

# Subject index

## A

abbreviations ..... 35  
ado-files ..... 288–291  
ado-path ..... 6  
adoupdate ..... 20  
analysis of variance ..... 139–140  
append ..... 83–84  
archiving data ..... 110–111  
arithmetic operators ..... 71–72  
audit trail ..... 93–94  
axes (graphs) ..... 237–241  
axis labels (graphs) ..... 237–238  
axis ticks (graphs) ..... 237–238

## B

backing up data ..... 108  
bar graphs ..... 264–269  
baselevels (option) ..... 152  
binreg ..... 167  
Bland–Altman plot ..... 199–201  
box-and-whisker plot ..... 251–252  
browse ..... 120  
by-graphs ..... 269–270  
by: (prefix) ..... 33

## C

c() (system settings and constants) ... 280  
capture ..... 220  
case–control data ..... 128–133  
    matched ..... 131–133, 167  
categorical predictors ..... 150–162  
cc ..... 128–130  
cd ..... 56  
centile ..... 135  
chi-squared test ..... 121  
ci ..... 141  
cii ..... 143

cisd ..... 146  
cisd1 ..... 204  
classification of diseases ..... 225  
clogit ..... 167  
clustered observations ..... 168–169  
codebook ..... 96–97  
codebook (command) ..... 117  
cohort data without censorings ... 125–127  
collapse ..... 90–91  
colors (graphs) ..... 247  
combining files ..... 83–87  
combining graphs ..... 270–271  
Command window ..... 10  
comments in do-files and ado-files .... 34  
comparing measurements ..... 199–202  
complex strings ..... 53–54  
compress ..... 54  
conditional logistic regression ..... 167  
confidence intervals ..... 141  
confidentiality ..... 108–109  
contract ..... 92  
copying graphs ..... 273  
copying tables to documents ..... 17  
correcting errors ..... 102–103  
count ..... 123  
Cox regression ..... 181–192  
    checking proportionality ... 185–186  
    delayed entry ..... 186  
    stratification ..... 184–185  
    time-varying coefficients ... 187–192  
    time-varying covariates ..... 187–192  
creturn list ..... 280  
cs ..... 125–127  
cut() (egen function) ..... 74–75

## D

Data Editor ..... 12–13, 59–60

data entry ..... 59–60, 97–98  
 data protection ..... 108–109  
 dataset label ..... 68  
 date formats ..... 47  
 date variables ..... 47–48  
 db ..... 14  
 debugging programs ..... 292  
 decimal periods and commas ..... 221–222  
 describe ..... 116  
 destring ..... 51  
 diagnostic tests ..... 206–212  
 diagt ..... 208–209  
 dialogs ..... 14, 21–22  
 direct standardization ..... 198  
 disease classification ..... 225  
 display ..... 144  
 do ..... 15  
*Documents* folder ..... 3  
 do-file ..... 14–15  
 Do-file Editor ..... 13  
 dotplot ..... 252–253  
 drop ..... 81

**E**

e() (returned results) ..... 278–280  
 egen ..... 73–75  
 encode ..... 52  
 encryption ..... 109  
 entering data ..... 59–60, 97–98  
 EpiData (program for entering data) ..... 60  
 epitab (family of commands) ..... 125–133  
 equivalence studies ..... 218  
 ereturn list ..... 278–279  
 error correction ..... 102–103  
 error finding ..... 101–102  
 error messages ..... 36–38  
 estat gof ..... 167  
 estat phtest ..... 186  
 Excel (Microsoft) ..... 62  
 exercises ..... 297–308  
 exit ..... 8  
 exlogistic ..... 167  
 expand ..... 92  
 explicit subscripting ..... 78–79  
 export delimited ..... 61

export excel ..... 62  
 exporting graphs ..... 273

**F**

factor variables ..... 150–162  
 FAQs ..... 23  
 file path ..... 56  
 file types ..... 25–26  
 filenames ..... 95  
 Fisher's exact test ..... 121  
 foreach ..... 284–286  
 format ..... 41–42  
 formats  
     date ..... 47  
     numeric ..... 41–42  
     string ..... 51  
 forvalues ..... 284  
 function plots ..... 262–263  
 functions  
     mathematical ..... 72  
     statistical ..... 72

**G**

generate ..... 69–71  
 gold standard ..... 206  
 graph area ..... 230  
 graph bar ..... 264–269  
 graph box ..... 251–252  
 graph combine ..... 270–271  
 graph command syntax ..... 231–232  
 graph display ..... 273  
 graph export ..... 273  
 graph matrix ..... 271–272  
 graph options ..... 232, 237–249  
 graph save ..... 273  
 graph schemes ..... 235–236  
 graph size ..... 232–234  
 graph use ..... 273  
 Graph window ..... 229–230  
 graphs ..... 229–273  
 grid lines (graphs) ..... 237–238  
 group() (egen function) ..... 74  
 groups ..... 124–125

**H**

help ..... 19–20  
 histogram ..... 249–251  
 Hosmer–Lemeshow test ..... 167

**I**

icd9 ..... 225  
 if (command) ..... 286–287  
 if (qualifier) ..... 30–31  
 immediate commands ..... 142–144  
 import delimited ..... 61  
 import excel ..... 62  
 imputation ..... 43  
 in (qualifier) ..... 31  
 indirect standardization ..... 195–198  
 infile ..... 61  
 infix ..... 62  
 input ..... 59  
 installing Stata ..... 3  
 interactions ..... 155–162  
 intraclass correlation coefficient ..... 205–206  
 ir ..... 179

**J**

joinby ..... 92

**K**

Kaplan–Meier curve ..... 175–178  
 keep ..... 81  
 kernel density curve ..... 250–251

**L**

label data ..... 68  
 label define ..... 65–67  
 label values ..... 65–67  
 label variable ..... 65  
 labels ..... 65–68  
 lag and lead functions ..... 77  
 legend (graph option) ..... 242–243  
 level() (option) ..... 141  
 lincom ..... 149–150  
 line patterns (graphs) ..... 248–249  
 line plots ..... 257–262  
 linear regression ..... 145–147  
 list ..... 118–120

local ..... 282–283  
 log book ..... 99  
 log file ..... 15–17, 223–224  
 log-rank test ..... 178  
 logical expressions ..... 70–71  
 logical operators ..... 30–31  
 logistic ..... 162–167  
 logistic regression ..... 162–167  
 logit ..... 162–167  
 log-scaled axes (graphs) ..... 239–240  
 long command lines ..... 34–35  
 longitudinal data ..... 225  
 lrtest ..... 167, 184  
 lvr2plot ..... 149

**M**

macros ..... 282–283  
     global ..... 283  
     local ..... 282–283  
 manuals ..... 23, 295–296  
 margins ..... 150  
 marker symbols (graphs) ..... 247–248  
 master do-file ..... 100  
 matched case–control data ..... 131–133, 167  
 matching datasets ..... 84–87  
 mathematical functions ..... 72  
 matrix graphs ..... 271–272  
 mcc ..... 131–133  
 mdy() (date function) ..... 47  
 measurement comparison ..... 199–202  
 measurement, reproducibility ..... 203–206  
 measurement variation ..... 203–204  
 memory considerations ..... 54  
 merge ..... 84–87  
 merging datasets ..... 84–87  
 meta-analysis ..... 225  
 mi ..... 43  
 missing (option) ..... 33  
 missing values ..... 42–44, 103  
 —more— ..... 10  
 multiple axes (graphs) ..... 240–241  
 multiple imputation ..... 43  
 mvencode ..... 43

- N**
- `_N` ..... 77
  - `_n` ..... 77
  - NetCourses ..... 23
  - `noconstant` (option) ..... 154
  - `nofvlabel` (option) ..... 153
  - `nolabel` (option) ..... 33
  - noninferiority studies ..... 218–220
  - nonparametric tests ..... 140
  - normal curve ..... 250–251
  - notes ..... 68
  - numbering observations ..... 77–78
  - numeric
    - formats ..... 41–42
    - lists ..... 29
    - ranges ..... 30
    - variables ..... 41–50
  - `numlabel` ..... 67
  - `numlist` ..... 29
- O**
- `oneway` ..... 139–140
  - online help ..... 19–22
  - operators
    - arithmetic ..... 71–72
    - logical ..... 30–31
    - relational ..... 30–31
  - options ..... 32–33
  - order ..... 82
  - output log ..... 15–17
- P**
- panel data ..... 225
  - PDF documentation ..... 23
  - pharmacokinetic data ..... 225
  - `pk` (family of commands) ..... 225
  - plot area ..... 230
  - `poisson` ..... 192–195
  - Poisson regression ..... 192–195
  - postestimation, regression ..... 148–150
  - power ..... 215–217
  - power estimation ..... 215–220
  - precision ..... 44–46
  - `predict` ..... 148
  - predictive values ..... 207–209
  - preferences, window ..... 8–13
  - prefixes ..... 33
  - `preserve` ..... 221
  - `probit` ..... 167
  - programs ..... 288–291
  - Properties window ..... 11
  - proportional hazards regression .. 181–192
  - protecting data ..... 108–109
  - `pwcompare` ..... 154
- Q**
- `qnorm` ..... 133
  - Q–Q plot ..... 133
  - qualifiers ..... 30–31
  - `quietly` ..... 220
  - quotes
    - double ..... 34
    - single ..... 282
- R**
- `r()` (returned results) ..... 277–278
  - random-effects models ..... 168
  - random numbers ..... 213
  - random sampling ..... 81, 213–214
  - randomization ..... 213
  - range plots ..... 263
  - `ranksum` ..... 140
  - rate ratio ..... 179
  - rates ..... 178–195
    - Poisson regression ..... 192–195
    - tabulating ..... 178
  - receiver operating characteristic .. 209–212
  - `recode` ..... 75–76
  - `regress` ..... 146–147
  - regression
    - postestimation ..... 148–150
    - regression analysis ..... 145–169
  - relational operators ..... 30–31
  - `rename` ..... 82
  - reordering variables ..... 82
  - `replace` ..... 69–71
  - reproducibility of measurements ..... 203–206
  - `reshape` ..... 87–89
  - reshaping data ..... 87–92

residuals ..... 148  
 restore ..... 221  
 Results window ..... 10–11  
 return list ..... 277–278  
 Review window ..... 11  
 risktable (option) ..... 176  
 rnormal() (function) ..... 213  
 robust standard errors ..... 168–169  
 ROC analysis ..... 209–212  
 roctab ..... 211–212  
 runiform() (function) ..... 213  
 rvfplot ..... 149  
 rvpplot ..... 149

**S**

sample ..... 81, 213  
 sample-size estimation ..... 215–220  
 save ..... 55–58  
 saveold ..... 58  
 saving graphs ..... 272–273  
 scalar ..... 283–284  
 scalars ..... 283–284  
 scatterplots ..... 253–257  
 schemes (graphs) ..... 235–236  
 sdtest ..... 139  
 search ..... 20–21  
 selecting observations ..... 81  
 selecting variables ..... 81  
 sensitivity ..... 207–209  
 separate ..... 92  
 set trace ..... 292  
 simulations ..... 214, 218–220  
 single quotes ..... 282  
 slist ..... 119  
 SMR ..... 195–198  
 sort ..... 82  
 specificity ..... 207–209  
 SSC archive ..... 20  
 standard populations ..... 198  
 standardization ..... 195–198  
   direct ..... 198  
   indirect ..... 195–198  
 standardized mortality ratio ..... 195–198  
 starting Stata ..... 7  
 Stata manuals ..... 23, 295–296

Stata program files ..... 5–6  
 Statalist ..... 24  
 statistical functions ..... 72  
 statsby: (prefix) ..... 280–281  
 Stat/Transfer ..... 63  
 stci ..... 177  
 stcox ..... 181–192  
 stcurve ..... 183  
 storage types ..... 44–46  
 stored results ..... 277–281  
 stphplot ..... 185  
 stpower ..... 218  
 stptime ..... 178  
 strate ..... 197  
 stratified analysis  
   case-control studies ..... 128–133  
   cohort studies ..... 125–127  
   Cox regression ..... 184–185  
   incidence-rate data ..... 179–180  
 string() (function) ..... 52  
 string formats ..... 51  
 string variables ..... 50–54  
 sts graph ..... 175, 176  
 sts list ..... 177  
 sts test ..... 178  
 stset ..... 173–175, 186  
 stsplot ..... 187  
 subscripting ..... 78–79  
 substr() (string function) ..... 52  
 summarize ..... 117  
 survey data ..... 225  
 survival analysis ..... 171–195  
 syntax ..... 288  
 syntax diagrams ..... 27–28  
 sysdir ..... 6  
 sysuse ..... 57

**T**

tab1 ..... 120–121  
 tab2 ..... 121–122  
 tabi ..... 142  
 table ..... 123, 136–137  
 tabm ..... 124  
 tabodds ..... 130–131  
 tabstat ..... 135–136

tabulate ..... 120–122  
 technical support ..... 24  
 testparm ..... 154–155  
 texp() (option) ..... 189–192  
 text data ..... 61–62  
 text (graphs) ..... 242–245  
 time-series data ..... 225  
 time-to-event data ..... 171–195  
 time variables ..... 48–50  
 title (graphs) ..... 242  
 tostring ..... 52  
 transformations ..... 135  
 translation programs ..... 63  
 ttest ..... 137–139  
 ttesti ..... 144  
 tvc() (option) ..... 189–192  
 twoway connected ..... 259–260  
 twoway function ..... 262–263  
 twoway line ..... 257–259  
 twoway rcap ..... 259–261  
 twoway scatter ..... 253–257

**U**

update ..... 3–4  
 updating Stata ..... 3–4  
 use ..... 55–57

**V**

value labels ..... 65–67  
 variable labels ..... 65  
 variable lists ..... 29  
 variable names ..... 95  
 variables ..... 41–54  
     date and time ..... 47–50  
     numeric ..... 41–50  
     string ..... 50–54  
 Variables window ..... 11  
 variation of measurements ..... 203–204  
 varlist ..... 29  
 vce() (option) ..... 169  
 version control ..... 35–36  
 view ..... 17  
 Viewer window ..... 12

**W**

webuse ..... 57  
 weights ..... 32  
 window preferences ..... 8–13

**X**

xi ..... 161–162  
 xpose ..... 92