

Indices

Separate indices are provided for subject (concept or task), SAS command, and R command. References to the examples are denoted in *italics*.

Subject index

- 3-D
 - histogram, 219
 - plot, 222
- 95% confidence interval
 - mean, 86
 - proportion, 86
- Abrams, Allyson, xxv
- absolute value, 59
- accelerated failure time model, 163
- access
 - elements in R, 365
 - files, 81
 - variables, 17
- add
 - lines to plot, 244
 - marginal rug plot, 245
 - matrices, 64
 - noise, 243
 - normal density, 245
 - straight line, 242
 - text, 246
 - variables, 19
- age variable, 107, 381
- agreement, 89
- AIC, 143, 170
- airline delays, 336
- Akaike information criterion (AIC), 143, 170
- alcohol abuse, 383
- alcoholic drinks
 - HELP dataset, 382
- Allaire, J.J., xxii
- Aloisio, Kathryn, xxii
- altitude, 325
- analysis of variance
 - interaction plot, 223
 - one-way, 117
 - two-way, 117, 139
- analytic power calculations, 95
- and operator, 47
- angular plot, 224
- annotating datasets, 43
- ANOVA
 - interaction plot, 223
 - one-way, 117
 - tables, 169
- Aotearoa (New Zealand), 357
- API (application programming interface), 334
- Apple R FAQ, 359
- application programming interface (API), 334
- arbitrary quantiles, 85
- area under the curve, 225
- ARIMA model, 162
- Arnold, Tim, 288
- arrays, 44, 74
 - extract elements, 367
- arrows, 247
- ASCII
 - datasets, 12
 - encoding, 27
- assertions, 76
- assignment operators in R, 365
- association plot, 224
- attach dataframes, 361
- attributable risk, 87
- attributes
 - R, 369
- AUC (area under the curve), 225
- Auckland, University of, 357
- automated report generation, 106, 287
- autoregressive model, 162
- available datasets in R, 377
- AvantGarde font, 252
- average
 - running, 318
- average number of drinks
 - HELP dataset, 382
- axes
 - labels, 254
 - multiple, 218

- omit, 256
- range, 253
- style, 253
- values, 254

- barchart
 - error bars, 216
- barplot, 211
- Base SAS, 341
- baseline interview, 379
- basic concepts
 - SAS, 347
- batch mode, 356, 362
- Bates, Douglas, 357
- Bayesian
 - external software, 291
 - inference task view, 290, 294, 373
 - information criterion, 170
 - logistic regression, 292
- BCA intervals, 303
- Bergstralh, Erik, 299
- best linear unbiased predictors, 158
- beta
 - distribution, 86
 - function, 60
- bias corrected and accelerated, 304
- bias-corrected and accelerated, 303
- BIC, 170
- big data, 5, 29, 114, 336
- Bike ride, 325
- binned scatterplot, 219
- binomial family, 150
- binomial probabilities
 - tabulation, 317
- bivariate
 - loess, 155
 - relationship, 101, 217–219
- Bland-Altman plot, 228
- BMDP files, 6, 13
- BMP export, 258
- book website, xxv
- Boolean
 - operations, 25, 31, 47, 238
 - R, 366
- bootstrapping, 31, 303
- box around plots, 252
- boxplot, 214
 - side-by-side, 193, 215
- Bradley Airport, 336
- Breslow estimator, 163
- Breslow–Day test, 91

- Breusch–Pagan test, 124
- “broken stick” models, 159
- bug reports, 377
- byte code compiler, 373

- c statistic, 151
- calculate derivatives, 62
- calculus, 62
- Call, Gregory, xxii
- calling functions from R, 369
- case
 - sensitivity, 356, 360
 - statement, 21
- case sensitivity, xxiv
- categorical data, 50
 - as predictor, 114
 - from continuous, 20
 - generation, 261
 - parameterization, 115, 296
 - plot, 224
 - tables, 103
- Cauchy
 - distribution, 86
 - link function, 150
- causal inference, 296
- censored data, 163, 227, 269
- Center for Epidemiologic Studies
 - Depression (CESD) scale, 381
- centering, 85
- Central Limit Theorem, 274
- CESD, 43, 44
- cesd variable, 43, 381
- cesdsv variable, 188
- chained equation models, 306
- Chambers, John, 357
- change working directory, 80
- character translations, 27
- character variable, *see* string variable
- characteristics, test, 87
- characters, plotting, 242
- chemometrics task view, 373
- chi-square
 - distribution, 86
 - statistic, 91
- Cholesky decomposition, 158
- choose function, 60
- choropleth maps, 223, 327
- circadian plot, 224
- circular plot, 224
- class methods, 369
- class variable, 50

- creating, 116
 - ordering of levels, 116
- classification, 166, 205
- clear graphics settings, 136
- clinical trial, 379
 - task view, 373
- closest values, 315
- closing a graphic device, 260
- cluster analysis
 - task view, 373
- clustering
 - hierarchical, 168
 - task view, 166, 168
- cocaine, 383
- Cochran–Mantel–Haenszel test, 91
- code completion, 357
- code examples
 - downloading, xxv
- coefficient
 - of determination, 127
 - of variation, 303
 - regression, 124
- coercing
 - character variable from numeric, 24
 - dataframes into matrices, 368
 - date from character, 4
 - date to numeric, 37
 - factor variable from numeric, 21
 - matrices into dataframes, 368
 - numeric from character, 4
 - numeric from integer, 23
 - string variable from numeric, 20
- collinearity, 156
- color
 - palettes, 255
 - selection, 255
- column width, 40
- combine matrices, 63
- Comic Sans font, 252
- comma separated value (CSV) files, 4, 12
- command history, 80
 - R, 359
 - SAS, 349
- comments
 - R, 367
 - SAS, 343
- comparison
 - floating point variables, 61
 - operators, 365
- compiler, 373
- complementary log-log link function, 150
- complex fixed format files, 3, 321
 - two lines, 331
- complex numbers, 61
- complex survey design, 168
- component-wise matrix multiplication, 66
- Comprehensive R archive network, 358
- computational economics task view, 373
- computational physics task view, 373
- concatenate, 286
 - datasets, 35
 - matrices, 63
 - strings, 24
- conditional execution, 72
- conditional logistic regression model, 151
- conditional probability, 278
- conditioning plot, 221, 232
- confidence interval, 78
 - for parameter estimates, 125
 - for predicted observations, 226
 - for the mean, 226
 - proportion, 86
- confidence level
 - default, 78
- confidence limits
 - for individual (new) observations, 126
 - for the mean, 125
 - plotting, 126
- confounding, 296
- constrained optimization, 337
- contingency table, 90
 - plot, 224
- contingency tables, 103
- contour plots, 222
- contrasts, 146
 - Helmert, 116
 - polynomial, 116
 - SAS, 116
 - treatment, 116
- control flow, 71
- control structures, 71, 362
- controlling graph size, 250
- controlling Type-I error rate, 119
- convergence diagnosis for MCMC, 290, 294
- convergence problems, 265
- converting covariance to correlation
 - matrix, 128
- converting datasets
 - long (tall) to wide format, 34

- wide to long (tall) format, 33
- Cook’s Distance, 122
- coordinate systems (maps), 325
- correlated data, 190
 - generating, 267
 - regression models, 157
 - residuals, 158
- correlation
 - Kendall, 89
 - matrix, 101, 128, 238
 - Pearson, 89
 - Spearman, 89
- cosine function, 60
- count models
 - goodness of fit, 171
 - negative binomial regression, 153, 180
 - Poisson regression, 153, 176
 - zero-inflated negative binomial, 154
 - zero-inflated Poisson regression, 154, 178
- Courier font, 252
- covariance matrix, 127, 128, 190
- covariate imbalance, 296
- Cowles, Kate, 290
- Cox proportional hazards model, 163, 200
 - frailty, 163
 - proportionality test, 164
 - simulate data, 269
 - time-varying covariate, 165
- CPU time, 78
- Cramer’s V, 91
- CRAN (Comprehensive R Archive Network), 358
- CRAN task views, *see* task views
- create
 - ASCII datasets, 12
 - categorical variable from continuous, 20
 - categorical variable using logic, 21
 - CSV (comma separate value) files, 12
 - datasets for other packages, 13
 - date variable, 37
 - factors, 116
 - files for other packages, 13
 - functions, 78
 - lagged variable, 28
 - matrix, 63
 - numeric variable from string, 22
 - observation number, 32
 - recode categorical variable, 21
 - string variable from numeric, 20
 - time variables, 39
- creating
 - dataset from counts, 87
- Cronbach’s α , 166, 201
- cross-classification table, 48, 90
- cross-validation
 - generalized, 155
- crosstabs, 90, 103
- CSV (comma separated value) files, 4, 12
- cumulative
 - density function, 53
 - hazard, 164
 - hazard plots, 228
 - product, 319
 - sum, 319
- Curran, James, 162
- curve plotting, 224
- custom graphic layouts, 251
- Dalgaard, Peter, 357
- dashed line, 254
- data
 - display, 18, 350
 - entry, 10
 - from R into SAS, 327
 - from Stata into SAS, 327
 - generation, 71
 - input, 39
 - output, 40
 - scraping, 330
- Data Expo 2009, 336
- data input
 - two lines, 331
- data step
 - range of variables, 74
 - repeat steps for a set of variables, 74
 - SAS, 347
- data structures in R, 365
- data technologies, 15
- data viewer, 357
- database system, 29, 336
- dataframes, 365
 - comparison with column bind, 368
 - comparison with matrix, 368
 - detaching, 17
 - R, 367
 - remove from workspace, 368
- dataset

- comments, 19
 - from counts, 87
 - HELP study, 381
 - other packages, 6
 - R, 377
 - subset, 350
- datasets, xxv
- date and time variables
 - create date, 37
 - create time, 39
 - extract month, 38
 - extract quarter, 38
 - extract weekday, 38
 - extract year, 38
 - reading, 3
- `dayslink` variable, 111, 381
- DBF files, 6, 13
- debugging, 76, 345
 - RStudio, 76
- decimal representation, 61
- Deducer, 359
- default confidence level, 78
- defining functions, 78
- delete objects, 365
- density
 - estimation, 213, 220
 - plot, 213, 220
- density functions, 53
 - generate random, 54
 - probability, 54
 - quantiles, 54
- density plot, 101, 109
 - overlapping, 216
- depressive symptoms, 43
- derivatives, 62
- derived variable, 19, 44, 47
- design matrix, 127, 143, 172
 - specification, 115, 296
- design of experiments task view, 373
- design weights, 168
- detach
 - dataframes, 17, 139, 368
 - packages, 17, 185, 368
- determinant, 67
- detoxification program, 379, 381
- deviance, 150
 - tables, 169
- dffits, 123
- diagnostic agreement, 87, 225
 - ROC curve, 235
- diagnostic plots, 123, 136
 - diagnostics from linear regression, 135
- diagonal elements, 66, 67
- Diedrich, Holger, 213
- difference in log-likelihoods, 169
- difference in sets, 24
- differences between SAS and R, xxiv
- differential equations
 - task view, 373
- dimension, 64
- diploma problem, 276
- directory delimiter, 2
- directory structure in R, 2
- directory structure in SAS, 2
- dispersion parameter, 181
- display data, 18, 42
- display format, 350
- display missing categories, 90
- displaying information about objects, 369
- displaying model results, 11
- distance metric, 25
- distribution
 - beta, 86
 - Cauchy, 86
 - chi-squared, 86
 - empirical probability density plot, 214
 - exponential, 86
 - F, 86
 - gamma, 86
 - geometric, 86
 - logistic, 86
 - lognormal, 86
 - negative binomial, 86
 - normal, 54, 86
 - parameters, 86
 - Poisson, 86
 - probability, 53
 - q-q plot, 225
 - quantile, 54
 - quantile-quantile plot, 225
 - stem plot, 212
 - t, 86
 - Weibull, 86
- DocBook document type definition, 14
- document type definition, 14
- documentation
 - R, 362
 - SAS, 346
- dotplot, 212
- downloading
 - code examples, xxv

- drinks of alcohol
 - HELP dataset, 382
- drinkstat variable, 47
- dropping variables, 30
- drugrisk variable, 238, 381
- DTD, 14
- duplicated values, 32
- dynamic graphics task view, 373
- dynamite plot, 216
- ecological data task view, 373
- econometrics task view, 373
- edit distance, 25
- editing data, 10
- efficiency
 - programming, 349
 - vector operations, 72
- Efron estimator, 163
- eigenvalues and eigenvectors, 67
- elapsed time, 39
- else statement, 362
- empirical
 - density plot, 109
 - estimation, 276
 - finance task view, 373
 - power calculations, 284
 - probability density plot, 214
 - variance, 161, 197
- encoding
 - ASCII, 27
- entering data, 10
- environment, 369
- environmental task view, 373
- Epi Info files, 6
- equal variance test, 93
- error bars
 - bar chart, 216
- error recovery, 76
- etiquette
 - R, 377
- evaluate integrals, 62
- Evans, Michael, 271
- exact
 - confidence intervals, 86
 - logistic regression model, 152
 - test of proportions, 92
- example code
 - downloading, xxv
 - R, 361
 - SAS, 343
- Excel
 - creating, 12
 - reading, 5
- excess
 - kurtosis, 84
 - zeroes, 154
- exchangeable working correlation, 161
- execution
 - conditional, 72
 - in operating system, 79
 - profiling, 76
- expected cell counts, 105
- expected values, 276
- experimental design task view, 373
- exponential
 - distribution, 86
 - random variables, 58, 274
- exponentiation, 59
- export
 - BMP, 258
 - datasets for other packages, 13
 - Excel, 12
 - graphs, 256
 - JPEG, 258
 - PDF, 256
 - PNG, 259
 - postscript, 256
 - TIFF, 259
 - WMF (windows metafile format), 258
- expressions
 - R, 365
- extensible markup language (XML), 9, 14
- extract characters from string, 23
- extract from objects, 89, 367
- F distribution, 86
- f1 variables, 45, 46, 201, 381
- factor
 - analysis, 166, 202
 - levels, 115, 296
 - reordering, 116
 - variable, 50, 114
- factorial function, 60
- failure time data, 163
- Falcon, Seth, 357
- false positive, 225
- family
 - binomial, 150
 - Gamma, 150
 - Gaussian, 150

- inverse Gaussian, 150
 - Poisson, 150
- FAQ
 - Apple R, 359
 - R, 364, 377
 - Windows R, 358
- female** variable, 47, 382
- Fibonacci sequence, 320
- file
 - browsing, 81
 - variable format, 7
- finance task view, 373
- find
 - approximate string, 25
 - closest values, 315
 - string within a string, 25
 - working directory, 80
- finite mixture models
 - task view, 313, 373
- Fisher’s exact test, 92, 103
- fit model separately by group, 138
- fixed format files, 2, 3
- flight delays, 336
- floating point representation, 61
- follow-up interviews, 379
- fonts in graphics, 252
- footnotes, 246
- for statement, 362
- foreign format, 43
- formatted
 - data, 12, 350
 - model results, 11
 - output, 287
 - variables, 28
- formula object, 90, 113
- forward stagewise regression, 171
- Foundation for Statistical Computing
 - R, 357
- Fox, John, 165
- fraction of missing information, 311
- frailty model, 163
- frequently asked questions
 - see FAQ, 364
- Friedman’s “super smoother”, 245
- Friendly, Michael, 224
- functions, 77
 - plotting, 224
 - R, 78, 369
- fuzzy search, 25
- G-rho family of Harrington and Fleming, 111
- g1b** variable, 232, 382
- g1btv** variable, 187, 197
- GAM, 155
- Gamma
 - distribution, 86
 - family, 150
 - function, 60, 271
 - regression, 149
- Gaussian
 - distribution, 53
 - family, 150
- Gelman, Andrew, 271, 273
- gender** variable, 50, 382
- general linear model for correlated data, 157, 190
- generalized additive model, 155, 185
- generalized cross-validation, 155
- generalized estimating equation, 197
 - exchangeable working correlation, 161
 - independence working correlation, 161
 - unstructured working correlation, 161
- generalized linear mixed model, 161, 199
- generalized linear model, 149, 172
 - correlated outcomes, 160
- generalized logit model, 152, 183
- generalized multinomial model, 152
- generate
 - arbitrary random variables, 58
 - categorical data, 261
 - correlated binary variables, 267
 - Cox model, 269
 - dataset from counts, 87
 - exponential random variables, 58
 - generalized linear model random effects, 264
 - grid of values, 75
 - logistic regression, 263
 - multinomial random variables, 56
 - multivariate normal random variables, 56
 - normal random variables, 56
 - other random variables, 58
 - pattern of repeated values, 73
 - predicted values, 120
 - random variables, 53
 - residuals, 121
 - sequence of values, 73

- truncated normal random variables, 58
- uniform random variables, 55
- genetics task view, 373
- genf** variable, 139
- Gentleman, Robert, 357
- geometric distribution, 86
- getting
 - data from R into SAS, 327
 - data from Stata into SAS, 327
 - help in R, 377
 - started with SAS, 346
- GitHub, 357
- goodness of fit, 171, 178
 - ROC curve, 235
- Google Maps, 325
- GPS coordinates, 325
- Grömping, Ulrike, 213
- graduation, 276
- grammar of graphics, 325
- graphical layouts, 251
- graphical models task view, 373
- graphical settings, 253
- graphical user interface
 - deducer, 359
 - R, 359
- graphics
 - boxplot, 214
 - choropleth, 327
 - exporting, 256
 - side-by-side boxplots, 215
 - size, 250
 - task view, 211, 373
- greater than operator, 47
- grid
 - graphics, 373
 - of values, 75
 - rectangular, 248
 - search, 337
- grouping variable
 - linear model, 283
 - summary statistics, 281
- growth curve models, 159
- Gruen, Bettina, 313
- guide to packages
 - R, 372
- guidelines
 - R-help postings, 377
- Hadoop, 29
- Hakim, Tanya, xxv
- hanging rootogram, 172
- Harrell, Frank, 39, 128, 216, 371
- Harrington and Fleming G-rho family, 111
- harvesting data, 330
- hat matrix, 122
- hat-check problem, 276
- hazard plots, 228
- Health Evaluation and Linkage to Primary Care (HELP) study, 379
- health survey
 - SF-36, 382
- Helmert contrasts, 116
- HELP study
 - clinic, 383
 - dataset, 381
 - introduction, 379
 - results, 379
- help system
 - other resources, 377
 - R, 361, 362
 - R packages, 373
 - SAS, 346
- Helvetica font, 252
- heroin, 383
- Hesterberg, Tim, 274
- heteroscedasticity test, 124
- hierarchical clustering, 168, 208
- high-performance computing task view, 373
- histogram, 100, 213
 - comparing, 215
- history
 - of commands, 80, 359
 - R, 357
 - SAS, 341
- homeless** variable, 103, 172, 382
- homelessness, 381
- homogeneity of odds ratio, 91
- honest significant difference, 119, 144
- Hornik, Kurt, 357
- Hosmer–Lemeshow test, 171
- hospitalization, 381
- Hotelling’s *t*, 162
- HTML files, 14
 - harvesting data, 330
 - reproducible output, 290
 - SAS, 351, 355
- HTTP/HTTPS, 9, 332
- Huber variance, 197

- hypertext markup language format (HTML), 14
- hypertext transport protocol (HTTP), 9
- i1 variable, 47, 176, 382
- i2 variable, 47, 382
- Iacus, Stefano, 357
- id number, 32
- id variable, 382
- identifying points, 249
- identity link function, 150
- if statement, 30, 72, 362
- Ihaka, Ross, xxv, 357
- ill-conditioned problems, 156
- image plot, 222
- imaginary numbers, 61
- imaging task view, 373
- import data, 6
- imputation, 306
- in statement, 362
- income inequality, 155
- incomplete data, 304, 306
- independence working correlation, 161
- index
 - R, xxiv
 - SAS, xxiv
 - subject, xxiv
- indexing
 - in R, 43, 323
 - lists, 366
 - matrix, 66
 - vector, 365
- indicator variable, 115, 296
- individual level data, 87
- indtot variable, 172, 231, 382
- InDUC (Inventory of Drug Use Consequences), 382
- infinite values, 305
- influence, 122
- information matrix, 127
- installing
 - packages in R, 371
 - R, 358
 - RStudio, 359
 - SAS, 341
- integer
 - functions, 60
 - problems, 340
- integration, 62
- interaction, 117
 - linear regression, 130
 - plot, 139, 223
 - testing, 141
 - two-way ANOVA, 139
- intercept
 - no, 116
- intersection, 24
- interval censored data, 227
- introduction
 - R, 357, 362
 - RStudio, 357
 - SAS, 341
- Inventory of Drug Use Consequences, *see* indtot variable
- inverse
 - Gaussian family, 150
 - link function, 150
 - matrix, 65
 - probability integral transform, 58
- iterative proportional fitting, 153
- JAGS, 291
- jitter points, 243
- joining datasets, 35
- joins, 29
- Jones, Albyn, xxv
- JPEG export, 258
- Kaplan, Danny, 224
- Kaplan–Meier plot, 227, 234
- Kappa, 89
- keeping variables, 30
- Kendall correlation, 89
- kernel smoother plot, 213, 220
- knapsack problem, 337
- knitr, 287
- Knuth, Donald, 287
- Kolmogorov–Smirnov test, 94, 108
- Kosanke, Jon, 299
- Kruskal–Wallis test, 94
- Kuhfeld, Warren, 288
- kurtosis, 84
- L1-constrained fitting, 170
- labels for variables, 18
- Laplace approximation, 265
- large data, 5, 29, 114, 336, 349
- large sample assumption, 274
- lasso method, 170
- latent class analysis, 167
- L^AT_EX output, 287
 - R, 134

- SAS, 351, 355
- Lavine, Michael, 273
- learning SAS, 346
- least absolute shrinkage and selection
 - operator, 170
- least angle regression, 171
- least squares
 - linear, 113
 - nonlinear, 155
- legend, 69, 248
- Leisch, Friedrich, 287, 313, 357
- length
 - of string, 23
 - of vector, 65
- Lenth, Russell, xxv, 287
- less than operator, 47
- Levene’s test for equal variances, 93
- Levenshtein edit distance, 25
- leverage, 122
- library
 - help, 373
 - R, 371
- license, SAS, 341
- Ligges, Uwe, 357
- likelihood ratio test, 141, 169
- line
 - on plot, 244
 - style, 254
 - types, 254
 - width, 255
- linear combinations of parameters, 120
- linear discriminant analysis, 167, 206
- linear models, 113
 - by grouping variable, 283
 - categorical predictor, 114
 - diagnostic plots, 123
 - diagnostics, 135
 - generalized, 149
 - interaction, 117, 130
 - no intercept, 116
 - parameterization, 115, 296
 - R object, 113
 - residuals, 121
 - standardized, 121
 - studentized, 121
 - standardized residuals, 121
 - stratified analysis, 283
 - studentized residuals, 121
 - test for heteroscedasticity, 124
- linear programming, 340
- link function
 - cauchit, 150
 - cloglog, 150
 - identity, 150
 - inverse, 150
 - log, 150
 - logit, 150
 - probit, 150
 - square root, 150
- linkage to primary care, 381
- linkstatus variable, 111, 382
- Linux installation
 - R, 358
- Lipsitz, Stuart, 267
- list files, 81
- lists, 366
 - extract elements, 89, 367
- literate programming, 287
- Little, Roderick, 306
- local polynomial regression, 245
- locating points, 249
- loess
 - bivariate, 155
- log
 - base 10, 59
 - base 2, 59
 - base e, 59
 - link function, 150
- log file
 - R, 80
 - SAS, 349
- log scale, 255
- log-likelihood, 170
- log-linear model, 153
- logic, 21
- logical expressions, 20, 21
- logical operator, 20, 365
- logistic
 - distribution, 86
 - generalized, 183
- logistic regression, 149, 172
 - Bayesian, 292
 - c statistic, 151
 - generating, 263
 - goodness of fit, 171
 - Nagelkerke R^2 , 151
 - ROC curve, 235
- logit link function, 150
- lognormal
 - distribution, 54, 86
 - regression, 149
- logrank test, 95, 111

- long (tall) to wide format conversion, 34
- longitudinal regression, 157
 - reshaping datasets, 187
- looping, 71, 188
- lower to upper case conversions, 27
- lowess, 155, 185, 244
- Lucida font, 252
- Lumley, Thomas, 169, 357

- M estimation, 156
- machine learning
 - task view, 167, 373
- machine precision, 61
- Macintosh R FAQ, 359
- macros, 77
 - SAS, 303
- MAD regression, 156
- Maechler, Martin, 357
- mailing list
 - R-help, 377
- make variables available, 17
- manipulate string variables, 23, 25, 26
 - remove spaces, 27
 - split, 26
- MANOVA, 162
- Mantel–Haenszel test, 91
- maps
 - choropleth, 223, 327
 - coordinate systems, 325
 - Google Maps, 325
 - plotting, 321
- margin specification, 253
- marginal
 - histograms, 232
 - plot, 245
- Markdown, 14, 287
- Markov Chain Monte Carlo, 271, 290, 294
- matching, 296
- mathematical constants, 60
- mathematical expressions, 69, 247
- mathematical functions
 - absolute value, 59
 - beta, 60
 - choose, 60
 - exponential, 59
 - factorial, 60
 - Fibonacci sequence, 320
 - gamma, 60
 - integer functions, 60
 - log, 59
 - maximum value, 59
 - mean value, 59
 - minimum value, 59
 - modulus, 59
 - natural log, 59
 - permute, 60
 - square root, 59
 - standard deviation, 59
 - sum, 59
 - trigonometric functions, 60
- mathematical symbols
 - adding, 247
- mathematics task view, 62, 373
- matrix
 - addition, 64
 - combine, 63
 - component-wise multiplication, 66
 - concatenate, 63
 - correlation, 128
 - covariance, 127, 128
 - creation, 63
 - design, 127
 - dimension, 64
 - extract elements, 367
 - graphs, 123
 - hat, 122
 - indexing, 66, 367
 - information, 127
 - inversion, 65
 - large, 63
 - multiplication, 57, 65, 127, 366
 - overview, 63
 - plots, 221
 - R, 367
 - sparse, 63
 - transposition, 64
- maximization, 265
- maximum likelihood estimation, 86
- maximum number of drinks
 - HELP dataset, 382
- maximum value, 59
- McArdle, Brian, xxv
- MCMC, 271, 290, 294
- McNemar’s test, 92
- mcs variable, 101, 382
- mean, 59, 83, 86
 - by group, 281
 - trimmed, 84
- mean-difference plot, 228
- median regression, 156
- medical imaging task view, 373

- medical problems, 381
- memory usage, 76
- merging datasets, 35
- meta analysis task view, 373
- metadata, 369
- methods, 369, 374
- metric for distance, 25
- Metropolis–Hastings algorithm, 271
- MICE (chained equations), 306
- Microsoft rtf format, 257
- Microsoft Word format, 257
- minimum absolute deviation regression, 156
- minimum value, 59
- Minitab files, 6
- missing data, 45, 304, 306
 - tables, 90
- missing information fraction, 311
- missing values, 96
 - recoding, 306
- mixed model, 158
 - generating, 264
 - logistic, 160, 161
 - logistic regression, 199
- model
 - comparisons, 143, 169
 - diagnostics, 135
 - selection, 143, 170
 - specification, 117, 130
- modeling language, 90, 113, 282
- modulus, 59
- moments, 84
- Mongo databases, 29
- month variable, 38
- Monty Hall problem, 278
- mosaic plot, 224
- Mosteller, Fred, 276
- motivational interview, 379
- moving average model, 162
- Mplus, 168
- multicollinearity, 156
- multilevel models, 160
- multinomial model
 - generalized, 152
 - logit, 183
 - nominal outcome, 152
 - ordered outcome, 152
- multinomial random variable, 56
- multiple comparisons, 119, 144
- multiple imputation, 306
- multiple plots per page, 250
- multiple y axes, 218, 230
- multiplication
 - matrix, 57, 65
- multivariate statistics
 - task view, 166, 373
- multivariate test, 162
- multiway tables, 91
- Murdoch, Duncan, 357
- Murrell, Paul, xxv, 15, 211, 230, 357
- Nagelkerke R^2 for logistic regression, 151
- named arguments in R, 78, 370
- named lists, 366
- names and variable types, 17
- native data files, 12
- native files, 1
- natural language processing task view, 373
- negative binomial distribution, 86
- negative binomial regression, 153, 180
 - zero-inflated, 154
- Nelson–Aalen estimate, 164
- nested models, 149
- nested quotes, 18
- New Century Schoolbook font, 252
- new users
 - R, 362
- New Zealand (Aotearoa), 357
- next statement, 362
- NIAAA, 379
- NIDA, 379
- NLP optimization, 62
- no intercept, 116
- noise
 - add to points, 243
- nonlinear least squares, 155
- nonparametric tests, 94, 108
- nonrandomized studies, 296
- normal density, 245
- normal distribution, 53, 54, 68, 84, 86
- normal random variables, 56
 - truncated, 58
- normality testing, 92
- normalizing, 85
 - constant, 271
 - residuals from linear model, 121
 - residuals from mixed model, 158
- not operator, 305
- notched boxplot, 216
- NP completeness, 337
- number of digits to display, 11

- numeric from string, 22
- numerical mathematics task view, 373
- object-oriented programming, 369
- objects
 - displaying, 369
 - R, 365
 - remove, 365
- observation number, 32
- observational studies, 296
- Octave files, 6
- ODBC, 29
- odds ratio, 87, 105
 - homogeneity, 91
- ODS, *see* output delivery system
- official statistics, 168
 - task view, 169, 373
- omit axes, 256
- OnDemand for Academics, 341
- one-way analysis of variance, 117
- open-source, xxiii
- OpenBUGS, 291
- operating system
 - change working directory, 80
 - execute command, 79
 - find working directory, 80
 - list files, 81
 - pause execution, 79
- optimization, 62
 - task view, 62, 373
 - with constraints, 337
- options
 - R, 369
 - SAS, 349
- OR (odds ratio), 87
- or operator, 47, 365
- order statistics, 84
- ordered factor, 114
- ordered logistic model, 152, 182
- ordered multinomial model, 152
- ordering of levels, 116
- ordinal logit, 152, 182
- orientation
 - axis labels, 254
 - boxplot, 214
- output data from analysis, 351
- output delivery system, 345, 351
- output file formats
 - R, 288
 - SAS, 351, 355
- overdispersion, 149, 150
- overplotting, 219
- packages
 - detaching, 17
 - help, 373
 - R, 371
 - remove from workspace, 368
- page, multiple plots per, 250
- pairs plot, 236
- pairwise differences, 119, 144
- Palatino font, 252
- palettes of colors, 255
- Parade magazine, 278
- parallel
 - boxplots, 193, 215
 - computation, 374
 - computing task view, 373
 - processing, 371
- parameter estimates
 - confidence interval, 125
 - standard errors, 125
 - used as data, 124
- parameter estimation
 - univariate distribution, 86
- parameterization of categorical variable,
 - 115, 172, 296
 - reference category, 143
- partial file read, 2
- pathological distribution
 - sampling, 271
- pause execution for a time interval, 79
- pcs variable, 101, 382
- pdf output
 - exporting, 256
 - SAS, 351, 355
- peakedness, 84
- Pearson correlation, 89
- Pearson’s χ^2 test, 91, 103, 171
- percentiles
 - probability density function, 54
- Perl
 - interface, 29
 - modules, 13
- permutation test, 94, 108
- permute function, 60
- permuted sample, 31
- pharmacokinetic task view, 373
- phi coefficient, 91
- pi (π), 60
- Pioneer Valley, 325
- plot

- adding arrows, 247
- adding footnotes, 246
- adding polygons, 247
- adding shapes, 247
- adding text, 246
- arbitrary function, 224
- characters, 242
- conditioning, 221
- curve, 224
- limits, 129
- maps, 321
- predicted lines, 226
- predicted values, 226
- regression diagnostics, 123
- rotating text, 246
- symbols, 242
- time series data, 333
- titles, 246
- Plummer, Martyn, 357
- PNG export, 259
- point size specification, 252
- point-and-click interface, 356
- points, 243
 - locating, 249
- Poisson distribution, 86
- Poisson family, 150
- Poisson regression, 149, 153, 176
 - Bayesian, 294
 - zero-inflated, 154, 178
- polygons, 247
- polynomial contrasts, 116
- polynomial regression, 155
- posterior probability, 290, 294
- posting guide (R-help), 377
- postscript, 252, 256
- power calculations
 - analytic, 95
 - empirical, 284
- practical extraction and report language (Perl), 13, 29
- predicted values
 - generating from linear model, 120
- presentations in RStudio, 290
- primary care
 - linkage, 381
 - visits, 382
- primary substance of abuse, 383
- printing model results, 11
- prior distribution, 290, 294
- probability density, 54, 214
- probability distributions, 68
 - parameter estimation, 86
 - quantiles, 53
 - random variables, 53
 - simulation, 261, 276
 - task view, 53, 373
- probability integral transform, 58
- probit link function, 150
- probit regression, 149
- profiling of execution, 76
- programming, 71
- projection, 325
- projects, 357
- propensity scores, 296
- proportion, 86
- proportional hazards model, 163, 200
 - frailty, 163
 - proportionality test, 164
 - simulate data, 269
 - time-varying covariate, 165
- proportional odds model, 152, 182
- proportionality test, 164
- Pruim, Randall, 216, 224
- pseudo R^2 , 151
- pseudo-likelihood, 265
- pseudo-random number
 - generation, 53
 - set seed, 55
- `pss_fr` variable, 238, 382
- psychometrics, 166, 201
 - task view, 166, 373
- QQ plot, 136, 225
- quadratic growth curve models, 159
- quantile regression, 156, 181
- quantile-quantile plot, 136, 225
- quantiles, 85
 - probability density function, 54
 - t distribution, 78
- quarter variable, 38
- quasi-complete separation, 294
- quitting R, 361
- quotes, nested, 18
- R
 - available datasets, 377
 - bug reports, 377
 - command history, 359
 - data structures, 365
 - detach packages, 185
 - Development Core Team, 357
 - differences from SAS, xxiv

- exiting, 360
 - export SAS dataset, 13
 - FAQ, 364, 377
 - Foundation for Statistical Computing, 357
 - graphical user interface, 359
 - help system, 361, 362
 - history, 357
 - installation, 358
 - introduction, 357
 - libraries, 371
 - Linux installation, 358
 - Markdown, 14, 290
 - objects, 365
 - packages, 371, 373
 - Project, 377
 - R Commander, 359
 - R-help mailing list, 377
 - reading SAS files, 6
 - resources for new users, 362
 - sample session, 360
 - starting, 360
 - support, 377
 - task views, 372
 - warranty, 361
 - Windows installation, 358
- R index, xxiv
- R^2
 - linear regression, 127
 - logistic regression, 151
- R-help mailing list, 377
- ragged data, 321
- rail trails, 325
- random coefficient model, 158, 159
- random effects model, 158, 193
 - estimate, 158
 - generating, 264
- random intercept model, 158
- random number
 - seed, 55, 319
- random slopes model, 158
- random variables
 - density, 54
 - generate, 54
 - generation, 53
 - probability, 54
 - quantiles, 54
- randomization group, 383
- randomized clinical trial, 379
- range
 - axes, 253
- rank sum test, 94
- reading
 - bytes, 8
 - comma separated value (CSV) files, 4
 - data, 39, 350
 - data with two lines per obs, 331
 - dates, 3
 - fixed format files, 2
 - HTTP from URL, 9
 - long lines, 3
 - more complex fixed format files, 3
 - native format files, 1
 - other files, 3
 - other packages, 6
 - R into SAS, 5
 - R objects, 1
 - SAS into R, 6
 - spreadsheets, 5
 - variable format files, 7, 321
 - XML files, 9
- receiver operating characteristic curve, 225, 235
- recoding missing values, 306
- recoding variables, 20, 21
- recover from error, 76
- rectangular grid, 248
- recursive partitioning, 166, 205
- reference category, 114, 115, 143, 172, 296
- regression, 113
 - categorical predictor, 114
 - coefficients, 124
 - diagnostic plots, 123
 - diagnostics, 135
 - forward stagewise, 171
 - Gamma, 149
 - interaction, 117, 130
 - least angle, 171
 - logistic, 149
 - lognormal, 149
 - no intercept, 116
 - overdispersed binomial, 149
 - overdispersed Poisson, 149
 - parameterization, 115, 296
 - Poisson, 149
 - probit, 149
 - residuals, 121
 - standardized coefficients, 124
 - standardized residuals, 121
 - stratified analysis, 283

- studentized residuals, 121
 - test for heteroscedasticity, 124
- regular expressions, 25, 31
- rejection sampling, 271
- relative risk, 87
- reliability measures, 166, 201
- remove
 - dataframe from workspace, 368
 - objects, 365
 - package from workspace, 368
 - spaces from a string, 27
- rename variables, 19
- repeat statement, 72, 362
- replace a string within a string, 26
- replicating examples from the book, 361
- report generation, 14, 106, 287
- reproducible analysis, 14, 106, 357
 - knitr, 287
 - rich text format, 257
 - Statweave, 287
 - tangle, 287
 - task view, 287, 288, 373
 - weave, 287
- resampling-based inference, 303
- reserved commands, 362
- reshaping datasets, 33, 187
- residuals, 121
 - analysis, 135
 - correlated, 158
 - plots, 136
 - standardized, 121
 - studentized, 121
- results from HELP study, 379
- rich text format (rtf), 257
- ridge regression, 156
- right censored data, 227
- Ripley, Brian, 357
- Risk Assessment Battery, 381
- robust statistical methods
 - empirical variance, 161, 197
 - regression, 156
 - task view, 156, 373
- ROC curve, 225, 235
- Rosenthal, Jeffrey, 271
- rotating
 - axis labels, 254
 - text, 246
- round results, 40, 60
- RR (relative risk), 87
- RSeek, 362
- RStudio, xxii, 357
 - exporting graphs, 256
 - installation, 359
 - presentations, 290
 - reproducible analysis, 290
- RTF (rich text format), 257, 351, 355
- Rubin, Donald, 306
- rug plot, 245
- running a script, 362
- running average, 318

- Samet, Jeffrey, 379
- sample size calculations
 - analytic, 95
- sampling
 - challenging distribution, 271
 - dataset, 31
- sampling distribution, 274
- sandwich variance, 161, 197
- Sarkar, Deepayan, 211, 230, 313, 357
- SAS
 - Base, 341
 - basic concepts, 347
 - differences from R, xxiv
 - files from R, 6
 - GUI, 341
 - Institute, 341
 - license file, 341
 - macros, 303
 - on-line help, 346
 - OnDemand, 341
- SAS index, xxiv
- SAS/ETS, 341
- SAS/GRAPH, 341
- SAS/IML, 341
- SAS/STAT, 341
- saving
 - data, 42
 - graphs, 256
 - parameter estimates as dataset, 175
 - printed result as SAS dataset, 132
 - R history, 359
 - SAS history, 345
- scale
 - log, 255
- scaling, 85
- scatterplot, 102, 129, 217
 - binned, 219
 - lines, 244
 - marginal histograms, 220, 232
 - matrix, 221
 - multiple y values, 218

- points, 243
- separate plotting characters per group, 242
- smoother, 129, 244
- Schoenfeld residuals, 164
- Schoenfeld, David, xxv
- Scott, Alastair, xxv
- scraping data, 330
- script file, 361, 362
- search for approximate string, 25
- seed, random number, 55, 275
- sensitivity, 87, 225
- sensitivity to case, xxiv
- separate model fitting by group, 138
- separate plotting characters per group, 242
- server version, 357
- set operations, 24
- setinit file, 341
- settings, graphical, 253
- sexrisk** variable, 172, 183, 383
- SF-36 short form health survey, 382
- shapes, 247
- short form (SF) health survey, 382
- shrinkage method, lasso, 170
- side-by-side boxplots, 193, 215
- sideways orientation
 - boxplot, 214
- significance stars in R, 113, 132
- simulate
 - categorical data, 261
 - Cox model, 269
 - generalized linear model random effects, 264
 - logistic regression, 263
 - power calculations, 284
- simulation studies, 263
- sine function, 60
- singular value decomposition, 68
- size of graph, 250
- skewness, 84
- slides in RStudio, 290
- Smith College, xxv, 276
- smoothing spline, 129, 155, 185, 213, 220, 244
- social sciences
 - task view, 113, 128, 149, 172, 373
- social supports, 382
- SOCR (Statistics Online Computational Resource), 359
- solve optimization problems, 62
- sorting, 35, 51
- sourcing commands, 361
- sparse matrices, 63
- spatial statistics
 - choropleth, 327
 - task view, 172, 325, 373
- spatio-temporal data
 - task view, 373
- Spearman correlation, 89
- specificity, 87, 225
- specifying
 - box around plots, 252
 - color, 255
 - design matrix, 115, 296
 - margin, 253
 - point size, 252
 - text size, 252
- splines, 374
- split string, 26
- spreadsheet, 5, 10
- SPSS files, 6, 13
- SQL, 29, 336
- square root, 59
 - link function, 150
- stagewise regression, 171
- standard deviation, 59, 83
- standard error, 76
- standardized regression coefficients, 124
- standardized residuals, 121
 - mixed model, 158
- Stata files, 6, 13, 327
- statements, SAS, 349
- statistical genetics task view, 373
- statistical learning task view, 373
- Statistics Online Computational Resource (SOCR), 359
- StatWeave, 287
- stem plot, 212
- straight line
 - adding, 242
- stratified analysis, 138, 283
- string variable
 - concatenating strings, 24
 - extract characters, 23
 - find a string, 25
 - find approximate string, 25
 - from numeric variable, 20
 - length, 23
 - remove spaces, 27
 - replace a string, 26
- structural equation modeling

- latent class analysis, 167
- structured query language (SQL), 29, 336
- Student’s *t* test, 93, 274
- studentized residuals, 121
- styles
 - axes, 253
 - line, 254
- sub variable, 129, 139
- subject index, xxiv
- submatrix, 66
- submitting code, 343
- subsetting, 30, 47, 51, 350
- substance abuse treatment, 382
- substance of abuse, 383
- substance variable, 102, 383
- sum, 59
- summary statistics, 97
 - mean, 83
 - separately by group, 52, 281
- sums of squares
 - cross products, 127
 - Type III, 130, 169
- support, 377
- survey methodology, 168
 - task view, 169, 373
- survival analysis, 163
 - accelerated failure time model, 163
 - Cox model, 200
 - frailty, 163
 - Kaplan–Meier plot, 227, 234
 - logrank test, 95, 111
 - proportional hazards model, 163
 - simulate data, 269
 - task view, 163, 228, 373
- suspend execution for a time interval, 79
- Sweave, 14, 287
- sweep operator, 85
- symbols
 - mathematical, 247
 - plot, 242
- syntax highlighting, 357
- Systat files, 6
- t* distribution, 68, 86
 - quantile, 78
- t* test, 93, 107, 274
- table, cross-classification, 90
- tabulate binomial probabilities, 317
- tangent function, 60
- tangle, 287, 290
- task view, 372
 - analysis of spatial data, 172
 - Bayesian inference, 290, 294
 - clustering, 166, 168
 - finite mixture models, 313
 - graphics, 211
 - machine learning, 167
 - multivariate statistics, 166
 - official statistics, 169
 - optimization and mathematical programming, 62
 - probability distributions, 53
 - psychometrics, 166
 - reproducible analysis, 287, 288
 - robust statistical methods, 156
 - social sciences, 113, 128, 149, 172
 - spatial statistics, 325
 - survival analysis, 163, 228
 - time series, 162
- Temple Lang, Duncan, xxv, 357
- temporal data
 - task view, 373
- test
 - characteristics, 87
 - heteroscedasticity, 124
 - interaction, 141
 - joint null hypotheses, 118, 119
 - normality, 92
- text
 - adding, 246
 - files, 12
 - rotating, 246
 - size specification, 252
- Tibshirani, Rob, 170
- tick marks, 254
- Tierney, Luke, 357
- TIFF export, 259
- time
 - elapsed, 39
 - variables, 39
- time series, 162
 - plotting, 333
 - task view, 162, 373
- time variable, 188
- time-to-event analysis, 163
- time-varying covariate, 165
- Times font, 252
- timing commands, 78
- titles, 246
- tolerance, 61
- topic index, xxiv

- tracing memory usage, 76
- transformed residuals, 158
- translations, character, 27
- transparent plot symbols, 219
- transposing
 - long (tall) to wide format, 34
 - matrix, 64
 - wide to long (tall) format, 33
- trap error, 76
- treat** variable, 111, 383
- treatment contrasts, 116
- trigonometric functions, 60
- trimmed mean, 84
- true positive, 225
- truncated normal random variables, 58
- truncation, 60
- Tufte, Edward, 216, 230
- Tukey, John, 230
 - honest significant differences, 119, 144
 - mean-difference plot, 228
 - notched boxplot, 216
- two line data input, 331
- two sample *t* test, 93, 107
- two-way ANOVA, 117, 139
 - interaction plot, 223
- two-way tables, 103
- Tyler, Kristin, xxv
- Type III sums of squares, 130, 169

- UCLA, 359
- uniform random variables, 55
- union, 24
- unique values, 32
- univariate distribution parameter
 - estimation, 86
- univariate loess, 155
- universal resource locator (URL), 9
- University of Auckland, 357
 - Department of Statistics, xxv
- unnamed function, 284
- unstructured covariance matrix, 190
- unstructured working correlation, 161
- upper to lower case conversions, 27
- Urbanek, Simon, 357
- URL, 9
 - harvesting data, 330
- user-defined functions, 77
- using the book, xxiv

- values of variables, 18

- van Buuren, Stef, 306
- Vanderbilt University, 216
- variable display, 18
- variable format files, 7, 321
- variable labels, 18
- variables
 - add, 19
 - rename, 19
- variance, 83, 276
- variance covariance matrix, 158
- variance equality test, 93
- varimax rotation, 166, 202
- vectors
 - efficiency, 72
 - extract elements, 367
 - from a matrix, 67
 - indexing, 365
 - recycling, 365
- version number, 372
- Verzani, John, 39
- violin plots, 215
- visualize correlation matrix, 238
- Vos Savant, Marilyn, 278

- warranty for R, 361
- weave, 287, 290
- web technologies, 9, 15
 - task view, 373
- website for book, xxv
- weekday variable, 38
- Weibull distribution, 86, 269
- weighted least squares, 156
- where to begin, xxiv
- while statement, 72, 362
- White variance, 197
- Wickham, Hadley, 29, 211, 230, 281, 284, 325
- wide to long (tall) format conversion, 33
- width of line, 255
- Wilcoxon test, 94, 108
- Wild, Chris, xxv
- wildcard expansion, 81
- Wilkinson dot plot, 212
- WinBUGS, 291
- Windows
 - installation of R, 358
 - metafile, 258
 - R FAQ, 358
- Word rtf format, 257
- workflow, 287
- working correlation matrix, 161, 197

- working directory, 80
- workspace, 369
- workspace browser, 357
- writing
 - CSV (comma separated value) files, 12
 - native format files, 12
 - other packages, 13
 - text files, 12
- $X'X$ matrix, 127
- x-y plot, *see* scatterplot
- Xie, Yihui, 288
- XML, 9, 13
 - create file, 13
 - DocBook DTD, 14
 - read file, 6
 - write files, 14
- year variable, 38
- Zaslavsky, Alan, xxv
- zero-inflated
 - negative binomial regression, 154
 - Poisson regression, 154, 178