

# Subject index

## A

Aalen–Nelson estimate . . . . . see  
Nelson–Aalen estimate

accrual . . . . . 340–350  
    exponential . . . . . 342, 343, 346–350  
    period . . . . . 341–346, 348, 349  
    uniform . . . . . 342–346, 350

additive-hazards model . . . . . 20

administrative censoring . . . . . 328, 331,  
    340, 341, 347

AFT metric  
    definition . . . . . 16, 19, 20, 224,  
        231–233  
    relation to PH metric . . . . . 233–236  
    specification . . . . . 225

Akaike information criterion . . . . . 270,  
    273–274, 309

allocation of subjects . . . . . 326–327, 330,  
    332

analysis time . . . . . 24–27, 52–55

at-risk table . . . . . 105–107

## B

bathtub hazard . . . . . 8

Bayesian information criterion . . . . . 274

Bonferroni adjustment . . . . . 198

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

categorical variables . . . . . 172–174, 301

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, 192–193,  
    317–318, 350

cnreg . . . . . 2, 225

concordance, measures . . . . . 216–217

conditional logistic regression . . . . . see  
    regression models, conditional  
    logistic

confidence bands, plot . . . . . 107

confidence interval for  
    cumulative hazard function . . . . . 109  
    estimated coefficients . . . . . 180  
    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 . . . . . 213–216, 276,  
    286–288

cubic splines . . . . . 179

cumulative  
  distribution function  
    conditional . . . . . 9  
    estimation . . . . . *see* empirical  
      distribution function  
    generalized gamma . . . . . 268  
    loglogistic . . . . . 265  
    lognormal . . . . . 261  
    relation to cumulative hazard . . . 9  
    relation to survivor function . . . 7  
    standard normal . . . . . 261  
    Weibull . . . . . 9, 257

hazard count data interpretation . .  
  . . . . . 13–15

hazard function  
  baseline . . . . . 135–137, 139, 143,  
  145, 155–156, 161  
  conditional . . . . . 9  
  estimation . . . . . 107–113, 135–137,  
  155–156, 161, 283–286  
  exponential . . . . . 239  
  Gompertz . . . . . 259  
  goodness of fit . . . . . 213–216  
  Nelson–Aalen estimate . . . . . *see*  
  Nelson–Aalen estimate  
  plot . . . . . 102, 142, 288–292  
  relation to survivor function . . . 8  
  Weibull . . . . . 249

**D**

date format . . . . . 48–50

delayed entry . . . . . 12, 35, 38, 62–64, 86,  
  97–99, 104, 105, 237, 285  
  generation of . . . . . 11–12

delta method . . . . . 133

density function . . . . . 16, 279, 280  
  conditional . . . . . 9  
  for censored data . . . . . 31  
  relation to cumulative hazard . . . 9  
  relation to likelihood . . . . . 21, 237  
  relation to survivor function . . . 7  
  Weibull . . . . . 9

deviance residuals . . . . . 206, 277

dispersion parameter . . . . . 257

dropout of subjects . . . . . *see* loss to  
  follow-up

**E**

effect size . . . . . 327, 329, 339, 348, 351  
  determination . . . . . 351–352  
  specification . . . . . 330, 348

efficient score residuals . . . . . 217–219

Efron approximation for tied failures . .  
  . . . . . 151

empirical distribution function . . . 99–101

enduring variables . . . . . 41, 47, 75

**estat concordance** . . . . . 216–217

**estat phtest** . . . . . 200–203

Euler’s constant . . . . . 247, 257, 280

exponential regression . . . *see* regression  
  models, exponential

extended mean . . . . . 120

extreme-value distribution . . . . . *see*  
  Gumbel distribution

**F**

failure function . . . . . *see* cumulative  
  distribution function

fixed effects . . . . . 193

follow-up period . . . . . 29, 341–345, 348,  
  349, 351

**fracpoly, stcox** . . . . . 177–179

fractional polynomials . . . . . 176–179, 200  
  multivariable . . . . . 179, 212–213

frailty models  
  shared . . . . . 156–164, 169, 193–194,  
  316–323  
  unshared . . . . . 303–316

**G**

gamma function . . . . . 16, 247, 257, 268,  
  279, 280

gaps . . . . . 36, 38, 39, 67–70, 77, 83,  
  86–87, 99–101, 237, 276, 285,  
  290–292

generalized gamma regression . . . . . *see*  
  regression models, generalized  
  gamma

Gompertz regression . . . . . see regression models, Gompertz  
 goodness of fit . . . . . 213–217  
 Greenwood's standard error . . . . . 96  
 Gumbel distribution . . . . . 247, 257

**H**

Harrell's *C* statistic . . . . . 216  
 hazard  
   contribution . . . 113, 135, 139, 141, 143, 145, 163  
   cumulative . . . . . see cumulative hazard function  
   function  
     baseline . . . . .  
       . . . 19, 129–131, 134–135, 139–141, 152, 163–164, 226–228, 231  
     bathtub . . . . . 8  
     conditional . . . . . 9  
     cumulative . . . . . see cumulative hazard function  
     definition . . . . . 7–8  
     estimation . . . 113–117, 135, 139–141, 163–164, 276, 283–286  
     examples of . . . . . 9  
     exponential . . . . . 239  
     Gompertz . . . . . 259  
     lognormal . . . . . 263  
     modeling . . 19–24, 156, 171–191, 293–323  
     plot . . . . . 102, 142, 163–164, 288–292  
     test . . . . . 124  
     Weibull . . . . . 9, 249  
   log relative . . . . . 135, 177–183  
   metric . . . . . see PH metric  
   rate . . . . . 8, 13, 15–16  
   ratio . . 130–134, 143, 160, 186, 226, 325, 330, 332, 336, 339, 348, 351, 353, 356  
   confidence interval for . . . . . 133  
   estimation . . . . . 276  
   interpretation in frailty models . . . . . 313  
   standard error of . . . . . 132–133

hazard ratio, *continued*  
   test . . . . . 133, 331  
   relative . . . . . 135, 177  
 heterogeneity  
   modeling . . . . . 302–304  
   test for . . . . . 309–316  
 hip-fracture data  
   description of . . . . . 82–89  
   sample size for . . . . . 325–327  
 hypothesis  
   alternative . . . . . 133, 337, 338  
   null . . 124, 133, 158, 201, 329, 331, 335, 337, 338

**I**

incomplete gamma function . . . . . 268  
 indicator variables . . . . . see categorical variables  
 influence diagnostics  
   Cox–Snell residuals . . see Cox–Snell residuals  
   deviance residuals . . . . . see deviance residuals  
   “dfbeta” . . . . . 217–219  
   efficient score residuals . . . . . see efficient score residuals  
 instantaneous variables . . 41, 47, 54, 63  
 interaction terms . . . . . 180–183  
 interval censoring . . . . . 32–33  
 interval regression . . . . . see regression models, interval  
 interval truncation . . . . . see gaps  
*intreg* . . . . . 225

**J**

jackknife, variance estimation . . . . . 165

**K**

Kaplan–Meier estimate . . . . . 5, 93–122, 126, 138–139, 141, 204–206, 222, 334  
 kernel smoothing . . . 113–117, 139–141  
   boundary bias . . . . . 115–116  
 Kidney data . . . . . 156–164, 304–308

**L**

left censoring ..... 34  
 left truncation ..... see delayed entry  
 likelihood function  
     for censored data ..... 31, 32  
     for parametric models .... 237–238  
     for truncated data ..... 35, 36  
     partial ..... see partial likelihood  
     penalized..see penalized likelihood  
 likelihood-ratio test ..... 159, 271–272,  
     313, 314, 319  
 linearization, variance estimation .. 165  
**linktest** ..... 197–198  
 log-rank test ..... 5, 123–124, 208, 326,  
     329, 337, 339, 353  
     power of ..... 330, 337, 356  
     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 ..... 340–341, 347–350,  
     see withdrawal  
**lrtest** ..... 271–272  
 lung cancer data ..... 166

**M**

Mantel–Cox test ..... see log-rank test  
 marginal effects ..... 182, 184  
 martingale residuals ..... 208–212, 277  
 maximum likelihood estimation .. 20–21,  
     31–35, 38, 146–147, 223, 237–  
     238  
 maximum pseudolikelihood estimators ..  
     ..... 165  
 mean survival time  
     definition ..... 16–17  
     estimation ... 91–92, 117, 119–122,  
         276–281  
 median survival time  
     definition ..... 16–17  
     estimation ... 93, 117–119, 276–283  
**mfp** ..... 179, 212  
 Mill's ratio ..... 7

**mkmat** ..... 218  
 multiple failures ... 59–62, 157–158, 308  
     generation of ..... 12  
 multiple-myeloma data .. 328–329, 331,  
     333, 334, 337, 351

**N**

Nelson–Aalen estimate ..... 5, 107–113,  
     138–139, 141, 214–216  
 nested models ..... 270–273  
 NHANES I ..... 166  
 NHEFS ..... 166  
 nonparametric analysis ..... 5, 91–128  
     censoring ..... 96–97  
     truncation ..... 97–101  
 null hypothesis .... see hypothesis, null  
 number of events ... 326–327, 330, 332,  
     337–339, 342, 353  
 number-at-risk table .. see at-risk table

**O**

OLS regression .. see regression models,  
     OLS  
 one-sided test ... 330, 332–333, 339, 351  
 outliers, identification .... see influence  
     diagnostics  
 overdispersion ..... see heterogeneity

**P**

parametric analysis ..... 2, 221–236  
     censoring ..... 31–32  
     likelihood function ..... 21, 31, 38  
     truncation ..... 35–36  
 partial likelihood ..... 146  
 penalized likelihood ..... 160–161  
 percentiles ..... 10, 117–119  
 PH metric  
     definition ..... 20, 224–231  
     relation to AFT metric ... 233–236  
     specification ..... 225  
 piecewise constant model .....  
     ..... see regression models,  
         piecewise constant

- piecewise exponential model . . . . .  
     . . . . . see regression models,  
     piecewise exponential  
 power . . . . . 327, 329, 330, 337  
     by simulation . . . . . 350  
     curve . . . . . 353, 356  
     definition . . . . . 329  
     determination . . . . . 335, 351–352  
     relation to number of events . . 327,  
     330  
 predict . . . . . 275–288  
 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 . . . . . 203–206  
     test of . . . . . 200–203  
 pweights . . . . . 165
- Q**
- quantile function . . . . . 9–12  
     Weibull . . . . . 11
- R**
- random effects . . . . . 156, 160, 193  
 random number generator . . . . . 11  
 random-effects models . . . . . see frailty  
     models, shared  
 recurrent events . . . . . see multiple failures  
**regress** . . . . . 1, 225, 339  
 regression models  
     censored normal . . . . . 2, 225  
     conditional logistic . . . . . 4, 191  
     Cox . . . . . 4, 6, 21–24, 128–219, 221,  
     226–230, 233, 246, 337  
     power analysis for . . . . . 351–352  
     sample size for . . . . . see sample size,  
     for Cox model  
     exponential . . . . . 20, 24, 239–248, 333  
     sample size for . . . . . see sample size,  
     for exponential survival  
     regression models, *continued*  
         generalized gamma . . . . . 268–270  
         Gompertz . . . . . 226, 258–261  
         interval . . . . . 225  
         logistic . . . . . 3  
         loglogistic . . . . . 265–267  
         lognormal . . . . . 261–264  
         OLS . . . . . 1–2, 21  
         piecewise constant . . . . . 246, 320, 323  
         piecewise exponential . . . . . 256, 288,  
         320–323, 346  
         Weibull . . . . . 226, 248–258  
     regression splines . . . . . 179  
     relative risk model . . . . . see regression  
     models, Cox  
     repeated failures . . . . . see multiple failures  
**reshape** . . . . . 41, 45  
 residuals  
     Cox–Snell residuals . . . . . 213–216, 276,  
     286–288  
     deviance residuals . . . . . 206, 277  
     efficient score residuals . . . . . 217–219  
     martingale residuals . . . . . 208–212, 277  
     scaled Schoenfeld residuals . . . . . 200–  
     203  
     Schoenfeld residuals . . . . . 200–203  
     restricted mean . . . . . 119  
     Reye’s syndrome data, description . . . . . 207  
     right censoring . . . . . 30–31, 52, 57–58,  
     96–97, 237  
     right truncation . . . . . 36  
     risk score . . . . . see hazard, log relative
- S**
- sample size  
     for complex survival study . . . . . 346–  
     351  
     for Cox model . . . . . 337–340  
     for exponential survival . . . . . 333–337,  
     345–346  
     for the log-rank test . . . . . 329–333,  
     344–345  
     notion of . . . . . 327

- sampling
  - complex survey . . . . . 164–166, 192
  - simple random . . . . . 164
- scaled Schoenfeld residuals . . . . . 200–203
- Schoenfeld residuals . . . . . 200–203
- semiparametric analysis . . . . . 3–5, 128–169
  - censoring . . . . . 31–33
  - likelihood function . . . . . 21, 39
  - truncation . . . . . 35–36
- significance level . . . . . 325, 329–330
- simulate . . . . . 14
- simulation . . . . . 12
- snapshot data . . . . . 40–44
- snapspan . . . . . 43–46
- Somers' *D* rank correlation . . . . . 217
- splines, cubic . . . . . 179
- standard error
  - adjusting for clustering . . . . . 157–158, 194
  - 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 . . . . . 257
  - of estimated coefficients . . . . . 132–133, 161
  - of hazard ratio . . . . . 132–133
  - of linear predictor . . . . . 276
  - of mean survival time . . . . . 119
  - of median survival time . . . . . 118
  - of survivor function
    - Kaplan–Meier estimate . . . . . 96
    - loglog-transformed estimate . . . . . 96
  - robust . . . . . 158, 271, 317–318
- stci . . . . . 117–122
- stcox . . . . . 337
  - basechazard() . . . . . 155
  - basesurv() . . . . . 155
  - efron . . . . . 151
  - exactm . . . . . 149, 187
  - exactp . . . . . 150, 187, 191
  - nohr . . . . . 130, 132, 133, 175
  - shared() . . . . . 159, 160, 162, 164, 193
- stcox, *continued*
  - strata() . . . . . 154, 155, 169, 194
  - tvf() . . . . . 185–187, 198–199
  - vce() . . . . . 158, 168, 192
- stcoxkm . . . . . 203–206
- stcurve . . . . . 142, 163, 288–292
- stdescribe . . . . . 73, 77–78
- stfill . . . . . 73, 80–82
- stjoin . . . . . 189
- stphplot . . . . . 203–206
- stpwr . . . . . 326–328, 330, 340–343, 352–354
  - stpwr cox . . . . . 339–341, 343, 351
  - stpwr exponential . . . . . 334–337, 340–343, 345, 350, 355
    - dialog box of . . . . . 347–350, 354–355
  - stpwr logrank . . . . . 326, 329, 331, 332, 335–337, 339–345, 356
  - table . . . . . 353–354
- stratification
  - Cox models . . . . . 152–156, 193–194
  - nonparametric tests . . . . . 126–128
  - parametric models . . . . . 299–302
  - survey data . . . . . see survey data, stratification
- streg
  - ancillary() . . . . . 293–298
  - anc2() . . . . . 293
  - dist() . . . . . 225
  - frailty() . . . . . 312, 314, 318, 319
  - nohr . . . . . 230, 239, 333
  - strata() . . . . . 299–302
  - time . . . . . 257
  - vce() . . . . . 317
- streset . . . . . 73
- sts . . . . . 91
  - sts generate . . . . . 110, 139, 197, 215, 344
  - sts graph . . . . . 102–107, 109, 122, 333
  - sts list . . . . . 95, 99, 101–102, 109, 344
  - sts test . . . . . 122–128, 208
- stset . . . . . 44, 47–71, 165, 167, 191, 328, 333, 337, 344

- stset*, *continued*
- 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`... 187–191, 199, 241, 248, 252, 307, 320
- `stvary` ..... 73, 78–80
- subjects-at-risk table .. see at-risk table
- survey data
- clustering ..... 164, 192
  - Cox model ..... 164–169
  - stratification ..... 164, 192
  - variance estimation ..... 165
- survival data
- declaring ..... see `stset`
  - modeling ..... 1–6, 91–93
  - power analysis for ..... 325–356
  - recording ..... 37–46
  - summaries ..... 77–78
- survivor function
- baseline ... 135, 137–139, 143, 155, 161–163
  - conditional ..... 9, 12, 276
  - definition ..... 7
  - estimation ..... 93–101, 135, 155, 161–163, 276, 283–286
  - exponential ..... 239, 247
  - generalized gamma ..... 269
  - Gompertz ..... 259
  - Kaplan–Meier estimate ..... see Kaplan–Meier estimate
  - loglogistic ..... 265
  - lognormal ..... 262
  - plot ..... 102–107, 142, 288–292
  - relation to cumulative hazard ... 8
  - test ..... 122–128, 208
  - Weibull ..... 9, 249, 257
- `svyset` ..... 165–166, 192, 194
- `svy: stcox` ..... 166–168, 192, 194
- 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, 148–151
- time at risk ..... see analysis time
- time-varying coefficients ..... 184
- time-varying covariates ..... 25, 84, 179, 183–191, 223, 233–236, 238, 240, 283, 292
- `tobit` ..... 225
- truncation ..... 34
- interval ..... see gaps
  - left ..... see delayed entry
  - right ..... 36
- two-sided test ..... 330, 332–333
- type I error
- probability of ..... see significance level
- type II error
- probability of ..... 329–330, 352
- U**
- uniform distribution ..... 10, 33, 346
- V**
- variables
- categorical ..... 172–174, 301
  - enduring ..... 41, 47, 75
  - indicator .. see categorical variables
  - instantaneous ..... 41, 47, 54, 63
- W**
- Wald test ..... 133, 176, 250, 270–272, 295–297, 302, 309, 338, 351

## Weibull

- failure times
  - generation of . . . . . 10–12
  - functions of . . . . . 9, 249
  - mean and median of . . . . . 16
  - regression model . . . see regression models, Weibull
- Wilcoxon test . . . . . 125–126
- withdrawal . . . . . 340–341, 347

**X**

- xi . . . . . 173, 302