

Subject index

Symbols

`%fmt` 106–110
* abbreviation character 374–377
* comment indicator 346
+ combining strings 124–125
- abbreviation character 374–377
. , .a, .b, ..., .z missing values .. 130–133, 370–374
/* */ comment indicator 347
// comment indicator 346–347
/// continuation comment indicator...
..... 347
? abbreviation character 374–377
interaction 137
~ abbreviation character 374–377
1., 2. prefix 133–137

A

abbreviating variable lists 374–377
accessing results of commands *see*
 saved results
ado-files 323–328
aggregate statistics, making dataset of
 274–276
`_all`, abbreviation 374–377
alphabetizing variables 166–172
`anycount()`, `egen` function ... 155–156
`anymatch()`, `egen` function 156
`append` command 174–188
appending datasets 174–178
 problems 178–188
ASCII data,
 reading
 comma-separated 18–20
 fixed column 22–26
 multiple lines per observation ..
 26–28

ASCII data, reading, *continued*

 space-separated 20–22
 tab-separated 18–20
 saving
 comma-separated 40–41
 space-separated 41–43
 tab-separated 40–41
`assert` command 289–292
`autocode()` function 129–130
automating data checking 289–296

B

backups,
 admonishments to perform... 279–
 282
 remote storage of 279–282
baseline group, selecting 133–137
`browse` command 73
by-groups .. 152–155, 318–322, 344–346
 change from last value ... 238–244
 computations across 224–226
 computations within 228–238
 filling in missing values ... 238–244
 first observation within ... 238–244
 last observation within ... 238–244
 previous value ... 228–238, 244–246
 repeating commands across .. 222–
 224
 singletons within 238–244
 subsequent value 228–238
by *varlist*: prefix ... 152–155, 222–226,
 228–246, 344–346
by *sort varlist*: prefix 152–155,
 222–226, 228–246, 344–346
by *sort* versus `tsset` 244–246
byte data type 347–357

C

- categorical variables 125–130, 133–137
 - by categorical variables, checking.. 54–56
 - by continuous variables, checking.. 56–60
 - checking 50–54
- cd 10
- centering variables 314–318
- cf command 49–50
- changing directories 10
- changing shape of data .. see *reshaping datasets*
- chasing your own tail.. see *tail, chasing your own*
- checking data,
 - automating 289–296
 - categorical by categorical variables 54–56
 - categorical by continuous variables 56–60
 - categorical variables 50–54
 - continuous by continuous variables 60–63
 - continuous variables 50–54
 - double data entry 47–50
- cleaning data..... see *correcting data*
- clear command 30–31
- clear option, with use command.. 30–31
- clock() function 146
- codebook command 80–82
- coding missing values 130–133
- collapse command 274–276
- collapsing datasets 274–276
- combining datasets
 - appending 174–178
 - crossing 218–219
 - many-to-many merge 216–218
 - merge options 206–211
 - merging multiple 199–203
 - one-to-many merge 195–199
 - one-to-one merge 189–194
- combining datasets, *continued*
 - problems appending 178–188
 - problems merging 211–215
 - update merge 203–206
- combining do-files 292–296
- commands,
 - accessing results.. see *saved results*
 - repeating across by-groups... 222–224
 - repeating across variables 303–310
 - repeating automatically... 344–346
 - repeating over anything... 312–314
 - repeating over numbers... 310–312
- commas,
 - reading data separated by... 18–20
 - saving data separated by... 40–41
- commenting
 - datasets 102–105
 - variables 102–105
- comments 346–347
- compress command 347–357
- computations across
 - observations 152–155, 224–228, 234–238
 - variables 150–152, 155–157
- continuation comment, /// 347
- continuous variables
 - by continuous variables, checking.. 60–63
 - checking 50–54
- converting variables
 - numeric to string 163
 - string to numeric 157–163
- correcting data 63–67
 - double data entry 47–50
- count(), egen function 224–228
- counting words 123–124
- counts, making dataset of 274–276
- creating dataset of estimation results.. 318–322
- creating variables 116–120
- cross command 218–219
- crossing datasets 218–219

- .csv files,
 - reading 18–20
 - saving 40–41
- D**
- D. prefix (difference) 246
- data
 - checking see checking data
 - cleaning see correcting data
 - correcting see correcting data
 - entry 33–39, 47–50
 - types 347–357
- data analysis project 292–296
- Data Editor 34–39
- dataset labels 97–102
- dataset of estimation results... 318–322
- datasets,
 - appending 174–178
 - changing the shape of see reshaping datasets
 - collapsing .. see collapsing datasets
 - commenting 102–105
 - crossing 218–219
 - describing 78–83
 - downloading, for this book 330
 - example datasets from Stata .. 14–16
 - labeling 84–86
 - large 31–32
 - long see long datasets
 - many-to-many merge 216–218
 - merge** options 206–211
 - merging multiple 199–203
 - multilevel .. see multilevel datasets
 - one-to-many merge 195–199
 - one-to-one merge 189–194
 - problems appending 178–188
 - problems merging 211–215
 - reading Stata 14–16
 - reshaping .. see reshaping datasets
 - saving Stata 16–18
 - update merge 203–206
 - wide see wide datasets
- date variables 137–144
- date()** function 138, 142–144
- date-and-time variables 144–150
- dates see date variables
- day()** function 141, 149–150
- decode** command 165–166
- deleting
 - observations 367–370
 - variables 367–370
- delimited files,
 - reading see ASCII data, reading
 - saving see ASCII data, saving
- describe** command 39, 79–80
- describing datasets 78–83
- descriptive statistics, making dataset of 274–276
- destring** command 158–161
- dichotomizing variables 116–120
- dictionary file
 - with **infile** command 22–28
 - with **infix** command 22–28
- diff()**, **egen** function 156
- digits, controlling number displayed ... 106–110
- directories, changing 10
- display formats 106–110
- documenting
 - do-files 346–347
 - project 279–282
- dofc()** function 149–150
- do-files,
 - automating data checking 289–292
 - checking 279–282
 - combining 292–296
 - commenting 346–347
 - documenting 346–347
 - introduction to 282–288
 - master 292–296
 - skeleton 288
 - version** command 279–282
- double data entry 47–50
- double data type 347–357
- dow()** function 141, 150
- downloading
 - datasets for this book 330
 - user-written programs 330–339

doy() function.....141, 150
 drop command.....367–370
 dropping
 observations.....367–370
 variables.....367–370
 dummy variables.....133–137
 duplicate observations,
 dropping.....48, 67–75
 identifying.....48, 67–75
 duplicates command.....67–75

E

e(), results stored in.....314–318
 edit command.....34
 editing data.....33–39
 egen command.....150–157, 224–228,
 373–374
 encode command.....162–163
 entering data.....33–39
 ereturn list command.....314–318
 errors in data,
 correcting..... see correcting data
 finding..... see checking data
 estimation results, making dataset of..
 318–322
 example datasets
 for this book.....330
 from Stata.....14–16
 exporting data, SAS XPORT files..43–44
 exporting raw data..... see ASCII data,
 saving
 expressions,
 numeric.....120–121
 string.....121–125

F

F. prefix (forward).....246
 factor variables.....133–137
 FAQs.....339
 FDA (SAS XPORT) files,
 reading.....29–30
 saving.....43–44
 fdasave command.....43–44
 fdause command.....29–30

filling in missing values, within
 by-groups.....238–244
 findit command.....330–339
 first observation within by-groups.....
 238–244
 fixed-column data, reading.....22–26
 multiple lines per observation..26–
 28
 float data type.....347–357
 foreach command.....303–314
 format command.....106–110
 formatted raw data,
 reading..... see ASCII data, reading
 saving..... see ASCII data, saving
 frequencies, making dataset of....274–
 276
 frequently asked questions..... see FAQs
 functions.....361–364
 numeric.....120–121, 361–364
 string.....121–125
 fvset command.....136

G

generate command.....116–120
 global command.....296–300
 global macros..... see macros

H

header variables.....110–113
 hh() function.....149

I

i. prefix.....133–137
 identifiable information.....279–282
 if *exp* modifier.....364–367
 SAS XPORT files.....29–30
 importing raw data..... see ASCII data,
 reading
 in *range* modifier.....364–367
 indicator variables.....133–137
 infile command.....20–22
 with dictionary.....22–28
 infix command.....22–26
 with dictionary.....22–28
 %*infmt*.....22–26

- inlist() function.....363
- inputting data interactively.....33–39
- inputting raw data.....see ASCII data,
reading
- inrange() function.....363
- insheet command.....18–20
- int() function.....120, 362
- int data type.....347–357
- interaction terms.....133–137
- intermediate files, pruning.....279–282
- irecode() function.....129
- isid command.....71–72

- J**
- joinby command.....216–218
- joining datasets.....216–218

- K**
- keep command.....367–370
- keeping
 - observations.....367–370
 - variables.....367–370

- L**
- L. prefix (lag).....245
- label
 - define command.....86–92
 - dir command.....92
 - language command.....81–82,
97–102
 - list command.....92–97
 - save command.....92–97
 - values command.....86–92
 - variable command.....84–86
- labelbook command.....92–97
- languages, multiple.....97–102
- large datasets.....31–32
- last observation within by-groups..238–
244
- leading spaces, removing.....123
- length() function.....124–125
- list command.....4–8
- listing
 - observations.....4–8
 - value labels.....92–97
- listserver for Stata.....340
- ln() function.....120
- loading saved data.....14–16
- local command.....296–303
- local macros.....see macros
- .log files.....282–288
- log files, introduction to.....282–288
- log using command.....282–288
- log10() function.....120
- logical expressions.....357–361
 - and missing values.....357–361
- long datasets,
 - advantages.....248–257
 - compared with multilevel datasets
.....271–274
 - compared with wide.....248–257
 - disadvantages.....248–257
 - reshaping to wide.....258–260
 - problems.....261–262
- long data type.....347–357
- lookfor command.....82
- looping
 - across variables.....303–310
 - over anything.....312–314
 - over numbers.....310–312
- lower() function.....122
- ltrim() function.....123

- M**
- macros,
 - expressions with.....300–303
 - functions with.....300–303
 - introducing.....296–300
 - local versus global.....296–300
 - manipulating.....300–303
 - quotes.....296–300
- many-to-many merge.....216–218
- master do-file.....292–296
- mathematical functions.....120
- max(), egen function.....152–155,
224–228
- maximums, making dataset of....274–
276
- mdy() function.....138, 142–144
- mdyhms() function.....145–146

- mean(), *egen* function 152–155, 224–228
 - means, making dataset of 274–276
 - memory, setting 31–32, 288
 - merge command 189–215
 - merge variable 189–203
 - merging datasets
 - one-to-many 195–199
 - crossing 218–219
 - many-to-many 216–218
 - multiple 199–203
 - one-to-many 271–274
 - one-to-one 189–194
 - options 206–211
 - problems 211–215
 - update 203–206
 - min(), *egen* function 152–155, 224–228
 - minimums, making dataset of . 274–276
 - missing values 130–133, 370–374
 - in logical expressions 357–361
 - missing() function 363, 371–373
 - mm() function 149
 - modifying variables 116–120
 - month() function 141, 149–150
 - multilevel datasets 271–274
 - multiple datasets, merging 199–203
 - multiple languages 97–102
 - multiple lines per observation, reading 26–28
 - mvdecode command 131–132
 - mvencode command 132–133
- N**
- _N (number of observations) .. 228–235, 238–244
 - _n (observation number) 228–244
 - note command 65, 102–105
 - notes command 66–67, 82–83, 102–105
 - numbers, repeating commands over 310–312
 - numeric
 - functions 120–121, 361–364
 - variable to string 163
 - numeric, *continued*
 - variables 347–357
 - numlabel command 86–92
- O**
- observations,
 - computations across 152–155, 224–238
 - computing differences between 234–235
 - deleting 367–370
 - dropping 367–370
 - dropping duplicates 67–75
 - identifying duplicates 67–75
 - keeping 367–370
 - listing 4–8
 - previous value 228–235
 - randomly sampling 361–364
 - running means across 236–238
 - running proportions across ... 236–238
 - running sums across 236–238
 - subsequent value 228–235
 - omitted group, selecting 133–137
 - one-to-many merge 195–199
 - one-to-one merge 189–194
 - online resources 339
 - order command 110–113, 164, 166–172
 - ordering variables 110–113, 166–172
 - out-of-range values,
 - correcting see correcting data finding see checking data
 - outfile command 41–43
 - outputting raw data see ASCII data, saving
 - outsheet command 40–41
- P**
- program 323–328
 - drop command 323–328
 - list command 323–328
 - programming Stata 323–328
 - programs, user-written 330–339
 - proper() function 122

Q

quarter() function.....141, 150
 quotes to expand macros.....296–300

R

r(), results stored in.....314–318
 random-number functions.....121,
 361–364

randomly sampling observations..361–
 364

raw data,

 reading....see ASCII data, reading
 saving.....see ASCII data, saving

rchi2() function.....121

reading files.....11–13

 common errors.....13

 Stata datasets.....14–16

 types of files.....11–13

reading raw data.....see ASCII data,
 reading

reading SAS XPORT files.....29–30

reading Stata datasets within SAS..43–
 44

recode command.....125–130

recoding variables.....125–130

regression coefficients, making dataset
 of.....318–322

rename command.....166–172

renpfix command.....166–172

reordering variables..110–113, 166–172

reorganizing datasets....see reshaping
 datasets

repeating commands.....344–346

 across by-groups.....222–224

 across variables.....303–310

 over anything.....312–314

 over numbers.....310–312

replace command.....65–66, 116–120

reshape

long command.....262–270

wide command.....258–262

reshaping datasets,

 long to wide.....258–260

 problems.....261–262

reshaping datasets, *continued*

 wide to long.....262–265

 problems.....266–270

return list command.....314–318

rnormal() function.....121

round() function.....120, 362

routines, benefits of.....279–282

rowmax(), **egen** function.....152

rowmean(), **egen** function....151–152,
 373

rowmin(), **egen** function.....152

rowmiss(), **egen** function.....152,
 373–374

rownonmiss(), **egen** function....152,
 373–374

R-squared, computing change in..314–
 318

runiform() function....121, 363–364

running means, across observations....
236–238

running proportions, across

 observations.....236–238

running sums, across observations....
236–238

S

SAS XPORT files,

 reading.....29–30

 saving.....43–44

save command.....16–18

saved results.....314–318

 creating dataset of.....318–322

saveold command.....16

saving

 files.....13

 raw data...see ASCII data, saving

 SAS XPORT files.....43–44

 Stata datasets.....16–18

 within SAS.....29–30

sd(), **egen** function.....224–228

set

memory command.....31–32, 288

more off command.....288

type command.....353–354

setting memory 31–32, 288
 Sheldon, J. 363
 singletons within by-groups ... 238–244
 .smcl files 282–288
 spaces,
 reading data separated by ... 20–22
 saving data separated by ... 41–43
 spreadsheets, transferring
 from Stata 40–41
 into Stata 18–20
sqrt() function 120, 361–362
ss() function 149
 SSC archive 330–339
ssc command 330–339
 standard deviations, making dataset of
 274–276
 Stat/Transfer 44
 Stata datasets,
 reading 14–16
 saving 16–18
Stata Journal 330–340
 Stata macros *see* macros
 Stata programs, writing 323–328
 Stata syntax 342–344
Stata Technical Bulletin 330–340
 Stata web site 339
 Stata, updating 3
 Statalist 340
statsby prefix command 318–322
 storage types 347–357
str# data type 347–357
 string functions 121–125
 string variable to numeric 157–163
 string variables 347–357
string() function 164
 subscripting observations 228–235
substr() function 123
sum() function 236–238
summarize command 52–54
 sums, making dataset of 274–276
 syntax of Stata commands ... 342–344
sysuse command 15–16

T

tabs,
 reading data separated by ... 18–20
 saving data separated by ... 40–41
tabulate command 51–54
 tail, chasing your own *see* chasing
 your own tail
%tC format 145
%tc format 145–147
tc() pseudofunction 148
%td format 138–139, 142–144
td() pseudofunction 140–141
tostring command 164–165
total(), **egen** function 224–228
 transferring data
 from Stata 40–44
 into Stata 18–30
translate command 282–288
tsset command 244–246
tsset versus **bysort** 244–246
 two-digit years 137

U

UCLA ATS web site 339
update command 3
 update merges 203–206
 updating Stata 3
upper() function 122
use command 14–15
 user-written programs, finding and
 downloading 330–339

V

validating data *see* checking data
 value labels 86–92
 listing 92–97
 multiple languages 97–102
 problems 92–97
 variable
 labels 84–86
 lists 374–377
 variables,
 1., 2. prefix 133–137
 alphabetizing 166–172
 categorical 133–137
 centering 314–318

variables, *continued*

- checking see checking data
- commenting 102–105
- computations across 150–152, 155–157
- converting numeric to string .. 163
- converting string to numeric..157–163
- correcting see correcting data
- creating 116–120
- date 137–144
- date and time 144–150
- deleting 367–370
- dichotomizing 116–120
- display formats 106–110
- dropping 367–370
- dummy 133–137
- factor 133–137
- i. prefix 133–137
- indicator 133–137
- keeping 367–370
- labeling 84–86
- modifying 116–120
- recoding 125–130
- reordering 110–113, 166–172
- repeating commands across .. 303–310
- types of 347–357
- Variables Manager 34–39
- varlist* 374–377
- version** command 279–282, 288

W

- web resources 339
- web site for this book 330
- webuse** command 15–16
- week()** function 141, 150
- wide datasets,
 - advantages 248–257
 - compared with long 248–257
 - compared with multilevel datasets 271–274
 - disadvantages 248–257
 - reshaping to long 262–265
 - problems 266–270

- word()** function 124
- wordcount()** function 123–124

Y

- year()** function 141, 149–150
- years, two digit 137