comma-separated) files, 8
- cumulative density function
  - area to the left, 45
- Dalgaard, Peter, 269
- dashed line, 204
- data
  - display, 16
  - entry, 11
  - input, 56
  - output, 56
- data display, 261
- data generation, 51
- data step
  - range of variables, 54
  - repeat steps for a set of variables, 54
  - SAS, 258
- data structures
  - R, 274
- dataframe
  - comparison with matrix, 278
  - remove from workspace, 278
- dataframes, 274
  - comparison with column bind, 278
- detaching, 15
  - R, 277
- dataset
  - comments, 17
  - HELP study, 287
  - other packages, 8
  - subset, 260
- datasets, 3
  - R, 284
- date and time variable
  - create date, 34
- date and time variables
  - create time, 35
  - extract month, 35
  - extract quarter, 35
  - extract weekday, 34
  - extract year, 35
- date variable
  - reading, 7
- dayslink variable, 95, 287
- DBF files, 8, 12
- debugging, 255
- decimal representation, 40
- default confidence level, 280
- defining functions
  - R, 280
- density functions, 45
  - generate random, 47
  - probability, 47
  - quantiles, 47
- density plot, 84, 93, 189
- depressive symptoms, 58
- derivatives, 41
- derived variable, 60, 61
- design matrix, 114, 129, 151
  - specification, 98
- design weights, 248
- detach
  - dataframes, 15, 124
  - datasets, 278
  - packages, 15, 163, 278
- determinant, 44
- detoxification, 285
- detoxification program, 287
- deviance, 136
- deviance tables, 102
- dffits, 109
- diagnostic agreement, 75, 190
  - ROC curve, 215
- diagnostic plots, 109, 122
- diagnostics from linear regression, 120
- diagonal elements, 43, 44
- difference in log-likelihoods, 102
- differences between SAS and R, 2
- directory delimiter, 6
- directory structure in R, 6
- directory structure in SAS, 6
- dispersion parameter, 159
- display data, 16, 57
- display format, 261
- displaying information about objects, 278
- displaying model results, 13
- distribution
  - empirical probability density plot, 189
  - normal, 47
  - q-q plot, 186
  - quantile, 46
  - quantile-quantile plot, 186
  - stem plot, 184
- distributions
  - probability, 45
- DocBook document type definition, 15
- document type definition, 15
- documentation
  - SAS, 257
- downloading
  - code examples, 3

- R, 270
- drinks of alcohol
  - HELP dataset, 288
- drinkstat variable, 61
- dropping variables, 33
- drugrisk variable, 219, 287
- DTD, 15
- editing data, 11
- efficiency of vector operations, 51
- efficient programming, 259
- Efron estimator, 147
- eigenvalues and eigenvectors, 44
- elapsed time, 36
- empirical density plot, 93
- empirical power calculations, 225
- empirical probability density plot, 189
- empirical variance, 145, 176
- entering data, 11
- environment variables
  - Windows, 273
- Epi Info files, 8
- estimated density plot, 93
- etiquette
  - R, 284
- exact logistic regression model, 137
- exact test of proportions, 79
- example code
  - downloading, 3
- excess zeroes, 138, 139
- exchangeable working correlation, 145
- execute command in operating system, 36
- execution
  - conditional, 51
- expected cell counts, 89
- exponentiation, 38
- export
  - BMP, 208
  - datasets for other packages, 12
  - JPEG, 207
  - PDF, 206
  - PNG, 209
  - postscript, 206
  - TIFF, 209
  - WMF (windows metafile format), 208
- exporting graphs, 206
- expressions
  - R, 274
- expressions in R, 279
- extensible markup language, 10
  - write files, 15
- extract characters from string, 18
- extract from objects, 76, 277
- f1 variables, 61, 239, 287
- factor analysis, 240
- factor levels, 98
- factor variable, 65
- factorial function, 39
- factors
  - ordering of levels, 99
- failure time data, 147
- Falcon, Seth, 269
- false positive, 190
- family
  - binomial, 136
  - Gamma, 136
  - Gaussian, 136
  - inverse Gaussian, 136
  - Poisson, 136
- FAQ
  - Apple R, 270
  - R, 274, 284
  - Windows R, 270
- female variable, 61, 288
- file browsing, 38
- find a string within a string, 19
- find working directory, 37
- Fisher's exact test, 79, 87
- fit model separately by group, 123
- five-number summary, 71
- fixed format files, 6, 7
- floating point representation, 40
- follow-up interviews, 285
- footnotes, 197
- foreign format, 58
- formatted data
  - reading, 261
- formatted data files, 11
- formatting model results, 13
- formatting values of variables, 22
- formula object, 97
- Foundation for Statistical Computing
  - R, 269
- seeFAQ, 274
- Friedman's 'super smoother', 196
- functions
  - defining in R, 280
  - R, 279
- G-rho family of Harrington and Fleming, 95

- g1b variable, 212, 288
- g1btv variable, 165, 176
- GAM, 148
- Gamma family, 136
- gamma function, 39
- Gamma regression, 135
- Gaussian distribution, 45
- Gaussian family, 136
- gender, 288
- gender** variable, 65
- general linear model for correlated data, 141, 169
- generalized additive model, 148, 163
- generalized cross-validation, 148
- generalized estimating equation
  - exchangeable working correlation, 145
  - independence working correlation, 145
  - unstructured working correlation, 145
- generalized estimating equations, 176
- generalized linear mixed model, 145, 178
- generalized linear model, 135, 151
- generalized multinomial model, 140
- generate
  - arbitrary random variables, 50
  - correlated binary variables, 229
  - exponential random variables, 49
  - generalized linear model random effects, 228
  - multinomial random variables, 46
  - multivariate normal random variables, 48
  - normal random variables, 48
  - other random variables, 50
  - pattern of repeated values, 52
  - predicted values, 106
  - random variables, 45
  - residuals, 107
  - sequence of values, 52
  - uniform random variables, 46
- genf** variable, 125
- Gentleman, Robert, 269
- getting help
  - R, 284
- Getting Started with SAS Software, 255
- goodness of fit, 138, 156
  - ROC curve, 215
- graphical settings, 202
- graphical user interface
  - R, 270
- graphics
  - boxplot, 185
  - side-by-side boxplots, 185
  - size, 201
  - task view, 181
- graphics in Unix
  - SAS, 265
- graphs
  - exporting, 206
- greater than operator, 61
- grouping variable
  - linear model, 100
- growth curve models, 144
- guide to packages
  - R, 282
- guidelines
  - R-help postings, 284
- hanging rootogram, 138
- Harrington and Fleming G-rho family, 95
- hat matrix, 108
- Health Evaluation and Linkage to Primary Care (HELP) study, 285
- health survey
  - SF-36, 288
- help
  - R packages, 282
- HELP clinic, 288
- HELP study
  - dataset, 287
  - introduction, 285
  - results, 285
- help system
  - R, 273
- help system in R, 272
- help, getting
  - R, 284
- heroin, 288
- hierarchical clustering, 247
- histogram, 84, 184
- history of commands
  - R, 270
- history of R, 269
- homeless** variable, 87, 151, 288
- homelessness
  - definition, 287
- homogeneity of odds ratio, 78
- honest significant difference, 105, 130
- Hornik, Kurt, 269
- hospitalization, 287
- HTML format, 14
- HTML output
  - SAS, 261, 265

- HTTP
  - reading from URL, 9
- Huber variance, 176
- hypertext markup language format (HTML), 14
- hypertext transport protocol (HTTP), 9
- i1 variable, 61, 155, 288
- i2 variable, 61, 288
- Iacus, Stefano, 269
- id number, 25
- id variable, 288
- identifying points, 200
- identity link function, 136
- if statement, 27, 51
- Ihaka, Ross, 269
- image plot, 188
- import data, 8
- income inequality, 147
- incomplete data, 23
- independence working correlation, 145
- index
  - R, 2
  - SAS, 2
  - subject, 2
- Index of Drug Abuse Consequences, *see* `indtot` variable
- indexing
  - in R, 59, 232
  - matrix, 43
  - vector, 276
- indexing in R, 59
- indicator variable, 98
- indicator variables, 98
- individual level data, 74
- `indtot` variable, 151, 212, 288
- InDUC (Index of Drug Abuse Consequences), 288
- influence, 108
- information matrix, 114
- installing libraries
  - R, 281
- installing R, 269
- integer functions, 40
- interaction, 100
  - linear regression, 116
  - plot, 124
  - testing, 126
  - two-way ANOVA, 124
- interaction plot, 187
- intercept
  - no, 100
- interquartile values, 71
- interval censored data, 191
- introduction to R, 269, 273
- inverse Gaussian family, 136
- inverse link function, 136
- inverse probability integral transform, 50
- invert matrix, 43
- iterative proportional fitting, 139
- jitter, 194
- JPEG
  - exporting, 207
- Kaplan–Meier plot, 191, 213
- Kappa, 77
- keeping variables, 33
- Kendall correlation, 76
- kernel smoother plot, 189
- Kolmogorov-Smirnov test, 80, 92
- Kruskal Wallis test, 80
- labels
  - variable, 23
- large data sets, 259
- LaTeX output
  - SAS, 261, 265
- Learning SAS Programming, 255
- least squares
  - non-linear, 147
- legend, 68
  - adding, 199
- Leisch, Friedrich, 269
- length of string, 19
- less than operator, 61
- leverage, 108
- libraries
  - R, 281
- library
  - help, 282
- likelihood ratio test, 102, 126
- line
  - style, 204
  - types, 204
  - width, 204
- linear combinations of parameters, 106
- linear discriminant analysis, 245
- linear models, 97
  - by grouping variable, 100
  - categorical predictor, 98
  - diagnostic plots, 109

- diagnostics, *120*
  - generalized, *135*
  - interaction, *100, 116*
  - no intercept, *100*
  - parameterization, *98*
  - residuals, *107*
    - studentized, *107*
  - stratified analysis, *100*
  - studentized residuals, *107*
- lines on plot, *195*
- link function
  - cauchit, *136*
  - cloglog, *136*
  - identity, *136*
  - inverse, *136*
  - log, *136*
  - logit, *136*
  - probit, *136*
  - square root, *136*
- linkage to primary care, *287*
- `linkstatus` variable, *95, 288*
- Linux installation
  - R, *270*
- list files, *38*
- lists
  - extract elements, *76, 277*
- local polynomial regression, *196*
- locating points, *200*
- loess
  - bivariate, *148*
- log
  - base 10, *38*
  - base 2, *38*
  - base e, *38*
- log file
  - SAS, *259*
- log link function, *136*
- log scale, *205*
- log-likelihood, *102*
- log-linear model, *139*
- log-rank test, *95*
- logic, *22*
- logical expressions, *21, 22*
- logical operator
  - R, *276*
- logical operators, *21*
- logistic regression, *135, 151*
  - c statistic, *137*
  - Nagelkerke  $R^2$ , *137*
  - ROC curve, *215*
- logit link function, *136*
- lognormal regression, *135*
- logrank test, *81*
- long (tall) to wide format conversion, *30*
- longitudinal regression
  - reshaping datasets, *165*
- looping, *51, 166*
- lower to upper case conversions, *20*
- lowess, *148, 163, 195*
- Lumley, Thomas, *269*
- M estimation, *149*
- machine precision, *40*
- Macintosh R FAQ, *270*
- macros
  - SAS, *73*
- MAD regression, *149*
- Maechler, Martin, *269*
- mailing list
  - R-help, *284*
- make variables available, *15*
- manipulate string variables, *18, 19*
  - remove spaces, *20*
- Mantel–Haenszel test, *78*
- map plotting, *230*
- maps
  - coordinate systems, *234*
- margin specification, *202*
- marginal plot, *197*
- Markov chain Monte Carlo, *237*
- mathematical constants, *39*
- mathematical expressions, *68, 198*
- mathematical functions
  - absolute value, *38*
  - beta, *39*
  - choose, *39*
  - exponential, *38*
  - factorial, *39*
  - gamma, *39*
  - integer functions, *40*
  - log, *38*
  - maximum value, *38*
  - mean value, *38*
  - minimum value, *38*
  - natural log, *38*
  - permute, *39*
  - square root, *38*
  - standard deviation, *38*
  - sum, *38*
  - trigonometric functions, *39*
- mathematical programming and optimization

- task view, 41
- mathematical symbols
  - adding, 198
- matrices
  - covariance, 114
  - create, 42
  - extract elements, 277
  - indexing, 43
  - inversion, 43
  - large, 42
  - overview, 42
  - R, 276
  - sparse, 42
  - transposition, 42
- matrix
  - design, 114
  - graphs, 110
  - hat, 108
  - indexing, 277
  - information, 114
  - multiplication, 276
- matrix multiplication, 49, 114
- matrix plots, 190
- maximum, 71
- maximum number of drinks
  - HELP dataset, 288
- maximum value, 38
- MCMC, 237
- McNemar's test, 79
- mcs variable, 85, 288
- mean, 69, 72
  - trimmed, 71
- mean value, 38
- median, 71
- median regression, 149
- medical problems, 287
- merging datasets, 31
- metadata
  - R, 279
- methods
  - R, 279
- Microsoft rtf format, 207
- minimum, 71
- minimum absolute deviation regression, 149
- minimum value, 38
- Minitab files, 8
- missing data, 23, 60
- missing values, 224
- mixed model, 142
  - logistic, 145
  - logistic regression, 178
- model
  - comparisons, 129
  - diagnostics, 120
  - selection, 129
  - specification, 100, 116
- model comparisons, 102
- month variable, 35
- mosaic plot, 187
- motivational interview, 285
- moving average autoregressive time series
  - model, 146
- multilevel models, 144
- multinomial logit, 161
- multinomial model
  - generalized, 140
  - nominal outcome, 140
  - ordered outcome, 140
- multinomial random variables, 46
- multiple comparisons, 105, 130
- multiple plots per page, 202
- multiple y axes, 211
  - scatterplot, 182
- multiplication
  - matrix, 49
- multivariate statistics
  - task view, 239
- multiway tables, 78
- Murdoch, Duncan, 269
- Murrell, Paul, 55, 269
- Nagelkerke  $R^2$  for logistic regression, 137
- named arguments in R, 279, 280
- names and variable types, 16
- native data files, 11
- native files, 5
- negative binomial model
  - zero-inflated, 139
- negative binomial regression, 139, 158
- nested quotes, 23
- new users
  - R, 273
- New Zealand (Aotearoa), 269
- NIAAA, 285
- NIDA, 285
- NLP optimization, 41
- no intercept, 100
- non-linear least squares, 147
- non-parametric tests, 80, 92
- normal density, 196
- normal distribution, 45, 47, 67
- normal random variables, 48

- normalized residuals
  - mixed model, 143
- normalizing, 72
- notched boxplot, 186
- number of digits to display, 14
- numeric from string, 18
- objects
  - displaying, 278
  - R, 274
- observation number, 25
- Octave files, 8
- odds ratio, 74, 89
  - homogeneity, 78
- ODS
  - parameter estimates as dataset, 154
- omit axes, 205
- one-way analysis of variance, 101
- open-source, 1
- operating system
  - change working directory, 37
  - execute command, 36
  - find working directory, 37
  - list files, 38
- optimization, 41
- optimization and mathematical programming
  - task view, 41
- options
  - SAS, 258
- OR (odds ratio), 74
- or operator, 61
- order statistics, 70
- ordered multinomial model, 140
- ordering of levels, 99
- ordinal logit, 140, 161
- orientation
  - axis labels, 203
  - boxplot, 185
- output data from analysis, 262
- output file formats
  - SAS, 261, 265
- overdispersed binomial regression, 135
- overdispersed Poisson regression, 135
- overdispersion, 136
- package
  - remove from workspace, 278
- packages
  - detaching, 15
  - help, 282
- R, 281
- page
  - multiple plots, 202
- pairs plot, 217
- pairwise differences, 105, 130
- palettes of colors, 204
- parallel boxplots, 172
- parameter estimates
  - standard errors, 111
  - used as data, 110
- parameterization of categorical variable,
  - 98, 151
  - reference category, 129
- partial file read, 6
- password
  - SAS, 251
- path variable
  - Windows, 273
- pcs variable, 85, 288
- PDF
  - exporting, 206
- pdf output
  - SAS, 261, 265
- Pearson chi-square statistic, 78
- Pearson correlation, 76
- Pearson's  $\chi^2$  test, 87, 138
- percentiles
  - probability density function, 46
- Perl interface, 27
- permutation test, 80, 92
- permute function, 39
- permuted sample, 28
- Pi ( $\pi$ ), 39
- plot
  - conditioning, 187
- plot characters, 193
- plot maps, 230
- plot symbols, 193
- plots
  - adding arrows, 199
  - adding footnotes, 197
  - adding polygons, 199
  - adding shapes, 199
  - adding text, 198
  - predicted lines, 113
  - predicted values, 113
  - regression diagnostics, 109
  - rotating text, 198
  - titles, 197
- plotting limits, 115
- Plummer, Martyn, 269

- PNG
  - exporting, 209
- point size specification, 201
- point-and-click interface, 267
- points, 194
  - locating, 200
- Poisson family, 136
- Poisson regression, 135, 138, 155
  - Bayesian, 237
  - zero-inflated, 138, 157
- polygons, 199
- polynomial regression, 148
- posterior probability, 237
- posting guide (R-help), 284
- postscript
  - exporting, 206
- power calculations
  - analytic, 223
  - empirical, 225
- practical extraction and report language (Perl), 27
- predicted values
  - generating from linear model, 106
- primary care
  - linkage, 287
- primary care visits, 288
- primary substance of abuse, 288
- printing model results, 13
- prior distribution, 237
- probability density function
  - quantile, 46
- probability density plot, 189
- probability distributions, 67
  - quantiles, 45
  - random variables, 45
  - task view, 45
- probability integral transform, 50
- probit link function, 136
- probit regression, 135
- programming, 51
- projection, 234
- properties
  - Windows, 273
- proportion, 74
- proportional hazards model, 147, 179
  - frailty, 147
- proportional odds model, 140, 161
- pseudo  $R^2$ , 137
- pseudo-random number
  - generation, 45
  - set seed, 50
- `pss_fr` variable, 219, 288
- psychometrics, 239
  - task view, 239
- q-q plot, 186
- QQ plot, 122
- quadratic growth curve models, 144
- quantile regression, 149, 160
- quantile-quantile plot, 122, 186
- quantiles, 71
  - probability density function, 46
  - t distribution, 280
- quarter variable, 35
- quitting R, 272
- quotes
  - nested, 23
- R
  - available datasets, 284
  - bug reports, 284
  - command history, 270
  - data structures, 274
  - detach packages, 163
  - Development Core Team, 269
  - differences from SAS, 2
  - exiting, 271
  - export SAS dataset, 12
  - FAQ, 274, 284
  - Foundation for Statistical Computing, 269
  - functions, 279
  - help system, 272, 273
  - history, 269
  - installation, 269
  - introduction, 269
  - libraries, 281
  - Linux installation, 270
  - objects, 274
  - packages, 281, 282
  - Project, 284
  - reading SAS files, 8
  - resources for new users, 273
  - sample session, 271
  - starting, 271
  - support, 284
  - task views, 282
  - warranty, 272
  - Windows installation, 270
- R Commander
  - graphical user interface, 270
- R graphical user interface, 270

- R index, 2
- $R^2$  for logistic regression, 137
- R-help mailing list, 284
- ragged data, 230
- random coefficient model, 142, 144
- random effects model, 143, 172
  - estimate, 143
  - generating, 228
- random intercept model, 142
- random number
  - seed, 50
- random slopes model, 143
- random variable generation, 45
- random variables
  - density, 47
  - generate, 47
  - probability, 47
  - quantiles, 47
- randomization group, 288
- randomized clinical trial, 285
- range
  - axes, 203
- rank sum test, 80
- reading
  - comma separated (CSV) files, 8
  - data, 56
  - dates, 7
  - fixed format files, 6
  - HTTP from URL, 9
  - more complex fixed format files, 7
  - native format files, 5
  - other packages, 8
  - R objects, 5
  - variable format files, 230
  - XML files, 10
- reading data, 261
- reading long lines, 7
- receiver operating characteristic curve, 190, 215
- recoding variables, 21
- recursive partitioning, 244
- reference category, 98, 129, 151
- regression, 97
  - categorical predictor, 98
  - diagnostic plots, 109
  - diagnostics, 120
  - Gamma, 135
  - interaction, 100, 116
  - logistic, 135
  - lognormal, 135
  - no intercept, 100
  - overdispersed binomial, 135
  - overdispersed Poisson, 135
  - parameterization, 98
  - Poisson, 135
  - probit, 135
  - residuals, 107
  - stratified analysis, 100
  - studentized residuals, 107
- regular expressions, 19, 34
- relative risk, 74
- reliability measures, 239
- remove
  - dataframe from workspace, 278
  - package from workspace, 278
  - spaces from a string, 20
- rename variables, 16
- replicating examples from the book, 273
- resampling-based inference, 73
- reshaping datasets, 29, 165
- residuals, 107
  - analysis, 120
  - correlated, 143
  - plots, 122
  - studentized, 107
- results from HELP study, 285
- Rich text format (rtf), 207
- ridge regression, 149
- right censored data, 191
- Ripley, Brian, 269
- Risk Assessment Battery, 287
- robust (empirical) variance, 145, 176
- robust statistical methods
  - regression, 148
  - task view, 149
- ROC curve, 190, 215
- rotating
  - axis labels, 203
  - text, 198
- round results, 40, 56
- RR (relative risk), 74
- RTF, 207
- rtf output
  - SAS, 261, 265
- rug plot, 197
- running script, 273
  
- Samet, Dr. Jeffrey, 285
- sample a dataset, 28
- sample size calculations
  - analytic, 223
- sandwich variance, 145, 176

- Sarkar, Deepayan, 269
- SAS
  - Base, 251
  - differences from R, 2
- SAS basic concepts, 257
- SAS files
  - from R, 8
- SAS GUI, 251
- SAS index, 2
- SAS Institute, 251
- SAS Learning edition, 251
- SAS macros, 73
- SAS on-line help, 257
- SAS OnDemand, 251
- SAS password file, 251
- SAS/ETS, 251
- SAS/GRAPH, 251
- SAS/IML, 251
- SAS/STAT, 251
- save parameter estimates as dataset, 154
- save printed result as SAS dataset, 118
- saving
  - R history of commands, 270
- saving data, 58
- saving graphs, 206
- scale
  - log, 205
- scaling, 72
- scatterplot, 86, 115, 182
  - lines, 195
  - matrix, 190
  - multiple y values, 182
  - points, 194
  - separate plotting characters per group, 193
  - smoothed line, 195
- scatterplot smoother, 115
- script file, 273
  - running, 273
- search path, 273
- seed
  - random numbers, 50
- sensitivity, 75, 190
- sensitivity to case, 2
- separate model fitting by group, 123
- separate plotting characters per group, 193
- settings
  - graphical, 202
- sexrisk variable, 151, 161, 288
- SF-36 short form health survey, 288
- shapes, 199
- short form (SF) health survey, 288
- side-by-side boxplots, 185
- sideways orientation
  - boxplot, 185
- significance stars in R, 97, 118
- simulation-based power calculations, 225
- sine function, 39
- singular value decomposition, 45
- size of graph, 201
- Smith College, 3
- smooth plot, 189
- smoothed line, 195
- smoothing, 163
- smoothing spline, 148
- social sciences
  - task view, 97, 114, 135
- social supports, 288
- solve optimization problems, 41
- sorting, 31, 65
- sourcing commands, 273
- spatial statistics
  - task view, 150, 234
- Spearman correlation, 76
- specification of design matrix, 98
- specification of margin, 202
- specificity, 75, 190
- specifying
  - box around plots, 201
  - color, 204
  - point size, 201
  - text size, 201
- spreadsheet, 11
- SPSS files, 8, 12
- SQL, 27
- square root, 38
- square root link function, 136
- standard deviation, 38
- standardized residuals
  - mixed model, 143
- Stata files, 8, 12
- statements
  - SAS, 259
- stem plot, 184
- straight line
  - adding, 192
- stratified analysis, 100, 123, 260
- string
  - from numeric variable, 17
- string length, 19
- string variable
  - concatenating strings, 19

- extract characters, 18
- find a string, 19
- remove spaces, 20
- structured query language (SQL), 27
- Student t-test, 79
- studentized residuals, 107
- style
  - axes, 203
- styles
  - line, 204
- sub variable, 115, 125
- subject index, 2
- submatrix, 43
- submitting code, 253
- subsetting, 27, 28, 62, 66, 260
- substance abuse treatment, 288
- substance of abuse, 288
- substance variable, 86, 288
- sum, 38
- summarize means by groups, 66
- summary statistics, 83
  - mean, 69
  - separately by group, 70
- sums of squares
  - Type III, 102, 116
- sums of squares and cross products, 114
- sunflower plot, 189
- support, 284
- survey design, 248
- survival analysis, 147
  - accelerated failure time model with frailty, 147
  - Cox model, 179
  - Kaplan–Meier plot, 191, 213
  - logrank test, 81, 95
  - proportional hazards model, 147
  - proportional hazards model with frailty, 147
  - task view, 147, 192
- sweep operator, 72
- symbols
  - mathematical, 198
  - plot, 193
- Systat files, 8
- t distribution, 67
  - quantile, 280
- t-test, 79, 90
- table
  - cross-classification, 77
- tangent function, 39
- task view
  - analysis of spatial data, 150
  - Bayesian inference, 237
  - cluster analysis, 239
  - clustering, 248
  - graphics, 181
  - multivariate statistics, 239
  - optimization and mathematical programming, 41
  - overview, 282
  - probability distributions, 45
  - psychometrics, 239
  - robust statistical methods, 149
  - social sciences, 97, 114, 135
  - spatial statistics, 234
  - survival analysis, 147, 192
  - time series, 146
- Temple Lang, Duncan, 269
- test characteristics, 75
- test joint null hypotheses, 103–105
- test of interaction, 126
- text
  - adding, 198
  - rotating, 198
- text size specification, 201
- tick marks, 203
- Tierney, Luke, 269
- TIFF
  - exporting, 209
- time
  - elapsed, 36
- time series, 146
  - task view, 146
- time variable, 166
- time variables, 35
- time-to-event analysis, 147
- timing commands, 36
- titles, 197
- tolerance
  - floating point comparisons, 40
- topic index, 2
- transformed residuals
  - mixed model, 143
- translations
  - character, 20
- transpose matrix, 42
- transposing datasets
  - long (tall) to wide format, 30
  - wide to long (tall) format, 29
- treat variable, 95, 288
- trigonometric functions, 39

- trimmed mean, 71
- true positive, 190
- truncation, 40
- Tufte, Edward, 210
- Tukey's HSD (honest significant differences), 105, *130*
- Tukey, John, 210
  - notched boxplot, 186
- two sample t-test, 79, *90*
- two-way ANOVA, 102, *124*
  - interaction plot, 187
- two-way tables, *86*
- Type III sums of squares, 102, *116*
  
- uniform random variables, 46
- unique values, 26
- univariate loess, 148
- universal resource locator (URL), 9
- University of Auckland, 269
  - Department of Statistics, 3
- unstructured covariance matrix, *169*
- unstructured working correlation, 145
- upper to lower case conversions, 20
- Urbanek, Simon, 269
- URL, 9
- user-defined functions, 266
- using the book, 2
  
- values of variables, 16
- variable display, 16
- variable format files, *230*
- variable labels, 23
- variables
  - rename, 16
- variance covariance matrix, 143
- varimax rotation, *240*
- vector
  - from a matrix, 44
  - indexing, 276
  - operations, 276
- vector operations
  - efficiency, 51
- vectors
  - extract elements, 277
- visualize correlation matrix, *219*
  
- warranty for R, 272
- website for book, 3
- weekday variable, 34
- weighted least squares, 148
- where to begin, 2
  
- White variance, *176*
- wide to long (tall) format conversion, 29
- width of line, 204
- wiki
  - R, 284
- Wilcoxon test, 80, *92*
- Windows
  - environment variables, 273
  - installation of R, 270
  - metafile, 208
  - path, 273
  - R FAQ, 270
- working correlation matrix, 145, *176*
- working directory, 37
- writing
  - native format files, 11
  - other packages, 12
  - text files, 13
  - unctions in R, 280
  
- X'X matrix, 114
- x-y plot, *see* scatterplot
- XML, 10
  - create file, 12
  - DocBook DTD, 15
  - read file, 8
  - write files, 15
- XML files, 12
  
- year variable, 35
  
- zero-inflated
  - negative binomial regression, 139
  - Poisson regression, 138, *157*