CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

Index of R symbols and functions

R symbols
!, 447
+ − * /, 2, 444
. . ., 439–40
< <= == != > >= >= 11, 14
<−, 2
[, 11, 16, 453, 460
[[, 453, 460
#, 2, 435
$, 4, 460
%%, 153
%in%, 23
&, &&, 443
|, ||, 443
ˆ, 2

NA, 12–13, 24, 446–7
NaN, 13
Inf, 13
NULL, 13, 19, 439, 446, 460
TRUE, FALSE, 8, 10, 11
break, 24
for, 24–5
if, 23
in, 25
repeat, 24
while, 24

Functions
Functions that are not otherwise identified are defined in the text.

.First, 433
.Last, 433
.M1, 373
.libPaths, 433

aaply (*plyr*)
adply (*plyr*)
abbreviate (*base*), 40, 41, 75
abline (*graphics*), 50, 78, 80, 110, 129, 145, 150, 157, 247, 248, 337, 370
abs (*base*), 129, 145, 168, 192, 194
accTrainTest (*hddplot*), 399
acf (*stats*), 138, 284, 289, 291, 293, 344
add1 (*stats*), 194, 195
additions, 465
addmargins (*stats*), 61–3
aes (*ggplot2*), 488, 490, 491
aggregate (*stats*), 17, 18, 63, 64, 247, 313, 444, 456
AIC (*stats*), 202
all (*base*), 19
all.vars (*base*), 230, 461, 462
anova (*lme4*), 121, 147, 172, 214, 226, 240, 241, 257, 265, 322, 329, 330, 342, 459
anova.lme (*nlme*), 446
any (*base*), 19, 446
aov (*stats*), 120, 158, 218, 221–3, 305, 308, 319, 321, 350
aovFbyrow (*hddplot*), 396
aperm (*base*), 61, 452
apply (*base*), 89, 139, 141, 157, 182, 247, 301, 323, 328, 344, 408, 409, 430, 456, 457, 470
apropos (*utils*), 7, 469
ar (*stats*), 286, 287
arcsin, 279
args (*base*), 19, 439, 440
ARIMA, 289, 290, 294, 296–8
arima (*stats*), 288, 289, 297, 302
arima.sim (*stats*), 243, 291
array (*base*), 60, 139
arrows (*graphics*), 29
as (*methods*), 450
asin, 279
as.character (*base*), 247, 444, 463, 486
as.data.frame (*lme4*), 199, 455
as.data.frame.table (*base*), 257, 456
as.Date (*base*), 54, 441–3, 458, 481
as.expression (*base*), 90, 475, 476, 484
as.factor (*base*), 376
as.integer (*base*), 101, 214, 274, 275, 442
as.layer (*latticeExtra*), 484
as.list (*multtest*), 437
as.matrix (*Matrix*), 409, 412, 450
as.numeric (*base*), 247, 317, 413, 444
as.vector (*Matrix*), 157, 184, 208, 317, 349, 449

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

attach (*base*), 17, 36, 39, 45, 75, 313, 393, 413, 430, 431
attr (*base*), 96, 312, 315, 443, 444
attributes (*base*), 349
auto.arima (*forecast*), 288–91, 293, 294, 296–8
axis (*graphics*), 25, 26, 48, 117, 293, 296, 317, 458, 475, 476
axis_flip (*ggplot2*), 491

barchart (*lattice*), 32
bestsetNoise (*DAAG*), 197
binomial (*stats*), 247, 280
bmp (*grDevices*), 472
boot (*boot*), 130–2, 140, 156, 157
boot.ci (*boot*), 131, 132, 140, 156
bounce (*DAAG*), 458
box (*graphics*), 117
Box.test (*stats*), 297, 298, 301
boxcox (*MASS*), 161, 168
boxplot (*graphics*), 33, 47, 137, 313, 353, 394, 489
bs (*splines*), 242
bwplot (*lattice*), 32, 47, 52, 74, 76, 477

c (*base*), 2, 5, 6, 10–19, *passim*
C (*stats*), 295
call (*base*), 462, 463, 488
callsmooth, 429
cast (*reshape*), 454, 455
cat (*base*), 19, 256, 438, 470
cbind (*base*), 19, 169, 177, 251, 296, 297, 323, 388, 390, 444, 455
chisq.test (*stats*), 114, 116, 118
CIcurves, 230, 232
citation (*utils*), 468
class (*base*), 15, 22, 459, 460
cloglog, 279
cloud (*lattice*), 32, 379
cm.colors (*grDevices*), 474
cmdscale (*stats*), 384, 385, 409, 423
coef (*lme4*), 81, 150, 156, 169, 182, 183, 208, 215, 337, 340, 341, 459, 460
coefficients (*stats*), 459
col (*base*), 388, 390, 408, 420
colnames (*base*), 18, 19, 316, 370, 449, 458
colors (*grDevices*), 27, 28, 474
Commander (*Rcmdr*), 428
compareTreecalcs (*DAAG*), 367, 370
complete.cases (*stats*), 24, 38, 381, 383, 412, 447
confusion, 408
contour (*graphics*), 474
contr.sum (*stats*), 445
contr.treatment (*stats*), 445
coord_equal (*ggplot2*), 491
cor (*stats*), 67, 68, 76, 132, 140, 172, 177, 201, 250, 251
cor.test (*stats*), 68, 113
cos (*base*), 27, 28
count.fields (*utils*), 435
cox.zph (*survival*), 278
coxph (*survival*), 277
crossprod (*Matrix*), 450
CrossTable (*gmodels*), 63
cumprod (*base*), 19
cumsum (*base*), 19
curve (*graphics*), 141, 464
cut (*base*), 101, 409, 423, 443, 475
CVbinary (*DAAG*), 255, 256, 388
cvdisc (*hddplot*), 402–4
CVlm (*DAAG*), 154
cvscores (*hddplot*), 404, 405

daisy (*cluster*), 384
data (*utils*), 9, 199, 302, 393, 408
data.frame (*base*), 3, 7, 88, 89, 101, 117, 119, 150, 156, 181, 228, 230, 232, 243, 270, 292, 328, 356, 360, 370, 383, 394, 395, 436, 437, 444, 450, 453, 484
datafile (*DAAG*), 433, 436
date (*base*), 443
dbConnect (*RSQLite*), 438
dbDisconnect (*RSQLite*), 438
dbDriver (*RSQLite*), 438
dbGetQuery (*RSQLite*), 438
dbinom (*stats*), 82
dbListTables (*RSQLite*), 438
dbWriteTable (*RSQLite*), 438
defectiveCVdisc (*hddplot*), 402
demo (*utils*), 25, 29, 474
density (*stats*), 45, 129, 250, 409, 462, 491
densityplot (*lattice*), 32, 53, 89, 94
deparse (*base*), 23, 476
detach (*base*), 17, 23, 39, 45, 75, 313, 393, 413
dev.copy (*grDevices*), 472
dev.off (*grDevices*), 31, 464, 472
dfbetas (*stats*), 185, 186, 191
dichromat (*dichromat*), 474
diff (*base*), 19, 288, 302, 441
dim (*base*), 24, 166, 375, 376, 402, 448–52
dimnames (*base*), 60, 197, 257, 291, 415, 449, 452
dir (*base*), 21, 431, 432
dist (*stats*), 384, 409
dist.dna (*ape*), 384, 409
divideUp (*hddplot*), 399, 400
dnorm (*stats*), 85, 87, 133
do.call (*base*), 462
dotchart (*graphics*), 39
dotplot (*lattice*), 32, 40, 64, 75, 123, 261
doubleYScale (*latticeExtra*), 486, 487
dpois (*stats*), 83

edit (*utils*), 15
eqscplot (*MASS*), 409
equal.count (*lattice*), 239
errorsINseveral (*DAAG*), 206–7

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

errorsINx (*DAAG*), 205, 206, 216
eval (*base*), 462, 463
example (*utils*), 7, 272
exp (*base*), 54, 89, 160, 177, 188, 195, 254, 259, 260, 263, 264, 271, 277, 278, 316, 375, 418, 419, 426, 481
expand.grid (*base*), 378, 381
expression (*base*), 29, 90, 137, 145, 336, 339, 475, 482–4, 487

factor (*base*), 13, 39, 40, 75, 144, 157, 214, 215, 218, 260, 265, 268, 344, 350, 356, 383, 416, 418, 444, 445, 447, 489
fig1, 464
fig2, 464
file.choose (*base*), 431
file.show (*base*), 433
filled.contour (*graphics*), 474
fisher.test (*stats*), 116
fitsmooth, 429
fitted (*stats*), 81, 145, 168, 192, 219, 254, 260, 311, 327, 333, 336, 338, 459, 483
fixef (*nlme*), 311
fligner.test (*stats*), 260
for (*base*), 42, 75, 87, 129, 250, 291, 367, 370, 375, 415, 493
forecast (*forecast*), 289, 290
format (*base*), 54, 441, 442, 458, 481
formula (*lme4*), 461, 484
ftable (*stats*), 61–3
function (*base*), 22, 23, 61, 62, 64, 89, 101, 123, 130–3, 139–41, 156, 169, 230, 261, 292, 328, 336, 338, 344, 374, 375, 380, 408, 409, 415, 416, 418, 421, 427, 429, 433, 441, 451, 461–5, 469, 475, 476, 483, 484
funlik, 133
funRel, 169

gam (*mgcv*), 235, 236, 240, 243, 282
garch (*tseries*), 299, 300
geom_point, etc. ((*ggplot2*), 488–91
getAnywhere (*utils*), 459
getwd (*base*), 4
ggplot (*ggplot2*), 488, 490, 491
glm (*stats*), 247, 252, 253, 255, 257–60, 262, 263, 265, 266, 268, 281, 333, 386–8, 424
gpar (*grid*), 61
gray (*grDevices*), 491
grep (*base*), 21
grid (*graphics*), 491
gui (*fgui*), 427, 429

hardcopy (*DAAG*), 464
hatvalues (*stats*), 184, 185
hcl (*grDevices*), 474
hcopy, 464
head (*Matrix*), 15, 18, 41, 64, 263, 454, 455
heat.colors (*grDevices*), 474
help (*utils*), 7, 8, 11, 16, 21, 26–9, 34–6, 48, 57, 140, 166, 221, 232, 235, 242, 288, 294, 312, 364, 384, 433, 434, 437, 441–3, 445, 448, 461, 468, 472, 492
help.search (*utils*), 7, 440, 469
help.start (*utils*), 8
hist (*graphics*), 7, 33, 45, 87
histogram (*lattice*), 32, 491
history (*utils*), 19
HoltWinters (*stats*), 300
houseprices.fn, 156
HPDinterval (*lme4*), 312, 316, 343

I (*base*), 40, 194, 215, 228–30, 240, 242, 243, 252, 253, 259, 342–4, 399, 424, 444, 453, 487, 488, 490
identify (*graphics*), 29, 41
identify3d (*Rcmdr*), 491, 492
if (*base*), 418, 439, 462, 475
ifelse (*base*), 443
image (*graphics*), 7, 474
importance (*randomForest*), 370
install.packages (*utils*), 5, 427, 485
interaction.plot (*stats*), 123, 331
intervals, intervals.lme (*nlme*), 467
invisible (*base*), 30, 31, 408, 418
is.character (*base*), 441
is.factor (*base*), 19
is.logical (*base*), 19
is.matrix (*base*), 19
is.na (*base*), 12, 19, 24, 36, 38, 374, 375, 415, 446, 447
is.null (*base*), 439
isoMDS (*MASS*), 384, 385, 423
italic (*grDevices*), 29, 475

jpeg (*grDevices*), 472
julian (*base*), 442, 443

lag.plot (*stats*), 284
lapply (*base*), 456, 457, 475, 476, 484
larrows (*lattice*), 32
lars, 199
latticist (*latticist*), 485
layer, layer_ (*latticeExtra*), 483
layout (*graphics*), 166, 477
layout.show (*graphics*), 477
lazyfoo, 464, 465
lda (*MASS*), 385, 387–90, 392, 394, 395, 399, 408, 409, 419–22
lda.formula, 389
legend (*graphics*), 413, 474
length (*base*), 19, 41, 101, 103, 104, 107, 111, 129, 144, 157, 344, 383, 409, 415, 453, 469, 470, 475, 476
levels (*base*), 13, 52, 218, 327, 330, 413, 418, 444, 447, 475, 479
leverage.plots (*car*), 186

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

library (*base*), 9, 12, 14, 18, 21, *passim*
lines (*monoProc*), 25, 26, 32, 45, 50, 75, 87, 145, 150, 192, 228, 230, 234, 237, 279, 293, 295, 296, 317, 360
list (*base*), 16, 18, 23, 31, 436, 437, 440, 442, 453, 456, 462, 463, 475, 476, 479, 481–4, *passim*
llines (*lattice*), 32
lm (*stats*), 80, 88, 127, 143, *passim*
lm.influence (*stats*), 185
lm.ridge (*MASS*), 203, 215
lme (*nlme*), 127, 303, 312, 446
lmer (*lme4*), 127, 303, 305, 310–12, 314–16, 318, 319, 326, 327, 329, 330, 338, 342, 349, 350
lmList (*nlme*), 337, 340, 341, 460, 461
load (*base*), 431
local (*base*), 463
locator (*graphics*), 29
locpoly (*monoProc*), 237
loess (*stats*), 168, 236, 237, 359, 360
log (*base*), 40, 42, 51, 54, *passim*
log10 (*base*), 48, 49, 51, 386
log2 (*base*), 39
logisticsim (*DAAG*), 281
logit (*car*), 245, 248, 279, 281, 421, 424
loglm (*MASS*), 272
lowess (*stats*), 50, 79, 145, 168, 192, 236, 237, 242, 279, 292, 293, 295, 296, 317
lpoints (*lattice*), 32
lqs (*MASS*), 149, 192, 194, 216
ls (*base*), 4, 19, 21, 430, 431
lsegments (*lattice*), 32
ltext (*lattice*), 32

mad (*stats*), 66
mantelhaen.test (*stats*), 139
margin.table (*base*), 61, 96
Markov, 100, 101
match (*base*), 24, 316, 332, 400
matplot (*graphics*), 88
matrix (*base*), 18, 89, 116, 117, 166, 197, 291, 301, 367, 370, 383, 402, 403, 415, 430, 448–50, 470, 492
max (*base*), 78, 380, 470
MCMCregress (*MCMCpack*), 165, 166
mcmcsamp (*lme4*), 312, 316, 343
MDSplot (*randomForest*), 406
mean (*Matrix*), 12, 13, 17, 20, 22, 23, 75, 100, 103, 107, 110, 111, 129, 144, 292, 293, 311, 323, 340, 375, 462, 463
mean.and.sd, 22, 140, 439
median (*stats*), 12, 17, 20, 23, 74, 104, 130, 131
median.fun, 130
melt (*reshape*), 454, 455, 484
merge (*base*), 455
methods (*utils*), 459, 469
min (*base*), 380, 470
model.matrix (*lme4*), 163, 172, 185, 218, 234, 243, 445
model.tables (*stats*), 223, 224, 322
monoproc (*monoProc*), 237
months (*base*), 442
mosaic (*vcd*), 60, 61
mosaicplot (*graphics*), 60, 61, 139
mt.maxT (*multtest*), 396, 397
mtext (*graphics*), 25, 26, 29, 49, 291, 474, 477
muhaz (*muhaz*), 277
multinom (*nnet*), 390

na.omit (*stats*), 24, 374, 375, 380, 381, 390, 412, 447
names (*base*), 15, 18, 19, 25, (*passim*)
nchar (*base*), 6, 440
new.env (*base*), 465
newtest, 463
nls (*stats*), 210, 211
normalizePath (*utils*), 432
nrow (*base*), 100
ns (*splines*), 231–4, 418, 420, 421, 424, 425, 488, 489
nswlm, 418, 426
numeric (*base*), 25, 100, 129, 299, 383, 415

object.size (*utils*), 432
objects (*base*), 465
onetPermutation (*DAAG*), 129
oneway.plot (*DAAG*), 120
onewayPlot (*DAAG*), 120
open3d (*rgl*), 492
options (*base*), 25, 34, 87, 88, 143, 218, 221, 387, 390, 445, 448
opts (*ggplot2*), 491
order (*base*), 20, 75, 194, 323, 403, 448
ordered (*base*), 14, 270, 444, 445
orderFeatures (*hddplot*), 395, 398
outer (*base*), 141, 451
overlapDensity (*DAAG*), 139, 421
oz (*oz*), 407

pacf (*stats*), 284, 293
package.skeleton (*utils*), 466
pairs (*graphics*), 30, 179, 199, 249, 251, 412
palette (*grDevices*), 27
panel.abline (*lattice*), 483, 486
panel.average (*lattice*), 64, 123, 261, 483
panel.curve (*lattice*), 483
panel.densityplot (*lattice*), 90
panel.dotplot (*lattice*), 64, 123, 261
panel.identify (*lattice*), 486
panel.lines (*lattice*), 483, 486
panel.points (*lattice*), 483
panel.rug (*lattice*), 483
panel.smooth (*graphics*), 145, 292
panel.superpose (*lattice*), 336, 339, 483
panel.text (*lattice*), 485, 486
panel.xyplot (*lattice*), 483

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

par (*graphics*), 7, 26–8, 40, 42, 44, 45, 49, 51, 87, 110, 137, 144, 148, 173, 233, 238, 239, 250, 255, 291, 293, 295, 317, 333, 365, 413, 473, 474, 476, 477, 493
par3d (*rgl*), 492
parallel (*lattice*), 32
paste (*base*), 40, 41, 48, 54, 144, 170, 177, 179, 188, 193, 197, 247, 248, 291, 293, 317, 323, 356, 375, 378, 416, 440, 451, 461, 464, 481, 484
pbinom (*stats*), 82, 83
pdf (*grDevices*), 464, 472
persp (*graphics*), 141
pexp (*stats*), 99
phantom (*grDevices*), 29
playwith (*playwith*), 485
plot (*graphics*), 2–4, 7, 14, 25–30, *passim*
plot.mcmc (*MCMCpack*), 166
plot.mtcars, 461
plot.ts (*stats*), 48
plotcp (*rpart*), 364, 366
plotTrainTest (*hddplot*), 400
png (*grDevices*), 472
pnorm (*stats*), 84, 85, 106, 193
points (*graphics*), 25–8, 32, 48, 407, 470
poissonsim (*DAAG*), 282
polr (*MASS*), 270, 271
poly (*stats*), 187, 188, 194, 195, 229, 232, 233, 242, 243, 298
polygon (*graphics*), 85, 473
postscript (*grDevices*), 472
ppoints (*stats*), 137, 397
ppois (*stats*), 84
prcomp (*stats*), 169
predict (*stats*), 149, 150, 156, 157, 178, 188, 200, 211, 219, 228, 230, 254, 263, 360, 389, 395, 408, 421, 459
predict.lm (*stats*), 157, 215
pretty (*base*), 54, 85, 87, 150, 232, 293, 296, 317, 475, 481
princomp (*stats*), 380, 381, 383, 412
print (*base*), 7, 20, 21, 23, 31, 34, 36, 52, 62, 101, 111, 197, 291, 349, 366, 369, 375, 383, 408, 418, 429, 451, 459, 460, 464, 477, 479, 480, 482
print.data.frame (*base*), 21
print.default (*base*), 22
print.factor (*base*), 21, 459
print.lm (*stats*), 459
print.summary.lm (*stats*), 459
printcp (*rpart*), 355, 364–6
prod (*base*), 133, 402
profile (*stats*), 312
prompt (*utils*), 467
prop.table (*base*), 96, 97
prune (*rpart*), 364, 367
prune.rpart (*rpart*), 367
pt (*stats*), 106, 107
q (*base*), 3, 5, 35
qbinom (*stats*), 83
qda (*MASS*), 389, 394, 408, 409
qf (*stats*), 396, 397
qnorm (*stats*), 85, 106
qplot (*ggplot2*), 408
qqmath (*lme4*), 32, 93, 94, 144, 327, 328
qqnorm (*stats*), 93, 132, 298, 317, 341
qqplot (*stats*), 137, 397
qqthin (*hddplot*), 397
qr.solve (*base*), 450
qreference (*DAAG*), 93, 94, 144
qt (*stats*), 106, 107, 140, 150, 171
qtukey (*stats*), 220
quantile (*stats*), 46, 423
quarters (*base*), 442
quartz (*grDevices*), 28
quickplot (*ggplot2*), 383, 408, 487–91
qunif (*stats*), 137

R.home (*base*), 432, 468
rainbow (*grDevices*), 474
randomForest (*randomForest*), 351, 369–72, 374–6, 408, 419–21, 423, 424
ranef (*nlme*), 316, 327, 328
range (*base*), 6, 12, 20, 45, 49, 234, 242, 263, 341, 370, 457, 475
rattle (*rattle*), 429
rbind (*base*), 114, 261, 400, 416, 418, 455
rbinom (*stats*), 86, 87, 215, 216
rchisq (*stats*), 99
read.csv (*utils*), 434
read.delim (*utils*), 434
read.dna (*ape*), 409
read.fwf (*utils*), 436
read.table (*utils*), 8, 34, 35, 434–7
readLines (*base*), 435, 436
rect (*graphics*), 476
refit (*lme4*), 328
regsubsets (*leaps*), 197
relevel (*stats*), 120, 257, 262, 263, 332
reorder (*stats*), 304
rep (*base*), 12, 13, 23, 39, 75, 90, 93, 117, 119, 137, 144, 157, 166, 169, 268, 270, 323, 349, 383, 396, 437, 445, 451, 458, 476
resid (*stats*), 81, 168, 260, 279, 289, 293, 296–8, 301, 333, 339, 344
residuals (*stats*), 118, 144, 145, 185, 192, 194, 211, 239, 260, 266, 317, 327, 459
residuals.glm (*stats*), 266
rev (*base*), 20, 453
rexp (*stats*), 88, 99, 140
rfun, 462
rgb (*grDevices*), 474
rlm (*MASS*), 149, 167, 241
rm (*base*), 4, 21, 33, 40, 430, 431, 440

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

rnorm (*stats*), 22, 23, 87–9, 93, 99, 137, 138, 141, 144, 169, 197, 215, 299, 402, 430, 439, 450, 462, 470, 471, 475, 484
rollmean (*zoo*), 101
round (*base*), 54, 61, 62, 81, 96, 97, 202, 248, 256, 375, 388, 408, 416, 423, 425, 476, 481
row (*base*), 291, 388, 390, 403, 408, 420
row.names (*base*), 26, 29, 38, 39, 407, 409, 486, 492
rownames (*base*), 15, 19, 75, 192, 340, 341, 449
rpart (*rpart*), 354–8, 360, 364–7, 369–72
rpois (*stats*), 87, 100
RSiteSearch (*utils*), 8, 406
rt (*stats*), 99, 137
rug (*graphics*), 49, 50, 313
runif (*stats*), 88, 99, 215, 439, 470

s (*mgcv*), 236, 238, 240, 242, 243
sammon (*MASS*), 384, 385, 409
sampdist, 90, 94
sample (*base*), 90, 99, 100, 129, 153, 256, 349, 353, 375, 376, 383, 394, 470
samplingDist, 89
sampvals, 89, 90
sapply (*base*), 20, 21, 35, 42, 53, 66, 223, 260, 261, 311, 322, 344, 374, 375, 380, 415, 416, 441, 456, 457, 469, 484, 493
save (*base*), 430, 431
save.image (*base*), 33, 431
scale (*base*), 328
scale_x_continuous, etc. (*ggplot2*), 328
scan (*base*), 434, 436, 437
scatter3d (*Rcmdr*), 428, 491, 492
scatterplot (*car*), 428
scoreplot (*hddplot*), 395, 396, 404, 405
sd (*stats*), 22, 23, 66, 103, 104, 107, 144, 339, 471
search (*base*), 430
seq (*base*), 12, 41, 54, 88, 90, 101, 117, 230, 323, 353, 360, 442, 458, 470, 481
sessionInfo (*utils*), 9, 431
set.seed (*base*), 86, 87, 90, 141, 156, 291, 312, 402
setwd (*base*), 431
show.colors (*DAAG*), 474
show.settings (*lattice*), 481
showMethods (*methods*), 460
showprop, 415
signif (*base*), 129, 316
simpleKey (*lattice*), 479
simpleTheme (*lattice*), 57, 58, 378, 379, 423, 479, 480, 482
simulate (*lme4*), 88, 328
simulate.distribution, 462
simulateLinear (*DAAG*), 159
simulateScores (*hddplot*), 396
sin (*base*), 27, 28
sink (*base*), 437, 438
slot (*methods*), 460
slotNames (*methods*), 460
smooth (*stats*), 488, 491
smooth.spline (*stats*), 235
solve (*Matrix*), 450
sort (*base*), 6, 20, 23, 192, 412, 416, 448
source (*base*), 125
spline (*stats*), 230
split (*base*), 66, 90, 104, 130, 137, 260, 261, 311, 313, 322, 344, 353, 457, 458, 469
splom (*lattice*), 32, 175, 179, 181, 191, 378, 412, 416
sqrt (*base*), 2, 34, 40, *passim*
stack (*utils*), 18, 41, 117, 454
stem (*graphics*), 46
str (*utils*), 15, 20, 96, 274
strheight (*graphics*), 29, 458
strip.custom (*lattice*), 90, 483, 484
stripplot (*lattice*), 32, 52, 120, 217, 304, 477, 483, 484, 486
strsplit (*base*), 436, 440, 441, 469, 475
StructTS (*stats*), 300
sub (*base*), 52
subset (*base*), 12, 16, 44, 114, 263, 274, 275, 415, 416, 418, 419, 482, 489
substitute (*base*), 23, 475, 476, 484
substring (*base*), 340, 341, 440
sum (*base*), 12, 39, 61, 62, 247, 260, 344, 374, 375, 388, 390, 408, 409, 415, 420, 441
summary (*multtest*), 6, 15, 34, 36, *passim*
summary.aov (*stats*), 218, 330
summary.lm (*stats*), 186, 218, 220–3
summary.rpart (*rpart*), 366
supsmu (*stats*), 486
Surv (*survival*), 274, 276, 277
survfit (*survival*), 274, 276
svm (*e1071*), 406
Sweave (*utils*), 467
switch (*base*), 415, 462
symbols (*graphics*), 29, 473
sys.call (*base*), 463, 464
Sys.Date (*base*), 443
sys.frame (*base*), 463, 465
Sys.getenv (*base*), 432
sys.nframe (*base*), 463
sys.parent (*base*), 463
system.file (*base*), 432, 466
system.time (*base*), 450, 471

t (*Matrix*), 88, 117, 312, 316, 344, 395, 449, 470
t.test (*stats*), 107, 109, 111, 137, 341
table (*base*), 20, 21, 39–41, (*passim*)
tail (*utils*), 15
tapply (*base*), 317, 456
termplot (*stats*), 179, 193, 194, 212, 224, 255, 419
test, 463
testfun, 476
text (*graphics*), 25, 26, 29, 32, 39, 41, 75, 117, 170, 192, 247, 248, 341, 354, 356, 360, 365, 367, 407, 409, 458, 474, 476

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

text.rpart (*rpart*), 356
textGrob (*grid*), 484, 486
theme_bw (*ggplot2*), 490
theme_gray (*ggplot2*), 491
theme_set (*ggplot2*), 491
tiff (*grDevices*), 472
title (*graphics*), 39, 474, 475
topo.colors (*grDevices*), 474
toupper.initial, 330
trellis.device (*lattice*), 31, 480, 481
trellis.focus (*lattice*), 485, 486
trellis.panelArgs (*lattice*), 485, 486
trellis.par.get (*lattice*), 480
trellis.par.set (*lattice*), 57, 480, 481
trellis.unfocus (*lattice*), 486
ts (*stats*), 14, 48, 292, 458
tuneRF (*randomForest*), 369, 370
twot.permutation (*DAAG*), 129

unclass (*base*), 40, 170, 297, 317, 397, 409, 444, 475
unique (*base*), 20, 144, 260, 311, 332, 416, 469
unit (*grid*), 484, 486
unlist (*base*), 16, 35, 54, 418, 454, 481
unstack (*utils*), 18, 41, 323
update (*stats*), 52, 54, 57, 58, 263, 331, 343, 479–83, 485
update_geom_defaults (*ggplot2*), 491
UseMethod (*base*), 459

var (*stats*), 66, 100, 104, 111, 121, 309
VarCorr (*lme4*), 315, 318, 349
vif (*DAAG*), 201–3
vignette (*utils*), 468
vis.gam (*mgcv*), 240

weekdays (*base*), 442, 443
which (*base*), 20, 192
which.max (*base*), 20
which.min (*base*), 20
white.test (*tseries*), 299
window (*stats*), 14, 48, 49, 458
windows (*grDevices*), 28
wireframe (*lattice*), 32
with (*base*), 17, 18, 20, 25, 26, 29, 36, *passim*
woolf, 139
write (*base*), 437
write.table (*utils*), 437

x11 (*grDevices*), 28
xlab (*ggplot2*), 383
xtable (*xtable*), 467
xtabs (*xtabs*), 20, 21, 61–3
xyplot (*lattice*), 30–2, 36, 53, 55–8, 101, 239, 327, 336, 339, 340, 381, 422, 423, 429, 442, 478–87

ylab (*ggplot2*), 383

z.inverse, 140
z.transform, 140

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

Index of terms

Akaike information criterion (AIC), *see* information criteria
analysis issues
  analysis of summary data, 20, 59, 61, 63, 65, 67
  analysis styles & traditions, xix, 43, 44, **71–3**, 77, 136
  assumptions, 43–5, 58–9, **91–8**, 115, 117–18, 221
  changes of plan, 73
  data snooping, 43
  deficiencies, 72
  independence, 73, 91, 92, 115, 133, 283, 302
  non-parametric methods, **95–6**, 98, 240, 352, 373
  planning, 72
  presentation issues, 111
  prior information, 135
  random sampling assumptions, 91
  robust & resistant methods, 91, 147, 149, 167, 183, 184, 191–4, 212, 236, 241, 295, 296
  significance tests, 171
  source/target differences, 153, 158, 189, 303, 307, 328, 361, 385, 387–8, 414, **493–4**
  strategies, 71, 72, 214
analysis of variance
  decomposition (anova), 158, 159, 321
  degrees of freedom, 121, 305–6
analysis of variance model (aov), 120–1, 158–9, 218–23, 305, 308, 319, 321–2, 330, 350
  as linear model (lm), 217–21
  categorized data *vs* linear fit, 125–6, 158–60
  multi-way, 127, 222–305, 319–32
  multiple comparisons, *see* inference
  one-way layout, 120–2, 127, 218, 305, 307, 309, 313
argument, *see* function
array, 60, 139, 256, 448, **451–2**, 455–6
  dimensions, 139, 256, 448, 451, 456
  permutation of dimensions, 61, 452
assignment, 2, 20, 35, 36, 446
  subscripted, 446

Bayesian Information Criterion (BIC), *see* information criteria
Bayesian methods, 8, 98, **132–6**, 165, 166, 312, 388
  posterior density or probability, 133, 134
  prior density or probability, 133–5, 141, 312, 386–7
bootstrap, *see* resampling methods

C_p statistic (Mallows), *see* information criteria
censoring, *see* survival analysis
classes & methods, 30, 443, 458–61
  S3, 458–61
  S4, 460, 461
coefficients
  GLM
    standard error, 258, 263–4, 267, 281
  linear model
    confidence intervals, 171–2
    correlation between estimates, 172–3, 181, 229
    interpretation, 174–83, 208–9
    regression spline coefficients, 234–5
    standard error, 220, 222, 226, 228, 231
commands
  comment character, 2
  continuation character, 2
concatenation, 3, 10
confidence interval (CI), 68, 107–9, 111–13, 124–5, 149–50, 156, 171, 278, 426
  1 & 2-sample means, 107–9, 111–12, 140, 301–2, 341
  autocorrelation, 289
  correlation, 68, 113, 131, 132, 140
  median, 130–2
  predicted values, 149–50, 171–2, 187–8, 215, 228–30, 232–3, 238, 275, 290, 426
  proportions, 107, 112, 113
  regression coefficients, 149–50, 171, 220, 312, 316, 426
confounding, 119
contrasts, *see* factor
correlation, 67–9, 76, 110, 147, 163, 172, 177, 310, 316, 334–6
  Kendall, 68
  linear (Pearson), 67, 68, 76, 113, 140, 163, 201, 207–8, 284–5

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

confidence interval, 68, 113, 140
confidence interval (bootstrap) 131–2
rank (Spearman), 67, 68
R^2, *see* regression, linear model
cross-validation
see resampling methods

data
database connection, 438
input, 4, 427, 433–5, 437
comment character, 2, 435, 437
tracking errors, 435
management strategy, 33, 35, 430, 431
manipulation
apply family of functions, 456–8
convert to/from tables or matrices, 257
combine data objects (cbind, rbind), 19, 169, 177, 251, 296–7, 323, 388, 390, **444–5**
computational efficiency, 450
count & identify NAs, 24, 38, 381, 383, 412, **447**
merge data frames, 455
omit NAs, 24, 316, 374–5, 380–1, 390, 412, 447–8
plyr package: aaply, daply & related functions, 450
reshape data frames, 454, 484
selection & matching, 23–4
split by levels, 64, 66, **457–8**, *passim*
subset, 11–12, 15–16, 23, 31, 44, *passim*
transpose matrix, data frame, 395, 449
measurement issues, 79, 81, 84, 169, 213, 334, 335, 426
output
database connections, 438
decimal places, 30
write to file, 437–8
patterned, 11, 12
summary, 20, 59, 61, 63, 65, 67
data analysis & commentary
(Data sets are in *DAAG*, unless otherwise indicated)
Antarctica 800KYr ice core (`edcCO2`, `edcT`), 486
aberrant crypt foci (`ACF1`), rat colons, 258
acupuncture, real *vs* sham, 138, 139
admission rates, contrived data, 140
AIDS (`Aids2`, *MASS*), survival, 274, 275, 277
alcohol consumed, by year, country (`grog`), 478
anesthetic, effect on pain (`anesthetic`), 246–7
animal body, brain weight (`Animals`, *MASS*), 51, 67
Antiguan corn yields (`ant111b`), 304, 310
apple taste (`appletaste`), 223, 224
Australian athletes, morphology & blood (`ais`), 31, 482, 489
blood pressure, rabbits (`Rabbit`, *MASS*), 41
book weight, dimensions (`allbacks`, `softbacks`), 170, 173, 185, 186, 208, 214
biased sample (`oddbooks`), 168, 181, 182
brain, body weight, litter size (`litters`), 179, 180, 215
Canadian city populations (`cities`), 214, 445
cancer, gene expression data (`Golub`, `golubInfo`, *hddplot*), 392–405
car data
1993 Consumer reports (`Cars93`, *MASS*; `Cars93.summary`), 14, 455, 469
fuel consumption, etc, 1974 data (`mtcar`, *datasets*), 461
mileage *vs* weight (`car.test.frame`, *rpart*), 360
mileage, etc (`table.b3`, *MPV*), 215
speed *vs* distance to stop (`cars`, *datasets*), 167
car window tinting (`tinting`), 56, 329
carbon emissions *vs* year (`fossilfuel`), 3–5
CO_2 level & leaf temp (`leaftemp`), 225
comparison of firing methods (`Gun`, *MEMSS*), 349
cricketer life spans (`cricketer`), 282
cuckoo egg sizes (`cuckoos`), 52, 70
dengue projections (`dengue`), 65
depression, lawn roller weight (`roller`), 78–80, 88, 143, 147, 150, 151, 163–5
dewpoint, min & max temp (`dewpoint`), 238
distance *vs* ramp angle, starting point (`toycars`), 241
drawings of dreams, 116
effort *vs* stool type (`ergoStool`, *MEMSS*), 350
egg dimensions
host *vs* cuckoo eggs (`cuckoohosts`), 70
elastic bands, distance *vs* stretch – heated *vs* unheated (`elastic1`, `elastic2`, `elasticband`)
paired (`pair65`), 94, 103, 105, 107, 110, 141
unpaired (`two65`), 139
electrical resistance, kiwifruit (`fruitohms`), 50, 69
email spam (`spam7`), 352–6, 366, 367, 374
Fisher's iris data (`iris`, *MASS*), 41
Food Frequency Questionnaire (FFQ), 204
frogs, spatial distribution (`frogs`), 256
gastric cancer screening, 119
geological substratum thickness *vs* distance (`geophones`), 241, 242
grain/head *vs* seed rate (`seedrates`), 228
head injury, simulated (`head.injury`), 374
height & weight of women (`women`, *datasets*), 204, 209, 214
hill race times (`nihills`, `hills2000`), 174, 175, 188, 189, 191, 214, 215, 438
house prices (`houseprices`), 157, 158, 312
human power, O_2 intake (`humanpower1`, `humanpower2`), 336, 338, 460
insurance claims (`Insurance`, *MASS*), 469

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

data analysis & commentary (*cont.*)
iron slag, magnetic *vs* chemical values (`ironslag`), 145
jobs, by Canadian region (`jobs`), 14, 18, 21, 32, 53–5, 442, 454, 455, 458, 481
kidney stone operations, 60, 61
kiwifruit shading (`kiwishade`), 69, 319, 320, 324, 325, 334, 345
labor training program (`nswdemo` & related datasets), 59–60, 71, 114–16, 136–8, 168, 374, 414–27
lake area, elevation (`Manitoba.lakes`), 38, 39
levels of Lake Huron (`LakeHuron`, *datasets*), 283–7, 289, 290
maths achievement (`MathAchieve`, *MEMSS*), 350
measles, deaths in London (`measles`), 47, 48
milk sweetness by additive (`milk`), 49
model cars (`modelcars`), 125
mortality, female heart attacks (`mifem`), 281, 364, 374
moths, occurrence (`moths`), 261–4, 266, 281, 332, 333
oil platform escape times (`ex01.36`, *Devore6*), 40
ozone levels (Antarctic), 1956–2000 (`ozone`), 242
plant architecture, leafshape (`leafshape`), 408
population growth
Australian states (`austpop`), 25
possums, morphometric data (`possum`), 44, 378–80
pressure of mercury vapour, *vs* temp (`pressure`, *datasets*), 168
primate body, brain weights (`primates`), 25, 26, 30
rare plants, habitat type (`rareplants`), 117, 118
record times, track, road (`worldRecords`), 241
rice variety 2-factor expt (`rice`), 123, 124
road accident mortality – US (`nassCDS`), 62, 63
science, school survey (`science`), 72, 313, 315
seal morphometrics (cross-sectional data) (`cfseal`), 17, 162
selfed *vs* crossed plants (`mignonette`, Darwin's data), 110, 140
skull growth in children (`Orthodont`, *MEMSS*), 340
social support survey (`socsupport`), 40, 71, 72, 410–13
Southern Oscillation Index, rainfall (`bomregions`), 291–2, 295–7, 302, 429, 487–8
space shuttle damage (`orings`), 38
sugar weight, wild type *vs* GM plants (`sugar`), 218
tomato yield, salinity (`ex10.22`, *Devore6*), 75
tree dimensions, biomass (`rainforest`), 12, 38, 457
UCB admissions, by sex & dept (`UCBAdmissions`, datasets), 96–7, 139, 210, 256–7, 272
wages, UK 19th C cotton workers (`wages1833`), 242
census data *vs* informal survey (`cottonworker`), 40
data frame, 3–4, 8, **14–19**, 20–1, 35, **452–8**, *passim*
see also data, manipulation
as database, 371, 430, 438
as list, 16–17, 453
attach & detach, 17, 35–6
names & numbers of rows & columns, 60, 257, 291
writing, 437, 440,
data mining, 158, 351, 406, 407, 429
dates, 441–3
degrees of freedom, **66–7**, 76, 105, 109–12, 115, 127, 147, 149–50, 229–30, 232, 246, 267, 305–7, 315, 319, 325, 330, 401
density
estimate, **44–6**, 76, 102, 426
plot, *see* plot
deviance, *see* model
discriminant analysis
linear, 387, 389, 390, 395, 408, 419, 420, 422–4
distance measure, 384, 385
binary data, 385
diagnostics & diagnostic plots
GLM, 265–6
linear models, 144, 148–9, 170, 173–6, 183, 185–7, 190–1, 194, 195, 214–15, 227, 232–3, 241–3
distribution
t-distribution, 105–8, 137, 264
degrees of freedom, 99, 105–8, 137, 264
Bernoulli, 86
binomial, 82, 83, 86, 245, 266, 267, 279, 425
chi-squared, 85, 99, 114–18, 260, 267, 271, 298
cumulative probability, 82–5, 107, 266
density, *see* density estimate
exponential, 85, 88, 99
heavy-tailed, 86, 105
normal, 45–7, 58, 68, **86–9**, 92–5, 99, **102–8**, *passim*
Poisson, 82–4, 245, 262, 264, 267, 279
quantile or percentile, 26, 46, 83, 85, 131, 132, 266
sampling distribution, 88, 89, 94, 95, 102, 103, 105, 155, 328
t-statistic, 105, 106
bootstrap estimate, 155
median, 130
simulation, 86–9, 99, 101, 134, 141, 165
skew, 45–6, **58**, 71, 89, 92, 94, 95, 113, 190, 417
uniform, 85, 88, 99, 137, 165

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

document preparation
  Sweave, 467–8

experimental design, 123, 222, 319, 325, 334
exploratory data analysis (EDA), 43, 44, 72–4, 77, 378, 379, 381, 383
expression, 1–2, 12, 39, 441, 446, **461–5**, 474–6, 482–4
  print on graph, 29, 474–6, 482–4, 487

factor, 10–12, **13–14**, 16–17, 434, 441, **441–6**, 447, 453–6, 459, 479, 482–4, 489, *passim*
  coding & contrasts, 195, 218, 220–1, 445–6
  columns in model matrix, **218–20**, 225
  in model formula, 372, 445
  levels, **13–14**, 17–18, 20, 31, 35, 39, 56, **444–5**, 447, 455–6, 475, 479, 482–4, *passim*
    order, 13–14, 444–5
  ordered, 14, 444–5
  reorder levels, 13, 120, 144, 257, 262–3, 332, 418, 484
  split by levels, *see* data, manipulation
file names, 4, 9, 431, 432, 433
fitted values, *see* prediction
function, 3, **19–25**, 429, 431–3, **438–44**, *passim*
  anonymous, 441, 451, 475
  argument, 4, **23**, 168, 439, 476
    abbreviated, 439
    lazy evaluation, 464–5
    the . . . argument, 439–40
    use list to pass arguments, 462–3
  common useful functions, 19–20, 431–2
  environment, 22, 36, 432–3, **463–5**, 469
    evaluation frame, 463
  generic, 21, 30, 36, **458–9**, 467, 474
  issues for writing & use, 439
  return value, 23, 31, 41, 254
  utility functions, 21, 431–2, 463, 466, 468

generalized linear model, *see* regression, generalized linear model (GLM)
generalized linear mixed model, *see* regression, generalized linear mixed model
ggplot2 graphics, 33, 383, 408, 472, 487, 488, 491, 492
  automatic generation of keys, 383
  functions return ggplot objects, 487–8
  layers, 488
  use of quickplot(), 487–90
graph
  see plot
graphics
  see also ggplot2; lattice; plot
  devices, 27, 30, 31, 57, 464, **472–4**, 477, 479, 480–1
  good practice, 32
  links with analysis, 44

image file, *see* R session
inference
  1 & 2-sample means, 103–14
  Bayesian, *see* Bayesian methods
  bootstrap methods, *see* resampling methods
  confidence interval, *see* confidence interval
  confidence interval *vs* hypothesis test, 113–14
  hypothesis test, 107–8, 114
  likelihood ratio test, 133, 265, 277, 278
  maximum likelihood, 133, 141, 286–7, 299, 329, 331, 343
  multiple comparison, 73, 119–21, **122–3**, 209, 220, 319, 329, 352, 373–4, 476
    Tukey's HSD, 120–2, 220
information criteria
  Akaike Information Criterion (AIC), **186–7**, 194–5, 202, 248, 253, 259, 262–4, 270, 271, 287–9, 294, 297, 298, 310, 314, **330**, 332, 338, 342, 343, 425
  Bayesian Information Criterion (BIC), 187, 289, 294, 297–8, 310, 314, 330, 332, 338, 342, 343
  C_p (Mallows), 187
interaction plot, 124

lattice graphics, 25, **30–3**, 36, 52, 56, 74, 93, 152, 412, 427, 428, 472, 475, **477–87**, 492
  add smooth curve, 239, 429, 486
  adding to plots, 52, 54, 57–8, 479–85
  box plot (bwplot), 32, 47, 52, 74, 76, 477
  built on grid package, 33, 477
  conditioning factor or variable, 31–2, 428, 479
  dotplot, 40, 64, 75, 123, 261
  functions return trellis objects, 30–1
  interaction with plots, 485–6
  keys & legends, 31, 53, 56, 58, 64, 261, 327, 378–9, 381, 416, 423, 442, **479**, 482
  layers (using functions from *latticeExtra*), 483–4
  layout of panels, 55, 64, 76, 89, 101, 144, 239, 327, 340, 422, 481
  panel function, 64, 89, 123, 261, 336, 483
  par.settings, 57–8, 378–9, 423, 479–80, 482
  point & text size, 412, 479–80, 492
  print from user functions, 31, 472, 477
  scaling of axes, 33, 475
  strip plot, 32, 52, 477, 483, 484, 486
library, 9, 30, 179, 310, 483, 485
  see also package
leverage
  GLM, 266, 268
  linear model, 149, 173, 176, 183–6, 195, 227, 233, 266, 419, 421
    plot residuals *vs* leverage, 173, **184–5**, 227, 233
list, 16–18, 23, 35, 81, 374, 440, **452–4**, 457–62, 463, *passim* in code
  concatenation, 36
  data frame as list, 16, 453

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

Markov chain, 100, 101
Markov Chain Monte Carlo (MCMC) estimation, 134, 165, 312
matrix, 18–20, 36, 89, 100, 225, 436, 444–5, **448–51**, 455–8, 470, 474, 492, *passim*
  see also data, manipulation
  arithmetic, 450
  combine (cbind, rbind), *see* data, manipulation
  convert to/from data frame, 450, 455–6
  convert to/from vector, 19, 448–9
  extract sub-matrix, 19, 449
  form as outer product, 451
  names & numbers of rows & columns, **18–19**, 449
  storage in memory, 35, **430–1**, 450, 471
  subscripts, 19, 449
  transpose, 395
mean
  trimmed, 75, 76
methods & classes
  see classes & methods
missing values, **12–13**, 20, **24**, 38, 242, 373, 381, 383, 390, **446–7**, 455 *see also* data, manipulation
model
  allometric growth, 161–2, 167
  classification, *see* discriminant
  discriminant, *see* discriminant
  GAM, *see* regression, generalized additive model (GAM)
  generalized linear mixed model, 332–4
    observation level random effects, 332
  GLM, *see* regression, generalized linear model (GLM)
  linear model, *see* regression, linear model
  model formula, **80**, 143, 305, 321, 360, 371–2, 445
  model object, 81, 459
    extractor function, 81, 143, 460
  multinomial (multinom), 390
  parameters, 79, 133, 134, 152, *passim*
  signal & noise, 78, 97, 98, 469, 470
  statistical *vs* deterministic, 77–9, 97
  survival analysis, *see* survival analysis
  time series, *see* time series
multi-dimensional scaling (MDS), *see* ordination
multi-level model, **302–50**
  fitted values, 311–12, 327, 331
  fit by restricted maximum likelihood (REML), 339, 331, 343
  random coefficients, 336–44
  repeated measures, 334–44
  residuals, 306–7, 311, 317, 325, 327–8
  variance components, 303, 308–9, 315–16, 318–19, 322, 323, 325, 326, 336, 344, 345, 349
multivariate analysis, 377–409

nonlinear model (nls), 210–11

object
  save, *see* R session
operator, 11, 14, 23, 444, 460, 465, 488
  arithmetic, 2, 444, 450
  assignment, *see* assignment
  logical, 443
  relational, 11, 14, 23
ordinal logistic model, *see* regression, ordinal logistic
ordination
  distance measure, *see* distance measure
  multi-dimensional scaling (MDS), 383
  principal components analysis (PCA), 169, 196, 198, 203, 347, **377–84**, 386, 407, 410–13, 426
    loadings, 382, 407, 411–13, 426
    regression on PCA scores, 410–13, 426
outliers, 132, 147–9, 169, 170, 183–6, 190, 192, 193, 211, 212, 215, 237, 242, 295, 426

package
  base, 9, 432, 465
  boot, 130, 156
  car, 186
  cluster, 384
  DAAG, 1, 5, 9, *passim*
  datasets, 6, 9, 30, 41, *passim*
  Deducer, 429
  Devore6, Devore7, 40, 75
  dichromat, 474
  dr, 191
  fgui, 427, 429
  forecast, 283, 288–90, 300–1
  foreign, 434
  fseries, 300
  gam, 236
  ggplot2, 33, 383, 408, 472, 491, 492
  golubEsets, 392
  grid, 33, 468, 477, 483, 485
  hddplot, 392, 395, 404
  KernSmooth, 237
  lars, 199
  latticeExtra, 486
  latticist, 427, 485
  lazy data mechanism, 9
  leaps, 197
  lme4, 127, 303, 305, 312, 337, 346, 460
  locfit, 240
  lqs, 192
  MASS, 9, 14, 40–1, *passim*
  MEMSS, 304, 340, 349–50
  mgcv, 235
  monoProc, 237
  muhaz, 277
  multtest, 396
  nlme, 303, 304, 312, 348, 446, 467
  playwith, 427, 485
  plyr, 450

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

pmg, 427
randomForest, 351, 352, 369, 371, 372, 374, 406, 408
Rcmdr, 427, 491
RColorBrewer, 474
reshape, 454, 484
RMySQL, ROracle, RSQLite, 438
rpart, 351, 352, 357, 371–4
stats, 9, 283, 287, 300
strucchange, 300
survival, 235, 256, 274, 280
tseries, 283, 299
xtable, 467
plot
see also ggplot2; graphics; lattice
aspect ratio, **28**, 31, 64, 93, 120, 123, 144, 217, 239, 328, 422–3, 491
axes, 33, 51, 78, 85, 117, 158, 475, 480–1, 488, 491
box-&-whisker (boxplot), 41, 47, 52, 58, 73, 75, 86, 122, 137, 313, 353, 394, 473, 477, 489
outlier criterion, 47
contour, 240, 408, 479, 489
dates as axis labels, 481
density plot, 32, 44–6, 52–3, 84–5, 89–90, 94, 99, 105–6, 122, 129–30, 133, 251, 408–9, 416, 421, 462, 480, 489–91
dynamic graphics, 484
expression, 29, 476
font family, 473
face, 473
histogram, 38, 44–6, 75, 487, 491
identify points, 29, 41, 486, 491–2
interaction plot, 124
legend, 58, 61, 413, 474, 479
mosaic (multi-way tables), 60, 61, 139
normal probability, 58, 88, **92–5**, 99, 144, 145, 227, 317, 327, 328
panel function, 64, 483
quantile-quantile (QQ), 397
shaded regions, 84, 85, 239
simulated data, 88, 93, 94, 159, 215, 281, 327, 328
size of points & text, 27, 28, 33, 451
stem-&-leaf, 46
strip plot, 52, 124, 483
transformation of scales, 40, 51, 58, 67, 160–1, 168, 175, 188, 214, 245, 250–1, 279, 292, 353
posterior density or probability, *see* Bayesian methods
prediction
predicted values
GLM, scale of linear predictor, 254, 263, 266
GLM, scale of response, 254, 333
linear model (lm), 79–80, 142, 145–6, 149, 150–1, 156, 157, 167, 172, 177, 182, 187–8, 192–3, 208, 214, 217
prediction interval (new y-value), 150
predictive accuracy, 152–3, 493–4
GLM, 255
linear model (lm), 153–8, 186–7, 189, 230
multi-level model, 303, 309, 319, 328
see also discriminant; survival analysis (survival estimate); time series; tree-based
principal components analysis (PCA), *see* ordination
printing, 20, 21, 366, 459, 463, 472, 477
digits, 36, 316, 408
prior density or probability, *see* Bayesian methods
propensity score, 410, 414–17, 419–23, 425, 426
use as regression covariate, 419–25

questionnaires & surveys, 71
quit session, *see* R session

R session
image file, 5, 430, 431, 466, 469
attachment of image files, 430–1
quit, 5, 430
search list, 17, 430, 464
database, 430
working directory, 4, 8, 33, 431, 433, 436, 466, 467
change, 431
workspace, 3–5, 8–9, 17, 21, 33, 35, 86, 428, 430–1, 433, 463, 465–6, 469, 471
image file, 3–5, 33, 428, 430, 431, 466, 469
management, 33, 35, 430, 431
random
coefficients, *see* multi-level model
numbers, 22, **86–9**, 93, 99, 155
seed, **86**, 93
sample, 71, 87, **90–1**, 99, **102**, 117, 137, 153, 186, 307, 353, 369, 420, 462, *see also* sampling
of permutations, 90, 128, 129, 199, 396, 470
regression
(Note the separate entries: regression, generalized additive model; regression, generalized linear model; regression, linear model)
see also discriminant; multi-level model; survival analysis; time series; tree-based
AIC, BIC, *see* information criteria
bootstrap, *see* resampling methods
C_p, *see* information criteria
cross-validation
see resampling methods
extrapolation, 230
fitted values, *see* prediction
functional relationship, 169
loglinear, 210, 257, 272
non-linear model, 210–13, 215, 217
observational data, 69–70, 183, 199, 212–13, 346, 385, 414, 426
normal probability plots, *see* plot, normal probability
ordinal logistic model, 268–72

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

regression (*cont.*)
resistant, 149, 183, 184, 191–4, 212, 236, 295, 296
robust, 147, 149, 167, 212, 241
spline smoothers
see regression, generalized additive model
strategies for fitting models, 189–91
regression, generalized additive model (GAM), 235, 240, 280
smooth terms, 232, 235, 236, 417
regression, generalized linear model (GLM), 210, 240, 244–68, 280–2, 332, 333
see also regression, linear model
coefficients
see also coefficients
Hauck-Donner effect, 254, 263
SEs & Wald ($z-$) statistics, 265, 267, 277–8
design, 268
deviance, 246, 248, 253, 254, 257, 259, 260, 262, 264–7, 270, 332, 424
dispersion, 260, 262, 264–7
family, 248, 253, 255, 258–60, 262, 264–6, 280
binomial, 248, 253, 255, 258–60, 262, 264–6, 280
poisson, 248, 253, 255, 258–60, 262, 264–6, 280
quasibinomial, 248, 253, 255, 258–60, 262, 264–6, 280
quasipoisson, 248, 253, 255, 258–60, 262, 264–6, 280
leverage, *see* leverage
link function, 245, 254, 255, 257, 258, 262, 266–8, 280
logistic regression, 244–6, 248, 253–7, 269, 270, 272, 279, 280, 387
predicted values, *see* prediction
residuals, *see* residuals
regression, linear model (lm), 59, 142–235, 246, 258, 263, 268, 280, 318
anova table, 226
assumptions, 183–4
bootstrap, *see* resampling methods
check for linearity, 59, 93, 152, 179, 183, 190, 193, 228
coefficients, *see* coefficients
Cook's distance, 148–9
cross-validation, *see* regression
design, 151–2
diagnostics & diagnostic plots, *see* diagnostics & diagnostic plots
errors in x, 195, 203, 208, 212
attenuation of coefficient, 195, 203, 204, 205
simulation, 204, 206
transfer of effect between variables, 206
fitted values, *see* prediction
hat matrix, 184
heterogeneity of variance, 146, 189, 210
influence, 147–9, 173, 183, 185, 419
leverage, *see* leverage
linearity between explanatory variables, 189–90
model formula, 80, 143
model matrix, 163–4, 167, 172, 184–5, 218, 201, 217–18, 220, 225, 228, 231, 234, 242, 243, 445
one term; several columns, 224, 226–7, 228–9, 231–2, 235–6, 238
model non-linear response, 228–35
multicollinearity, 199, 203, 251
remedies, 203
variance inflation factor (VIF), 201–3, 251
omission of intercept, 172
on principal component scores, *see* ordination, regression on PCA scores
on propensity scores, *see* propensity scores, use as regression covariates
outliers, *see* outliers
polynomial, 167, 194, 195, 228, 229, 231, 232, 235, 241–3, 298
predicted values, *see* prediction
R^2 & adjusted R^2, 147, 186
regression splines, 231–5
residual degrees of freedom, 121–2, 147, 150, 171, 223, 226
residuals, *see* residuals
robust & resistant methods, *see* regression
shrinkage methods, 198
straight line model, 228, 231
strategies for fitting models, *see* regression
term plots, 224, 226, 228, 230, 231, 236, 240
terms, 59, 193, 224, 226, 228, 231, 232, 235, 238, 240
variable selection, *see* variable selection
variance inflation factor (VIF), *see* multicollinearity
x on y *vs* y on x, 217, 224, 228, 230, 231, 234
replication, 140, 267
resampling methods
bootstrap, 128, 130–2, 140, 155–8, 189, 215, 351, 362, 369, 373, 375, 376, 382, 383, 416, 420
cross-validation, 153–5, 157–8, 187, 189, 198, 215, 236, 255, 256, 356, 361–3, 367, 374, 385, 387–9, 399, 401–4, 406, 420
validity of error estimate, 158
permutation, 128, 129, 396, 397
simulation, *see* simulation
residuals,
GAM, 239, 243
GLM, 254, 260, 266, 267, 278, 279
deviance, 266
Pearson, 266
working, 266
linear model, **80–1**, 143–6, 147, 148–51, 155, 164, 167–8, 173–4, 176, 178, 179, 183–5, 188, 191–5, 211–12, 215, 217–19, 227, 232, 233, 236, 242

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

loess smooths, 237
multi-level models, *see* multi-level model
time series, *see* time series
rounding, 34, 92, 344

sampling
see also random sample
cluster sampling, 91
with replacement, 90, 155, 382, *see also* resampling methods, bootstrap
search list, *see* R session
selection bias, 396–8, 407
session, *see* R session, 5, 430
simulation, 86–9, 91, 99–101, 134, 141, 158, 165, 197, 206, 207, 290, 291
smoother
see also regression, generalized additive model, spline
loess, 44, 80, 146, 168, 236, 237, 278, 359–61, 487
lowess, 50, 79, 96, 145–6, 168, 192, 236–7, 242, 279, 292–3, 295–6, 317
spline, 231–7, 239–42, 347, 418, 420, 489
knots, 231–3, 235, 242
regression spline, *see* regression, linear model
standard deviation, 22, 47, 65–7, *passim*
vs inter-quartile range, 65, 66
degrees of freedom, 66, 67, 76, 111, 133, 137, 143, 171
maximum likelihood estimate, 133, 141
pooled, 67, 76, 104, 110, 111, 141
standard error (SE), 102, 105, 109–11, 113, 130, 156, 185, 263
binomial distribution
proportions, 112, 267, 281, 425
difference of means (SED), 66, 103, 110, 111, 124, 133, 137, 141, 143
of mean (SEM), 102, 103, 105, 111, 130
of median, 102, 104, 130, 131, 143, 219, 425
survival analysis, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272–8, 280, 282, 352
censoring, 273, 275, 276, 279
non-informative, 275, 276, 278
right, 282
Cox proportional hazard, 277, 278, 280, 282
data collection, 273
frailty model (multiple levels of variation), 280
hazard rate, 275–8, 282
confidence interval, 278
residuals
martingale, 278, 279
Schoenfeld, 278
survival estimate, 276–7, 372
tree-based, 352, 372, 374

table
margins, 96, 97
of frequencies, 20, 21, 61–3
adding across tables (Simpson's paradox), 62, 96–7, 140
table formula, 62, 63
test
2-way table
chi-square statistic, 114, 116–18
Fisher exact, 116
residuals from row/column independence model, 118
3-way table
Mantel-Haenzel, 139
Woolf, 139
correlation, 113
homogeneity of variance (Fligner-Killeen), 260
likelihood ratio, 265, 277, 278
nonparametric, 999
permutation
one- & two-sample, 128–9
proportion(s), 112–13
sequential correlation, 172, 298
t-statistic
one- & two- sample, 109, 110
two-sample, unequal variances, 110–11
type III, 446
Wald, 265, 267, 277–8
white noise (Ljung-Box), 297–8, 301
time series, 10, 14, 47, 48, 53, 55, 158, 165, 283–90, 292–4, 296, 298–303
ARCH or GARCH model, 298–300
ARMA or ARIMA model, 287–98
automated selection, 289–91, 293, 294, 296–8
autocorrelation, 165, 284–90, 293–5, 298, 302
autoregressive (AR) model, 165, 283, 285–7, 298–300, 301
Ljung-Box test (white noise), 297–8, 301
moving average (MA) model, 283, 287–90
partial autocorrelation function, 284–9, 293, 294
predicted values, 290
residuals, 289, 293–8, 301
transformation, 40, 51, 58, 59, 67, 73, **160–1**, 168, 175, 189–90, 196, 210, 279, 286, 292, 354, 416–19, 424
Box-Cox, 161
count data
angular, 279
complementary log-log, 279
logit, 245, 254, 279, 424
probit, 279
square root, 160, 250, 279
cube root, 160, 292–3, 296
logarithmic, 51, 55, 58, 59, 160, 161, 175, 188, 190, 210, 214, 249, 250, 286, 354, 416, 418
logit, 245, 254, 279, 424
power, 160, 161, 196
tree-based, random forests (randomForest), 369, 371, 375, 420, 422

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

tree-based, decision trees (rpart)
  classification, 351, 352, 354, 358–60, 362, 364, 366, 368, 369, 372–4
  cost-complexity (CP), 361–2
  error or impurity measures, 359
  information on each split, 354, 357, 359, 366, 374
  output, 354, 357, 359, 366, 374
  predicted values, 360, 366
  predictive accuracy, 355, 356, 361
  pruning, 362, 372
  splitting criterion, 356, 359, 362, 372
    1 SE rule, 365–6
  table of accuracies, 355, 364, 366
  tree diagram, 354, 356, 360, 362, 363, 370
  regression with a continuous outcome, 348, 357, 359–61

variability
  heterogeneity, *see* model, homogeneity of variance
variable selection
  discriminant analysis, 398, 401
    unbiased accuracy estimates, 398–404
    simulation, 398–404
  linear models, 196–9, 212, 411
    realistic SEs, 197–9
    simulation, 197–9
  pruning decision trees as variable selection, 372
  survival analysis, 273
vector
  atomic, 452, 454
  character, 6, 13, 16, 19, 21, 36, 434, 435, 440–2, 444, 448, 452, 454, 455, 469, 474, 475
    number of characters, 6, 440
    splitting, 436, 440
  complex, 10, 434, 448, 452
  concatenation, 10
  logical, 10, 11, 19, 20, 24, 36, 434, 443, 446, 448, 452, 455
  numeric, 10, 13, 15, 16, 19, 30, 36, 217, 434, 436, 444, 448, 451, 452, 454, 455
  patterned, 11, 12
  recursive (list), 457
  subset, 11–12, 14, 23, 371, 393, 452

working directory, *see* R session
workspace, *see* R session

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

Index of authors

Agresti, 139
Aitchison, 199
Aldrich, 97
Ambroise and McLachlan, 392, 406
Andersen, 74

Barnett, 91
Bartholemew, 345
Bates, 305
Bates and DebRoy, 461
Bates and Watts, 213
Belson, 351
Berk, 203, 373, 406
Bickel *et al.*, 96
Blake and Merz, 353
Bland and Altman, 209, 280
Bolker, 136, 212
Boot and Maindonald, 242
Box and Cox, 161
Braun and Murdoch, 468
Breiman, 98
Brockwell and Davis, 300
Burns *et al.*, 56
Bussolari, 336

Canty, 130
Carroll, 204
Carroll *et al.*, 206
Chalmers and Altman, 348
Chambers, 37, 461, 468
Chanter, 74
Charig, 60
Chatfield, 74, 300
Christie, 242
Chu *et al.*, 141
Clarke, 98
Cleveland, 33 37, 74, 237
Clutton-Brock *et al.*, 242
Cochran and Cox, 346
Cohen, 209
Collett, 280
Cook and Swayne, 406, 492
Cook and Weisberg, 184, 190, 212
Cox, 136, 348
Cox and Cox, 385
Cox and Reid, 136, 348
Cox and Wermuth, 68, 213

Dalgaard, 36
Daniels *et al.*, 345
Davison and Hinkley, 132
Dehejia and Wahba, 414, 417
Diggle, 300
Diggle *et al.*, 331, 335, 348
Dobson, 280
Donner and Klar, 91

Edwards, 213
Efron and Tibshirani, 131, 132
Efron *et al.*, 199
Eubank, 240
Ezzet and Whitehead, 269

Fan and Gijbels, 237
Faraway, 212, 240, 280
Farmer, 62, 63
Finney, 280
Fisher, 348
Fox, 36, 186, 212

Gardner *et al.*, 74
Gaver *et al.*, 347, 348
Gelman and Hill, 348
Gelman *et al.*, 136
Gentleman, 37, 461, 468
Gentleman and Lang, 467
Gentleman *et al.*, 407
Gigerenzer, 135, 136
Gigerenzer *et al.*, xix, 136
Gihr and Pilleri, 162
Goldstein, 348
Golub *et al.*, 392
Gordon, 70, 385
Gordon *et al.*, 208
Gourieroux, 300

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

Grasso *et al*., 123
Guy, 47

Hales *et al*., 65
Hall, 98, 237, 390
Harker and Maindonald, 232
Harlow *et al*., 136
Harrell, 24, 196, 197, 203, 212, 280
Hastie *et al*., 196, 198, 203, 212, 330, 373, 406
Hauck and Donner, 268
Hoaglin, 43
Hobson, 116
Hunter, 249
Hyndman and Khandakar, 300
Hyndman *et al*., 300

Ihaka and Gentleman, 468
Izenman, 385

Johnson, 96, 98

King, 386
King and Maindonald, 386
Krantz, 136
Krzanowski, 406

Lalonde, 59, 115, 414, 417
Latter, 52, 70
Leavitt and Dubner, 212, 344
Leisch, 468
Liaw and Wiener, 352
Lim and Loh, 373
Linacre, 238
Linacre and Geerts, 238
Linde *et al*., 138
Lindenmayer *et al*., 44, 390
Lumley, 461

Maindonald, 37, 74, 136, 158, 213, 319, 392, 404
Maindonald and Burden, 392, 404
Maindonald and Cox, 74
Maindonald *et al*., 266, 280
Manly, 406
Marland *et al*., 3
McCullagh and Nelder, 266, 267, 280
McLellan *et al*., 258
McLeod, 228
Meyer, 406
Meyer and Finney, 62
Miller, 111, 116, 136, 285
Mitchell, 47
Muenchen, 468
Murrell, 33, 37, 492
Myers, 203, 213

Nadel and Bussolari, 336
Nelder, 74
Newton, 70
Nicholls, 136
Nicholls *et al*., 291

Ord *et al*., 300

Payne *et al*., 346, 348
Pinheiro and Bates, 335, 348
Pollard *et al*., 396

R Development Core Team, 468
Ramsay and Silverman, 347
Rao and Wu, 196
Ripley, 362, 373, 406
Rosenbaum, 212, 426
Rosenbaum and Rubin, 414, 426

Sammon, 384
Sarkar, 33, 492
Schatzkin *et al*., 204, 205
Schmidt-Nielsen, 162
Senn, 47, 136
Shanklin, 242
Sharp *et al*., 208
Shumway and Stoffer, 300
Simpson, 97
Smyth, 123
Snelgar *et al*., 319
Snijders and Bosker, 348
Spiegelhalter *et al*., 330
Sprent, 163
Steel *et al*., 126
Stewart *et al*., 78
Stidd, 292
Stiell *et al*., 281
Stocks, 47
Streiner and Norman, 72, 411, 426

Talbot, 308, 348
Therneau and Atkinson, 351, 372, 373
Therneau and Grambsch, 280
Tippett, 52
Trapletti and Hornik, 299
Tufte, 37
Tukey, 113
Turner *et al*., 347, 348

Vaida and Blanchard, 330
Venables, 212, 446
Venables and Ripley, 9, 36, 196, 212, 237, 240, 280, 300, 348, 351, 372, 373, 385, 406

Würtz, 300
Wainer, 37
Weisberg, 212
Welch, 111
Wickham, 492

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76293-9 - Data Analysis and Graphics Using R – an Example-Based Approach: Third Edition
John Maindonald and W. John Braun
Index
More information

Wilkinson, 487
Wilkinson and Task Force on Statistical Inference, 37, 74, 136
Williams, 208
Williams *et al*., 348
Wonnacott and Wonnacott, 136
Wood, 240, 280
Xie and Cheng, 91
Young and Smith, 136
Zeger *et al*., 207, 208
Zhang and Singer, 374
Zhu *et al*., 392

© in this web service Cambridge University Press www.cambridge.org