Index

Applications
ANCOVA, 130
approval, 403
binomial regression, 335, 345
cattle feed, 405
coins, 303
crime, 125
differential invalidation, 345
geography, 207, 208
GLM, 276, 279
heteroskedasticity, 210
imputation, 410
lifetime, 249
logistic regression, 335, 345
median regression, 229
nominal regression, 395
occupation, 395
ordinal regression, 403, 405, 410
physics, 210
Poisson regression, 373
property crime, 229
South Sudan, 176
terrorism, 373
time series, 206
voting, 138, 164, 176, 212, 276
wealth, 168, 279
R data
cattleData, 405
coinflips, 304
cows, 138, 139, 164, 276
crime, 125, 228, 365
fakepoisson, 359
gdpcap, 279
gdp, 168
gssocc, 395, 403
ocanada, 335
rur2013parl, 214
sri2010pres, 345
summary.aov, 145
suvr, 410
terrorism, 373
xsd2011referendum, 176
R functions
– (drop), 178
AIC, 318
BIC, 320
I, 378
ROC, 312
&&, 105, 106
accuracy, 307, 310
anova, 340
aov, 132, 341
attach, 125
no, 176
autocor.test, 206
barplot, 477
binom.test, 90
bptest, 103, 104, 126

cbind, 336
col, 207
confint, 105, 106, 129, 137, 211, 212
cor.test, 140, 405
c, 112, 211, 212
data.frame, **147**, 166, 170, 280, 366, 406, 411
exp, 359
fligner.test, 133
for-loop, 90, 92, 105, 106, 126, 477
function, 477
glm.nb, 372, 374
glm.nb,offset, 374
glm, 276, **304**, 336, 366, 374
glm,family, 278, 279
glm,link, 278, 279
glm,offset, 374
glm,quasipoisson, 370
head, 85
hetero.test, 132
hist, 83, 94
ks.test, 91
legend, 408
length, 95
library, 98, 228
lines, 180
lm, 89, 90, 94, 103, 108, 112, 125, 134, 169, 178, 199
lm,weights, 212, 215
logistic, 165, 166
logit.inv, 165
logit, 165
make.link, 278–280, 301
matrix, 310
mean, 108, 153
multinom, 398, 406
ordered, **410**
overlay, 83, 125, 131
plot, 95, 103, 126
points, 129
polr, 403, 411
predict, 108, 129, 137, 147, 148, 166, 170, 179, 280, 305, 366, 398, 404, 406, 411
predict,se.fit, **179**
qchisq, 368
qqline, 82
qqnorm, 82
quantile, 171
rcauchy, 93
rchisq, 93
read.csv, **125**, 139, 228, 405
residuals, 95, 125, 131, 206, 341
rexp, 83, 92
rnorm, 81, 89, 90, 94, 95, 98, 103, 105, 106, 477
row, 207
rq, 227, 228, **228**, 229, 231, 232
rt, 87
runif, 98

runs.test, 98, 126, 131, 341, 480
sample, 126, 310
seq, 95, 103, 105, 106, 112
set.base, 134, 399
set.seed, 85, 90, 95, 98, 108, 112, 359
shapiroTest, 85–87, 125, 131
sort, 359
source, **81**, 98, 125, 139
subset, 366, 372
summary.aov, 130, 133
summaryHCE, 182
summary, 89, 90, 94, 125, 132, 276
tail, 86
vif, 113, 146
which, 178
R packages
Epi, 312
MASS, 372, 403, 410
RFS, 85, 98, 165, 182, 307, 399, 480, **492**
car, 113, 146
lawstat, 98, 125
lmtest, 103, 125
nnet, 398, 405
quantreg, 227

accuracy, 307, 311, 400, 404
AUC, 312
maximum, 309
rate, 307
relative, 308
additive model, 112, 131, 145
AIC, 277, 278, **317**, 340, 359, 400, 411
alpha-testing, 89, 92, 93
ANCOVA, 130
ANOVA, 45, 132, 466
effects model, 50, 51
means model, 49
table, 131
toy example, 24
ANOVA table, 144
appropriate test, **89**
AR(1), 207
AUC, 312
autocorrelation, 206, 207

back-transform, 166, 174, 305
base category, 134, 399
bias, 28, 33, 61, 99, 255, 446, 451
bias-variance trade-off, 373, **379**
BIC, **320**, 340, 359
bijection, 252
Binomial test, 90
boolean, 105
bounds, **161**
one, 168
two, 162, **163**
others, **173**
Breusch-Pagan test, **103**, 126, 133

canonical link, 270, 272, 274, 299, 364

CDF, 300, 455
center of gravity, 26
centering, 35
Central Limit Theorem, *see* CLT
classical linear model, *see* CLM
CLM, 124, 160, 199, 205, 251, 266, 267
CLT, 88, 93, 458, **480**, 488
coefficient of variation, 139
conditional distribution, **269**, 298, 333, 336, 363
confidence band, 75, 179, 310
confidence interval, 64, 70, 71, 74, 107, 129, 137, 228, 302
confusion matrix, 311
correlation, 113, 140, 450, 455
  population, 455
  sample, 450
  symmetric, 450
covariance, 33, 62, 439, 449, 455
  bilinear, 449
  population, 455
  sample, 449
  symmetric, 449
covariance matrix, 267
coverage, **105**, 108
cumulative distribution function, *see* CDF

data matrix, *see* design matrix
degrees of freedom, 435
design matrix, 35, 80
deviance, 447
differential invalidation, **176**, 212, 345
directional hypotheses, 138

electoral forensics, 176, 212, 345
estimator, **31**
exponential class, **271**, 333
exposure, *see* offset
extrapolation, 23

factor analysis, 115
Fligner-Killeen test, 133
function
  absolute value, 224
  beta, 467
  exponential, 170
  gamma, 462, 464, 467
  identity, 370
  logarithm, 168, 361
  logistic, **163**, 166, 295
  logit, **163**, **272**, 294
functional form, 95, 96, 101, 126, 131, 272

Gaussian process, *see* normality
generalized least squares, *see* GLS
generalized linear model, *see* GLM
geographic regression, 208
GLM, **267**, 268
  vs CLM, 268
  assumptions, 272
  components, 268, 290, 296, 298, 304, 361
  distribution, **269**

exponential class, **271**
linear predictor, **268**
link function, **270**, 275
GLS, **205**
grammar of forumlas, **143**
graphics, 128, 135, **148**, 151, **407**, 411
R philosophy, **179**

hat matrix, **37**, **203**
heteroskedasticity, 105, 182, 267, 292, 359
adjustment, 181
bulge, **106**
effects, 99, 105, 108
funnel, **106**
modeling, 198
none, *see* homoskedasticity
trumpet, **105**, 108, 362
heteroskedasticity test, 132, 133
Hildebrand Rule, 127, 395, 472
homoskedasticity, 58, **101**, 126, 132, 197, 297
Huber-White, 181

iid, 58, 60, 197, 462
independent, 60, 198, 252
independent and identically distributed, *see* iid
influential point, 229, 366
inner product, 428
interaction model, 130, 131, 143, 144
interaction term, 378
interpolation, 23, 128
interpretation, 108, 151, 176, 215, 277, 400, 408
of logarithm, **171**
of logit, **165**

Kingdom of Ruritania, *see* Ruritania
Kolmogorov-Smirnov test, **91**, 92–94
kurtosis, **457**, 459, 463, 482

latent variable, 294
likelihood, 245, 246, 249
log-likelihood, 247, 249
likelihood ratio test, **320**, 340
line of "best" fit, 16, 242
linear predictor, **268**, 294, 295, 298, 306, 333, 335
link function, **270**, 278, 298, 314, 334, 336, 363, 394
canonical, 270, 299, 333, 364
cauchit, 301
complementary log-log, 301, **314**
identity, 274, 370
log-log, 301, **315**
logarithm, 363
logit, 272, 294, 298, 301, 303, 394, 398, 404
probit, 301
logistic regression, **290**

machine epsilon, 110
matrix, 422
addition, 424
additive identity, 424

additive inverse, 425
adjacency, 209
associative, 425, 426, 434
commensurate, 424, 426
commutative, 425, 431
confusion, 311
covariance, 197, 206, 267, 437, 439
determinant, 429
diagonalization, 434
dimension, 422
distributive, 426
eigenvalue, 435
eigenvector, 435
hat, **203**
idempotent, 38, 435
inner product, 428
inverse, 48, 54, 428
invertible, 428
multiplication, 426
multiplicative identity, 428
multiplicative inverse, 428, 436
orthogonal, 38, 41, 435
positive definite, 206, 436
projection, 38, 435
rank, 48, 428, 435, 437
representation, 423
sample covariance, 439
sample mean, 438
sample variance, 439
singular, 110, 428
square, 422
symmetric, 38, 209
symmetrize, 434
trace, 434
transpose, 54, 434, 436
maximum likelihood estimation, *see* MLE
mean, 271, 362, 438, 447
distribution of, 467
linear, 447
population, 451
sample, **447**
mean squared error, *see* MSE
median regression, 17, **228**, 229
MGF, 481–486
minimum-width interval, 72
vs central, 72, 73, **73**
MLE, 16, 245–247, 251, 270, 298, 394
bias, 250
function of, 250
of β_0, 251
of β_1, 253
of σ^2, 255
uniqueness, 250
model selection, 143, 145, 164, **317**, 376
AIC, **317**
BIC, **320**
likelihood ratio test, **320**
model specification, 101
model stability, 394
moment, 481

central, 481
raw, 481
standardized, 482
moment generating function, *see* MGF
Monte Carlo, 150, 166, 171, 310
MSE, 29, 63, 72, 204
multicollinearity, 32, 45, **110**
approximate, 110
CS, 110
fix, **114**
indications, 112
logic, 111
super, 110, 111
test, 113
multiple comparisons, 472
Bonferroni adjustment, **474**

nominal regression, 396
norm, 227
L_1, 227
L_2, 227
normality, 58, 61, 81, 125, 131
nuisance parameter, **272**
null model, 43, 411

Occam's Razor, 130
odds ratio, 165, 167
offset variable, **374**
OLS, 17, 160, 197
assumptions, 28, **32**, 58, 124, 145, 160, 197
b_0, 20, 62, 99
b_1, 20, 27, 28, 62, 99
derivation, matrix, 30–31
derivation, scalar, 18–20
matrix, 197
matrix model, **30**
scalar, 197
scalar model, **18**
toy example, 21, 35
OLS assumptions, 95
ordinary least squares, *see* OLS
overdispersion, 337, 363, **368**, 375
adjustment, 369
causes, 368
effects, 368
test, 368

p-value, 89
parameter, 446, **451**, 456
point estimate, 152
Poisson regression, 359
PRE, 339
definition, **42**, 308
R^2, 42, 113, 142
$\overline{R}^2$, 43
pseudo-R^2, 44, 276, 307
preamble, 139
precision, 65
prediction interval, 67, 74, **74**, 129, 137
principal component analysis, 115
probability distribution
Bernoulli, 296, **456**, 483