Appendix D

Indices

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

D.1 Subject index

3-D
- histogram, 128
- plot, 130

95% confidence interval
- mean, 52
- proportion, 53

absolute value, 36

accelerated failure time model, 99

access
- Dropbox files, 6
- elements in R, 221
- files, 50
- variables, 11

add
- lines to plot, 146
- marginal rug plot, 147
- matrices, 39
- noise, 146
- normal density, 147
- straight line, 145
- text, 147
- variables, 13

age variable, 64, 239

agreement, 54

AIC, 86, 102

airline delays, 207

Akaike information criterion (AIC), 86, 102

alcohol abuse, 241

alcoholic drinks
- HELP dataset, 240

Allaire, J.J., xxii

altitude, 193

Amazon sales rank, 195

analysis of variance
- interaction plot, 130
- one-way, 70
- two-way, 70, 84

analytic power calculations, 58

and operator, 28

angular plot, 131

annotating datasets, 26

ANOVA
- interaction plot, 130
- one-way, 70
- tables, 102

Aotearoa (New Zealand), 211

API (application programming interface), 199, 200, 202

Apple R FAQ, 213

application programming interface (API), 199, 200, 202

arbitrary quantiles, 52

area under the curve, 132

ARIMA model, 98

arrays, 27, 46
- extract elements, 223

arrows, 148
D.1 Subject index

ArXiv.org, 202
ASCII
datasets, 5, 8
encoding, 17
assertions, 47
assignment operators in R, 221
association plot, 131
attributable risk, 53
attributes
R, 226
AUC (area under the curve), 132
Auckland, University of, 211
automated report generation, 63, 171
autoregressive model, 98
available datasets in R, 236
AvantGarde font, 150
average
running, 188
average number of drinks
HELP dataset, 240
axes
labels, 151
multiple, 127
omit, 152
range, 151
style, 151
values, 151
barchart
error bars, 126
barplot, 123
baseline interview, 237
batch mode, 216
Bates, Douglas, 211
Bayesian
external software, 174
inference task view, 173, 176, 232
information criterion, 102
logistic regression, 175
methods, 186
BCA intervals, 181
Beatles, 199
best linear unbiased predictors, 96
beta
distribution, 33, 53
function, 37
beta-binomial distribution, 33
beta-normal distribution, 33
bias corrected and accelerated, 182
bias-corrected and accelerated, 181
BIC, 102
big data, 2, 18, 207
regression, 69
Bike ride, 193
binned scatterplot, 128
binomial distribution, 33
binomial family, 91
binomial probabilities
tabulation, 188
bitmap image file, 153
bivariate
loess, 94
relationship, 60, 127, 128
Bland–Altman plot, 133
BMDP files, 3, 8
BMP export, 153
Bonferroni correction, 71
book website, xxii
Boolean
operations, 16, 19, 28, 140
R, 222
bootstrapping, 20, 181
box around plots, 150
boxplot, 125
side-by-side, 113, 125
Bradley International Airport, 207
break lines, 202
Breslow estimator, 98
Breslow–Day test, 55
Breusch–Pagan test, 73
“broken stick” models, 97
bug reports, 236
byte code compiler, 231
c statistic, 91
calculate derivatives, 38
calculus, 38
calling functions from R, 226
capture output, 50
cartoon guide, 195
case
sensitivity, 214
statement, 14
categorical data, 30
as predictor, 68
from continuous, 13
generation, 155
parameterization, 68, 177
plot, 131
tables, 61
Cauchy
distribution, 53
D.1 Subject index

link function, 91
causal inference, 177
censored data, 98, 133, 165
simulate data, 158
Center for Epidemiologic Studies Depression (CESD) scale, 239
centering, 52
Central Limit Theorem, 161
CESD, 27
cesd variable, 27, 239
chained equation models, 183, 186
Chambers, John, 211
change working directory, 50
character translations, 17
character variable, see string variable
characteristics, test, 54
characters, plotting, 145
chemometrics task view, 232
chi-square
distribution, 53
statistic, 55
Cholesky decomposition, 96
choose function, 37
choropleth maps, 130, 193
circadian plot, 131
circular plot, 131
class methods, 226
class variable, 30
creating, 68
ordering of levels, 68
classification, 100, 119
cleaning data, 219
clinical trial, 237
task view, 232
clinical trials, 186
clock
system, 34
closest values, 187
closing a graphic device, 153
cluster analysis
task view, 232
clustering
hierarchical, 101
task view, 100, 101
cocaine, 241
Cochran–Mantel–Haenszel test, 55
code completion, 211
code examples
downloading, xxii
coding numbers, 7
coefficient
of determination, 75
of variation, 181
regression, 73
coercing
character variable from numeric, 15
dataframes into matrices, 224
date from character, 4
factor variable from numeric, 14
matrices into dataframes, 224
numeric from character, 4
string variable from numeric, 13
collinearity, 95
color
palettes, 151
selection, 151
column width, 25
combine matrices, 39
Comic Sans font, 150
comma-separated value (CSV) files, 2, 8
command history, 49
R, 213
comments, 223
comparison
floating-point variables, 38
operators, 221
compiler, 231
complementary log-log link function, 91
complex fixed format files, 3, 190
two lines, 196
complex numbers, 38
complex survey design, 101
component-wise matrix multiplication, 40
Comprehensive R archive network, 212
computational economics task view, 232
computational physics task view, 232
concatenation, 170
data sets, 22
matrices, 39
strings, 15
conditional execution, 45
conditional logistic regression, 92
conditional logistic regression model, 91
conditional probability, 163
conditioning plot, 129, 155
confidence interval, 48
for parameter estimates, 74
for predicted observations, 132
for the mean, 132
proportion, 53
confidence level
D.1 Subject index

default, 48
confidence limits
  for individual (new) observations, 75
  for the mean, 74
plotting, 74
conflicts, 224, 230
confounding, 177
constrained optimization, 208
contingency table, 55, 61
  plot, 131
contour plots, 130
contrasts, 68, 88
  Helmert, 68
  polynomial, 68
  SAS, 68
  treatment, 68
control flow, 45
control structures, 45, 217
control widgets, 205
controlling graph size, 149
controlling Type-I error rate, 71
convergence diagnosis for MCMC, 173,
  174, 176
converting characters, 17
correlation
covariance matrix, 75, 76, 112
covariate imbalance, 177
Cox, Kate, 173
Cox proportional hazards model, 98, 117
  frailty, 99
  proportionality test, 99
  simulate data, 158
  time-varying covariate, 100
CPU time, 49
Cramer’s V, 56
CRAN (Comprehensive R Archive Network), 212
CRAN task views, see task views
create
  ASCII datasets, 8
courses
  swirl, 217
categorical variable from continuous, 13
categorical variable using logic, 14
CSV (comma-separated value) files, 8
dataset from counts, 53
datasets for other packages, 8
date variable, 23
factors, 68iles for other packages, 8
generating, 157
functions, 48
lagged variable, 17
matrix, 39
numeric variable from string, 15
observation number, 20
recode categorical variable, 14
string variable from numeric, 13
time variables, 24
Cronbach’s α, 100, 117
correlation
cross-classification table, 29, 55
crosstabs, 55, 61
cumulative
csv (comma-separated value) files, 2, 8
density function, 33
hazard, 99
hazard plots, 133
product, 189
sum, 189
curated guide to learning R, 217
Curran, James, 98
curve plotting, 131
custom graphic layouts, 149
Dalgaard, Peter, 211
Cook’s distance, 72
corpus, 202
correlated data, 112
generating, 157
  regression models, 96
residuals, 96
correlation
  Kendall, 54
  matrix, 60, 76, 141
  Pearson, 54
  Spearman, 54
cosine function, 37
count models
goodness of fit, 103
  negative binomial regression, 93, 107
  Poisson regression, 93, 105
  zero-inflated negative binomial, 94
  zero-inflated Poisson regression, 93,
  106
Courier font, 150
D.1 Subject index

<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>dashed line</td>
<td>151</td>
</tr>
<tr>
<td>data</td>
<td></td>
</tr>
<tr>
<td>display, 12</td>
<td></td>
</tr>
<tr>
<td>entry, 7</td>
<td></td>
</tr>
<tr>
<td>generation, 45</td>
<td></td>
</tr>
<tr>
<td>input, 25</td>
<td></td>
</tr>
<tr>
<td>mining, 202</td>
<td></td>
</tr>
<tr>
<td>scraping, 195</td>
<td></td>
</tr>
<tr>
<td>Data Expo 2009, 207</td>
<td></td>
</tr>
<tr>
<td>data input, 1</td>
<td></td>
</tr>
<tr>
<td>two lines, 196</td>
<td></td>
</tr>
<tr>
<td>data step</td>
<td></td>
</tr>
<tr>
<td>repeat steps for a set of variables</td>
<td>46</td>
</tr>
<tr>
<td>data structures in R, 220</td>
<td></td>
</tr>
<tr>
<td>data technologies, 9</td>
<td></td>
</tr>
<tr>
<td>data viewer, 211</td>
<td></td>
</tr>
<tr>
<td>database system, 18, 69, 207</td>
<td></td>
</tr>
<tr>
<td>dataframes, 221</td>
<td></td>
</tr>
<tr>
<td>comparison with column bind, 224</td>
<td></td>
</tr>
<tr>
<td>comparison with matrix, 224</td>
<td></td>
</tr>
<tr>
<td>detaching, 11</td>
<td></td>
</tr>
<tr>
<td>R, 223</td>
<td></td>
</tr>
<tr>
<td>remove from workspace, 224</td>
<td></td>
</tr>
<tr>
<td>dataset</td>
<td></td>
</tr>
<tr>
<td>comments, 12</td>
<td></td>
</tr>
<tr>
<td>from counts, 53</td>
<td></td>
</tr>
<tr>
<td>HELP study, 239</td>
<td></td>
</tr>
<tr>
<td>in book, xxii</td>
<td></td>
</tr>
<tr>
<td>other packages, 3</td>
<td></td>
</tr>
<tr>
<td>R, 236</td>
<td></td>
</tr>
<tr>
<td>date and time variables</td>
<td></td>
</tr>
<tr>
<td>create date, 23</td>
<td></td>
</tr>
<tr>
<td>create time, 24</td>
<td></td>
</tr>
<tr>
<td>extract month, 24</td>
<td></td>
</tr>
<tr>
<td>extract quarter, 24</td>
<td></td>
</tr>
<tr>
<td>extract weekday, 24</td>
<td></td>
</tr>
<tr>
<td>extract year, 24</td>
<td></td>
</tr>
<tr>
<td>reading, 3</td>
<td></td>
</tr>
<tr>
<td>dayalink variable, 66, 239</td>
<td></td>
</tr>
<tr>
<td>DBF files, 3, 8</td>
<td></td>
</tr>
<tr>
<td>debugging, 47</td>
<td></td>
</tr>
<tr>
<td>RStudio, 47</td>
<td></td>
</tr>
<tr>
<td>decimal representation, 38</td>
<td></td>
</tr>
<tr>
<td>decomposition</td>
<td></td>
</tr>
<tr>
<td>singular value, 41</td>
<td></td>
</tr>
<tr>
<td>Deducer, 213</td>
<td></td>
</tr>
<tr>
<td>default confidence level, 48</td>
<td></td>
</tr>
<tr>
<td>defining functions, 48</td>
<td></td>
</tr>
<tr>
<td>delete objects, 221</td>
<td></td>
</tr>
<tr>
<td>density</td>
<td></td>
</tr>
<tr>
<td>estimation, 124, 128</td>
<td></td>
</tr>
</tbody>
</table>

overlapping, 126
plot, 60, 65, 124, 128
density functions, 33
generate random, 33
probability, 33
quantiles, 33
dependency management, 231
depressive symptoms, 27
derivatives, 38
derived variable, 13, 27, 28
design matrix, 75, 87
specification, 68, 177
design of experiments task view, 232
design weights, 101
detach
dataframes, 11, 83, 224
packages, 11, 109, 225
determinant, 41
detoxification program, 237, 239
deviance
tables, 102
DFFITs, 73
diagnostic
diagnostic agreement, 132
plots, 73
tests, 73
diagnostic agreement, 54
ROC curve, 138
diagnostic plots, 82
diagnostics from linear regression, 81
diagonal elements, 40, 41
difference in log-likelihoods, 102
difference in sets, 16
differential equations
task view, 232
dimension, 40
diploma problem, 162
directory delimiter, 1
directory structure, 1
dispersion parameter, 107
display missing categories, 55
displaying
data, 12, 26
model results, 7
objects, 226
scientific notation, 12
distance metric, 16
distribution
beta, 53
Cauchy, 53
chi-squared, 53
D.1 Subject index

empirical probability density plot, 125
exponential, 53
F, 53
gamma, 53
geometric, 53
logistic, 53
lognormal, 53
negative binomial, 53
normal, 33, 53
parameters, 53
Poisson, 53
probability, 33
q-q plot, 131
quantile, 33
quantile–quantile plot, 131
stem plot, 124
t, 53
Weibull, 53
divert output, 50
DocBook document type definition, 9
document mining, 202
document term matrix, 203
document type definition, 9
documentation
   R, 216
dotplot, 124
downloading
   code examples, xxii
dplyr, see library(dplyr) in R index
drinks of alcohol
   HELP dataset, 240
drinkstat variable, 28
Dropbox, 6
dropping variables, 19
drugrisk variable, 141, 239
DTD, 9
duplicated values, 20
dynamic
   web applications, 205, 211
dynamic graphics task view, 232
dynamite plot, 126

ecological data task view, 232
econometrics task view, 232
edit distance, 16
editing data, 7
efficiency
   vector operations, 45
Efron, Bradley, 202
   Efron estimator, 98
eigenvalues and eigenvectors, 41
elapsed time, 24
else statement, 217
empirical
   density plot, 65
   estimation, 162
   finance task view, 232
   power calculations, 169
   probability density plot, 125
   variance, 97, 115
encoding
   ASCII, 17
entering data, 7
environment, 226, 230
environmental task view, 232
Epi Info files, 3
equal variance test, 57
error bars
   bar chart, 126
error recovery, 47
etiquette
   R, 236
evaluate integrals, 38
Evans, Michael, 159
exact
   confidence intervals, 53
   logistic regression model, 92
test of proportions, 56
example code
   downloading, xxii
   R, 215
Excel
   creating, 8
   reading, 2
excess
   kurtosis, 52
   zeroes, 93, 94
exchangeable working correlation, 97
execution
   conditional, 45
   in operating system, 49
   profiling, 47
expansion
   wildcard, 50
expected cell counts, 63
expected values, 162
experimental design task view, 232
exponential
   distribution, 53
   random variables, 36, 161, 165
scientific notation, 12
D.1 Subject index

exponentiation, 36
export
  BMP, 153
datasets for other packages, 8
Excel, 8
graphs, 152
JPEG, 153
PDF, 152
PNG, 153
postscript, 152
TIFF, 153
WMF (Windows metafile format), 153
expressions
  R, 221
extensible markup language (XML), 6, 9, 202
extract characters from string, 15
extract from objects, 54, 223
F distribution, 53
f1 variables, 27, 28, 117, 239
factor
  analysis, 100, 118
  levels, 68, 177
  reordering, 68
  variable, 30, 68
factor object, 68
factorial function, 37
failure time data, 98
Falcon, Seth, 211
false discovery rate correction, 71
false positive, 132
family
  binomial, 91
  Gamma, 91
  Gaussian, 91
  inverse Gaussian, 91
  Poisson, 91
FAQ
  Apple R, 213
  R, 217, 236
  Windows R, 212
female variable, 28, 240
Fibonacci sequence, 189
file
  browsing, 50
temporary, 50
  variable format, 4
filtering, 19
finance task view, 232
find
  approximate string, 16
  closest values, 187
  string within a string, 16
  working directory, 49
finite mixture models
  task view, 186, 232
finite population correction, 101
Fisher’s exact test, 56, 61
fit model separately by group, 83
fixed format files, 1
fixed width files, 2, 3
flight delays, 207
floating-point representation, 38
follow-up interviews, 237
fonts in graphics, 150
footnotes, 147
for statement, 217
foreign format, 26
formatted
  data, 8
  model results, 7
  output, 171
  variables, 18
formula object, 55, 67
forward stagewise regression, 103
Foundation for Statistical Computing
  R, 211
Fox, John, 100
fraction of missing information, 185
frailty model, 99
frequently asked questions
  see FAQ, 217
Friedman’s super smoother, 146
functions, 48
  plotting, 131
  R, 48, 226
fuzzy search, 16
G-rho family of Harrington and Fleming, 65
g1b variable, 135, 240
g1btv variable, 110, 112, 115
GAM, 94
Gamma
  distribution, 53
  family, 91
  function, 37, 159
  gamma distribution, 33
  regression, 91
Gaussian
D.1 Subject index

- distribution, 33
- family, 91
- Gelman, Andrew, 159, 160
- gender variable, 36, 240
- general linear model for correlated data, 96, 112
- generalized additive model, 94, 109
- generalized estimating equation, 115
  - exchangeable working correlation, 97
  - independence working correlation, 97
  - unstructured working correlation, 97
- generalized linear mixed model, 97, 116
- generalized linear model, 91, 104
- big data, 69
- generalized logit model, 93, 108
- generalized multinomial model, 93
- generate
  - arbitrary random variables, 36
  - categorical data, 155
  - correlated binary variables, 157
  - Cox model, 158
  - dataset from counts, 53
  - exponential random variables, 36
  - generalized linear model random effects, 156
  - grid of values, 47
  - logistic regression, 156
  - multinomial random variables, 35
  - multivariate normal random variables, 35
  - normal random variables, 35
  - other random variables, 36
  - pattern of repeated values, 46
  - predicted values, 72
  - random variables, 33
  - residuals, 72
  - sequence of values, 46
  - truncated normal random variables, 36
  - uniform random variables, 34
- genetics task view, 232
- genf variable, 84
- Gentleman, Robert, 211
- geometric distribution, 33, 53
- getting
  - and cleaning data, 219
  - help in R, 236
- ggplot2, see library(ggplot2) in R index
- GitHub, 211, 230
- goodness of fit, 103, 106
- ROC curve, 138
- Google Maps, 193
- GPS coordinates, 193
- graduation, 162
- grammar of graphics, 193
- graphical layouts, 149
- graphical models task view, 232
- graphical reporting, 186
- graphical settings, 150
- graphical user interface
  - deducer, 213
  - R, 213
  - RStudio, 211
- graphics
  - boxplot, 125
  - choropleth, 193
  - exporting, 152
  - side-by-side boxplots, 125
  - size, 149
  - task view, 123, 232
- greater than operator, 28
- grid
  - graphics, 232
  - of values, 47
  - rectangular, 148
  - search, 208
- grouping variable
  - linear model, 168
  - summary statistics, 167
- growth curve models, 97
- Gruen, Bettina, 186
- guide to packages
  - R, 231
- guidelines
  - R-help postings, 236
- Hadoop, 19
- hanging rootogram, 103
- Harrell, Frank, 76, 126, 186, 229
- Harrington and Fleming G-rho family, 65
- harvesting data, 195
- hat matrix, 72
- hat-check problem, 162
- hazard plots, 133
- Health Evaluation and Linkage to Primary Care (HELP) study, 237
- health survey
  - SF-36, 240
- Helmert contrasts, 68
- HELP study
D.1 Subject index

clinic, 241
dataset, 239
introduction, 237
results, 237
help system
other resources, 236
R, 215, 216
R packages, 231
Helvetica font, 150
heroin, 241
Hesterberg, Tim, 161
heteroscedasticity test, 73
hierarchical clustering, 101, 121
high-performance computing task view, 232
histogram, 124
comparing, 125
history
of commands, 49, 213
R, 211
Hochberg correction, 71
Holm correction, 71
homeless variable, 61, 104, 240
homelessness, 239
homogeneity of odds ratio, 55
honest significant difference, 71, 87
Hornik, Kurt, 211
Hosmer–Lemeshow test, 103
hospitalization, 239
Hotelling’s t, 98
HSD (honest significant difference), 87
HTML files, 8
harvesting data, 195
reproducible output, 172
table, 6, 198
HTTP/HTTPS, 5, 197
Huber variance, 115
hypergeometric distribution, 33
hypertext markup language format
(HTML), 8
hypertext transport protocol (HTTP), 5
i1 variable, 28, 105, 240
i2 variable, 28, 240
Iacus, Stefano, 211
id number, 20
id variable, 240
identifying points, 148
identity link function, 91
if statement, 19, 45, 217
Ihaka, Ross, 211
ill-conditioned problems, 95
image plot, 130
imaginary numbers, 38
imaging task view, 232
import data, 3
imputation, 183
in statement, 217
income inequality, 94
incomplete data, 182, 183
independence working correlation, 97
indexing, 191
in R, 27
lists, 222
matrix, 40
vector, 221
indicator variable, 68, 177
individual level data, 53
indtot variable, 104, 135, 240
InDUC (Inventory of Drug Use Consequences), 240
infinite values, 182
influence, 72
information criterion (AIC), 86
information matrix, 75
inner join, 23
installing
packages in R, 229
R, 212
RStudio, 213
integer
functions, 37
problems, 210
integration, 38
interaction, 69
linear regression, 77
plot, 84, 130
testing, 85
two-way ANOVA, 84
interactive
courses in swirl, 217
visualization, 203
web applications, 205
intercept
no, 69
intersection, 16
interval censored data, 133
introduction
R, 211, 216
RStudio, 211
invalid locale, 5
Inventory of Drug Use Consequences, see indtot variable

inverse
  Gaussian family, 91
  link function, 91
  matrix, 40
  probability integral transform, 36
iterative proportional fitting, 93

JAGS, 174
JavaScript Object Notation (JSON) format, 6
jitter points, 146
joining datasets, 22
joins, 19
JPEG export, 153
JSON format, 6

Kaplan, Danny, xxii, 131
Kaplan–Meier plot, 133, 137
Kappa, 54
keeping variables, 19
Kendall correlation, 54
kernel smoother plot, 124, 128
knapsack problem, 208
knitr, 171
Knuth, Donald, 171
Kolmogorov–Smirnov test, 57, 64
Kruskal–Wallis test, 57
kurtosis, 52

L1-constrained fitting, 102
labels for variables, 12
Laplace distribution, 33
large data, 2, 18, 207
large sample assumption, 161
lasso method, 102
latent class analysis, 101
\LaTeX output, 171
R, 80
Lavine, Michael, 160
Lawrence, Michael, 211
learning R, 217
least absolute shrinkage and selection operator, 102
least angle regression, 103
least squares
  linear, 67
  nonlinear, 94
legend, 42, 148
Leisch, Friedrich, 171, 186, 211
length
  of string, 15
  of vector, 40
less than operator, 28
Levene’s test for equal variances, 57
Levenshtein edit distance, 16
leverage, 72
library
  help, 231
  R, 229
Ligges, Uwe, 211
likelihood ratio test, 85, 102
line
  on plot, 146
  style, 151
  types, 151
  width, 151
line wrap, 202
linear combinations of parameters, 71
linear discriminant analysis, 100, 120
linear models, 67
  big data, 69
  by grouping variable, 168
  categorical predictor, 68
  diagnostic plots, 73
  diagnostic tests, 73
diagnostics, 81
generalized, 91
interaction, 69, 77
no intercept, 69
parameterization, 68, 177
R object, 67
residuals, 72
  standardized, 72
  studentized, 72
stratified analysis, 168
  stratified residuals, 72
  stratified analysis, 168
  studentized residuals, 72
test for heteroscedasticity, 73
linear programming, 210
link function
  cauchit, 91
cloglog, 91
density, 91
inverse, 91
log, 91
logit, 91
probit, 91
square root, 91
linkage to primary care, 239
linkstatus variable, 66, 240
D.1 Subject index

Linux installation  R, 212
Lipsitz, Stuart, 157
list files, 50
lists, 222
  extract elements, 54, 223
literate programming, 171
Little, Roderick, 183
Liverpool, England, 198
local polynomial regression, 146
locating points, 148
loess
  bivariate, 94
log
  base 10, 36
  base 2, 36
  base e, 36
  link function, 91
log file
  R, 49
log scale, 152
log-likelihood, 102
log-linear model, 93
log-normal distribution, 33
logic, 14
logical expressions, 13, 14
logical operator, 13, 221
logistic
  distribution, 53
  generalized, 108
logistic regression, 91, 104
  Bayesian, 175
  c statistic, 91
  generating, 156
  goodness of fit, 103
  Nagelkerke $R^2$, 91
  ROC curve, 138
logit link function, 91
lognormal
  distribution, 33, 53
  regression, 91
logrank test, 58, 65
long (tall) to wide format conversion, 21
longitudinal regression, 96
  reshaping datasets, 110
looping, 45
lower to upper case conversions, 17
lowess, 94, 109, 146
lubridate, see library(lubridate) in R
Lucida font, 150
Lumley, Thomas, 101, 211
M estimation, 95
machine learning
  task view, 100, 232
machine precision, 38
Macintosh R FAQ, 213
macros, 48
MAD regression, 95
Maechler, Martin, 211
mailing list
  R-help, 236
make variables available, 11
manipulate string variables, 15–17
  remove spaces, 17
  split, 17
MANOVA, 98
Mantel–Haenszel test, 55
maps
  choropleth, 130, 193
  coordinate systems, 192
  Google Maps, 193
  plotting, 190
margin specification, 150
marginal
  histograms, 135
  plot, 147
Markdown, 8, 171
  in Shiny, 205
Markov Chain Monte Carlo, 92, 159, 173, 176
Masarotto, Guido, 211
masking, 224, 230
matching, 177
mathematical constants, 37
mathematical expressions, 42, 148
mathematical functions
  absolute value, 36
  beta, 37
  choose, 37
  exponential, 36
  factorial, 37
  Fibonacci sequence, 189
  gamma, 37
  integer functions, 37
  log, 36
  maximum value, 36
  mean value, 36
  minimum value, 36
  modulus, 36
  natural log, 36
D.1 Subject index

permute, 37
square root, 36
standard deviation, 36
sum, 36
trigonometric functions, 37
mathematical symbols
adding, 148
mathematics task view, 39, 232
matrix
addition, 39
combine, 39
component-wise multiplication, 40
concatenate, 39
correlation, 76
covariance, 75, 76
creation, 39
design, 75
dimension, 40
document term, 203
extract elements, 223
graphs, 73
hat, 72
indexing, 40, 223
information, 75
inverse, 40
large, 39
multiplication, 35, 40, 75, 222
overview, 39
plots, 129
R, 223
structured, 7
transposition, 40
maximum likelihood estimation, 53
maximum number of drinks
HELP dataset, 240
maximum value, 36
MCMC, 92, 159, 173, 176
McNemar’s test, 56
mcs variable, 60, 240
mean, 36, 51, 52
by group, 167
trimmed, 52
weighted, 51
mean–difference plot, 133
median regression, 95
medical imaging task view, 232
medical problems, 239
memory usage, 47
merging datasets, 22
meta analysis task view, 232
metadata, 226
methods, 226, 232
metric for distance, 16
Metropolis–Hastings algorithm, 159
MICE (chained equations), 183
Microsoft rtf format, 152
Microsoft Word format, 152, 171, 172
minimum absolute deviation regression, 95
minimum value, 36
mining
text, 202
Minitab files, 3
missing data, 27, 182, 183, 186
tables, 55
missing information fraction, 185
missing values
recoding, 183
mixed model, 96
generating, 156
logistic, 97
logistic regression, 116
mode of storage, 226
model
comparisons, 86, 102
diagnostics, 81
selection, 86, 102
specification, 69, 77
modeling language, 55, 67, 167
modulus, 36
moments, 52
Mongo databases, 19
month variable, 24
Monty Hall problem, 163
Morgan, Martin, 211
mosaic plot, 131
Mosteller, Fred, 162
motivational interview, 237
movies in Liverpool, 198
moving average model, 98
Mplus, 101
multicollinearity, 95
multilevel models, 97
multinomial model
generalized, 93
logit, 108
nominal outcome, 93
ordered outcome, 92
multinomial random variable, 35
multiple comparisons, 71, 87
multiple imputation, 183, 186
multiple plots per page, 149
D.1 Subject index

multiple y axes, 127, 134
multiplication
  matrix, 35, 40
multivariate statistics
  task view, 100, 232
multivariate test, 98
multiway tables, 55
Murdoch, Duncan, 211
Murrell, Paul, 9, 123, 134, 211

Nagelkerke $R^2$ for logistic regression, 91
name conflicts, 224, 230
named arguments in R, 48, 227
named lists, 222
names and variable types, 11
native data files, 8
native files, 1
natural language processing, 202
  task view, 203
natural language processing task view, 232
negative binomial distribution, 53
negative binomial regression, 93, 107
  zero-inflated, 94
negative-binomial distribution, 33
Nelson–Aalen estimate, 99
nested models, 91
nested quotes, 12
New Century Schoolbook font, 150
new users
  R, 216
New Zealand (Aotearoa), 211
next statement, 217
NIAAA, 237
NIDA, 237
NLP optimization, 39
no intercept, 69
noise
  add to points, 146
non-ASCII, 5
non-randomized studies, 177
nonlinear least squares, 94
nonparametric tests, 57, 64
normal density, 147
normal distribution, 33, 42, 52, 53
normal random variables, 35
  truncated, 36
normality testing, 56
normalizing, 52
  constant, 159
  residuals from linear model, 72
  residuals from mixed model, 96
not operator, 182
notched boxplot, 125
NP completeness, 208
number coding, 7
number of digits to display, 7
numeric from string, 15
numerical mathematics task view, 232
object-oriented programming, 226
objects
  displaying, 226
  R, 220, 221
  remove, 221
observation number, 20
observational studies, 177
Octave files, 3
ODBC, 19
odds ratio, 53, 62
  homogeneity, 55
official statistics, 101
  task view, 101, 232
Omegahat, 6, 230
omit axes, 152
one-way analysis of variance, 70
open-source, xxiii
OpenBUGS, 174
operating system
  change working directory, 50
  execute command, 49
  find working directory, 49
  list files, 50
  pause execution, 49
  temporary files, 50
optimization, 39
  task view, 39, 232
  with constraints, 208
options
  R, 226
  scientific notation, 12
OR (odds ratio), 53
or operator, 28, 221
order statistics, 51
ordered factor, 68
ordered logistic model, 92, 108
ordered multinomial model, 92
ordering of levels, 68
ordinal logit, 92, 108
orientation
  axis labels, 151
  boxplot, 125
D.1 Subject index

outer join, 23
output file formats
   R, 171
overdispersion, 91
overplotting, 128

packages
   conflicts, 230
detaching, 11
help, 231
   R, 229
      remove from workspace, 225
Packrat projects, 231
page, multiple plots per, 149
pairs plot, 138
pairwise differences, 87
Palatino font, 150
palettes of colors, 151
Pandoc, 152, 171
Parade magazine, 163
parallel
   boxplots, 113, 125
computation, 232
   computing task view, 232
processing, 228
parameter estimates
   confidence interval, 74
   standard errors, 74
   univariate distribution, 53
   used as data, 73
parameterization of categorical variable,
   68, 177
   reference category, 87
Parel, Daniel, xxii
partial file read, 1
pathological distribution
   sampling, 159
pause execution for a time interval, 49
pdf output
   creating, 171, 172
   exporting, 152
peakedness, 52
Pearson correlation, 54
Pearson’s χ² test, 55, 61, 103
percentiles
   probability density function, 33
Perl
   interface, 18
   modules, 8
permutation test, 57, 64
permute function, 37
permuted sample, 20
pharmacokinetic task view, 232
phi coefficient, 56
phylogenetics
   task view, 232
Pi (π), 37
Pioneer Valley, 193
pipe operator, 21, 111, 228
plot
   adding arrows, 148
   adding footnotes, 147
   adding polygons, 148
   adding shapes, 148
   adding text, 147
   arbitrary function, 131
   characters, 145
   conditioning, 129
   curve, 131
   limits, 76
   maps, 190
   predicted lines, 132
   predicted values, 132
   regression diagnostics, 73
   rotating text, 147
   symbols, 145
   time series data, 197
   titles, 147
Plummer, Martyn, 211
PNG export, 153
point size specification, 150
points, 146
   locating, 148
Poisson distribution, 53
Poisson family, 91
Poisson regression, 91, 93, 105
   Bayesian, 176
      zero-inflated, 93, 106
polygons, 148
polynomial contrasts, 68
polynomial regression, 94
posterior probability, 173, 176
posting guide (R-help), 236
postscript, 150, 152
power calculations
   analytic, 58
   empirical, 169
practical extraction and report language
   (Perl), 8, 18
predicted values, 71
   generating from linear model, 72
D.1 Subject index

preprints, 202
presentations in RStudio, 172
primary care
  linkage, 239
  visits, 240
primary sampling unit, 101
primary substance of abuse, 241
printing model results, 7
prior distribution, 173, 176
probability density, 33, 125
probability distributions, 42
  parameter estimation, 53
  quantiles, 33
  random variables, 33
  simulation, 155, 162
  task view, 33, 232
probability integral transform, 36
probit link function, 91
probit regression, 91
productivity, xxi
profiling of execution, 47
programming, 45
projection, 192
projects, 211
propensity scores, 177
proportion, 53
proportional hazards model, 98, 117
  frailty, 99
  proportionality test, 99
  simulate data, 158
  time-varying covariate, 100
proportional odds model, 92, 108
proportional odds test, 99
Pruim, Randall, xxii, 126, 131
pseudo $R^2$, 91
pseudo-random number
  generation, 33
  set seed, 34
pss.fr variable, 141, 240
psychometrics, 100, 117
  task view, 100, 232
punctuation, 203

QQ plot, 82, 131
quadratic growth curve models, 97
quantile regression, 95, 107
quantile-quantile plot, 131
quantile-quantile plot, 82
quantiles, 52
  probability density function, 33
t distribution, 48
quart variable, 24
quasi-complete separation, 176
quitting R, 215
quotes, nested, 12

R
available datasets, 236
bug reports, 236
command history, 213
data structures, 220
detach packages, 109
Development Core Team, 211
exiting, 214
export SAS dataset, 8
FAQ, 217, 236
Foundation for Statistical Computing, 211
  graphical user interface, 213
help system, 215, 216
history, 211
installation, 212
introduction, 211
libraries, 229
Linux installation, 212
Mark down, 8, 172
Markdown in Shiny, 205
objects, 221
packages, 229, 231
programming, 219
Project, 236
questions, 200, 217
R Commander, 213
R-help mailing list, 236
reading SAS files, 3
resources for new users, 216
sample session, 214
starting, 214
support, 236
task views, 231
warranty, 215
Windows installation, 212

$R^2$
  linear regression, 75
  logistic regression, 91
R-help mailing list, 236
ragged data, 190
rail trails, 193
random coefficient model, 96, 97
random effects model, 96, 113
  estimate, 96
generating, 156
# Subject index

<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>random intercept model</td>
<td>96</td>
</tr>
<tr>
<td>random number seed</td>
<td>34, 189</td>
</tr>
<tr>
<td>random slopes model</td>
<td>96</td>
</tr>
<tr>
<td>random variables density</td>
<td>33</td>
</tr>
<tr>
<td>generate</td>
<td>33</td>
</tr>
<tr>
<td>generation</td>
<td>33</td>
</tr>
<tr>
<td>probability</td>
<td>33</td>
</tr>
<tr>
<td>quantiles</td>
<td>33</td>
</tr>
<tr>
<td>randomization group</td>
<td>241</td>
</tr>
<tr>
<td>randomized clinical trial</td>
<td>237</td>
</tr>
<tr>
<td>range axes</td>
<td>151</td>
</tr>
<tr>
<td>rank sum test</td>
<td>57</td>
</tr>
<tr>
<td>reading bytes</td>
<td>5</td>
</tr>
<tr>
<td>comma-separated value (CSV) files</td>
<td>2</td>
</tr>
<tr>
<td>data</td>
<td>25</td>
</tr>
<tr>
<td>data with two lines per obs</td>
<td>196</td>
</tr>
<tr>
<td>dates</td>
<td>3</td>
</tr>
<tr>
<td>fixed format files</td>
<td>1</td>
</tr>
<tr>
<td>HTML table</td>
<td>6, 198</td>
</tr>
<tr>
<td>HTTP from URL</td>
<td>5</td>
</tr>
<tr>
<td>long lines</td>
<td>3</td>
</tr>
<tr>
<td>more complex fixed format files</td>
<td>3</td>
</tr>
<tr>
<td>native format files</td>
<td>1</td>
</tr>
<tr>
<td>other files</td>
<td>2</td>
</tr>
<tr>
<td>other packages</td>
<td>3</td>
</tr>
<tr>
<td>R into SAS</td>
<td>2</td>
</tr>
<tr>
<td>R objects</td>
<td>1</td>
</tr>
<tr>
<td>SAS into R</td>
<td>3</td>
</tr>
<tr>
<td>spreadsheets</td>
<td>2</td>
</tr>
<tr>
<td>variable format files</td>
<td>4, 190</td>
</tr>
<tr>
<td>XML files</td>
<td>6</td>
</tr>
<tr>
<td>receiver operating characteristic curve</td>
<td>132, 138</td>
</tr>
<tr>
<td>recoding missing values</td>
<td>183</td>
</tr>
<tr>
<td>variables</td>
<td>13, 14</td>
</tr>
<tr>
<td>recover from error</td>
<td>47</td>
</tr>
<tr>
<td>rectangular grid</td>
<td>148</td>
</tr>
<tr>
<td>recursive partitioning</td>
<td>100, 119</td>
</tr>
<tr>
<td>redirect output</td>
<td>50</td>
</tr>
<tr>
<td>reference category</td>
<td>68, 87, 177</td>
</tr>
<tr>
<td>regression big data</td>
<td>69</td>
</tr>
<tr>
<td>categorical predictor</td>
<td>68</td>
</tr>
<tr>
<td>coefficients</td>
<td>73</td>
</tr>
<tr>
<td>diagnostic tests</td>
<td>73</td>
</tr>
<tr>
<td>diagnostics</td>
<td>71, 73, 81</td>
</tr>
<tr>
<td>forward stagewise</td>
<td>103</td>
</tr>
<tr>
<td>Gamma</td>
<td>91</td>
</tr>
<tr>
<td>interaction</td>
<td>69, 77</td>
</tr>
<tr>
<td>least angle</td>
<td>103</td>
</tr>
<tr>
<td>linear</td>
<td>46, 67</td>
</tr>
<tr>
<td>logistic</td>
<td>91</td>
</tr>
<tr>
<td>lognormal</td>
<td>91</td>
</tr>
<tr>
<td>no intercept</td>
<td>69</td>
</tr>
<tr>
<td>overdispersed binomial</td>
<td>91</td>
</tr>
<tr>
<td>overdispersed Poisson</td>
<td>91</td>
</tr>
<tr>
<td>parameterization</td>
<td>68, 177</td>
</tr>
<tr>
<td>Poisson</td>
<td>91</td>
</tr>
<tr>
<td>probit</td>
<td>91</td>
</tr>
<tr>
<td>residuals</td>
<td>72</td>
</tr>
<tr>
<td>standardized coefficients</td>
<td>73</td>
</tr>
<tr>
<td>standardized residuals</td>
<td>72</td>
</tr>
<tr>
<td>stratified analysis</td>
<td>168</td>
</tr>
<tr>
<td>studentized residuals</td>
<td>72</td>
</tr>
<tr>
<td>test for heteroscedasticity</td>
<td>73</td>
</tr>
<tr>
<td>regular expressions</td>
<td>16, 17, 19</td>
</tr>
<tr>
<td>rejection sampling</td>
<td>159</td>
</tr>
<tr>
<td>relative risk</td>
<td>53</td>
</tr>
<tr>
<td>reliability measures</td>
<td>100, 117</td>
</tr>
<tr>
<td>remove dataframe from workspace</td>
<td>224</td>
</tr>
<tr>
<td>numbers</td>
<td>203</td>
</tr>
<tr>
<td>objects</td>
<td>221</td>
</tr>
<tr>
<td>package from workspace</td>
<td>225</td>
</tr>
<tr>
<td>punctuation</td>
<td>203</td>
</tr>
<tr>
<td>spaces from a string</td>
<td>17</td>
</tr>
<tr>
<td>whitespace</td>
<td>203</td>
</tr>
<tr>
<td>rename variables</td>
<td>13</td>
</tr>
<tr>
<td>repeat statement</td>
<td>45, 217</td>
</tr>
<tr>
<td>replace a string within a string</td>
<td>17</td>
</tr>
<tr>
<td>replicable variates</td>
<td>34</td>
</tr>
<tr>
<td>replicating examples from the book</td>
<td>215</td>
</tr>
<tr>
<td>report generation</td>
<td>8, 63, 171</td>
</tr>
<tr>
<td>repository of preprints</td>
<td>202</td>
</tr>
<tr>
<td>reproducible analysis</td>
<td>8, 63, 186, 211</td>
</tr>
<tr>
<td>knitr</td>
<td>171</td>
</tr>
<tr>
<td>packages</td>
<td>231</td>
</tr>
<tr>
<td>random numbers</td>
<td>34</td>
</tr>
<tr>
<td>rich text format</td>
<td>152</td>
</tr>
<tr>
<td>Statweave</td>
<td>171</td>
</tr>
<tr>
<td>tangle</td>
<td>171</td>
</tr>
<tr>
<td>task view</td>
<td>171, 232</td>
</tr>
<tr>
<td>weave</td>
<td>171</td>
</tr>
<tr>
<td>resampling-based inference</td>
<td>181</td>
</tr>
<tr>
<td>reserved commands</td>
<td>217</td>
</tr>
<tr>
<td>reshaping datasets</td>
<td>21, 110</td>
</tr>
</tbody>
</table>
D.1 Subject index

residuals, 72
  analysis, 81
  correlated, 96
  plots, 82
  standardized, 72
  studentized, 72
results from HELP study, 237
rich text format (rtf), 152
ridge regression, 95
right censored data, 133
Ripley, Brian, 211
Risk Assessment Battery, 239
robust statistical methods
  empirical variance, 97, 115
  regression, 95
  task view, 95, 232
ROC curve, 132, 138
RODBC, 69
Rosenthal, Jeffrey, 159
rotating
  axis labels, 151
  text, 147
round results, 25, 37
RR (relative risk), 53
RSeek, 217
RStudio, xxi, xxii, 211
  curated guide to learning R, 217
  exporting graphs, 152
  installation, 213
  Packrat projects, 231
  presentations, 172
  reproducible analysis, 172
RTF (rich text format), 152
Rubin, Donald, 183
rug plot, 147
running a script, 216
running average, 188
sales rank, 195
Samet, Jeffrey, 237
sample size calculations
  analytic, 58
  empirical, 169
sampling
  challenging distribution, 159
  dataset, 20
sampling distribution, 161
sandwich variance, 97, 115
Sarkar, Deepayan, 123, 134, 186, 211
SAS
  files from R, 3
saving
  data, 26
  graphs, 152
  R history, 213
scale
  log, 152
scaling, 52
scatterplot, 61, 76, 127
  binned, 128
  lines, 146
  marginal histograms, 129, 135
  matrix, 129
  multiple y values, 127
  points, 146
  separate plotting characters per group, 145
  smoother, 76, 146
Schoenfeld residuals, 99
Schwarte, Heiner, 211
scientific notation, 12
scraping data, 195
script file, 215, 216
search for approximate string, 16
seed, random number, 34, 161
sensitivity, 54, 132
separate model fitting by group, 83
separate plotting characters per group, 145
server version, 211
session information, 224
set names, 18
set operations, 16
settings, graphical, 150
sexrisk variable, 104, 108, 241
SF-36 short form health survey, 240
shapes, 148
Shiny, 205, 211
short form (SF) health survey, 240
shrinkage method, lasso, 102
side-by-side boxplots, 113, 125
sideways orientation
  boxplot, 125
significance stars in R, 67, 77
simulate
  categorical data, 155
  Cox model, 158
  generalized linear model random effects, 156
  linear regression, 46
  logistic regression, 156
  power calculations, 169
272

D.1 Subject index

simulation studies, 156
sine function, 37
singular value decomposition, 41
sink output, 50
size of graph, 149
skewness, 52
slides in RStudio, 172
Smith College, 162
smoothing spline, 76, 94, 109, 124, 128, 146
social sciences
  task view, 67, 76, 91, 103, 232
social supports, 240
SOCR (Statistics Online Computational Resource), 213
solve optimization problems, 39
sorting, 22, 31
sourcing commands, 215
sparse matrices, 39
spatial statistics
  choropleth, 193
  task view, 103, 192, 232
spatio-temporal data
  task view, 232
Spearman correlation, 54
specificity, 54, 132
specifying
  box around plots, 150
  color, 151
  design matrix, 68, 177
  margin, 150
  point size, 150
  text size, 150
splines, 232
split string, 17
spreadsheet, 2, 7
SPSS files, 3, 8
SQL, 18, 207
square root, 36
  link function, 91
stack exchange, 200
stack overflow, 217
stagewise regression, 103
standard deviation, 36, 51
standard error, 47
standardized regression coefficients, 73
standardized residuals, 72
  mixed model, 96
Stata files, 3, 8
statistical genetics task view, 232
statistical learning task view, 232
Statistics Online Computational Resource (SOCR), 213
status codes, 201
stem plot, 124
stop words, 203
storage mode, 226
straight line
  adding, 145
stratification, 101
stratified analysis, 83, 168
string variable
  concatenating strings, 15
  extract characters, 15
  find a string, 16
  find approximate string, 16
  from numeric variable, 13
  length, 15
  remove spaces, 17
  replace a string, 17
structural equation modeling
  latent class analysis, 101
structured matrices, 7
structured query language (SQL), 18, 207
Student's t-test, 56, 161
studentized residuals, 72
styles
  axes, 151
  line, 151
sub variable, 76, 84
submatrix, 40
subsetting, 19, 29, 31
substance abuse treatment, 240
substance of abuse, 241
substance variable, 61, 241
sum, 36
summary statistics, 59
  mean, 51
  separately by group, 31, 167
  weighted mean, 51
sums of squares
  cross products, 75
  Type III, 77, 102
support, 236
survey methodology, 101
  task view, 101, 232
  weighted mean, 51
survival analysis, 98, 165
  accelerated failure time model, 99
  Cox model, 117
  frailty, 99
### D.1 Subject index

<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaplan–Meier plot</td>
<td>133, 137</td>
</tr>
<tr>
<td>logrank test</td>
<td>58, 65</td>
</tr>
<tr>
<td>proportional hazards model</td>
<td>98, 99</td>
</tr>
<tr>
<td>simulate data</td>
<td>158</td>
</tr>
<tr>
<td>task view</td>
<td>98, 133, 232</td>
</tr>
<tr>
<td>suspend execution for a time interval</td>
<td>49</td>
</tr>
<tr>
<td>Sweave</td>
<td>8, 171</td>
</tr>
<tr>
<td>sweep operator</td>
<td>52</td>
</tr>
<tr>
<td>swirl interactive courses</td>
<td>217</td>
</tr>
<tr>
<td>symbolic numbers</td>
<td>7</td>
</tr>
<tr>
<td>symbols</td>
<td>148</td>
</tr>
<tr>
<td>plot</td>
<td>145</td>
</tr>
<tr>
<td>syntax highlighting</td>
<td>211</td>
</tr>
<tr>
<td>Systat files</td>
<td>3</td>
</tr>
<tr>
<td>system clock</td>
<td>34</td>
</tr>
<tr>
<td>t distribution</td>
<td>42, 53</td>
</tr>
<tr>
<td>quantile</td>
<td>48</td>
</tr>
<tr>
<td>t-test</td>
<td>56</td>
</tr>
<tr>
<td>t-test</td>
<td>64, 161</td>
</tr>
<tr>
<td>table</td>
<td>55</td>
</tr>
<tr>
<td>reading HTML</td>
<td>6, 198</td>
</tr>
<tr>
<td>tabulate binomial probabilities</td>
<td>188</td>
</tr>
<tr>
<td>tagged image file format</td>
<td>153</td>
</tr>
<tr>
<td>tangent function</td>
<td>37</td>
</tr>
<tr>
<td>tangle</td>
<td>171, 172</td>
</tr>
<tr>
<td>task view</td>
<td>231</td>
</tr>
<tr>
<td>analysis of spatial data</td>
<td>103</td>
</tr>
<tr>
<td>Bayesian inference</td>
<td>173, 176</td>
</tr>
<tr>
<td>clustering</td>
<td>100, 101</td>
</tr>
<tr>
<td>finite mixture models</td>
<td>186</td>
</tr>
<tr>
<td>graphics</td>
<td>123</td>
</tr>
<tr>
<td>machine learning</td>
<td>100</td>
</tr>
<tr>
<td>multivariate statistics</td>
<td>100</td>
</tr>
<tr>
<td>natural language processing</td>
<td>203</td>
</tr>
<tr>
<td>official statistics</td>
<td>101</td>
</tr>
<tr>
<td>optimization and mathematical programming</td>
<td>39</td>
</tr>
<tr>
<td>probability distributions</td>
<td>33</td>
</tr>
<tr>
<td>psychometrics</td>
<td>100</td>
</tr>
<tr>
<td>reproducible analysis</td>
<td>171</td>
</tr>
<tr>
<td>robust statistical methods</td>
<td>95</td>
</tr>
<tr>
<td>social sciences</td>
<td>67, 76, 91, 103</td>
</tr>
<tr>
<td>spatial statistics</td>
<td>192</td>
</tr>
<tr>
<td>survival analysis</td>
<td>98, 133</td>
</tr>
<tr>
<td>time series</td>
<td>98</td>
</tr>
<tr>
<td>transient files</td>
<td>50</td>
</tr>
<tr>
<td>test</td>
<td>54</td>
</tr>
<tr>
<td>characteristics</td>
<td>73</td>
</tr>
<tr>
<td>heteroscedasticity</td>
<td>73</td>
</tr>
<tr>
<td>interaction</td>
<td>85</td>
</tr>
<tr>
<td>joint null hypotheses</td>
<td>70</td>
</tr>
<tr>
<td>normality</td>
<td>56</td>
</tr>
<tr>
<td>proportionality</td>
<td>99</td>
</tr>
<tr>
<td>text</td>
<td>147</td>
</tr>
<tr>
<td>adding</td>
<td>147</td>
</tr>
<tr>
<td>analytics</td>
<td>202</td>
</tr>
<tr>
<td>files</td>
<td>8</td>
</tr>
<tr>
<td>mining</td>
<td>202</td>
</tr>
<tr>
<td>rotating</td>
<td>147</td>
</tr>
<tr>
<td>size specification</td>
<td>150</td>
</tr>
<tr>
<td>Tibshirani, Rob</td>
<td>102</td>
</tr>
<tr>
<td>tick marks</td>
<td>151</td>
</tr>
<tr>
<td>tidy, see library(tidyr) in R index</td>
<td>198</td>
</tr>
<tr>
<td>Tierney, Luke</td>
<td>211</td>
</tr>
<tr>
<td>TIFF export</td>
<td>153</td>
</tr>
<tr>
<td>time</td>
<td>24</td>
</tr>
<tr>
<td>elapsed</td>
<td>24</td>
</tr>
<tr>
<td>variables</td>
<td>24</td>
</tr>
<tr>
<td>time series</td>
<td>98</td>
</tr>
<tr>
<td>plotting</td>
<td>197</td>
</tr>
<tr>
<td>task view</td>
<td>98, 232</td>
</tr>
<tr>
<td>time variable</td>
<td>112</td>
</tr>
<tr>
<td>time-to-event analysis</td>
<td>98</td>
</tr>
<tr>
<td>time-varying covariate</td>
<td>100</td>
</tr>
<tr>
<td>Times font</td>
<td>150</td>
</tr>
<tr>
<td>timing commands</td>
<td>49</td>
</tr>
<tr>
<td>titles</td>
<td>147</td>
</tr>
<tr>
<td>tolerance</td>
<td>38</td>
</tr>
<tr>
<td>tracing memory usage</td>
<td>47</td>
</tr>
<tr>
<td>transformed residuals</td>
<td>96</td>
</tr>
<tr>
<td>translations, character</td>
<td>17</td>
</tr>
<tr>
<td>transparent plot symbols</td>
<td>128</td>
</tr>
<tr>
<td>transposing</td>
<td>21</td>
</tr>
<tr>
<td>long (tall) to wide format</td>
<td>21</td>
</tr>
<tr>
<td>matrix</td>
<td>40</td>
</tr>
<tr>
<td>wide to long (tall) format</td>
<td>21</td>
</tr>
<tr>
<td>trap error</td>
<td>47</td>
</tr>
<tr>
<td>treat variable</td>
<td>66, 241</td>
</tr>
<tr>
<td>treatment contrasts</td>
<td>68</td>
</tr>
<tr>
<td>trigonometric functions</td>
<td>37</td>
</tr>
<tr>
<td>trimmed mean</td>
<td>52</td>
</tr>
<tr>
<td>true positive</td>
<td>132</td>
</tr>
<tr>
<td>truncated normal random variables</td>
<td>36</td>
</tr>
<tr>
<td>truncation</td>
<td>37</td>
</tr>
<tr>
<td>Tuft, Edward</td>
<td>126, 134</td>
</tr>
<tr>
<td>Tukey, John</td>
<td>134</td>
</tr>
</tbody>
</table>
honest significant differences, 71, 87
mean–difference plot, 133
notched boxplot, 125
two line data input, 196
two sample t-test, 56, 64
two-way ANOVA, 70, 84
interaction plot, 130
two-way tables, 61
Type III sums of squares, 77, 102

UCLA, 213
uniform random variables, 34
union, 16
unique filename, 50
unique values, 20
univariate distribution parameter estimation, 53
univariate loess, 94
universal resource identifier (URI), 202
universal resource locator (URL), 5
University of Auckland, 211
unnamed function, 169
unstructured covariance matrix, 112
unstructured working correlation, 97
upper to lower case conversions, 17, 203
Urbanek, Simon, 211
URI (universal resource identifier), 202
URL, 5
harvesting data, 195

values of variables, 12
van Buuren, Stef, 183
Vanderbilt University, 126
variable display, 12
variable format files, 4, 190
variable labels, 12
variables
add, 13
rename, 13
variance, 51, 162
weighted, 51
variance equality test, 57
variance–covariance matrix, 96
varimax rotation, 100, 118
vectors
efficiency, 45
extract elements, 223
from a matrix, 41
indexing, 221
recycling, 221
version number, 224, 231
Verzani, John, 25
violin plots, 125
visualization
interactive, 203
matrices, 7
visualize correlation matrix, 141
vos Savant, Marilyn, 163
warranty for R, 215
weave, 171, 172
web applications, 211
in Shiny, 205
web technologies, 6, 9, 198
task view, 232
website for book, xxii
weekday variable, 24
Weibull distribution, 33, 53, 158
weighted least squares, 95
weighted mean, 51
weighted variance, 51
where to begin, xxiv
while statement, 45, 217
White variance, 115
whitespace, 203
Wickham, Hadley, xxii, 19, 25, 123, 134, 167, 169, 193, 228
wide-to-long (tall) format conversion, 21
widgets
control, 205
width of line, 151
Wikipedia, 198
Wilcoxon test, 57, 64
wildcard, 16, 17, 19
wildcard expansion, 50
Wilkinson dotplot, 124
WinBUGS, 174
Windows
installation of R, 212
metafile, 153
R FAQ, 212
word boundaries, 202
Word format, 152, 172
workflow, xxi, 171
working correlation matrix, 97, 115
working directory, 49, 50
workspace, 226, 230
browser, 211
conflicts, 224, 230
wrap strings, 202
writing
D.1 Subject index

CSV (comma-separated value) files, 8
   native format files, 8
   other packages, 8
   text files, 8

X'X matrix, 75
x-y plot, see scatterplot
Xie, Yihui, 171
XML, 6, 8, 202
   create file, 8
   DocBook DTD, 9
   read file, 3
   write files, 9

year variable, 24

zero-inflated
   negative binomial regression, 94
   Poisson regression, 93, 106