

Index

Entries in **this font** are names of S-PLUS functions and SAS procedures. Page numbers in **bold** are to the most comprehensive treatment of the topic.

- D_{xy} , 70, 113, 223, **223**, 225, 238, 319, 484, 510
 - censored data, 485, 498
- R^2 , 87, 88, 176, 177, 242, 356, 357
 - adjusted, 64
 - generalized, 177
 - significant difference in, 184
- c index, 223, **223**, 225, 319, 485, 498
 - censored data, 484
 - generalized, 319, 484

- Aalen survival function estimator, *see* survival function
- abbreviate, 245
- accelerated failure time, *see* model
- accuracy, 87, 90, 181, 350, 417
 - apparent, 90, 239, 510
 - approximation, 95, 273, 346, 445
 - bias-corrected, 86, 90, 91, 112, 357, 510
 - calibration, **60–64**, 73, 77, 78, 87, 91, 112, 203, 224, 225, 228, 238, 239, 276, 292, 322, 417, 418, 445, 485
 - discrimination, 77, 78, 87, **87**, 223, 224, 238, 319, 330, 343, 445, 484, 485, 487
 - future, 182
 - index, 97, 98, 112
- ACE, 68, 143, 149, **356–357**, 357, 358
- ace, 143, 358
- actuarial survival, 383
- add1, 241

- AIC, 23, 57, 64, 73, 139, 141, **173**, 175, 181, 184, 208, 209, 236, 331
- Akaike information criterion, *see* AIC
- analysis of covariance, *see* ANOCOVA
- ANOCOVA, 14, 191, 199, 418
- ANOVA, 11, 26, 62, 199, 202, 418, 458, 516
- anova, 99, 105, 107, 108, 141, 143, 241, 250, 251, 253–255, 258, 264–266, 293, 298, 303, 334, 340, 343, 344, 369, 442
- anova.Design, 102
- apply, 129, 360
- areg.boot, 358, 359, 361
- Arjas plot, 472
- asis, 104, 105
- assumptions
 - accelerated failure time, 409, 437
 - additivity, 32, 215
 - continuation ratio, **320**, 321, 338
 - distributional, 34, 80, 418, 506
 - linearity, 17–22
 - ordinality, **314**, 319, 332, 338
 - proportional hazards, 400, **472–483**
 - proportional odds, **315**, 317, 334
- AVAS, 357, 358
 - case study, 359–369
- avas, 358, 360, 361, 363–365

- B-spline, *see* spline function
- bastbw, 261
- battery reduction, 72

- Bayesian modeling, 59, 184
- BIC, 175, 182, 184, 236
- binary response, *see* response
- `bj`, 103, 418, 424
- `bootcov`, 106, 107, 166, 167, 169–171
- `bootkm`, 391
- `bootplot`, 169
- bootstrap, 83–86, 90–92
 - .632, 91, 97, 98
 - adjusting for imputation, 48
 - approximate Bayesian, 46, 70, 145
 - cluster, 106, 166, 167, 183
 - conditional, 90, 97, 166
 - confidence limits, *see* confidence limits
 - covariance matrix, 106, 166, 173
 - density, 85, 106, 169, 170
 - estimating shrinkage, 64, 91
 - model uncertainty, 9, 89
 - overfitting correction, 88, 90, 91, 221, 357
 - variable selection, 58, 81, 89, 148, 225
- Breslow survival function estimator, *see* survival function
- Brier score, 113, 224, 239, 270, 319
- CABG, 462
- `calibrate`, 107, **112**, 239, 276, 291, 319, 323, 350, 445
- `calibrate.cph`, 498
- calibration, *see* accuracy
- `cancor`, 101
- canonical correlation, 101
- canonical variate, 68, 69, 101, 131
- CART, *see* recursive partitioning
- casewise deletion, *see* missing data
- categorical predictor, *see* predictor
- categorical response, 360
- categorization of continuous variable, 6
- `catg`, 104, 105
- cause removal, 386
- censoring, **373–374**, 378, 379, 385, 396
 - informative, 374, 386, 387, 392
 - interval, 373, 390, 392
 - left, 373
 - right, 374, 390
 - type I, 374
 - type II, 374
- `ciapower`, 490, 492
- clustered data, 166, 389
- clustering
 - hierarchical, 101, 123, 327
 - variable, **66–67**, 101, 130, 351
- `coef`, 107
- collinearity, 65–66
- concordance probability, *see* *c* index
- conditional logistic model, *see* logistic model
- conditional probability, 320, 376, 454, 455, 461
- confidence limits, 8, 25, 30, 54, 55, 80, 108, 154, 167, 236, 254, 260, 357
 - bootstrap, 84, 85, 94, 97, 106, 167, 168, 170, 172, 173
 - coverage, 30, 167, 168, 356
 - simultaneous, 106, 167–169, 392, 499
- confounding, 26, 200
- `confplot`, 169
- contingency table, 164, 196, 199, 202
- `contrast`, 107, 108
- convergence, 162, 227
- `cor`, 125
- coronary artery disease, 43, 178, 205, 207, 211, 218, 470, 475
- correspondence analysis, 67, 101
- cost-effectiveness, 4
- Cox model, 453–498
 - case study, 501–514
 - data reduction example, 133–143
- `cox.zph`, 476, 497, 498, 507
- `coxph`, 103, 394, 492

- cph, 103, 105–107, 109, 133, 136, 138, 139, 141, 147, 394, 419, 492, **492**, 493–495, 497
- cpower, 490
- cr.setup, 323, 338, 350
- cross-validation, *see* validation of model
- cubic spline, *see* spline function
- cumcategory, 353
- cumulative hazard function, *see* hazard function
- cut2, 101, 105, 270, 332, 391
- data reduction, 66–75
 - case study 1, 119–148
 - case study 2, 247–252
 - case study 3, 327–331
- data-splitting, *see* validation of model
- data.frame, 306
- data.restore, 517
- datadist, 102, **102**, 110, 116, 255, 281, 297, 441
- dataRep, 307
- datasets, 517
 - meningitis, 230, 233
 - prostate, 119, 245, 501
 - SUPPORT, 51, 427
 - Titanic, 279
- degrees of freedom
 - effective, 24, 36, 64, 79, 106, 181, 236
 - generalized, 8
 - phantom, 30, 87
- delayed entry, 373, 374
- describe, 101, 120, 279, 361
- Design library, **102–112**, 517
- deviance, 203, 424, 464, 497
- deviance, 241
- DFBETA, 76
- DFBETAS, 76
- DFFIT, 76
- DFFITS, 76
- diag, 254
- diagnosis, 360
- Dialog, 107, 308
- dimensionality, 73
- discriminant analysis, 189, 199
- discrimination, *see* accuracy
- display library, 518
- distribution
 - binomial, 61, 151, 202
 - exponential, 113, 380, 397, 399, 425
 - extreme value, 399, 410
 - log-logistic, 7, 396, 399, 414, 483
 - log-normal, 7, 83, 358, 395, 399, 414, 441
 - Weibull, 34, **380**, 381, 392, 398, 404–406, 408–410, 416, 419
- dose-response, 503
- Dotplot, 283
- doubly nonlinear, 103
- drop-in, 492
- drop1, 241
- dummy variable, 2, 14, 15, 33, 34, 62, 101, 102, 180
- economists, 59
- effective.df, 106, 107, 343
- Emax, 350
- epidemiology, 32
- estimator
 - Buckley-James, 418, 424
 - maximum likelihood, 151
 - penalized, *see* maximum likelihood
 - self-consistent, 504
 - smearing, 358, 359
- expand.grid, 367
- explained variation, 242
- exponential distribution, *see* distribution
- external validation, *see* validation of model
- factor, 104, 105
- failure time, 371
- fastbw, 105, 107, 108, 348, 445

- fillin, 125, 127
- financial data, 3
- fit.mult.impute, 48, 70, **70**, 107, 147, 302
- fitted, 241, 362
- Fleming-Harrington estimator, 384, 462
- formula, 107, 241
- Function, 107, 109, 116, 255, 307, 362
- functions, generating S-PLUS code, 362

- GAM, *see* generalized additive model
- gam, 113, 356
- GDF, *see* degrees of freedom
- Gehan-Wilcoxon test, *see* hypothesis test
- gendata, 107, 108, 261
- generalized additive model, 24, 35, 109, 113, 356
 - case study, 359–369
- glm, 102, 103, 113, 241
- goodness-of-fit, 203, 238, 399, 412, 437
- Greenwood's formula, *see* survival function
- groupkm, 391

- hare, 424
- hat matrix, 76
- Hazard, 107, 421
- hazard function, 107, 372, **375**, 377, 382, 399, 453, 454
 - bathtub, 381
 - cause-specific, 386, 387
 - cumulative, 375–382
- hazard ratio, **401–404**, 405, 456, 458
 - interval-specific, 473, 474, 483
- hazard.ratio.plot, 497
- hclust, 101
- heft, 391
- heterogeneity, unexplained, 4, 200, 372
- Hmisc library, 101, 105, 109, 517
- hoeffd, 101
- Hoeffding D , 101, 124, 433
- Hosmer-Lemeshow test, 203

- Hotelling test, *see* hypothesis test
- hypothesis test, 16, 27
 - additivity, 32, 215
 - association, 2, 16, 27, 38, 55, 101, 202, 337, 464
 - equal slopes, 317, 321, 322, 337, 338, 437, 438, 472
 - exponentiality, 381, 398
 - Gehan-Wilcoxon, 484
 - global, 57, 80, 116, **159**, 176, 199, 201, 340, 507
 - Hotelling, 199, 200
 - independence, 101, 124
 - Kruskal-Wallis, 2, 101
 - linearity, 16, 27, 30, 31, 34, 37, 55, 75, 206
 - log-rank, 36, 394, 453, 464, 490, 499
 - Mantel-Haenszel, 464
 - partial, 159
 - Pearson χ^2 , 164, 202
 - robust, 7, 66, 67, 313
 - Wilcoxon, 2, 61, 101, 114, 199, 223, 313, 315

- I, 105, 359, 361
- ignorable non-response, *see* missing data
- imbalances, baseline, 372
- imputation, **43–50**, 69, 360
 - model for, 44, **45**, 46, 47, 49, 51, 69, 70, 101
 - multiple, 43, 48, **48**, 49, 79, 101, 145, 301, 518
 - single, **47**, 49, 50, 109, 125
- impute, 101, 107, 109, 116, 129, 133, 245, 246, 297, 334, 360, 361, 440
- incidence
 - crude, 388
 - cumulative, 387
- infinite regression coefficient, 202
- influential observations, **75–76**, 92, 221, 238, 483

- information function, 152, 153
- information matrix, 65, **158**, 159, 161, 165, 179, 181, 201, 344
- informative missing, *see* missing data
- inter-quartile-range effect, 93, 106
- interaction, 14, 31
- intra-cluster correlation, 106, 112, 165, 166, 389
- `is.imputed`, 116, 360
- jackknife, 89, 483
- Kalbfleisch-Prentice estimator, *see* survival function
- Kaplan-Meier estimator, *see* survival function
- Key, 283
- Kullback-Leibler information, 184
- `labcurve`, 367
- lasso, 59, 81, 352
- L^AT_EX, 101, 518
- `latex`, 107, 109, 267, 281, 334, 340, 344, 447, 518
- leave-out-one, *see* validation of model
- left truncation, 373, 392
- life expectancy, 4, 380, 450
- LIFEREG, 478
- likelihood function, **152**, 157, 158, 160, 163, 170, 454
- likelihood ratio test, **155–156**, 159–164, 167, 175, 176, 178, 196, **207**
- linear spline, *see* spline function
- link function, 13
- `l1ist`, 440
- `lm`, 103
- local regression, *see* nonparametric
- loess, *see* nonparametric
- `loess`, 24, 113, 213, 241, 471
- log-rank, *see* hypothesis test
- logistic model
 - binary, 187–200
 - case study 1, 245–278
 - case study 2, 279–308
- conditional, 460
- continuation ratio, 320–323
 - case study, 338
- extended continuation ratio, 321–322
 - case study, 338–351
- ordinal, 313
- proportional odds, 61, 314, **315–319**, 332, 360
 - case study, 332–338
- lowess, *see* nonparametric
- `lowess`, 113, 284
- `lrm`, 103, 106, 109, 171, 236, **236**, 237, 238, 247–249, 251, 252, 256, 257, 263, 265, 266, 285, 287, 293, 297, 302, 319, 323, 334, 337, 338, 340, 342, 368, 419, 493
- `lsp`, 105
- Mallows' C_p , 57
- Mantel-Haenszel test, *see* hypothesis test
- marginal distribution, 22, 389, 456
- martingale residual, 464, 471, 496, 497
- `matrix`, 105
- `matrx`, 105
- maximum generalized variance, 67–69
- maximum likelihood
 - estimation, 151
 - penalized, 8, 64, 91, 106, **179–183**, 236, 325, 326
 - case study, 342–351
 - weighted, 178, 179
- maximum total variance, 67
- `Mean`, 107, 421, 450, 493
- `Mean.cph`, 493
- meningitis, *see* datasets
- `merge.levels`, 245
- MGV, *see* maximum generalized variance

- missing data, 293
 - casewise deletion, 43, 67, 285, 298
 - describing patterns, *see* `naclus`, `naplot`
 - imputation, *see* imputation
 - informative, 42, 396
 - random, 42
- model
 - accelerated failure time, **409–418**, 427
 - case study, 437
 - Andersen-Gill, 492
 - approximate, **94–98**, 346, 348, 349, 352
 - Buckley-James, 418, 424
 - Cox, *see* Cox model
 - logistic, *see* logistic model
 - nonparametric additive, *see* generalized additive model
- model approximation, *see* model
- model validation, *see* validation of model
- `model.frame.default`, 102, 105
- monotone, 359, 361
- monotonicity, 49, 54, 69, 70, 79, 101, 124, 143, 355–357, 359, 433
- MTV, *see* maximum total variance

- `na.action`, 103
- `na.delete`, 103, 104
- `na.detail.response`, 103
- `na.fail`, 104
- `na.fun.response`, 103
- `na.include`, 129
- `na.omit`, 104
- `na.tree.replace`, 104, 113
- `naclus`, 42, 113, 293, 433, 440
- `naplot`, 42, 293, 440
- `naprint`, 107
- `naresid`, 104, 107
- nearest neighbor, 46
- Nelson estimator, *see* survival function, 394
- `Newlabels`, 450

- Newton-Raphson algorithm, 162, 164, 180, 201, 398
- noise, 29, 57, 60, 179, 273, 465, 503
- nomogram, **93**, 233, 234, 267, 349, 495, 514
- `nomogram`, 107, 109, 237, 267, 307, 319, 348, 450, 493, 494
- non-compliance, 374, 490, 492
- non-ignorable non-response, *see* missing data
- non-proportional hazards, 61, 424, 472
- nonparametric
 - correlation, 54
 - censored data, 498
 - independence test, 101, 124
 - regression, 23, 24, 36, 113, 211
 - test, 2, 101
- `ns`, 104, 105
- nuisance parameter, 159–161

- object-oriented program, viii, 99, 105
- observational study, 3, 50, 200, 372
- odds ratio, 190
- `ols`, 103, 106, 108, 109, 348, 419, 445, 447
- optimism, 86, 87, 90, 276, 357
- `ordered`, 105
- ordinal response, *see* response
- ordinality, *see* assumptions
- outlier, 92, 284
- over-adjustment, 2
- overfitting, **86**

- parsimony, 72, 80, 95
- partial residual, *see* residual
- partial test, *see* hypothesis test
- PC, *see* principal component
- penalized maximum likelihood, *see* maximum likelihood
- `pentrace`, 106, 107, 236, 323, 342
- person-years, 380, 397
- `plclust`, 101
- `plogis`, 237

- plot, 107, 124, 141, 241, 259, 267, 270, 281, 288, 291, 293, 298, 303, 327, 343, 345, 350, 360, 362, 433, 441, 442, 445, 493, 496
- plot.anova, 107, 108
- plot.Design, 102, 166, 421
- plot.lm, 117
- plot.lrm.partial, 338
- plot.summary, 107
- plot.survfit, 107, 391
- plot.xmean.ordinally, 319, 324, 332
- plsmo, 284, 309, 353
- pol, 105
- poly, 104, 105
- polynomial, 17
- popower, 319
- posamsize, 319
- post, 125
- power calculation, *see* cpower, spower, ciapower, popower
- pphsm, 419
- prcomp, 101
- predab.resample, 112, 238, 323
- predict, 99, 104, 107, 108, 112, 116, 125, 129, 241, 261, 270, 306, 348, 360, 367, 369, 445, 498, 507
- predict.Design, 102
- predictor
 - continuous, 17
 - nominal, **13–14**, 180
 - ordinal, 33
- principal component, 66, 72
- princomp, 101, 133, 136–138, 249
- PRINQUAL, 68, 69, 130
- print, 106, 107, 241
- print.lrm, 106
- product-limit estimator, *see* survival function
- propensity score, 3, 50, 200
- proportional odds model, *see* logistic model
- prostate, *see* datasets
- prune.tree, 270
- psm, 103, 106, 107, 109, 419, **419**, 420, 421, 437, 441
- Q-R decomposition, 19
- qqnorm, 117, 362
- Quantile, 107, 421, 450, 493
- Quantile.cph, 493
- rank correlation, *see* nonparametric
- Rao score test, **157**, 161, 162, 164, 167
- rcorr, 124
- rcorr.cens, 498
- rccs, 105, 285, 287
- rccspline.eval, 101
- rccspline.plot, 236, 496
- rccspline.restate, 101
- receiver operating characteristic curve
 - area, 77, 87, 223, 343
 - area, generalized, 319, 484
- recursive partitioning, 8, **25**, 26, 36, 42, 43, 46, 47, 49, 69, 72, 95, 113, 129, 130, 273, 293, 294, 346, 360
- regression to the mean, 61, 512
- resampling, 83, 89
- resid, 107, 117, 334, 337, 362, 437, 497
- residual, 362
 - logistic score, 316, 334
 - martingale, 464, 471, 496, 497
 - partial, 29, 242, 316, 317, 321, 336
 - Schoenfeld score, 316, **464**, 476, 497, 498, 506, 507
- residuals, 104, 107, 238, 241
- residuals.areg.boot, 362
- residuals.coxph, 497
- residuals.cph, 497
- residuals.lrm, 238, 334, 337
- residuals.psm, 437
- response
 - binary, 187–189
 - censored or truncated, 373

- continuous, 127, **355–369**
- ordinal, 360
- polytomous, 129, 360
- restricted cubic spline, *see* spline function
- ridge regression, 64, 91, 179, 180
- `rm.boot`, 168
- `rm.impute`, 107
- `rnorm`, 348
- `robcov`, 106, 107, 166, 254
- ROC, *see* receiver operating characteristic curve
- `rpart`, 113, 125, 127, 129, 131, 293, 294, 360
- `runif`, 437
- sample survey, 106, 165, 166, 179, 389
- SAS, 69, 124, 125, 130, 143, 317, 321, 476, 478, 519
- `sas.get`, 101
- `sascode`, 109
- `scat1d`, 259, 266, 267
- score function, 152, 153, 157
- score test, *see* Rao score test
- `score.binary`, 72
- `scored`, 104, 105
- scoring, hierarchical, 72
- scree plot, 133
- `sensuc`, 107
- shrinkage, 61–64, 72, 73, 179–183, 342–346
- similarity measure, 67, 294, 329, 433
- smearing estimator, *see* estimator
- smoother, 357, 362
- `solve`, 447
- Somers' rank correlation, *see* D_{xy}
- `somers2`, 343
- `spearman`, 125, 127
- Spearman rank correlation, *see* non-parametric
- `spearman2`, 101, 114, 311, 440
- `specs`, 106, 107, 257
- spline function, 17, 25
- B-spline, 19, 36, 104, 478
- cubic, 19
- linear, 17, 18, 105
- normalization, 22
- restricted cubic, 19–23
- `spower`, 490, 492
- standardized regression coefficient, 93
- state transition, 388
- `step`, 107, 241
- step halving, 164
- `strat`, 105
- `strata`, 105
- stratification, 195, 204, **219**, 390, 391, **458–460**, 465
- subgroup estimates, 29, 208, 372
- `summary`, 99, 106, 107, 130, 137–139, 171, 241, 258, 281, 293, 296, 301, 303, 362, 365, 442
- `summary.areg.boot`, 362
- `summary.Design`, 102, 106, 166
- `summary.formula`, 114, 293, 319
- super smoother, 24
- SUPPORT study, *see* datasets
- `supsmu`, 113, 237, 357
- `Surv`, 133, 390, 394, 437, 497
- `survdiff`, 498
- `survest`, 107, 419, 495
- `survest.psm`, 421
- `survfit`, 107, 390, 391, 420
- `survfit.coxph`, 495
- `survfit.cph`, 495
- `Survival`, 107, 421, 493
- survival function
 - Aalen estimator, 384, 385
 - Breslow estimator, 462, 464
 - crude, 388
 - Kalbfleisch-Prentice estimator, 462, 464
 - Kaplan-Meier estimator, **382–385**, 386–388
 - multiple state estimator, 388
 - Nelson estimator, 384, 385, 390, 462

- standard error, 384
- Survival.cph, 493
- Survival.psm, 493
- survplot, 107, 391, 419, 421, 437, 495, 496
- survplot.Design, 421
- survplot.survfit, 391
- survreg, 103, 419
- survreg.distributions, 424

- table, 360, 361
- tapply, 297
- test of linearity, *see* hypothesis test
- test statistic, *see* hypothesis test
- text, 270, 345, 360
- text.anova.Design, 108
- tile.tree, 270
- time to event, 371
- time-dependent covariable, 322, 390, 476, 478, 481, 482, 492, 499, 509
- Titanic, *see* datasets
- training sample, 87–89, 97
- transace, 143
- transcan, 46, 48, 49, 69, **69**, 70, 71, 101, 107, 109, 125, 127, 129–133, 136, 140, 143, 145–147, 270, 296, 301, 302, 329, 334, 501, 502, 504
- transformation, 362
 - post, 105
 - pre, 149
- tree, 69, 113, 129–131, 270, 360
- tree model, *see* recursive partitioning
- tree.screens, 270

- un-conditioning, 94
- univariable screening, 59
- update, 100, 241, 291, 338, 343, 445

- val.prob, 86, 107, 239
- val.surv, 86, 424, 498

- validate, 107, **112**, 225, 238, 239, 270, 273, 274, 291, 292, 319, 323, 350, 445
- validate.cph, 498
- validate.tree, 113
- validation of model, **86–92**, 224, 273, 291, 319, 322, 350, 417, 445, 485, 510
 - bootstrap, 90–92
 - cross, 89, 91, 181
 - data-splitting, **87**, 88
 - external, 86, 239, 424, 498
 - leave-out-one, 89, 97, 184, 220
 - randomization, 89, 273
- VARCLUS, 65
- varclus, 65, 101, 124, 141, 327, 433
- Varcov, 447
- variable selection, **55–60**
 - step-down, 57, 108
- variance inflation factors, 65, 107, 109, 220
- variance stabilization, 357, 362
- variance-covariance matrix, 47, 48, 95, 101, 159, 161, 162, 165–167, 179, 181, 184
 - cluster sandwich, 165
 - sandwich, 181
- vif, 107, 109

- waiting time, 374
- Wald statistic, **156**, 159, 161, 162, 165–167, 177, **211**
- weighted analysis, *see* maximum likelihood
- which.influence, 107–109, 238
- working independence model, 166

- xYplot, 367