Subject index

A

Aalen–Nelson estimate . . . . . *see* Nelson–Aalen estimate
accrual 348–358
  exponential 350, 351, 354–358
  period 349–354, 356, 357
  uniform 350–354, 358
additive-hazards model 20
administrative censoring 336, 339, 348, 349, 355
AFT metric
  definition 16, 19, 20, 232, 239–241
  relation to PH metric 241–244
  specification 233
Akaike information criterion 278, 281–282, 317
allocation of subjects 334–335, 338, 340
analysis time 24–27, 52–55
at-risk table 105–107

B

bathtub hazard 8
Bayesian information criterion 282
Bonferroni adjustment 204
bootstrap, variance estimation 165
boundary effects . . . . . *see* kernel smoothing, boundary bias
boundary kernels method 115
Breslow approximation for tied failures 150–151
Breslow test . . . . . *see* Wilcoxon test
BRR, variance estimation 165

C

cancer data 170–171
casewise deletion . . *see* listwise deletion
categorical variables 178–180, 309
censored-normal regression . . . . . *see* regression models, censored normal
censoring 29–30
  administrative . . *see* administrative censoring
  interval 32–33
  left 34
  loss to follow-up . . . . . *see* loss to follow-up
  right . . . . . *see* right-censoring
`ci` 92
clustered data 157–158, 198–199, 325–326, 358
competing risks 365–391
complete case analysis . . . . . *see* listwise deletion
concordance, measures 222–223
conditional logistic regression . . . . . *see* regression models, conditional logistic
confidence bands, plot 107
confidence interval for
  cumulative hazard function 109
  estimated coefficients 186
  hazard ratio 133
  mean survival time 119–120
  median survival time 118–119
  survivor function 96
contingency table 123–124
Cox regression . . *see* regression models, Cox
Cox–Snell residuals 219–222, 284, 294–295
cubic splines 185

cumulative
distribution function
conditional 9
estimation . . . . . *see* empirical distribution function
generalized gamma 276
loglogistic 273
lognormal 269
relation to cumulative hazard . . 9
relation to survivor function . . . 7
standard normal 269
Weibull 9, 265
hazard count data interpretation 13–15
hazard function
baseline 135–137, 139, 143, 145, 155–156, 161, 172
conditional 9
estimation 107–113, 135–137, 155–156, 161, 291–294
exponential 247
Gompertz 267
goodness of fit 219–222
Nelson–Aalen estimate . . . . . *see* Nelson–Aalen estimate
plot 102, 142, 295–300
relation to survivor function . . . 8
Weibull 257
incidence function 367–368, 372–375, 382–389
subhazard function 383

D

date format 48–50
delayed entry 12, 34–35, 38, 62–64, 86, 97–99, 104, 105, 245, 293
generation of 11–12
delta method 133
density function 16, 287, 288
conditional 9
for censored data 31
relation to cumulative hazard 9
relation to likelihood 21, 245
relation to survivor function 7
Weibull 9
deviance residuals 212, 285
DFBETA . . . . . *see* influence diagnostics, DFBETA
diagnostic measures 213, 223–228
dispersion parameter 265
dropout of subjects . . . . . *see* loss to follow-up

E

effect size 335, 337, 347, 356, 359
determination 359–360
specification 338, 356
efficient score residuals 223–228
Efron approximation for tied failures 151
empirical distribution function . . 99–101
enduring variables 41, 47, 75
`estat concordance` 222–223
`estat phtest` 206–209
Euler's constant 255, 265, 288
exponential regression . . . *see* regression models, exponential
extended mean 120
extreme-value distribution . . . . . *see* Gumbel distribution

F

failure function . . . . . *see* cumulative distribution function
fixed effects 199
FMI 174–175
follow-up period 29, 349–353, 356, 357, 359
`fracpoly`, `stcox` 183–185
fraction missing information . . . *see* FMI
fractional polynomials 182–185, 206
multivariable 185, 218–219
frailty models
shared 156–164, 169, 199–200, 324–331
unshared 311–324

G

gamma function 16, 255, 265, 276, 287, 288

gaps 35–36, 38, 39, 67–70, 77, 83, 86–87, 99–101, 245, 284, 293, 298–300
generalized gamma regression . . . . . *see* regression models, generalized gamma
Gompertz regression . . . . . *see* regression models, Gompertz
goodness of fit 219–223
Greenwood's standard error 96
Gumbel distribution 255, 265

H

Harrell's C statistic 222
hazard
  cause-specific . . . 366–367, 369–372, 375–381
  contribution . . . 113, 135, 139, 141, 143, 145
  cumulative . . . . . *see* cumulative hazard function
  function
    baseline 19, 129–131, 134–135, 139–141, 152, 163–164, 234–236, 239
    bathtub 8
    conditional 9
    cumulative . . . . . *see* cumulative hazard function
    definition 7–8
    estimation . . . 113–117, 135, 139–141, 163–164, 284, 291–294
    examples of 9
    exponential 247
    Gompertz 267
    lognormal 271
    modeling . . 19–24, 156, 177–197, 301–331
    plot 102, 142, 163–164, 295–300
    test 124
    Weibull 9, 257
  log relative 135, 183–189
  metric . . . . . *see* PH metric
hazard, *continued*
  rate 8, 13, 15–16
  ratio . . 130–134, 143, 160, 192, 234, 333, 338, 340, 344, 347, 356, 359, 361, 364
    confidence interval for 133
    estimation 284
    interpretation in frailty models 321
    standard error of 132–133
    test 133, 339
  relative 135, 183
heterogeneity
  modeling 310–312
  test for 317–324
hip-fracture data
  description of 82–89
  sample size for 333–335
hypothesis
  alternative 133, 345, 346
  null . . 124, 133, 158, 207, 337, 339, 343, 345, 346

I

imputation . . . . *see* multiple imputation
incomplete gamma function 276
indicator variables . . . . . *see* categorical variables
influence diagnostics
  Cox–Snell residuals . . *see* Cox-Snell residuals
  deviance residuals . . . . *see* deviance residuals
  DFBETA 223–225
  efficient score residuals . . . . . *see* efficient score residuals
  likelihood displacement values 225–226
  LMAX 225–227
instantaneous variables . . 41, 47, 54, 63
interaction terms 186–189
interval regression . . . . . *see* regression models, interval
interval-censoring 32–33

interval-truncation . . . . . *see* gaps
`intreg` 2, 233

J

jackknife, variance estimation 165

K

Kaplan–Meier estimate 5, 93–122, 126, 138–139, 141, 210–212, 230, 342
kernel smoothing 113–117, 139–141
  boundary bias 115–116
Kidney data 156–164, 312–316

L

left-censoring 34
left-truncation . . . . . *see* delayed entry
likelihood displacement values . . . . . *see* influence diagnostics, likelihood displacement values
likelihood function
  for censored data 31, 32
  for parametric models 245–246
  for truncated data 35, 36
  partial . . . . . *see* partial likelihood
  penalized . . *see* penalized likelihood
likelihood-ratio test 159, 279–280, 321, 322, 327
linearization, variance estimation . . 165
`linktest` 203–204
listwise deletion 169–170
LMAX . . *see* influence diagnostics, LMAX
log-rank test 5, 123–124, 214, 334, 337, 345, 347, 361
  power of 338, 345, 364
  sample size for . . . . *see* sample size, for the log-rank test
log-time metric . . . . . *see* AFT metric
`logistic` 3
logistic regression . . . . . *see* regression models, logistic
loss to follow-up . . . . . 348–349, 355–358, *see* withdrawal
`lrtest` 279–280
lung cancer data 166

M

Mantel–Cox test . . . . . *see* log-rank test
MAR 170–172
marginal effects 188, 190
martingale residuals 214–218, 285
maximum likelihood estimation . . 20–21, 31–35, 38, 146–147, 231, 245–246
maximum pseudolikelihood estimators 165
MCAR 170–172
mean survival time
  definition 16–17
  estimation . . . 91–92, 117, 119–122, 284–289
median survival time
  definition 16–17
  estimation . . . 93, 117–119, 284–291
`mfp` 185, 218
`mi` 170–175
  `mi estimate` 170, 173–175
  `mi impute` 170–173
  `mi register` 172
  `mi set` 172
Mill's ratio 7
missing at random . . . . . *see* MAR
missing completely at random . . . . . *see* MCAR
missing data . . *see* multiple imputation
missing not at random . . . . . *see* MNAR
MNAR 171
multiple failures . . . 59–62, 157–158, 316
  generation of 12
multiple imputation
  Cox model 169–175
multiple-myeloma data . . 336–337, 339, 341, 342, 345, 359

N

Nelson–Aalen estimate 5, 107–113, 138–139, 141, 172, 220–222
nested models 278–281
NHANES I 166
NHEFS 166

nonparametric analysis 5, 91–128
censoring 96–97
truncation 97–101
null hypothesis . . . . *see* hypothesis, null
number of events . . . 334–335, 338, 340, 345–347, 350, 361
number-at-risk table . . *see* at-risk table

O

OLS regression . . *see* regression models, OLS
one-sided test . . . 338, 340–341, 347, 359
outliers, identification . . . . *see* influence diagnostics
overdispersion . . . . . *see* heterogeneity

P

parametric analysis 2, 229–244
censoring 31–32
likelihood function 21, 31, 38
truncation 35–36
partial likelihood 146
penalized likelihood 161
percentiles 10, 117–119
PH metric
definition 20, 232–239
relation to AFT metric 241–244
specification 233
piecewise constant model . . . . . *see* regression models, piecewise constant
piecewise exponential model . . . . . *see* regression models, piecewise exponential
power 335, 337, 338, 345
by simulation 358
curve 360, 364
definition 337
determination 343, 359–360
relation to number of events . . 335, 338
`predict` . . . 135–145, 155–156, 160–162, 283–295
primary sampling units 166
Probability Integral Transform 10
probability weights 164
product limit estimate . . . . . *see* Kaplan–Meier estimate
proportional hazards regression . . . . . *see* regression models, Cox
proportional-hazards assumption
graphical assessment 209–212
test of 206–209
`pweights` 165

Q

quantile function 9–12
Weibull 11

R

random effects 156, 160, 199
random number generator 11
random-effects models . . . . . *see* frailty models, shared
recurrent events . . . *see* multiple failures
`regress` 1, 233, 347
regression models
censored normal 2, 233
conditional logistic 4, 197
Cox 4, 6, 21–24, 128–229, 234–238, 241, 254, 345
power analysis for 359–360
sample size for . . *see* sample size, for Cox model
exponential . . . 20, 24, 247–256, 341
sample size for . . *see* sample size, for exponential survival
generalized gamma 276–278
Gompertz 234, 266–269
interval 233
logistic 3
loglogistic 273–275
lognormal 269–272
OLS 1–2, 21
piecewise constant . . . 254, 328, 331
piecewise exponential 264, 296, 328–331, 354
Weibull 234, 256–266
regression splines 185

relative risk model . . . . . *see* regression models, Cox
repeated failures . . *see* multiple failures
`reshape` 41, 45
residuals
  Cox–Snell residuals . . 219–222, 284, 294–295
  deviance residuals 212, 285
  efficient score residuals . . . 223–228
  martingale residuals . . 214–218, 285
  scaled Schoenfeld residuals . . . 206–209
  Schoenfeld residuals 206–209
restricted mean 119
Reye's syndrome data, description . . 213
right-censoring 30–31, 52, 57–58, 96–97, 245
right-truncation 36
risk score . . . . . *see* hazard, log relative

S

sample size
  for complex survival study . . . 354–359
  for Cox model 345–348
  for exponential survival . . 341–345, 353–354
  for the log-rank test 337–341, 352–353
  notion of 335
sampling
  complex survey 164–166, 198
  simple random 164
scaled Schoenfeld residuals 206–209
Schoenfeld residuals 206–209
semiparametric analysis . . 3–5, 128–175
  censoring 31–33
  likelihood function 21, 39
  truncation 35–36
significance level 333, 337–338
`simulate` 14
simulation 12
snapshot data 40–44
`snapspan` 43–46
Somers' D rank correlation 223
splines, cubic 185
stacked cumulative incidence plot . . 387–388
standard error
  adjusting for clustering . . 157–158, 200
  adjusting for survey design . . . 165, 167, 168
  delta method 132–133
  of cumulative hazard function
    log-transformed estimate 109
    Nelson–Aalen estimate 109
  of dispersion parameter 265
  of estimated coefficients . . 132–133, 161
  of hazard ratio 132–133
  of linear predictor 284
  of mean survival time 119
  of median survival time 118
  of survivor function
    Kaplan–Meier estimate 96
    loglog-transformed estimate . . 96
  robust 158, 279, 325–326
`stci` 117–122
`stcompet` 373
`stcox` 345, 389
  `efron` 151
  `exactm` 149, 193
  `exactp` 150, 193, 197
  `nohr` 130, 132, 133, 181
  `shared()` . . 159, 160, 162, 163, 199
  `strata()` 154, 155, 169, 200
  `tvc()` 191–193, 204–205
  `vce()` 158, 168, 198, 381
`stcoxkm` 209–212
`stcrreg` 382–388
  `compete()` 384
  `noshr` 384
`stcurve` . . . 142, 163, 295–300, 376–378
  `cif` 384–385
`stdescribe` 73, 77–78
`stfill` 73, 80–82
`stjoin` 195
`stpepemori` 373–375
`stphplot` 209–212

`stpower` 334–336, 338, 348–351, 360–362
`cox` 347–349, 351, 359
`exponential` . . . 342–345, 348–351, 353, 358, 363
dialog box of . . 355–358, 362–363
`logrank` 334, 337, 339, 340, 343–345, 347–353, 364
table 361–362
stratification
Cox models 152–156, 199–200
nonparametric tests 126–128
parametric models 307–310
survey data . . . . . *see* survey data, stratification
`streg`
`ancillary()` 301–306
`anc2()` 301
`dist()` 233
`frailty()` 320, 322, 326, 327
`nohr` 238, 247, 341
`strata()` 307–310
`time` 265
`vce()` 325
`streset` 73
`sts` 91
`sts generate` 172
`sts generate` 110, 139, 203, 221, 352
`sts graph` 102–107, 109, 122, 341
`sts list` 95, 99, 101–102, 108, 352
`sts test` 122–128, 214
`stset` 44, 47–71, 165, 167, 171, 197, 336, 341, 345, 352
analysis time 52–55
`enter()` 62–64
`exit()` 59–62
`failure()` 57–58
`id()` 65–66
`PROBABLE ERROR` 65, 73–76
`time0()` 67–70
variables defined by 55–56
`stsplit` . . . 193–197, 205, 249, 256, 260, 315, 328
`stvary` 73, 78–80
subhazard function 382–383
subjects-at-risk table . . *see* at-risk table
successive difference replication, variance estimation 165
survey data
clustering 164, 198
Cox model 164–169
stratification 164, 198
variance estimation 165
survival data
declaring . . . . . *see* `stset`
modeling 1–6, 91–93
power analysis for 333–364
recording 37–46
summaries 77–78
survivor function
baseline . . . 135, 137–139, 143, 155, 161–163
conditional 9, 12, 284
definition 7
estimation 93–101, 135, 155, 161–163, 284, 291–294
exponential 247, 255
generalized gamma 277
Gompertz 267
Kaplan–Meier estimate . . . . . *see* Kaplan–Meier estimate
loglogistic 273
lognormal 270
plot 102–107, 142, 295–300
relation to cumulative hazard 8
test 122–128, 214
Weibull 9, 257, 265
`svyset` 165–166, 198, 200
`svy: stcox` 166–168, 198, 200

T

test
likelihood-ratio . . . . . *see* likelihood-ratio test
log-rank . . . . . *see* log-rank test

test, *continued*
nonparametric stratification ..... *see* stratification, nonparametric tests
proportional-hazards assumption ..... *see* proportional-hazards assumption
Wald ..... *see* Wald test
Wilcoxon ..... *see* Wilcoxon test
tied failure times 33, 148–151
time at risk ..... *see* analysis time
time-varying coefficients 190
time-varying covariates 25, 84, 185, 189–197, 231, 241–244, 246, 248, 291, 300
`tobit` 233
truncation 34
interval ..... *see* gaps
left ..... *see* delayed entry
right 36
two-sided test 338, 340–341
type I error
probability of ..... *see* significance level
type II error
probability of 337–338, 360

U
uniform distribution 10, 33, 354

V
variables
categorical 178–180, 309
enduring 41, 47, 75
indicator ..... *see* categorical variables
instantaneous 41, 47, 54, 63

W
Wald test 133, 182, 258, 278–280, 303–305, 310, 317, 346, 359
Weibull
failure times
generation of 10–12
functions of 9, 257
mean and median of 16
Weibull, *continued*
regression model ..... *see* regression models, Weibull
Wilcoxon test 125–126
withdrawal 348–349, 355