

# Subject Index

- $\Delta P$  226, 232, 234, *see also*  
association measures
- $\phi$ -coefficient 209–210, 217
- $\chi^2$ -distance 370–371
- $\chi^2$ -statistic 212–213, 218, 221, 234
- $\chi^2$ -test 29, 210, 212–214, 217
- 3D plot 340, 371–372
- A**
- accuracy 284, 287, 298–299
- additive effect 144, *see also*  
interaction
- add-on package 22–23
- adjusted  $R^2$  *see*  $R^2$
- aesthetics 53, 424
- agglomerative clustering *see*  
cluster analysis
- AIC *see* Akaike Information  
Criterion
- Akaike Information  
Criterion 149
- alternative hypothesis *see*  
hypothesis
- American English 241–248
- Analysis of Variance *see*  
ANOVA
- ANOVA 163, 171–172, 175,  
267–268, 286, *see also*  
post hoc test
- between-group design 172
- between-group variable 172
- factorial ANOVA
- independent factorial  
ANOVA 172, 182–189
- mixed factorial  
ANOVA 172, 191–196
- one-way ANOVA
- independent one-way  
ANOVA 172, 176–179
- Kruskal-Wallis one-way  
ANOVA by ranks 178
- non-parametric one-way  
ANOVA 178–179
- parametric one-way  
ANOVA 176–178
- repeated-measures one-  
way ANOVA 172, 198
- two-way ANOVA *see* factorial  
ANOVA
- area under the ROC-curve  
*see* C-index
- arithmetic operation 24–26
- association measures 200,  
223–239
- bidirectional (symmetric)  
association  
measures 224–225
- contingency-based  
association  
measures 225–226
- non-contingency-  
based association  
measures 225–226
- (uni)directional (asymmetric)  
association  
measures 224–225
- association plot 220
- Attraction 224–225, 228–231,  
234, 236, 238, *see*  
*also* association  
measures
- autocorrelation 126, 128, 161
- average *see* mean
- average method of  
agglomerative clustering  
*see* cluster analysis
- average silhouette width  
311–312, 319–320, 330, *see*  
*also* cluster analysis
- B**
- backward selection *see* variable  
selection
- bag-of-words model 323–324,  
*see also* Semantic Vector  
Space
- bar plot 74–76, 203–207,  
414–417, 423
- bar plot with 95% confidence  
intervals 100–104
- Basic Colour Terms 80–81
- Bayesian Information  
Criterion 194
- be going to* 387–393
- Behavioural Profiles 301,  
303–306, 384
- Belgian Dutch 7–8, 242, 256,  
260–262, 269
- Best Linear Unbiased  
Predictors 193
- best subsets method 149,  
151, *see also* variable  
selection
- between-group design *see*  
ANOVA
- between-group variable *see*  
ANOVA
- BIC *see* Bayesian Information  
Criterion
- bidirectional selection *see*  
variable selection
- bimodal distribution 45
- binary recursive  
partitioning 291, *see also*  
conditional inference  
tree
- binary variable *see* variable
- binomial logistic regression 253,  
*see also* regression
- bivariate normal  
distribution 126–127
- BLUP *see* Best Linear Unbiased  
Predictors
- BNC *see* British National  
Corpus
- Bonferroni correction 181,  
*see also* post hoc test
- Boolean values 71
- bootstrap 166–169
- multiscale bootstrap 315–316
- Bootstrap Probability 316
- bootstrap validation 166–167,  
274, 315–316
- box plot *see* box-and-whisker  
plot
- box-and-whisker plot 57–58,  
91–93
- Box-Cox plot 158
- British English 243–248
- British National Corpus 352
- Brown corpus 90
- C**
- C-index 259, 299
- Canadian English 248–251
- Canberra distance *see* distance  
metric

- categorical variable *see* variable  
causality 15–16  
Causee 255  
Causer 255  
character vector 30, 399–400  
chi-squared test *see*  $\chi^2$ -test  
CI *see* confidence interval  
C-index 273, 299  
city-block distance *see* distance  
metric  
classical MDS *see*  
Multidimensional  
Scaling  
Cleveland dot chart 78–79, 418  
cluster centroid 317, *see also*  
*k*-means  
cluster dendrogram 309–310  
cluster analysis  
agglomerative cluster  
analysis 310  
agglomerative clustering  
methods 310–311  
average clustering  
method 310–311  
complete clustering  
method 310–311  
single clustering  
method 310–311  
Ward clustering  
method 309, 311  
divisive clustering 310, 321  
fuzzy clustering 321  
hierarchical cluster  
analysis 309  
*k*-means 317–319  
*k*-medoids 317, 319–320  
partitioning around  
medoids 317, 319–320  
COCA *see* Corpus of  
Contemporary  
American English  
COHA *see* Corpus of Historical  
American English  
Cohen  $\kappa$  201  
collostructional analysis 239,  
241, *see also* distinctive  
collexeme analysis  
collostructional  
strength 225, 238  
comma-separated values file 34  
complete method of  
agglomerative clustering  
*see* cluster analysis  
complete separation 273  
component-residual  
plot 156, 158  
conceptual metaphor 199  
concordance index *see* C-index  
conditional inference tree  
291–292, 294–297  
confidence ellipses 359–361,  
380–381  
confidence interval 98–100, 103  
confidence region 118, 159  
context window 324  
contingency information  
225–226  
contingency table 199–200,  
202–203  
continuity correction 110, 212  
continuous variable *see* variable  
contrasts 146, 185–186  
Helmert contrasts 146  
orthogonal contrasts  
185–186  
planned contrasts 196  
sum contrasts 146, 186  
treatment contrasts 146–147,  
185, 254, 259  
convenience sampling 8  
Cook distance 154, 270  
Corpus of Contemporary  
American  
English 81–82, 281  
Corpus of Global Web-based  
English 243, 248  
Corpus of Historical American  
English 387–388  
correction for continuity 110  
correlation 115–120, 122–125  
correlation analysis vs.  
*t*-test 121–122  
correlation coefficient 116,  
129–130  
Kendall  $\tau$  116, 132–133  
Pearson  $r$  116, 119, 122,  
124–126, 142  
Spearman  $\rho$  116, 132–133  
correlation matrix 134–135  
correlation test 116  
Mantel test for distance  
matrices 348–349  
non-parametric correlation  
test 132–134  
parametric correlation  
test 126–129, 134  
negative (inverse)  
correlation 116  
positive correlation 115–116  
correlogram 135–137, 236–237  
Correspondence Analysis (CA)  
CA with supplementary  
variables 378–381  
interpretation of CA  
symmetric maps 371  
Multiple Correspondence  
Analysis 375–384  
Simple Correspondence  
Analysis 369–372  
cosine similarity 308, 328–330  
Cramér  $V$  209–210, 217, 221–222  
CRAN 21–22  
critical value 10, 12  
Cronbach  $\alpha$  364–365  
cross-over interaction 164,  
*see also* interaction  
cross-validation 166  
D  
data frame 31–32, 404–408  
exporting a data frame 38  
‘long’ format data frames 279  
merging data frames 244  
‘short’ format data frames 279  
sorting a data frame 94  
splitting a data frame 305  
subsetting a data frame 32,  
84, 107–108, 405–407  
DCA *see* distinctive collexeme  
analysis  
degenerate MDS *see*  
Multidimensional  
Scaling  
degree modifier 243  
degrees of freedom 12, 213  
delta  $P$  *see*  $\Delta P$   
dendrogram 309–310  
density plot 51–54, 423, *see also*  
probability density  
dependent variable *see* variable  
descriptive statistics 7–8  
deviance 258  
Deviation of Proportions 82–85  
Dice coefficient 235, *see also*  
association measures  
directional alternative  
hypothesis *see*  
hypothesis  
discrete variable *see* variable

- distance matrix 306  
 Mantel test for distance matrices 348
- distance metric 306–308  
 Canberra distance 307–308  
 Euclidean  
   distance 306–308, 344  
 Gower distance 308, 343–344  
 Levenshtein distance 308  
 Manhattan (city-block)  
   distance 307–308
- distinctive collexeme analysis (DCA) 241–251  
 simple DCA 244–248  
 multiple DCA 248–251
- distribution 10–11, *see also*  
 normal distribution
- distributional semantics 323–324
- ditransitive construction 226
- divisive clustering *see* cluster analysis
- dot chart *see* Cleveland dot chart
- double equal sign 26
- dummy variable *see* variable
- Durbin-Watson test 128
- Dutch causative auxiliaries 253–254
- D-W statistic *see* Durbin-Watson test
- E**
- effect size vs. statistical significance 103, 129–130
- effected predicate 256
- eigenvalue 338, 354–355, 381
- Electronic World Atlas of Varieties of English 333, 342–343
- English causative constructions 293–294, 302
- English verbs of cooking 323
- English Lexicon Project 41–42, 142
- English permissive constructions 277–278
- entrenchment 82
- equal sign 26
- error message 29
- estimated coefficient 145–146, *see also* regression
- Euclidean distance *see* distance metric
- eWAVE *see* Electronic World Atlas of Varieties of English
- Exemplar Theory 373
- expected frequency 210–211, 218, 228
- expected proportion *see* Deviation of Proportions
- explanatory variable *see* variable
- extreme adjectives 243, 247, 250
- F**
- F*-ratio 149, 178–179, *see also* ANOVA
- factor 30, 76–78, 400–402
- Factor Analysis 361–365  
 factor loadings 362  
 factor rotation 361, 363–366
- Firth penalized regression 273
- Fisher exact test 29, 200, 214, 232–234, 245
- fitted value 120–121, 140–141, 147, 153–154, 157, 264–266, 282
- Fligner-Killeen test 88, 177, 184
- forward selection *see* variable selection
- functions in R 27–29
- future marker 387
- fuzzy clustering *see* cluster analysis
- G**
- Generalized Linear Model 170, 192–193, 257
- geographic coordinates 333
- geometric mean 235, *see also* association measures
- geoms 423
- GLM *see* Generalized Linear Model
- GloWbE corpus *see* Corpus of Global Web-based English
- goodness of fit 142, 148, 151, 170, 194, 259, 274–276, 280, 283, 286, 289, 299–300, 317, 333, 337, 341
- Gower distance *see* distance metric
- grammatical development 130
- Graphical User Interface 22
- H**
- hapax legomena 64
- hat-value 153–154, *see also* influential observation
- Helmert coding *see* contrasts
- heterogeneous variance 88, 187
- heteroscedasticity 115, 127–128, 157–159, 176
- hierarchical cluster analysis *see* cluster analysis
- histogram 50–56, 64–67, 105, 423
- Holm method 181, *see also* post hoc test
- homogeneous variance 179–180
- homoscedasticity 126–128, 157, 176
- hypothesis  
 alternative hypothesis 8–9  
 directional alternative hypothesis 9, 13–14  
 non-directional alternative hypothesis 9, 13–14  
 null hypothesis 8–9
- I**
- image schema 200
- independent variable *see* variable
- inertia 382, 385, *see also* Correspondence Analysis
- inferential statistics 7–8
- inflation of surprise 175
- influential observation 153, 155, 270
- integer vector 90
- interaction 144, 162–166, 172, 183–184, 194–196, 268–270
- interaction plot 163–164, 183–184, 269
- intercept 140–141, 147, 254, 259, 264  
 random intercept 193
- interquartile range 49, 58
- inter-rater agreement 201
- interval variable *see* variable
- inverse correlation *see* correlation
- IQR *see* interquartile range

## J

Jaccard coefficient 235, *see also*  
association measures  
jitter 105, 424

## K

Kaiser criterion 355–356, *see also*  
Principal Components  
Analysis  
Kendall  $\tau$  *see* correlation  
*k*-means *see* cluster analysis  
*k*-medoids *see* cluster analysis  
Kruskal non-metric MDS  
*see* Multidimensional  
Scaling

## L

Latent Semantic  
Analysis 324, 332  
latent variable *see* variable  
Law of Abbreviation 91, 116, 129  
least squares method 121, *see*  
*also* regression  
legend 73  
Levenshtein distance *see*  
distance metric  
leverage 124, 154, 270  
Levene test 88, 176–177  
lexical field 324  
Liddell difference of  
proportions 235, *see also*  
association measures  
limit adjectives 243, 247–248  
line chart 63, 412–414  
linear regression *see* regression  
linguistic relativity 171  
local MI 234, *see also*  
association measures  
LOESS 136  
log odds 253–254, 259–265, 276,  
284–285  
log odds ratio 235, 260–263  
logical vector 400  
logistic regression *see*  
regression  
logit *see* log odds  
log-likelihood ratio 233,  
*see also* association  
measures  
log-linear analysis 170, 222  
log-linear regression *see*  
regression  
long format *see* data frame

## M

MAD *see* median absolute  
deviation  
MAD-scores 57, 59–61  
Manhattan distance *see* distance  
metric  
Mantel test for distance  
matrices 348  
marginal frequencies 200, 211,  
221  
masked function 159–160  
Matt-Whitney test 87  
maximum 46, 48  
maximum likelihood 193, 254,  
*see also* restricted  
maximum likelihood  
MCA *see* Correspondence  
Analysis  
McFadden  $R^2$  *see*  $R^2$   
MDS *see* Multidimensional  
Scaling  
mean 43–44, 53, *see also*  
measures of central  
tendency  
measures of association *see*  
association measures  
measures of central  
tendency 43–46, 50  
measures of  
dispersion 48–49, 50  
median 43–46, 53  
median absolute  
deviation 49, 59  
medoid 317, 319, *see also* cluster  
analysis  
merge *see* data frame  
metric MDS *see*  
Multidimensional  
Scaling  
MI2 234, *see also* association  
measures  
MI3 234, *see also* association  
measures  
minimum sensitivity 235,  
*see also* association  
measures  
minimum 46, 48  
missing data 34, 46–47, 125–126,  
143, 343  
mixed Generalized Linear  
Model 192–196, 271, *see*  
*also* regression, ANOVA  
ML *see* maximum likelihood

mode 45–46, 56  
monotonic relationship 122–123,  
132–133  
Monte-Carlo simulation 296  
mosaic plot 219–220, 368–369  
motion chart 389–393  
multicollinearity 155, 159–161,  
263, 271–273, 353–354,  
365, 383  
multidimensional  
analysis of register  
variation 351–352  
Multidimensional Scaling  
(MDS) 336–342,  
345–348  
classical MDS 338–342  
degenerate MDS  
solution 347  
Kruskal non-metric  
MDS 345–347  
metric MDS 337  
non-metric MDS 337  
non-metric MDS with  
ties 348  
stress 337, 341, 347  
multifactorial grammar 253  
multinomial logistic regression  
*see* logistic regression  
multiple comparisons 175, *see*  
*also* post hoc test  
Multiple Correspondence  
Analysis *see*  
Correspondence  
Analysis  
multiple distinctive collexeme  
analysis *see* distinctive  
collexeme analysis  
multiple regression *see*  
regression  
multiscale bootstrap  
resampling 315–316  
  
N  
Nagelkerke  $R^2$  *see*  $R^2$   
negative correlation *see*  
correlation  
negatively skewed  
distribution 53, 61, 66  
neighbour-joining  
algorithm 321  
nested variable *see* variable  
Netherlandic Dutch 7–8, 242,  
256, 259–262, 269

- Nicaraguan Sign Language  
173–175, 176–180, 182,  
184, 189
- nominal variable *see* variable
- non-constant variance test 128,  
157–158
- non-directional alternative  
hypothesis *see*  
hypothesis
- non-metric MDS *see*  
Multidimensional  
Scaling
- non-parametric regression *see*  
regression
- non-parametric test 18
- normal distribution 10–11,  
53, 56
- normalized PMI 234, *see also*  
association measures
- NSL *see* Nicaraguan Sign  
Language
- null hypothesis *see* hypothesis
- numeric vector 29–30, 397–399
- O**
- oblique rotation 363
- observed frequency 210,  
212–213, 218–220, 232,  
245, 327
- Occam's razor 149
- odds 71–72, 208–210, 259–262,  
264–265, 285
- odds ratio 200, 208–210, 238,  
259–264, 281
- omnibus test 179, 258
- one-tailed test 13–14, 89, 95, 98,  
112, 128–129, 133, 148
- optimism 166–167, 274–275
- ordered factor 343, 401
- ordinal variable *see* variable
- orthogonal contrasts *see*  
contrasts
- orthogonal rotation 363
- outlier 44, 49, 57–62, 122,  
153–155, 270–271
- overfitting 143–144, 166–167,  
274–275
- P**
- paired *t*-test *see t*-test
- parameter 7–8
- parametric test 18
- parsimony 149, 337
- partial-residual plot *see*  
component-residual plot
- partitioning around medoids *see*  
cluster analysis
- paste 73, 400, 419
- PCA *see* Principal Components  
Analysis
- Pearson contingency  
coefficient 209
- Pearson product-moment  
correlation coefficient  
*see* correlation
- Pearson residuals 218–220
- Pearson *r* *see* correlation
- penalization in regression 167,  
275
- percentile 44–45
- permutation 67–68, 155, 292
- p*-hacking 130, 148
- pie chart 72–74, 418–420
- planned contrasts *see* contrasts
- PMI *see* Pointwise Mutual  
Information
- Pointwise Mutual  
Information 234,  
238, 326–328, *see also*  
association measures
- Poisson regression 170
- polytomous logistic regression  
*see* regression
- population 7–8
- positive correlation *see*  
correlation
- Positive Pointwise Mutual  
Information 326–328
- positively skewed  
distribution 53
- post hoc test 179–181, 188–189,  
196, *see also* ANOVA
- power transformation 66,  
158–159, 168
- PPMI *see* Positive Pointwise  
Mutual Information
- precision 284, *see also* recall
- predicted value *see* fitted value
- Principal Components Analysis  
(PCA) 353–361  
PCA with supplementary  
variables 354, 357–361
- Principal Coordinates  
Analysis 337, *see also*  
Multidimensional  
Scaling
- principal inertia 369, 381–382,  
*see also* Correspondence  
Analysis
- Principle of Economic  
Versatility 91
- probability density 10–11, *see*  
*also* density plot
- probability distribution 10–11
- Promax 363–364, *see also* Factor  
Analysis
- prompt line 24
- proportion 70–72, *see also* ratio
- Prototype Theory 373
- p*-value 11–12, 98, 103–104,  
129–130  
Approximately Unbiased  
(AU) *p*-value 316
- Q**
- Q–Q plot 53–56
- qualitative variable *see* variable
- quantile 44–45, 53
- quantitative variable *see* variable
- quartile 44–45, 49, 58
- quasi-complete separation 273
- quite* + ADJ 242–251
- R**
- R console 22, 24, 28, 32
- $R^2$  142, 148, 151, 162, *see also*  
goodness of fit  
adjusted  $R^2$  148–149, 151  
McFadden  $R^2$  280, 284  
Nagelkerke  $R^2$  259
- random effects 193–196
- random forest 291–292, 297–299
- random sampling 8
- range 48, *see also* measures of  
dispersion
- ratio 71–72, *see also* odds ratio
- ratio-scaled variable *see* variable
- recall 284, *see also* precision
- reference level 78, 97, 146, *see*  
*also* factor
- regression 120, 139–170, 253–289  
Firth penalized  
regression 273  
linear regression 139–170, 266  
linear regression  
assumptions 155  
least squares method 121, 142  
logistic regression 222,  
253–289, 383–384



- binomial (dichotomous)
  - logistic regression 253–276
- logistic regression
  - assumptions 271
- multinomial (polytomous)
  - logistic regression 253, 277–289
- multiple regression 129, 139–142
- non-parametric regression 155, 167–169, 187
- regression coefficient 140, 145, 148, 254, 357–358, 379
- regression hyperplane 141
- regression line 117–119, 121, 124–125, 140–141
- simple regression 139–142
- Reliance 225, 228–232, 234, 236, 238, *see also* association measures
- REML *see* restricted maximum likelihood
- representative sampling 8
- resampling methods 166, 292, *see also* bootstrap, permutation
- residuals 120–121, 126, 141–142, 145, 149, 153–162, 178, 197, 218–221, 266
  - Pearson residuals 218–221
  - standardized Pearson residuals 220–221
  - studentized residuals 153–154
- residuals vs. fitted plot 157
- response variable *see* variable
- restricted maximum likelihood 193
- RStudio 33, 39
- rug plot 105–106
- S**
- sample 7–8
  - dependent samples 87–88
  - independent samples 87–88
- sampling distribution 88, 99
- sampling error 8, 13
- scalar adjectives 243, 247
- scale of measurement 16–18, 134
- scatter plot 39, 63, 117–119, 229–231, 409–412, 423
- scientific notation 56–57
- Score test 263
- scree plot 318–319, 338–339, 355–356, 382
- SE *see* standard error
- Semantic Vector Space 323–332
- Sessel 373–384
- Shapiro-Wilk normality test 56, 66, 95, 109, 112, 162, 176, 184
- shared reference 182–184
- Shepard plot 341–342, 347, *see also* Multidimensional Scaling
- significance level 12–13
- simple distinctive collexeme analysis *see* distinctive collexeme analysis
- simple regression *see* regression
- single method of agglomerative clustering *see* cluster analysis
- slope 119, 125, 140, 145, 153, 166–167, 274–275, *see also* regression
  - random slopes 193
- small *n* large *p* 292
- snake plot 313–315
- sorting a data frame *see* data frame
- Spearman *p* *see* correlation
- splitting a data frame *see* data frame
- standard deviation 48–50, *see also* measures of dispersion
- standard error 98–100
- standard scores *see* z-scores
- standardized Pearson residual *see* residuals
- statistical power 13, 134
- statistical significance 12, 103
- statistics 7–8, 15–16
  - descriptive statistics 7–8
  - inferential statistics 7–8
- stepwise selection *see* variable selection
- stress 337, 341–342, 345, 347, 349, *see also* Multidimensional Scaling
- strip chart 105–106
- studentized residuals *see* residuals
- Stuhl 373–384
- subcategorization
  - frames 323–324
- subsetting a data frame *see* data frame
- sum contrasts *see* contrasts
- supplementary variable *see* variable
- symmetric plot 370, *see also* Correspondence Analysis
- T**
- tab delimited text file 34
- table 30
- test of independence 199–200, 210
- ties 133, 348
- time series analysis 126
- token frequency 30
- trail 391
- transpose 315, 404
- treatment contrasts *see* contrasts
- t*-score 99–100, 234, 238
- t*-statistic 96–97
- t*-test 87–89, 95–98, 110–113, 121–122
  - assumptions of *t*-test 88
  - dependent (paired) *t*-test 110–112, 121–122
  - independent (unpaired) *t*-test 95–98
- Tukey Honest Significant Differences test 179–180, 188, *see also* post hoc test
- two-tailed test 13–14, 89, 95–96, 98, 108–110, 128, 148
- type frequency 75, 79
- Type I error 13, 110, 175
- Type II error 13
- Type III sums of squares 185, 187, *see also* ANOVA
- U**
- unweighted average 146
- U*-test 87, 109, 179
- V**
- variable 16
  - binary variable 16
  - categorical variable 18
  - continuous variable 18, 42
  - dependent variable 16
  - discrete variable 18, 42

- dummy variable 185  
 explanatory variable 16  
 independent variable 16  
 interval variable 17–18  
 nominal variable 16–17, 30  
 ordinal variable 17–18, 146, 343–344  
 qualitative variable 18  
 quantitative variable 18  
 ratio-scaled variable 17–18  
 response variable 16, 120, 137, 139–144, 155, 158–159, 176, 198, 253, 291, 299  
 response variable  
     transformation 158–159  
 supplementary variable 354, 357–361, 378–381  
 variable importance 297–298  
 variable selection  
     best subsets method 149, 151  
     stepwise selection 149–152, 266–267  
     backward stepwise selection 149–150, 152, 266  
     bidirectional stepwise selection 149, 150, 267  
     forward stepwise selection 149–150, 152, 266  
 variance 48–50, *see also* measures of dispersion  
 Variance Inflation Factors 160, 272–273  
 Varimax 363–364, *see also* Factor Analysis  
 VIF-scores *see* Variance Inflation Factors  
 VU Amsterdam Metaphor Corpus 201–202, 215
- W**  
 Wald test 263, 282  
 Ward method of agglomerative clustering *see* cluster analysis  
 warning message 6, 29  
 WCSS *see* within-cluster sum of squares  
 Welch correction for *t*-test 88
- White adjustment for ANOVA 187  
 wide format *see* data frame  
 Wilcoxon test 87–88, 108–110  
 will 387–393  
 within-cluster sum of squares 317–319  
 within-group variability 178  
 within-subject design 172, *see also* ANOVA  
 within-subject variable 172, 191, *see also* ANOVA  
 word length 42–43, 89, 116–118, 120, 122, 125, 134–135, 137–139, 160, 163–165, 169
- Y**  
 Yates continuity correction 212
- Z**  
 Zipf's law 64  
 z-score 57, 59–62, 99, 234





# Index of R functions and packages

- ? 28
- ?? 29
- A**
- abline() 118, 124
- abs() 59–60, 83, 85
- adist() 308
- aes() 53, 424
- aggregate() 183–184
- alpha() 365
- anova() 152, 163, 194, 267–268, 286
- Anova() 187, 197
- aov() 178, 186, 197
- ape (package) 321
- apply() 49, 235, 403–404
- as.character() 314, 401
- as.dist() 330, 338
- as.factor() 314, 400
- as.matrix() 37, 328
- as.numeric() 71, 299, 334, 401
- assoc() 220
- assocstats() 209–210, 217
- attach() 43, 117
- axis() 413
- B**
- barplot() 74, 203, 205, 216, 338, 355, 415–416, 420–421
- barplot2() 101
- bmp() 39
- boot (package) 142, 167
- boot() 167
- boot.ci() 168–169
- box() 413
- boxCox() 158–159
- boxplot() 57–58, 93, 192, 423
- boxplot.stats() 59, 93
- bp() 304, 306
- C**
- c() 30–31, 73–74, 78, 89, 92, 101–102, 105, 116, 142, 146, 173, 185–186, 189, 191, 200, 202–205, 216, 226, 228, 254, 277, 293, 302, 335, 338, 342–343, 352, 367, 373, 399–404, 406, 411–413, 416, 419–420
- ca (package) 367, 369, 373, 382
- ca() 369
- car (package) 116, 128, 142, 153, 156–160, 173, 176, 182, 187, 254, 270, 272
- cat() 39
- cbind() 31, 127, 185, 201–202, 213, 235, 365
- cforest() 297
- chisq.test() 29, 211–213, 217–218, 221, 328, 375
- citation() 23
- ckappa() 201
- class() 397
- cluster (package) 22, 302, 308, 311–312, 319, 321, 324, 330, 342, 344
- cmdscale() 338, 340
- coef() 168
- coin (package) 173, 179
- colMeans() 313
- colnames() 202, 204, 216, 403, 412–413
- colors() 73, 420
- colSums() 84
- confint() 147, 263, 281
- contrasts() 146, 185–186
- contr.helmert() 146
- contr.sum() 146
- contr.treatment() 146
- coord\_flip() 79, 106, 424
- coord\_polar() 74, 424
- cor() 28, 119, 132, 134, 235, 238, 353
- cor.test() 128, 132
- corrgram (package) 134–135, 226, 236
- corrgram() 135–136, 236–237
- cortest.bartlett() 354
- cossim() 329
- crossprod() 329
- crPlot() 156
- ctree() 294, 296
- cutree() 311–312
- D**
- daisy() 308, 344
- data() 23, 279
- data.frame() 31, 92, 102, 106, 136, 230, 343, 404–405
- density() 51, 66, 105
- detach() 43, 129
- dev.off() 39
- diana() 321
- diff() 53
- dim() 61, 305, 402
- dimdesc() 357, 377
- dist() 306, 308, 328, 330, 341
- do.call() 306
- dotchart() 78, 298, 418
- drop1() 152, 267
- durbinWatsonTest() 128, 161
- E**
- energy (package) 116, 127
- exp() 259, 261, 263–265, 281, 286
- F**
- fa() 366
- factanal() 361, 363–364
- FactoMineR (package) 352, 354, 360, 373, 375, 377, 382, 385
- factor() 30, 74, 77–78, 146, 293, 400–402, 404
- fields (package) 338
- file.choose() 38–39
- fisher.test() 214, 232
- fitted() 120, 147, 282
- fix() 408
- fligner.test() 177, 184
- friedman.test() 198
- function() 168–169, 176, 306, 312, 318, 320, 343
- funny() 321
- G**
- geom\_bar() 74–75, 102, 205–206, 423–424
- geom\_boxplot() 58, 93, 192, 423
- geom\_density() 423–424
- geom\_errorbar() 102, 423
- geom\_histogram() 54, 423
- geom\_line() 54, 63, 423
- geom\_point() 63, 79, 106, 118, 423
- geom\_text() 230, 423

getmap() 334  
getwd() 28  
ggmap (package) 336  
ggplot() 53–54, 58, 63, 74–75, 79,  
93, 102, 106, 118, 192, 205–206,  
230, 423  
ggplot2 (package) 5, 50, 53,  
57–58, 62–64, 69, 74–76, 79,  
89, 92–93, 97, 102, 104, 106,  
116, 118–119, 174, 189, 192, 200,  
204–207, 226, 230–231, 336,  
342, 346, 421, 424, 431  
ggsave() 424  
ggtitle() 424  
glht() 181  
glm() 257–258, 263, 266–267,  
271–272  
googleVis (package) 387, 389  
gplots (package) 89, 101, 189  
grep() 399  
gsub() 400  
gvisMotionChart() 389

## H

hclust() 309–310  
head() 35, 42, 76–77, 90, 92, 97,  
111, 120–121, 136, 174, 183,  
227–229, 232–233, 235, 243–244,  
256, 279, 282, 326, 355, 388  
help() 28, 39, 275, 296, 321, 420  
help.search() 29  
hist() 51  
Hmisc (package) 299

## I

I() 159  
identical() 306  
ifelse() 232–234, 245, 249,  
327–328  
influencePlot() 153, 270  
install.packages() 22, 41, 50, 62,  
69, 104, 134, 182, 201, 215, 292,  
324, 333, 348, 387  
interaction() 184  
interaction.plot() 183  
intervals() 195  
IQR() 49  
is() 32, 397  
is.na() 47, 245, 248, 344  
isoMDS() 345, 348

## J

jpeg() 39

## K

kmeans() 317–318  
kruskal.test() 179

## L

labs() 106, 192, 424  
lapply() 306, 343  
leaps (package) 142, 151  
legend() 191, 203, 205, 216, 335,  
411, 413, 416, 420  
length() 51, 100, 124, 235, 305,  
313, 398  
levels() 73, 77–78, 97, 259, 293,  
303, 305, 335, 400–401  
leveneTest() 176  
library() 22–23, 41, 50, 62,  
69, 80, 89, 104, 111, 116, 134,  
142, 173, 182, 189, 200, 201, 215,  
226, 242, 254, 277, 292, 302,  
324, 333, 338, 342, 348, 352, 367,  
373, 387, 409, 414  
lines() 131, 341, 413  
list() 105  
lkappa() 201  
LL.collostr() 233–234  
lm() 118, 124, 144, 147, 150, 154,  
157, 159, 160, 162–163, 166,  
168–169, 177  
lme() 193, 197  
load() 28, 39  
log() 25, 66, 144, 151, 154,  
158–159, 163, 166, 168, 234–235,  
259–261, 265, 399, 409–411  
log10() 233–234  
log1p() 66, 136  
log2() 234  
lowess() 131, 136  
lrm() 257–258, 262, 272, 274, 383  
ls() 27

## M

mad() 49  
mantel() 348  
MASS (package) 338,  
341–342, 345  
matrix() 402  
max() 46, 51  
MCA() 376, 379  
mean() 43, 47, 61, 71, 100–101,  
174, 183, 344  
median() 44  
melt() 204  
merge() 244, 248

min() 46, 51, 84–85, 235  
mjca() 382  
mlogit (package) 277, 279–280,  
282–283, 286, 289  
mlogit() 279, 282  
mlogit.data() 279  
mlv() 45  
modeest (package) 41, 45  
model.tables() 188  
month.name 25  
mosaic() 219, 368  
mosaicplot() 368  
multcomp (package) 181  
multinom() 289  
mvnorm.etest() 127

## N

names() 229, 313, 415  
ncvTest() 128, 157–159  
nj() 321  
nlme (package) 189, 193, 195  
normalize() 59–62  
npcomp (package) 173, 180  
npcomp() 180  
nrow() 244, 247, 279, 355, 363

## O

ols() 166  
oneway.test() 178, 187  
oneway\_test() 179  
options() 57  
order() 94, 245–246, 249–250, 407  
ordered() 343, 401

## P

pairwise.t.test() 181  
pam() 319–320, 330  
par() 51, 156, 421  
party (package) 264, 292, 294,  
371  
paste() 73, 400, 419  
PCA() 354  
pdf() 39  
pie() 72–73, 420  
plot() 39, 51, 62–63, 66, 105, 117,  
131, 157, 162, 180, 230–231, 294,  
309, 313–314, 316, 318, 334, 339,  
341, 356, 358–359, 363, 369, 376,  
379, 389, 409, 411, 413  
plot.lm() 157, 162  
plot.philo() 321  
plot3d() 340, 371–372  
plotellipses() 359, 380

png() 39  
 points() 334, 411  
 polytomous (package) 253, 277,  
   283, 286, 289  
 polytomous() 283, 286  
 position\_dodge() 424  
 position\_jitter() 424  
 predict() 266, 297, 299  
 princomp() 366  
 print() 285, 286  
 prop.table() 70, 82–85, 207,  
   304, 419  
 psy (package) 201  
 psych (package) 352, 354,  
   365–366  
 pv.Fisher.collostr() 232, 245  
 pvclust (package) 302, 315–316  
 pvclust() 315–316  
 pvrect() 316

## Q

q() 28  
 qnorm() 99  
 qqplot() 346, 423  
 qqline() 53, 66, 105  
 qqnorm() 53, 66, 105  
 qt() 100  
 quantile() 45

## R

rainbow() 420  
 range() 48, 53, 413  
 rank() 132  
 rbind() 31, 202, 216,  
   305–306, 412  
 rdist.earth() 338  
 read.csv() 35, 37, 408  
 read.csv2() 37  
 read.table() 35–38, 408  
 rect.hclust() 312  
 regsubsets() 151  
 relevel() 78, 146, 401  
 rep() 92, 106, 160, 191, 203, 397  
 reshape (package) 200, 204  
 residuals() 121, 162  
 return() 168–169  
 rgl (package) 338, 340, 367, 372  
 rm() 27  
 rms (package) 142, 159–160, 166,  
   254, 257, 272, 274, 373  
 rnorm() 111

round() 111, 151, 232–233, 235, 238,  
   285, 298, 308–309, 353, 399, 419  
 rowMeans() 403  
 rownames() 202, 229–230, 305,  
   334–335, 340, 346, 353, 363, 412  
 rowSums() 326–327, 403, 415  
 rug() 105  
 rworldmap  
   (package) 333–334, 336

## S

sample() 235, 274  
 sapply() 312, 318, 320  
 save() 39  
 scale\_fill\_grey() 192, 205–206  
 scale\_fill\_manual() 74  
 scale\_x\_continuous() 58  
 scale\_x\_discrete() 74  
 scan() 39  
 sd() 48, 100, 196, 403  
 seq() 51, 159, 397  
 set.seed() 294, 298  
 shapiro.test() 56, 66, 109, 112,  
   162, 176, 184  
 Shepard() 341  
 silhouette() 311–312  
 smacof (package) 348, 350  
 smacofSym() 348  
 somers2() 299  
 sort() 44, 62–63, 229,  
   231, 298, 313, 398, 415, 418, 420  
 split() 305  
 sqrt() 25, 27, 48, 66, 100, 159,  
   218, 234–235, 329, 341  
 stat\_bin() 423  
 stat\_boxplot() 423  
 stat\_density() 424  
 stat\_identity() 424  
 stat\_smooth() 118  
 stat\_qq() 54  
 step() 149–150, 266  
 str() 31, 35–36, 42, 69, 90, 143,  
   174, 190, 248, 256, 278, 302,  
   305, 333, 343, 352, 374, 388, 405  
 stripchart() 105  
 sum() 70, 83, 85, 244–245, 341  
 summary() 46, 53, 62, 69–70,  
   77–78, 91, 104, 117, 144, 151,  
   154, 161, 165, 169, 177–178, 194,  
   257–258, 268, 279, 283–287, 293,  
   312, 344, 369, 382, 398, 401, 405

## T

t() 315, 329, 416  
 t.test() 96–97, 112  
 table() 30, 45, 70, 203, 207,  
   262, 283, 297, 299, 304, 315,  
   401–402, 408, 423  
 tail() 97  
 tapply() 174–175  
 text() 230, 313, 335, 339, 363,  
   410–411  
 text3d() 340  
 theme() 58, 79  
 theme\_bw() 79  
 tiff() 39  
 treeresponse() 299  
 TukeyHSD() 179–180, 188

## U

unlist() 299  
 update() 194

## V

validate() 166, 274  
 var() 48, 404  
 varimp() 298  
 vcd (package) 200, 209,  
   215, 217, 219–220, 302, 315,  
   367–368  
 vegan (package) 348  
 vif() 159–160, 272  
 vignette() 289  
 visreg (package) 142, 163,  
   254, 269  
 visreg() 163, 269

## W

waldtest() 282  
 which() 60, 398–400, 407, 411  
 wilcox.test() 110  
 write() 39  
 write.csv() 38  
 write.table() 38, 408

## X

xlab() 63, 75, 79, 93, 102  
 xlim() 335, 424  
 xtabs() 203, 375, 402

## Y

ylab() 63, 75, 79, 93, 102  
 ylim() 335, 424