
Index to R Functions, Packages, and Data Sets

->, 735
<-, 735
=, 735
[], 3, 725, 728
[[]], 728
#, 705
\$, 3
%>%, 716
%%, 28
%in%, 589
' , 725

abs(), 23, 297
ACTgpa, 547
adaptIntegrate(), 184, 213
aggregate(), 15
AIC(), 576
AirPollution, 608, 622, 626
alr3, 2, 649, 700
anova(), 488, 569, 603, 638
anscombe, 546
anti_join(), 723
any(), 732
aov(), 623, 626
apply(), 737
apropos(), 704
args(), 704
arrange(), 722
as.character(), 730
as.data.frame(), 236
as.integer(), 730
as.matrix(), 474, 776
as.numeric(), 727, 730
as.vector(), 709, 730
attr(), 725, 729
attributes(), 725, 729
avPlots(), 580

backsolve(), 685
BallDrop, 507, 545, 673
Batting, 30
bbmle, 357
beta.mom(), 233
binom.test(), 68–70, 111, 251, 252,
281, 282, 285, 294, 330
BradleyTerry2, 419
broom, 700
BTm(), 443

c(), 27, 710, 726
car, 2, 580, 700
cat(), 736
cats, 343
cbind(), 99, 100, 409, 410, 443, 474,
731
cdata(), 430
chickwts, 341
chisq.test(), 397, 443
cholesterol, 628
choose(), 47, 111
ci(), 257
CIsim(), 257, 273
cloud(), 577

- coagulation, 599, 600, 602, 619
 coda, 430
 coef(), 443, 466
 colnames(), 443
 colorspace, 25
 Concrete, 558
 confint(), 485, 513, 572, 598, 626, 659
 constrOptim(), 357
 contains(), 716
 coord_flip(), 21
 cor(), 556, 652
 Corn, 344, 515, 549
 cornnit, 546
 cubature, 184
 cummax(), 732
 cummin(), 732
 cumprod(), 732
 cumsum(), 732
 cut(), 9, 27

 DAAG, 2, 477, 543, 700
 data(), 3, 27, 702, 706
 data.frame(), 176, 213, 293, 710
 data.table, 707, 713
 datasets, 706
 dbeta(), 213
 dbinom(), 62, 63, 111
 dchisq(), 330
 density(), 172
 deparse(), 336
 derivedFactor(), 715
 derivedVariable(), 715
 deviance(), 530
 devtools, 698, 699
 dexp(), 213
 df(), 330
 df_stats(), 15
 dgamma(), 213
 dhyper(), 111
 diag(), 494
 diff(), 732
 dim(), 3, 729, 774
 Dimes, 301, 317
 dlogis(), 551
 dmultinom(), 105
 dnbinom(), 111
 dnorm(), 213
 do(), 235, 236, 238, 378, 535
 domedata, 649
 dot(), 560, 763
 dplyr, 2, 15, 700, 713
 dpois(), 111, 143, 146

 drag, 547, 548
 dt(), 330
 dunif(), 213
 dweibull(), 213

 Effect(), 582
 effects, 2, 582
 elasticband, 477, 543
 ends_with(), 716
 Endurance, 343
 example(), 705
 expand.grid(), 443
 extractAIC(), 576

 factor(), 9
 factorial(), 58
 faithful, 14, 362, 718
 FamilySmoking, 671
 faraway, 2, 502, 523, 546, 599, 602, 619, 651, 652, 700
 fastR2, 2
 favstats(), 600, 735, 737
 filter(), 12, 719
 fisher.test(), 100, 101, 111, 294
 fitted(), 467, 560, 577
 fivenum(), 29
 floor(), 128
 foreign, 708
 formatR, 704
 fractions(), 133, 213
 function(), 133, 136, 168, 184, 213, 233, 236, 257, 735
 FUSION1, 722

 gamma(), 213
 gather(), 724
 gaussian(), 552
 geyser, 14
 gf_abline(), 177, 178
 gf_boxplot(), 21, 27, 29
 gf_dens(), 172
 gf_density(), 172
 gf_dhistogram(), 10, 235
 gf_dist(), 59, 63
 gf_facet_grid(), 7, 740
 gf_facet_wrap(), 7, 740
 gf_histogram(), 10, 11, 14, 27, 710
 gf_jitter(), 642, 715
 gf_line(), 58, 73, 713
 gf_linerange(), 747
 gf_lm(), 466, 476, 509, 579
 gf_point(), 7, 27, 58, 476

- gf_qq(), 175, 177–179, 213, 248
gf_qqline(), 177
gf_refine(), 21, 743
gf_ribbon(), 491
gf_segment(), 58
gf_summary(), 642
GGally, 652, 744
ggformula, 6, 7
ggpairs(), 652, 744
ggplot2, 2, 6, 7, 700
ggvis, 698
glht(), 623, 626, 631, 659
glimpse(), 2
glm(), 524, 540, 589, 626, 662
golfballs, 397
googlesheets, 700, 708
grid, 25, 700
group_by(), 15, 721

hatvalues(), 498
haven, 700, 708
head(), 3, 27, 31, 702
help(), 703
hessian(), 377
HPDI(), 430
HPDinterval(), 430

I(), 460, 569, 589
Ice, 657
ilogit(), 443
inner_join(), 722
install.packages(), 698
install_github(), 698
integrate(), 133, 136, 213
iris, 3, 12, 271, 296, 340
is.array(), 730
is.character(), 730
is.integer(), 730
is.logical(), 730
is.matrix(), 730
is.numeric(), 730

kde(), 168
knitr, 697

lapply(), 737
lattice, 2, 6, 7, 700
left_join(), 722
length(), 23, 725
letters[], 729
library(), 698
list(), 727
lm(), 459, 466, 476, 555, 558, 589, 626
lmPerm, 670
load(), 712
log(), 128, 731
logit(), 443, 702
lubridate, 700

mad(), 30
magrittr, 716
makeFun(), 490, 525
mapply(), 367, 437, 737
MASS, 2, 25, 343, 700
match.arg(), 336
matches(), 716
MathNoise, 641
matrix(), 728, 775
max(), 747
maxLik, 357, 443
maxLik(), 357, 358, 362, 364, 376
mcp(), 623, 626
mean(), 15, 17, 27, 179, 235, 731
median(), 15, 27, 731
merge(), 443, 450, 661, 722
MTAA05, 332, 341
MLB2004, 526
mle2(), 357
mode(), 725
model.matrix(), 474, 607
mosaic, 2
mosaic(), 25, 26, 443
mosaicData, 2
mosaicplot(), 25
msummary(), 466, 485, 489, 528, 571
multcomp, 2, 623, 626, 628, 659
mutate(), 713

na.omit(), 732
names(), 718, 726, 727, 729
ncol(), 443, 729
nlm(), 357
nlmax(), 357, 430
nrow(), 443, 729
ntiles(), 12, 599
number_range(), 716

offset(), 589
optim(), 357
optimx(), 357
options(), 358, 443
order(), 732
orings, 523
outer(), 777

Pallets, 646

- paste(), 73, 732, 736
 pbeta(), 213
 pbinom(), 62, 63, 72, 111
 pc(), 654
 pchisq(), 330
 Pendulum, 546
 pexp(), 143, 146, 213
 pf(), 330
 pgamma(), 213
 Pheno, 597, 722
 phyper(), 99, 100, 111
 plogis(), 551
 plot(), 376, 500, 502, 577, 623, 644
 pmax(), 732
 pmin(), 732
 pnbinom(), 64, 111
 pnorm(), 159, 160, 213
 poison, 642
 poly(), 460, 589, 594
 polym(), 594
 power.t.test(), 519, 520, 549
 ppoints(), 175
 ppois(), 111
 predict(), 490, 525, 560
 prod(), 111, 732
 project(), 560, 765
 prop.test(), 251, 252, 281, 294, 330, 343
 prop1(), 291, 669
 pt(), 330
 punif(), 213
 puntling, 555
 pval(), 111
 qweibull(), 213

 qbeta(), 213
 qbinom(), 72, 73, 111
 qchisq(), 330
 qdata(), 20
 qexp(), 179, 213
 qf(), 330
 qgamma(), 213
 qhyper(), 111
 qlogis(), 551
 qnbinom(), 111
 qnorm(), 174, 213
 qpois(), 111
 qr(), 567, 685
 qt(), 330
 qtukekey(), 622
 quantile(), 19, 27, 29
 qunif(), 213

 qweibull(), 178, 213

 R2cuba, 184
 rank(), 732
 rapply(), 737
 RatPoison, 665
 rbeta(), 213, 233, 236
 rbind(), 71, 100, 176, 409, 410, 443, 731
 rbinom(), 62, 111
 rchisq(), 330
 read.csv(), 707, 708
 read.file(), 707
 read.table(), 25, 707
 read_csv(), 708
 readr, 700, 707, 708
 readRDS(), 712
 readxl, 700, 708
 rename(), 718
 rep(), 73, 177, 710
 replicate(), 235, 236, 238, 248, 293, 330, 535
 require(), 27, 698
 resample(), 238, 711, *see also* sample()
 resid(), 577
 rethinking, 430
 return(), 737
 rev(), 732
 rexp(), 213
 rf(), 330
 rgamma(), 213
 rhyper(), 111
 rlogis(), 551
 rmultinom(), 105, 289, 290
 rnbinom(), 111
 rnorm(), 213, 534, 710
 round(), 731
 row.names(), 718, 729
 rownames(), 443
 rpois(), 111
 rsquared, 489
 rstandard(), 498
 rt(), 330
 RubberBand, 543
 runif(), 213
 rvest, 700
 rweibull(), 213

 sample(), 238, 330, 710, 711, *see also* resample()
 sample_frac(), 711

sample_n(), 711
sapply(), 213, 737
save(), 712
saveRDS(), 712
scale_color_gradientn(), 747
scale_color_manual(), 743
scale_fill_brewer(), 743
scan(), 708
Scent, 344
sd(), 23, 27, 235, 731
seatpos, 651
select(), 715, 719
seq(), 73, 174, 710
set.seed(), 333
setdiff(), 723
setNames(), 726
shuffle(), 293
signif(), 731
snippet(), 4
soap, 511
solve(), 474, 566, 778
sort(), 732
source(), 705
Spheres, 548
splom(), 652
spread(), 724
sqrt(), 232
star, 502
starts_with(), 716
statTally(), 289–291, 377
stdEr(), 377
stem(), 16, 27
step(), 577, 649, 665
str(), 5, 27, 706
structable(), 25, 26
structure(), 336
substitute(), 336
sum(), 23, 27, 76, 111, 732
summarise(), 720
summary(), 15, 27, 443, 466, 489, 528,
570, 571, 605, 626, 636, 638

t(), 236, 474, 565, 566, 776
t.test(), 271, 303, 330, 337
table(), 732
tail(), 3, 27, 31
tally(), 9, 10, 25, 27, 62, 293, 294
tapply(), 737
TasteTest, 514, 632, 636
tibble, 713
tidyr, 700, 713, 724
tidyverse, 700

tolower(), 734
toupper(), 734
treadwear, 549
Trebuchet2, 476, 485, 488, 489
trunc(), 128
TukeyHSD(), 623

unique(), 732
uniroot(), 233, 236, 330, 443
usage(), 704
Utilities, 548, 589

vanBort, 238
var(), 23, 27, 731
vcd, 25, 26, 238
vcov(), 377, 486, 494
vif(), 652
vlength(), 763, 764
VonBort, 238

Wald test, 375
which(), 732
which.max(), 747
with(), 3, 297, 560, 611
write.csv(), 712
write.table(), 712

xchisq.test(), 443

z.test(), 336
zscore(), 480

Index

Page numbers for references to exercises are underlined. Page numbers for references to *examples* are italicized. Page numbers for references to **definitions** are in bold. In *see also* entries, items referred to from Index to R Functions, Packages, and Data Sets are in **typewriter type**.

- 1.5 IQR rule, **21**
- 68-95-99.7 rule, 160

- A^c , 45, **751**
- $A \cup B$, **751**
- $A \times B$, **752**
- $A \cap B$, **751**
- F_X , 59, 134
- f_X , 57, 59, 134
- $f_{X|Y}(x | y)$, 185
- $\binom{n}{k}$, **47**
- \mathbb{R} , 750
- \mathbf{x} , 183, 185
- $\llbracket x \in A \rrbracket$, 753
- \mathbb{Z} , 750
- z_α , **259**
- z_* , **259**
- $\Gamma(\alpha)$, 162
- $\hat{\theta}$, 228

- acceptance sampling, 117
- ACT, 547, *595*
- added variable plot, *see also* partial regression plot, 686
- additive model, 581
- adjusted r^2 , 575, *see also* correlation coefficient
- age, *459*
- AIC, *see* Akaike's information criterion
- Akaike's information criterion, 576, 649

- algebraic argument, 48
- allele, 56, 662
- analysis of variance, *see* ANOVA
- ANOVA, 465, 513, 547, 599
 - for regression, 487
 - one-way, 600
 - repeated measures, 657
 - two-way, 601, 634, 645
- Anscombe, F., 501, 546
- anti-conservative, 253, 402
- argument of an R function, 9, 735

- bandwidth, 171
- baseball, 30, *357*
- basis
 - for a vector space, 761, **761**
- Bayes's Theorem, 117, 299, 424
- Bayesian, **424**
- Benford's Law, 102, 128
- Benford, Frank, 102
- Bernoulli distribution, **81**, 82, 87, 89, 96, 112
- beta distribution, **165**, *357*
- $\text{Beta}(\alpha, \beta)$, *see* beta distribution
- BFGS algorithm, 366
- BHHH algorithm, 366
- bias
 - bootstrap estimate, 307
- biased estimator, 262
- bijection, 40, **41**

- bin, 13
Binom(n, π), *see* binomial distribution
 binomial coefficient, **47**
 binomial distribution, **60, 61**, 97, 112, 154, 177, 249, 350
 binomial test, 68, 281, *see also* `binom.test()`
 Bioconductor, 698
 biology, 353, 354
 blocking, 646, 646, 693
 Bonferroni correction, 621
 Bonferroni's Inequality, 114
 bootstrap, 301, 324
 bootstrap distribution, 301
 bootstrap sample, 301
 bootstrap standard error, 301
 Box, G., 455
 boxplot, **21**, *see also* `gf.boxplot()`
 Bradley-Terry model, 541
 Bristol, Muriel, 98
 bulge rule, 506
- cardinality, 40, **752**
 cards, 44, 48, 53, 113, 117, 118, 123
 carrier, 553
 Cartesian product, **43, 752**
 case-control study, 413, 414
 cat, 121
 categorical variable, 8
 Cauchy distribution, **215**, 215, 273
 cdf, 108, *see* cumulative distribution function
 cdf method, **138**, 140, 142, 158, 353
 Central Limit Theorem, 245, 249, 277
 chaining syntax, 716
 characteristic function, 247, **753**
 Chebyshev's Inequality, 242
 Chevalier de Mere, 119
 chi-squared distribution, 267, **267**, 330, 376, 394, 402
 chi-squared statistic
 likelihood ratio (G), 392, 394
 Pearson (X^2), 394, 395, 397
Chisq(ν), *see* chi-squared distribution
 cholesterol, 459, 628
 clinical trial, 411
 Clopper-Pearson interval, 281
 cloud plot, 577, *see also* `cCloud()`
 coagulation, 599, 602
 codomain, **752**
 coefficient of determination, 489, 509, 544, 572, 575, *see also* correlation coefficient
 coin toss, 36, 45, 57, 66, 75, 117, 118
 collinearity, 597, 651
 column matrix, 772
 combinatorial argument, 48
 comma separated values, 708
 comment character in R (`#`), 705
 complement, 45, **751**
 complement rule, **45**
 Comprehensive R Archive Network (CRAN), 697
 concrete, 558, 569, 668
 conditional distribution, **83**, 185
 conditional probability, **51**
 confidence ellipse, *see* confidence region
 confidence interval, 255, **260**, 301, *see also* confidence region
 for mean, 260
 for new response, 489
 for parameters of a linear model, 485, 485, 486
 for proportion, 279
 likelihood, 529
 Wald, 279, 528, 572
 Wilson, 281
 confidence level, 259
 confidence region, **442**
 conjugate gradient optimization, 366
 conjugate prior, 436
 conservative, 98, 253, 403, 450, 621
 consistent estimator, **242**
 continuity correction, 250, 251, 251, 253
 continuous random variable, **132**
 contrast, **610**, 618, 692
 orthogonal, 617
 sample, 616
 convergence in distribution, **245**
 Cook's distance, 501
 Cook, R. D., 501
 correlation coefficient, **90**, 195, 470, 552, 572, 590
 counting, 40, 108
 covariance, **89**, 90
 covariance matrix, *see* variance-covariance matrix
 coverage rate, 256
 CRAN, *see* Comprehensive R Archive Network
 credible interval, **430**, 431

- critical values, 72
cross-moment, 192
CSV, *see* comma separated values
cumulative distribution function, **59**,
134
- data, xvii
data cleaning, 279
data frame, 1, 2
De Morgan's Laws, **114**, **114**
decile, **20**
degrees of freedom, 267, 269, 763
delta method, 317
density histogram, 10
dependent variable, 456
deviance, 529, 663
DFBETA, 502
DFBETAS, 502
DFFITS, 502
diabetes, *see* FUSION
Diaconis, Persi, 104
diagnostics, **494**
dice, *53*, *54*, *70*, **117**, 119, *347*
digit
 leading, *see* Benford's Law
 $\dim(\Omega)$, *see* dimension of a model
dimension
 of a model, 568
 of a vector space, **762**
disease, **117**
distribution, 5
division principle, **46**
domain, **752**
donut, **113**
dot product, 265, **763**, *see also* `dot()`
double exponential distribution, *see*
 Laplace distribution
Dunnet contrasts, 631, 681
- effect plot, 582
empirical distribution, 301
empirical rule, 160
equal in distribution, **135**, 154, **215**
equally likely rule, **39**
estimand, **234**
estimate, **234**
 interval, 234
 point, 234
estimator, 234, **234**
 consistent, *see* consistent estimator
 unbiased, *see* unbiased estimator
event, **34**, 107, 751
- exact test, *see* permutation test
expected value, **75**, 109
explanatory variable, 456
exponential distribution, **141**, 142, *153*,
178, *333*, *404*
exponential family, **180**
exponential generating function, *see*
 generating function
extended real number, 149
- $F(m, n)$, *see* F -distribution
 F -distribution, **269**, 330
 F -statistic, 488, **543**
facet, 7, **11**, 740
factor, 9, 599
factorial, 754
fair coin, *see* coin toss
false positive, **117**
family of distributions, 109, 227
family-wise error, 620
family-wise error rate, 620, 631
first digit, *see* Benford's Law
Fisher's exact test, 98, 413, *see also*
 `fisher.test()`
Fisher, R. A., xviii, xx, 292, **449**, 553
fitting a model, 327
five-number summary, 20
flush, *44*, *48*
Foo Foo, 716
football, *284*, 555
formula
 for a linear model, 459, 589
four-step method, 66, 370
free parameter, 387, 393
frequentist, **424**
frequentist approach, 433
Frisch-Waugh-Lovell Theorem, **686**
full house, **113**
function, **752**
functions in R, *see* `function()`, 735
FUSION, *597*, *661*
- G -statistic, *see* chi-squared statistic
Galton, Francis, 478
game, *70*
gamma distribution, 163, **163**, *333*
gamma function, **162**
 $\text{Gamma}(\lambda, \alpha)$, *see* gamma distribution
Gaussian distribution, *see* normal
 distribution
generalized linear model, 522, 523, 540
generating function, **150**

- genetics, [56](#), [354](#), [449](#)
Geom(π), *see* geometric distribution
geometric distribution, 65, **65**, 79, 87, 89, 97, 112, [121](#)
geometric series, 150, **754**
ghost, 70
GitHub, 698
golf, [288](#), [392](#), [395](#), [397](#)
goodness of fit testing, 167, 395
Gosset, W. S., 269, 515, 549
grade point average, [461](#), 547, [595](#)
gradient, 366
Gram-Schmidt orthonormalization, 771
greedy algorithm, 656
grid search, 367
group, 599
- Hardy-Weinberg equilibrium, [354](#)
hat matrix, 475, 498
hat notation, 228
hat values, 498, 500
heavy tails, [273](#)
height, 459
helium, [284](#)
Hessian, 366
heteroskedasticity, 506
highest posterior density interval, 430
Hill, Theodore P., 104
hinge, [29](#)
histogram, **10**, *see also* `gf_histogram()`
homoskedasticity, 506
HPDI, *see* highest posterior density interval
Hyper(m, n, k), *see* hypergeometric distribution
hypergeometric distribution, 95, **95**, [96](#), 97, 112
hypothesis, 66
hypothesis test, 66, 110, 254
- idempotent, 544, 781
identity matrix, 772
iid random sample, 238
improper prior, 427
inclusion-exclusion, [114](#)
independence, **54**, **186**
 conditional, [125](#)
 test for, 295
independent random variables, **84**, 186
independent variable, 456
indicator function, **753**
indicator variable, 459
- inference, 253
information, 372
 observed, 372
injection, 41
inner product, *see* dot product
interaction, 584, 587, 638
interaction plot, 641
intercept, 458
 fitting a model without, [543](#)
interquartile range, **20**
intersection, **751**
inverse
 of a matrix, **773**
IQR, *see* interquartile range
Iverson bracket, **753**
- joint cdf, 185
joint distribution, **82**, 183
- kernel, **141**, 156, 161, 164, 427
kernel density estimation, 167
Knuth, Donald, 103
kurtosis, 148
- ladder of re-expression, 506
lady tasting tea, *xviii*, 68, 98, 682
Lagrange multipliers, [356](#), [445](#)
Laplace distribution, **447**, [447](#)
Law of Large Numbers, 243, [243](#), [244](#)
leading digit, *see* Benford's Law
least squares estimation, 288, 464, 465, 468
leave-one-out, **501**
leptokurtic, 149
level, 9, 599
leverage, **500**
lifetime models, 142, 144
likelihood
 normalized, 372
likelihood function, 347, 426
likelihood ratio test, 369, 663
line search, 366
linear algebra, 264, 759
linear model, 456, *see also* `lm()`
linearly dependent, **761**
linearly independent, **761**
link function, 523, 540
linkage, [449](#)
location-scale family, 210
log odds, 522
logistic distribution, [551](#)
logistic regression, 522, 540, 661

- logit, **523**
 lurking variable, 25, 496
- margin of error, **256**, 340, 621
 marginal distribution, **83**, 185
 Mastermind, 117
 matrix, 771
 maximum, 59, 118
 maximum a posteriori estimate, 430
 maximum likelihood estimation, 288
 maximum likelihood estimator, 347, 470
 MCMC algorithms, 437
 mean, **14**, 30
 of a random variable, *see* expected value
 mean absolute deviation, **22**, 30
 mean integrated squared error, **172**, 219
 mean squared error, 484
 median, **14**, 30
 medical testing, 55
 memoryless, 121, **144**, 165, 180
 Mendelian genetics, 54, 56
 method of moments, 227, 327
 mgf, *see* moment generating function
 MISE, *see* mean integrated squared error
 missing data, 279
 mixture distribution, 362, **446**, 446
 MLE, *see* maximum likelihood estimator
 model, 225
 model comparison test, 530, 567, 610, 639
 model formula, 227, 589
 model matrix, 458
 model space, **464**
 model utility test, 484, 530, 572, 602
 moment
 of a distribution, 147
 sample, 230
 moment generating function, **152**, 153
 monotone function, 351
 Monte Carlo error, 292
 mosaic plot, **25**, 408
 MSE, *see* mean squared error
 multicollinearity, 597
 multinomial coefficient, **104**
 multinomial distribution, **104**, 129, 289, 391
 multiple comparisons, 620, 692
 multiple regression, 465, 553
 multiplication principle, 43
 multiplication rule, 43
 multivariate normal distribution, **199**
- natural parameter, **180**
 $\text{NBinom}(s, \pi)$, *see* negative binomial distribution
 negative binomial distribution, **64**, 65, 79, 87, 89, 112
 Nelder, J. A., 523
 Nelder-Mead algorithm, 365
 Newcomb, Simon, 102
 Newton's Law of Cooling, 676, 678, 694
 Newton-Raphson algorithm, 366
 non-linear least squares, 672
 non-negative definite matrix, **199**, 223
 $\text{Norm}(\mu, \sigma)$, *see* normal distribution
 normal distribution, 149, **156**, **158**, 231, *see* multivariate normal distribution
 bivariate, **195**
 multivariate, **195**
 normal equation, 474
 normal-quantile plot, **173**
 nuisance parameter, 98
 null hypothesis, **66**
 null hypothesis significance test, *see* hypothesis test
- observational unit, 1
 odds, 101, 522
 odds ratio, **101**, **102**, 529, 550, 662
 Old Faithful, 362
 omnibus test, 602
 one-to-one, **41**
 onto, **41**
 optimization, 365
 order effect, 419
 order statistic, 222
 ordinary generating function, *see* generating function
 orthogonal design, 560
 orthogonal vectors, 265, **764**
 orthonormal basis, **770**
 outcome, **34**, 107
 outlier, 17, **21**, 278, 500
- p-value, **67**, 110
 empirical, 288
 paired *t*-test, 283, 343, 549, 552
 pairwise comparisons, 619
 parameter, 60, 109, 253
 partial regression plot, 579

- partitioned regression, [686](#)
- pdf, *see* probability density function
- Pearson chi-square statistic, 395
- Pearson, Karl, 395
- penetrance, [56](#), [115](#)
- percentile, 18
- permutation test, 668
- physics, [507](#), [510](#), [546–548](#)
- pivot, *see* pivotal quantity
- pivotal quantity, 261, [261](#), 269, 280, 381
- plant density, [353](#)
- platykurtic, 149
- pmf, 108, *see* probability mass function
- Poisson distribution, 93, **93**, 142, 150, [217](#)
- pollution, [608](#), [622](#), [626](#)
- polynomial regression, 594
- population, xix, 253
- positive definite matrix, **199**
- positive semi-definite matrix, *see* non-negative definite matrix
- posterior, **425**
- power, [72](#), **73**, [74](#), [345](#), 393, 519
- power series, 149
- precision, 434
- prediction interval, 489
- principle components, 654
- prior, **424**, **425**
- uninformative, 426
- probability, 33, 107
- axioms, 37, 57, 107
 - empirical, 34, 107
 - theoretical, 34, 37, 107
- probability density function, **132**
- probability mass function, **57**
- probability table, 40
- probit, [551](#)
- propagation of error, *see* propagation of uncertainty
- propagation of uncertainty, 317, 320
- proportion, [348](#), [350](#)
- pseudo-random number, *see* random number generation
- Punnett square, [449](#)
- Pythagorean decomposition, 465, **609**
- Pythagorean Identity, 264
- QR-decomposition, 564
- qualitative variable, *see* categorical variable
- quantile, **18**, 178
- theoretical, 174
- quantile-quantile plot, **173**
- quantitative variable, 8
- quartile, **20**
- quasi-Newton optimization algorithms, 366
- R Markdown, 697
- r^2 , *see* correlation coefficient
- radius of convergence, 149
- Raimi, Ralph, 104
- random number generation, 137
- random variable, **34**, **38**, 108
- continuous, **38**
 - discrete, **38**
 - distribution of, **57**
- random vector, *see* joint distribution
- randomization test, *see* permutation test
- randomized response, [334](#)
- randomness, xviii, 33
- range, **752**
- rdf, *see* residual degrees of freedom
- recentering, 607
- rectangular data, 1
- recycling, 733
- regression, 456, 553
- regression to mediocrity, 479
- rejection region, 72
- relative risk, 101, 529, [550](#)
- relative uncertainty, 325
- repeated measures ANOVA, 657
- replication of statistical results, 300
- resample, *see* bootstrap sample
- resampling, 301
- resampling distribution, *see* bootstrap distribution
- researcher degrees of freedom, 495
- residual, **462**, 465, *see also* `resid()`
- Pearson, 407–409
- residual degrees of freedom, **568**
- residual plot, 495
- residuals
- checking normality, 500
 - externally studentized, **499**
 - internally studentized, **498**
 - partial, 583
 - semistudentized, **497**
 - standardized, **498**
 - variance of, 497
- resistant, 17
- response variable, 456
- Riemann sums, 132

- risk, **101**
 robust, 272, 273, 328
 row matrix, 772
RSS, 464, **568**
 RStudio, 697
 rule of thumb, 250
 for *t*-procedures, 277
 ruler method, **18**, 174, 359
 Russel, B., 347
- sample, xix, 253
 sample space, **34**, 751
 sampling distribution, 67, **234**
 sampling with replacement, 97
 sampling without replacement, 97
 SAT, 461, 595
 Satterthwaite, F. E., 517
 saturated model, 663
 scale parameter, 163, 164
 scaling constant, 141, 164
 scatterplot, 7
 scatterplot matrix, *see* `ggpairs()`, *see*
 `splom()`
 score function, 372
 score interval, 280
 sensitivity, 55
 series, **754**
 set, **750**
 shape parameter, 163
 sign test, 284, 285, 297, 343
 significance level, 110
 simple hypothesis, **398**
 simple linear regression, 465
 simple random sample, **238**, 335
 Simpson's paradox, **25**
 simulated annealing, 365
 size
 of a set, **752**
 skewness, **13**, 148
 smoking, 405, 408, 410, 671
 soap, 511
 space shuttle, 523
 span, **761**
 specificity, 55
 spring constant, 458
 SRS, *see* simple random sample
SSE, 464, 488, 568, 605
SSM, 488, 568, 605
SST, 488, 568
 St. Petersburg Paradox, 125
 standard deviation, **22**, **80**
 standard error, **247**, 268, 484, 547
 standardized score, 159, 480
 stars, 502
 statistic, 67, 253
 stepwise regression, 649, *see also* `step()`
 straight
 large, 54
 Student, *see* Gosset
 studentized range distribution, 622
 subjective approach, 433
 subset, **750**
 subspace, **762**
 sufficient statistic, 351
 sum of squares, 30
 supernormality of residuals, 500
 superset, **750**
 support
 of a random variable, **38**, **132**, 136,
 371, **371**
 supremum, 59, 118
 surjection, **41**
 systematic component, 540
- $t(\nu)$, *see* *t*-distribution
t-distribution, **269**
t-statistic, 543
 Taylor polynomial, 150
 Taylor series, 150, 396
 Taylor's Theorem, 246
 tea, *see* lady tasting tea
 terminal velocity, 548
 test statistic, 67, 110
 tidyverse, 700
 time series plot, **31**
 time until failure, *see* lifetime models
 tolerance, 453
 total deviation from the mean, 30
 transformation
 of a random variable, **76**
 of data, 505, 545
 transpose
 of a matrix, 565, **774**
 treatment, 599
 trebuchet, 476, 485, *see also*
 Trebuchet2
 trimmed mean, **17**
 Tukey bulge, 507, 546
 Tukey contrasts, 631, 660
 Tukey's honest significant differences,
 621, 681
 Tukey, J., 506
 twins, 292, 413
 2-sample *t*-test, 517

- two-sided alternative, 69, 255, 291, 293, 296
- type I error, 72, 110
- type II error, 72, 110
- unbalanced design, 624
- unbiased estimator, **241**, 471, 482
- uncertainty, **321**
- Unif(a, b), *see* uniform distribution
- uniform distribution, 124, 136, **136**, 136, 228, 229, 232, 351, 355
- discrete, 217
- unimodal, **14**
- union, **751**
- universal set, **751**
- upper triangular matrix, 564
- urn model, 95, 97
- variance, **22**, **80**, 231
- of a random variable, 109
- unbiased estimate for, 262
- variance-covariance matrix, **196**
- vector, 183
- random, *see* joint distribution
- vector decomposition, 766
- Venn diagram, 751
- wear, 180
- Wedderburn, R. W. M., 523
- Weibull distribution, **164**, 178, 353
- weighted least squares, 496
- weighted mean, 333
- weighted sum, 333
- Wickham, Hadley, 700, 716
- wine
- gentleman tasting, 118
- Yahtzee, 54
- Yates's continuity correction, 412
- z -score, *see* `zscore()`, *see* standardized score
- z -test, 253, 336