

# Subject index

## A

Aalen–Nelson estimate ..... see  
Nelson–Aalen estimate  
accrual ..... 365–376  
    exponential ..... 367, 368, 371–375  
    period ..... 366–371, 373, 374  
    uniform ..... 367–371, 375  
additive-hazards model ..... 20  
adjusted predictions ..... see predictive  
    margins  
administrative censoring ..... 352, 356,  
    365, 366, 372  
AFT metric  
    definition ..... 17, 19, 20, 234,  
        241–243  
    relation to PH metric ..... 243–246  
    specification ..... 235  
Akaike information criterion ..... 280,  
    283–284, 333  
allocation of subjects ..... 350–351,  
    354–355, 357  
analysis time ..... 24–27, 52–55  
at-risk table ..... 105–107

## B

bathtub hazard ..... 8  
Bayesian information criterion ..... 284  
Bonferroni adjustment ..... 206  
bootstrap, variance estimation ..... 168  
boundary effects ..... see kernel  
    smoothing, boundary bias  
boundary kernels method ..... 115–116  
Breslow approximation for tied failures  
    ..... 152–153  
Breslow test ..... see Wilcoxon test  
BRR, variance estimation ..... 168

## C

cancer data ..... 173–174, 303  
casewise deletion .. see listwise deletion  
categorical variables ..... 180–182, 325  
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 ..... 160–161, 200–201,  
    342–343, 375  
competing risks ..... 381–408  
complete case analysis ..... see listwise  
    deletion  
concordance, measures ..... 224–225  
conditional logistic regression ..... see  
    regression models, conditional  
    logistic  
confidence bands, plot ..... 107  
confidence interval for  
    cumulative hazard function ... 109  
    estimated coefficients ..... 188  
    hazard ratio ..... 135  
    mean survival time ..... 120  
    median survival time ..... 119  
    survivor function ..... 96  
contingency table ..... 123–124  
Cox regression .. see regression models,  
    Cox

- Cox–Snell residuals . . . . . 221–224, 286, 296–298
- cubic splines . . . . . 187
- cumulative
- distribution function
    - conditional . . . . . 9
    - estimation . . . . . see empirical distribution function
    - generalized gamma . . . . . 278
    - loglogistic . . . . . 275
    - lognormal . . . . . 271
    - relation to cumulative hazard . . . . . 9
    - relation to survivor function . . . . . 7
    - standard normal . . . . . 271
    - Weibull . . . . . 9, 267
  - hazard count data interpretation . . . . . 14–15
  - hazard function
    - baseline . . . . . 137–139, 141, 145, 147, 157–158, 164, 175
    - conditional . . . . . 9
    - estimation . . . . . 107–113, 137–139, 157–158, 164, 293–296
    - exponential . . . . . 249
    - Gompertz . . . . . 269
    - goodness of fit . . . . . 221–224
    - Nelson–Aalen estimate . . . . . see Nelson–Aalen estimate
    - plot . . . . . 102, 144, 298–302
    - relation to survivor function . . . . . 8
    - Weibull . . . . . 259
  - incidence function . . . . . 383–384, 389–392, 399–407
  - subhazard function . . . . . 400
- D**
- date format . . . . . 48–50
- delayed entry . . . . . 12, 34–35, 38, 62–64, 87, 97–99, 104, 105, 247, 295
- generation of . . . . . 11–12
- delta method . . . . . 135
- density function . . . . . 16, 289, 290
- conditional . . . . . 9
  - for censored data . . . . . 31
  - relation to cumulative hazard . . . . . 9
- density function, *continued*
- relation to likelihood . . . . . 21, 247
  - relation to survivor function . . . . . 7
  - Weibull . . . . . 9
- deviance residuals . . . . . 214, 287
- DFBETA . . . . . see influence diagnostics, DFBETA
- diagnostic measures . . . . . 215, 225–230
- dispersion parameter . . . . . 267
- dropout of subjects . . . . . see loss to follow-up
- E**
- effect size . . . . . 351, 354, 364, 373, 376
- determination . . . . . 376–377
  - specification . . . . . 354, 373
- efficient score residuals . . . . . 225–230
- Efron approximation for tied failures . . . . . 153
- empirical distribution function . . . . . 99–101
- enduring variables . . . . . 41, 47, 75
- estat concordance . . . . . 224–225
- estat phtest . . . . . 209–211
- Euler’s constant . . . . . 257, 267, 290
- exponential regression . . . . . see regression models, exponential
- extended mean . . . . . 120
- extreme-value distribution . . . . . see Gumbel distribution
- F**
- failure function . . . . . see cumulative distribution function
- fixed effects . . . . . 201
- FMI . . . . . 177–178
- follow-up period . . . . . 29, 366–370, 373, 374, 376
- fp, stcox . . . . . 185–187
- fraction missing information . . . . . see FMI
- fractional polynomials . . . . . 184–187, 208
- multivariable . . . . . 187, 220–221
- frailty models
- shared . . . . . 158–167, 172, 201–202, 341–348
  - unshared . . . . . 327–341

**G**

- gamma function . . . . . 16, 257, 267, 278,  
289, 290
- gaps . . . . . 36, 38, 39,  
67–70, 77, 83, 87, 99–101, 247,  
286, 295, 300–302
- generalized gamma regression . . . . . see  
regression models, generalized  
gamma
- Gompertz regression . . . . . see regression  
models, Gompertz
- goodness of fit . . . . . 221–225
- Greenwood's standard error . . . . . 96
- Gumbel distribution . . . . . 257, 267

**H**

- Harrell's *C* statistic . . . . . 224
- hazard
  - cause-specific . . . 382–383, 386–389,  
392–399
  - contribution . . . 114, 137, 141, 143,  
145, 147
  - cumulative . . . . . see cumulative  
hazard function
  - function
    - baseline . . . . .  
. . . 19, 131–133, 136–137, 141–  
143, 155, 166–167, 236–238,  
241
    - bathtub . . . . . 8
    - conditional . . . . . 9
    - cumulative . . . . . see cumulative  
hazard function
    - definition . . . . . 7–8
    - estimation . . . 113–117, 137, 141–  
143, 166–167, 286, 293–296
    - examples of . . . . . 9
    - exponential . . . . . 249
    - Gompertz . . . . . 269
    - lognormal . . . . . 273
    - modeling . . . . . 19–24, 158–159,  
179–199, 317–348
    - plot . . . . . 102, 144, 166–167,  
298–302

hazard function, *continued*

- test . . . . . 125
    - Weibull . . . . . 9, 259
  - log relative . . . . . 137, 185–191
  - metric . . . . . see PH metric
  - rate . . . . . 8, 13, 15–16
  - ratio . . 132–136, 145, 163, 194, 236,  
349, 354, 355, 357, 361, 364,  
373, 376, 378, 380
    - confidence interval for . . . . . 135
    - estimation . . . . . 286
    - interpretation in frailty models . .  
. . . . . 337
    - standard error of . . . . . 134–135
    - test . . . . . 135, 356
  - relative . . . . . 137, 185
- heterogeneity
    - modeling . . . . . 326–328
    - test for . . . . . 333–341
  - hip-fracture data
    - description of . . . . . 83–89
    - sample size for . . . . . 349–351
  - hypothesis
    - alternative . . . . . 135, 362, 363
    - null . . 124, 125, 135, 161, 209, 354,  
356, 359, 362, 363

**I**

- imputation . . . . . see multiple imputation
  - Cox model
- incomplete gamma function . . . . . 278
- indicator variables . . . . . see categorical  
variables
- influence diagnostics
  - Cox–Snell residuals . . . . . see  
Cox–Snell residuals
  - deviance residuals . . . . . see deviance  
residuals
  - DFBETA . . . . . 225–227
  - efficient score residuals . . . . . see  
efficient score residuals
  - likelihood displacement values . . . .  
. . . . . 227–228
  - LMAX . . . . . 227–229
- instantaneous variables . . 41, 47, 54, 63

interaction terms.....188–191  
 interval regression..... see regression  
   models, interval  
 interval-censoring.....32–33  
**intreg**.....2, 235

**J**

jackknife, variance estimation..... 168

**K**

Kaplan–Meier estimate.....5, 93–  
 122, 126, 140–141, 143, 212–  
 214, 232, 358  
 kernel smoothing.... 113–117, 141–143  
   boundary bias..... 115–117  
 Kidney data..... 159–167, 328–332

**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.... 247–248  
   for truncated data..... 35, 36  
   partial..... see partial likelihood  
   penalized.. see penalized likelihood  
 likelihood-ratio test..... 162, 281–282,  
   337, 338, 344  
 linearization, variance estimation .. 168  
**linktest**..... 205–206  
 listwise deletion..... 172–173  
 LMAX.. see influence diagnostics, LMAX  
 log-rank test..... 5, 123–125, 216, 350,  
   354, 362, 364, 378  
   power of..... 355, 362, 380  
   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..... 365–366, 372–375,  
   see withdrawal

**lrtest**.....281–282  
 lung cancer data.....169

**M**

Mantel–Cox test..... see log-rank test  
 MAR..... 173–175  
 marginal effects.... 190, 192, 311–315  
 marginal predictions..... see predictive  
   margins  
**margins**..... 302–315  
**marginsplot**.....306–307  
 martingale residuals..... 216–220, 287  
 maximum likelihood estimation.. 20–21,  
   31–35, 38, 148–149, 233, 247–  
   248  
 maximum pseudolikelihood  
   estimators..... 168  
 MCAR..... 173–175  
 mean survival time  
   definition..... 16–17  
   estimation..... 91–92, 117–122,  
   286–291  
 median survival time  
   definition..... 16–17  
   estimation... 93, 117–119, 286–293  
**mfp**..... 187, 220  
**mi**..... 173–178  
   **estimate**..... 173, 176–178  
   **impute**..... 173–176  
   **register**..... 175  
   **set**..... 175  
 Mill’s ratio.....7  
 missing at random..... see MAR  
 missing completely at random..... see  
   MCAR  
 missing data .. see multiple imputation  
   Cox model  
 missing not at random..... see MNAR  
 MNAR..... 174  
 multiple failures... 59–62, 160–161, 332  
   generation of..... 12–13  
 multiple imputation Cox model... 172–  
   178

- multiple-myeloma data .. 352–353, 356, 358, 362, 376
- N**
- Nelson–Aalen estimate ..... 5, 107–113, 140–141, 143, 175, 222–224
- nested models ..... 280–283
- NHANES I ..... 169
- NHEFS ..... 169
- nonparametric analysis ..... 5, 91–129
- censoring ..... 96–97
- truncation ..... 97–101
- null hypothesis .... see hypothesis, null
- number of events .... 351–352, 355–357, 362–364, 367, 378
- number-at-risk table .. see at-risk table
- O**
- OLS regression .. see regression models, OLS
- one-sided test ..... 355, 357, 364, 376
- outliers, identification .... see influence diagnostics
- overdispersion ..... see heterogeneity
- P**
- parametric analysis ..... 2, 231–246
- censoring ..... 31–32
- likelihood function ..... 21, 31, 38
- truncation ..... 35–36
- partial likelihood ..... 148
- penalized likelihood ..... 164
- percentiles ..... 10, 118–119
- PH metric
- definition ..... 20, 234, 236–241
- relation to AFT metric .... 243–246
- specification ..... 235
- piecewise constant model .....  
    ..... see regression models,  
    piecewise constant
- piecewise exponential model .....  
    ..... see regression models,  
    piecewise exponential
- population-averaged predictions .... see predictive margins
- power ..... 351–352, 354, 362
- by simulation ..... 375
- curve ..... 377, 380
- definition ..... 354
- determination ..... 359, 376–377
- relation to number of events .. 352, 355
- power ..... 350–352, 355, 365–368, 377–378
- cox ..... 364–366, 368, 376
- exponential ... 358–362, 365–368, 371, 375, 380
- dialog box of .. 372–375, 378–379
- logrank ... 351, 354–357, 359, 360, 362, 364–370, 380
- table ..... 378
- predict ... 137–147, 157–158, 163–165, 285–298
- predictive margins ..... 302–311
- primary sampling units ..... 169
- probability integral transform ..... 11
- probability weights ..... 167
- product limit estimate ..... see Kaplan–Meier estimate
- proportional hazards regression ..... see regression models, Cox
- proportional-hazards assumption
- graphical assessment ..... 211–214
- test of ..... 209–211
- pweights ..... 168
- Q**
- quantile function ..... 9–13
- Weibull ..... 11
- R**
- random effects ..... 158, 163, 201
- random-effects models ..... see frailty models, shared
- random-number generator ..... 11
- recurrent events ... see multiple failures
- regress ..... 1, 235, 364

- regression models
- censored normal . . . . . 2, 235
  - conditional logistic . . . . . 4, 199
  - Cox . . . . . 4, 6, 21–24, 129–231, 236–240, 243, 256, 362
    - power analysis for . . . . . 376–377
    - sample size for . . . see sample size, for Cox model
  - exponential . . . 20, 24, 249–259, 358
    - sample size for . . . see sample size, for exponential survival
  - generalized gamma . . . . . 278–280
  - Gompertz . . . . . 236, 268–271
  - interval . . . . . 235
  - logistic . . . . . 3
  - loglogistic . . . . . 275–277
  - lognormal . . . . . 271–274
  - OLS . . . . . 1–2, 21
  - piecewise constant . . . 256, 345, 348
  - piecewise exponential . . . 266, 298, 345–348, 371
  - Weibull . . . . . 236, 259–268
- regression splines . . . . . 187
- relative risk model . . . . . see regression models, Cox
- repeated failures . . . see multiple failures
- reshape** . . . . . 41, 45
- residuals
- Cox–Snell residuals . . 221–224, 286, 296–298
  - deviance residuals . . . . . 214, 287
  - efficient score residuals . . . 225–230
  - martingale residuals . . 216–220, 287
  - scaled Schoenfeld residuals . . . 209–211
  - Schoenfeld residuals . . . . . 209–211
- restricted mean . . . . . 120
- Reye’s syndrome data, description . . 215
- right-censoring . . . . . 30–31, 52, 57–58, 96–97, 247
- right-truncation . . . . . 35–36
- risk score . . . . . see hazard, log relative
- S**
- sample size
- for complex survival study . . . 371–376
  - for Cox model . . . . . 362–365
  - for exponential survival . . 358–362, 370–371
  - for the log-rank test . . . . . 354–357, 369–370
  - notion of . . . . . 352
- sampling
- complex survey . . . . . 167–169, 200
  - simple random . . . . . 167
- scaled Schoenfeld residuals . . . . 209–211
- Schoenfeld residuals . . . . . 209–211
- semiparametric analysis . . 2–5, 129–178
  - censoring . . . . . 31–33
  - likelihood function . . . . . 21, 39
  - truncation . . . . . 35–36
- significance level . . . . . 349, 354–355
- simulate** . . . . . 14
- simulation . . . . . 13
- snapshot data . . . . . 40–44
- snapspan** . . . . . 43–46
- Somers’ *D* rank correlation . . . . . 225
- splines, cubic . . . . . 187
- stacked cumulative incidence plot . . 404–405
- standard error
- adjusting for clustering . . 160–161, 202
  - adjusting for survey design . . . 168, 170, 171
  - delta method . . . . . 134–135
  - of cumulative hazard function
    - log-transformed estimate . . . . 109
    - Nelson–Aalen estimate . . . . . 109
  - of dispersion parameter . . . . . 267
  - of estimated coefficients . . 134–135, 164
  - of hazard ratio . . . . . 134–135
  - of linear predictor . . . . . 286
  - of mean survival time . . . . . 120
  - of median survival time . . . . . 118

- standard error, *continued*  
of survivor function  
    Kaplan–Meier estimate . . . . . 96  
    loglog-transformed estimate . . 96  
    robust . . . . . 161, 281, 342–343
- stci . . . . . 118–122
- stcompete . . . . . 389–390
- stcox . . . . . 362, 406–407  
    efron . . . . . 153  
    exactm . . . . . 151, 195  
    exactp . . . . . 152, 195, 199  
    nohr . . . . . 132, 134, 135, 183  
    shared() . . 162, 163, 165, 166, 201  
    strata() . . . . . 156, 157, 172, 202  
    tvc() . . . . . 193–195, 206–207  
    vce() . . . . . 161, 171, 200, 399
- stcoxkm . . . . . 211–214
- stcrreg . . . . . 399–406  
    compete() . . . . . 401  
    nohr . . . . . 401
- stcurve . . 144, 166, 298–302, 393–395  
    cif . . . . . 402
- stdescribe . . . . . 73, 77–78
- stfill . . . . . 73, 81–83
- stjoin . . . . . 197
- stpepemori . . . . . 389–392
- stphplot . . . . . 211–214
- stratification  
    Cox models . . . . . 155–158, 201–202  
    nonparametric tests . . . . . 126–129  
    parametric models . . . . . 323–326  
    survey data . . . . . *see* survey data,  
    stratification
- streg  
    ancillary() . . . . . 317–322  
    anc2() . . . . . 317  
    dist() . . . . . 235  
    frailty() . . . . . 336, 338, 343, 344  
    nohr . . . . . 240, 249, 358  
    strata() . . . . . 323–326  
    time . . . . . 267  
    vce() . . . . . 342
- streset . . . . . 73
- sts . . . . . 91  
    generate . . . . . 110, 141, 175, 205,  
    223, 369  
    graph . . . . . 102–107, 109, 122, 358  
    list . . . . . 95, 99, 101–102, 109, 370  
    test . . . . . 122–129, 216
- stset . . . 44, 47–72, 168, 170, 174, 199,  
    353, 358, 362, 369  
    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 . . 195–199, 207, 251, 258, 262,  
    331, 345
- stvary . . . . . 73, 78–81
- subhazard function . . . . . 400
- subjects-at-risk table . . *see* at-risk table
- successive difference replication,  
    variance estimation . . . . . 168
- survey data  
    clustering . . . . . 167, 200  
    Cox model . . . . . 167–172  
    stratification . . . . . 167, 200  
    variance estimation . . . . . 168
- survival data  
    declaring . . . . . *see* stset  
    modeling . . . . . 1–6, 91–93  
    power analysis for . . . . . 349–380  
    recording . . . . . 37–46  
    summaries . . . . . 77–78
- survivor function  
    baseline . . 137, 139–141, 145, 157,  
    164–166  
    conditional . . . . . 9, 12, 286  
    definition . . . . . 7  
    estimation . . . . . 93–101, 137, 157,  
    164–166, 286, 293–296  
    exponential . . . . . 249, 257  
    generalized gamma . . . . . 279  
    Gompertz . . . . . 269

survivor function, *continued*  
 Kaplan–Meier estimate . . . . . see  
   Kaplan–Meier estimate  
 loglogistic . . . . . 275  
 lognormal . . . . . 272  
 plot . . . . . 102–107, 144, 298–302  
 relation to cumulative hazard . . . . 8  
 test . . . . . 122–129, 216  
 Weibull . . . . . 9, 259, 267  
 svyset . . . . . 168–169, 200, 202  
 svy: stcox . . . . . 169–171, 200, 202

**T**

test  
 likelihood-ratio . . . . . see  
   likelihood-ratio test  
 log-rank . . . . . see log-rank test  
 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, 150–153  
 time at risk . . . . . see analysis time  
 time-varying coefficients . . . . . 192  
 time-varying covariates . . . . .  
   . . . . . 25, 85, 187, 191–199, 233,  
   243–246, 248, 250, 293, 302  
 tobit . . . . . 235  
 truncation . . . . . 34  
   left . . . . . see delayed entry  
   right . . . . . 35–36  
 two-sided test . . . . . 355, 357  
 type I error, probability of . . . . . see  
   significance level  
 type II error, probability of . . . 354–355,  
   377

**U**

uniform distribution . . . . . 11, 33, 372

**V**

variables  
 categorical . . . . . 180–182, 325  
 enduring . . . . . 41, 47, 75  
 indicator . . . . . see categorical variables  
 instantaneous . . . . . 41, 47, 54, 63

**W**

Wald test . . . . . 135, 184, 260, 280–282,  
   319–321, 326, 333, 363, 376  
 Weibull  
   failure times  
     generation of . . . . . 11–13  
     functions of . . . . . 9, 259  
     mean and median of . . . . . 16  
     regression model . . . . . see regression  
     models, Weibull  
 Wilcoxon test . . . . . 125–126  
 withdrawal . . . . . 365–366, 372