Subject index

Symbols

%*fmt* 136–140
* abbreviation character 487–490
* comment indicator 458–459
+ combining strings 158–160
- abbreviation character 487–490
., .a, .b, . . . , .z missing values 166–168, 483–487
/* */ comment indicator 459
// comment indicator 459
/// continuation comment indicator 459
? abbreviation character 487–490
interaction 171
~ abbreviation character 487–490
1., 2., . . . prefix 168–171

A

abbreviating variable lists 487–490
accessing results of commands . . . . . *see* stored results
ado-files 359–451
aggregate statistics, making dataset of 314–316
_all, abbreviation 487–490
alphabetizing variables 201–207
anycount(), egen function 191
anymatch(), egen function 191
append command 210–225
appending datasets 210–214
  problems 215–225
ASCII data,
  importing
    comma-separated 30–32
    fixed column 35–39
    multiple lines per observation 39–41
ASCII data, importing, *continued*
    space-separated 33–35
    tab-separated 30–32
  saving
    comma-separated 62–65
    space-separated 65–66
    tab-separated 62–65
assert command 328–332
autocode() function 164–165
automating data checking 328–336

B

backups,
  admonishments to perform 318–321
  remote storage of 318–321
baseline group, selecting 168–171
browse command 104
by-groups 188–190, 456–458
  change from last value 276–282
  computations across 262–266
  computations within 266–276
  filling missing values 276–282
  first observation within . . . 276–282
  last observation within . . . 276–282
  previous value . . . 266–276, 282–285
  repeating commands across 260–262
  singletons within 276–282
  subsequent value 266–276
by *varlist*: prefix 188–190, 260–262, 266–285, 456–458
bysort *varlist*: prefix 260–262, 266–285, 456–458
bysort versus tsset 282–285
byte data type 459–469

C

categorical variables 160–165, 168–171
  by categorical variables, checking 85–87
  by continuous variables, checking 87–91
  checking 81–85
`cd` command 12–13
`cf` command 79–80
changing directories 12–13
changing shape of data . . *see* reshaping datasets
chasing your own tail . . *see* tail, chasing your own
checking data,
  automating 328–336
  categorical by categorical variables 85–87
  categorical by continuous variables 87–91
  categorical variables 81–85
  continuous by continuous variables 91–94
  continuous variables 81–85
  double data entry 77–81
checking for unlabeled values . . 439–445
checking value labels using a Stata program 415–418
checking variable labels using a Stata program 410–415
cleaning data . . . . . *see* correcting data
`clear` command 42–43
`clear` option with `use` command 42–43
`clock()` function 181–182
`codebook` command 110–112
coding missing values 166–168
`collapse` command 314–316
collapsing datasets 314–316
combining datasets
  appending 210–214
  crossing 255–257
  `merge` options 243–248
  merging multiple 236–240
combining datasets, *continued*
  one-to-many merge 231–236
  one-to-one merge 225–231
  problems appending 215–225
  problems merging 249–253
  update merge 240–243
combining do-files 332–336
commands,
  accessing results . . . . . *see* stored results
  repeating across by-groups 260–262
  repeating across variables 344–351
  repeating automatically . . . 456–458
  repeating over anything . . . 353–355
  repeating over numbers . . . 351–353
commas,
  importing data separated by 30–32
  saving data separated by 62–65
commenting
  datasets 132–135
  variables 132–135
comments 458–459
`compress` command 459–469
computations across
  observations 188–190, 262–266, 272–276
  variables 186–188, 190–192
continuation comment, `///` 459
continuous variables
  by continuous variables, checking 91–94
  checking 81–85
converting variables
  numeric to string 199
  string to numeric 192–198
correcting data 94–98
  double data entry 77–81
`count()`, `egen` function 262–266
counting words 154–158
counts, making dataset of 314–316
creating variables 146–150
`cross` command 255–257

crossing datasets 255–257
`.csv files`,
importing 30–32
saving 62–65
customized `describe` command 418–431
customized `summarize` command 432–439

D

`D.` prefix (difference) 284–285
data
checking . . . . . *see* checking data
cleaning . . . . . *see* correcting data
correcting . . . . . *see* correcting data
entry 43–50, 77–81
types 459–469
data analysis project 332–336
Data Editor 44–50
dataset labels 127–132
datasets,
appending 210–214
changing the shape of . . . . . *see* reshaping datasets
collapsing . . *see* collapsing datasets
commenting 132–135
crossing 255–257
describing 108–114
downloading, for this book 2–3
example datasets from Stata 14–16
labeling 114–116
long . . . . . *see* long datasets
`merge` options 243–248
merging multiple 236–240
multilevel . . *see* multilevel datasets
one-to-many merge 231–236
one-to-one merge 225–231
problems appending 215–225
problems merging 249–253
reading Stata 14–16
reshaping . . . *see* reshaping datasets
saving Stata 52–55
update merge 240–243
wide . . . . . *see* wide datasets
date variables 172–179
`date()` function 172–173, 177–179
date-and-time variables 179–186
dates . . . . . *see* date variables
`day()` function 176, 185
dBase files,
exporting 61–62
importing 28
debugging Stata programs 445–450
`decode` command 200–201
deleting
observations 480–483
variables 480–483
delimited files,
reading . . . . *see* ASCII data, reading
saving . . . . . *see* ASCII data, saving
`describe` command 50, 109–110
describing datasets 108–114
descriptive statistics, making dataset of 314–316
`destring` command 193–197
dichotomizing variables 146–150
dictionary file
with `infile` command 35–41
with `infix` command 35–41
`diff()`, `egen` function 192
digits, controlling number displayed 136–140
directories, changing 12–13
display formats 136–140
documenting
do-files 458–459
project 318–321
`dofc()` function 184–186
do-files,
automating data checking 328–332
checking 318–321
combining 332–336
commenting 458–459
documenting 458–459
introduction to 322–328
master 332–336
skeleton 327
`version` command 318–321

double data entry 77–81
`double` data type 459–469
`dow()` function 176, 185–186
downloading datasets for this book 2–3
`doy()` function 176, 185–186
`drop` command 480–483
dropping
observations 480–483
variables 480–483
dummy variables 168–171
duplicate observations,
dropping 78–79, 98–106
identifying 78–79, 98–106
`duplicates` command 98–106

E

`edit` command 45
editing data 43–50
`egen` command 186–192, 262–266, 486–487
`encode` command 197–198
entering data 43–50
errors in data,
correcting . . . . . *see* correcting data
finding . . . . . *see* checking data
example datasets
for this book 2–3
from Stata 14–16
Excel file,
exporting 55–73
Excel files,
importing 16–20
`export dbase` command 61–62
`export delimited` command 62–65
`export excel` command 55–73
`export sasxport5` command 60–61
`export sasxport8` command 59–60
exporting 52
dBase files 61–62
Excel files 55–73
raw data . . . *see* ASCII data, saving
SAS XPORT Version 5 files . . . 60–61
SAS XPORT Version 8 files . . . 59–60
expressions,
numeric 150–152
string 152–160

F

`F.` prefix (forward) 284–285
factor variables 168–171
FAQs 9
filling missing values, within by-groups 276–282
first observation within by-groups 276–282
fixed-column data, importing 35–39
multiple lines per observation 39–41
`float` data type 459–469
`foreach` command 344–355
`format` command 136–140
formatted raw data,
reading . . . . *see* ASCII data, reading
saving . . . . . *see* ASCII data, saving
frames 490–502
frequencies, making dataset of 314–316
frequently asked questions . . . . *see* FAQs
functions 474–477
numeric 150–152, 474–477
string 152–160
Unicode string 152–160

G

`generate` command 146–150
`global` command 336–340
global macros . . . . . *see* macros

H

header variables 140–144
`hh()` function 184

I

`i.` prefix 168–171
IBM SPSS `.sav` files, importing . . . 25–27
identifiable information 318–321
`if` condition
with a Stata program 371–375

`if` condition with a Stata program 390–391
`if` *exp* modifier 477–480
`import dbase` command 28
`import delimited` command 30–32
`import sas` command 20–22
`import sasxport5` command 23–24
`import sasxport8` command 24
`import spss` command 25–27
importing
dBase `.dbf` files 28
Excel files 16–20
IBM SPSS `.sav` files 25–27
SAS files 20–22
SAS XPORT Version 5 files . . . 23–24
SAS XPORT Version 8 files 24
types of files 13–14
importing raw data . . . . *see* ASCII data, reading
`in` *range* modifier 477–480
indicator variables 168–171
`infile` command 33–35
with dictionary 35–41
`infix` command 35–39
with dictionary 35–41
`%`*infmt* 35–39
`inlist()` function 476
inputting data interactively 43–50
inputting raw data . . . . . *see* ASCII data, reading
`inrange()` function 476
`int` data type 459–469
`int()` function 150–151, 474–475
interaction terms 168–171
intermediate files, pruning 318–321
`irecode()` function 164
`isid` command 102–103

J

`joinby` command 253–255
joining datasets 253–255

K

`keep` command 480–483
keeping
observations 480–483
variables 480–483

L

`L.` prefix (lag) 283
`label`
`define` command 116–122
`dir` command 122
`language` command 111–112, 127–132
`list` command 122–127
`save` command 122–127
`values` command 116–122
`variable` command 114–116
`labelbook` command 122–127
languages, multiple 127–132
last observation within by-groups 276–282
leading spaces, removing 153–154
`list` command 5–8
listing
observations 5–8
value labels 122–127
listserver for Stata 9
`ln()` function 150–151
loading saved data 14–16
`local` command 336–343
local macros . . . . . *see* macros
`.log` files 322–328
log files, introduction to 322–328
`log using` command 322–328
`log10()` function 150–151
logical expressions 469–474
and missing values 469–474
`long` data type 459–469
long datasets,
advantages 288–297
compared with multilevel datasets 311–314
compared with wide 288–297
disadvantages 288–297
reshaping to wide 298–300
problems 301–302
`lookfor` command 112–113

looping
across variables 344–351
over anything 353–355
over numbers 351–353

M

macros,
expressions with 341–343
functions with 341–343
introducing 336–340
local versus global 336–340
manipulating 341–343
quotes 336–340
master do-file 332–336
mathematical functions 150–151
`max()`, `egen` function 188–190, 262–266
maximums, making dataset of 314–316
`mdy()` function 172, 175, 177–179
`mdyhms()` function 180–181
`mean()`, `egen` function 188–190, 262–266
means, making dataset of 314–316
`merge` command 225–253
`_merge` variable 225–240
merging datasets
one-to-many 231–236
crossing 255–257
multiple 236–240
one-to-many 311–314
one-to-one 225–231
options 243–248
problems 249–253
update 240–243
`min()`, `egen` function 188–190, 262–266
minimums, making dataset of . . 314–316
missing values 166–168, 483–487
in logical expressions 469–474
`missing()` function 476, 484–486
`mm()` function 184
modifying variables 146–150
`month()` function 176, 185
multilevel counting with a Stata program 377–386
multilevel datasets 311–314
multiple datasets, merging 236–240
multiple languages 127–132
multiple lines per observation, importing 39–41
`mvdecode` command 167
`mvencode` command 167–168

N

`_N` (number of observations) . . 266–273, 276–282
`_n` (observation number) 266–282
`note` command 97, 132–135
`notes` command . . 97–98, 113, 132–135
numbers, repeating commands over 351–353
numeric
functions 150–152, 474–477
variable to string 199
variables 459–469
`numlabel` command 116–122

O

observations,
computations across 188–190, 262–276
computing differences between 272–273
deleting 480–483
dropping 480–483
dropping duplicates 98–106
identifying duplicates 98–106
keeping 480–483
listing 5–8
previous value 266–273
randomly sampling 474–477
running means across 273–276
running proportions across 273–276
running sums across 273–276
subsequent value 266–273
omitted group, selecting 168–171
one-to-many merge 231–236

one-to-one merge 225–231
online resources 8–9
options
adding to Stata programs 372–375
options, adding to Stata programs 391–393
`order` command 140–144, 201–207
ordering variables 140–144, 201–207
out-of-range values,
correcting . . . . . *see* correcting data
finding . . . . . *see* checking data
`outfile` command 65–66
outputting raw data . . . . *see* ASCII data, saving

P

`program` 359–451
`drop` command 366–367
`list` command 366–367
programming Stata 359–451

Q

`quarter()` function 176, 185–186
quotes to expand macros 336–340

R

`r()`, results stored in 355–357
random-number functions 151–152, 474–477
randomly sampling observations 474–477
raw data,
reading . . . . *see* ASCII data, reading
saving . . . . . *see* ASCII data, saving
`rchi2()` function 152
reading files 13–14
common errors 14, 41–43
Stata datasets 14–16
types of files 13–14
reading raw data . . . . . *see* ASCII data, reading
`recode` command 160–165
recoding variables 160–165
`rename` command 201–207
reordering variables . . 140–144, 201–207
reorganizing datasets . . . . *see* reshaping datasets
repeating commands 456–458
across by-groups 260–262
across variables 344–351
over anything 353–355
over numbers 351–353
`replace` command 96–97, 146–150
`reshape`
`long` command 303–310
`wide` command 298–302
reshaping datasets,
long to wide 298–300
problems 301–302
wide to long 303–305
problems 306–310
`return list` command 355–357
`rnormal()` function 151
`round()` function 150–151, 474–475
routines, benefits of 318–321
`rowmax()`, `egen` function 187
`rowmean()`, `egen` function 186–187, 486
`rowmin()`, `egen` function 187
`rowmiss()`, `egen` function 187–188, 486–487
`rownonmiss()`, `egen` function 187–188, 486–487
`runiform()` function 151, 476
running
means, across observations 273–276
proportions, across observations 273–276
sums, across observations 273–276

S

SAS XPORT Version 5 files,
exporting 60–61
importing 23–24
saving 60–61

SAS XPORT Version 8 files,
exporting . . . 59–60
importing . . . 24
saving . . . 59–60
`save` command . . . 52–55
`saveold` command . . . 53
saving
files . . . 52
raw data . . . *see* ASCII data, saving
SAS XPORT Version 5 files . . . 60–61
SAS XPORT Version 8 files . . . 59–60
Stata datasets . . . 52–55
Stata programs . . . 375–377
scoring a scale using a Stata program . . . 394–402
`sd()`, `egen` function . . . 262–266
`set trace` command . . . 445–447
Sheldon, J. . . . 475
singletons within by-groups . . . 276–282
`.smcl` files . . . 322–328
spaces,
importing data separated by . . . 33–35
saving data separated by . . . 65–66
spreadsheets, transferring
from Stata . . . 62–65
into Stata . . . 30–32
`sqrt()` function . . . 150–151, 474
`ss()` function . . . 184
standard deviations, making dataset of . . . 314–316
standardizing variables . . . 355–357
using a Stata program . . . 402–410
Stata Blog . . . 9
Stata datasets,
reading . . . 14–16
saving . . . 52–55
Stata Journal . . . 9
Stata macros . . . *see* macros
Stata program
basics . . . 363–367
checking for unlabeled values . . . 439–445
checking value labels . . . 415–418
checking variable labels . . . 410–415
Stata program, *continued*
customized `describe` command . . . 418–431
customized `summarize` command . . . 432–439
debugging . . . 445–450
how to save . . . 363–367
multilevel counting . . . 377–386
options . . . 372–375, 391–393
pros and cons . . . 359–360
scoring a scale . . . 394–402
standardizing variables . . . 402–410
strategies for writing . . . 360–361
tabulations . . . 386–393
`tempvar` command . . . 385–386
using temporary variables . . . 385–386
where to save . . . 375–377
with a variable list . . . 368–371
with an `if` condition . . . 371–375, 390–391
wrapper program . . . 406, 409–410
writing . . . 359–451
Stata syntax . . . 454–456
Stata video tutorials . . . 9
Stata website . . . 8–9
Statalist . . . 9
storage types . . . 459–469
stored results . . . 355–357
`str#` data type . . . 459–469
strategies for writing Stata programs . . . 360–361
string functions . . . 152–160
string variable to numeric . . . 192–198
string variables . . . 459–469
`string()` function . . . 199
strings, Unicode . . . 152–160
subscripting observations . . . 266–273
`sum()` function . . . 273–276
`summarize` command . . . 83–85
sums, making dataset of . . . 314–316
syntax of Stata commands . . . 454–456
`sysuse` command . . . 15–16

T
tabs,
importing data separated by 30–32
saving data separated by 62–65
`tabulate` command 82–85
tail, chasing your own . . . . . *see* chasing your own tail
`%tc` format 180–183
`tc()` pseudofunction 183–184
`%td` format 173–174, 177–179
temporary variables in a Stata program 385–386
`tempvar` command 385–386
`tostring` command 200
`total()`, `egen` function 262–266
transferring data
from Stata 62–66
into Stata 30–41
`translate` command 322–328
`tsset` command 282–285
`tsset` versus `bysort` 282–285
two-digit years 172

U
UCLA IDRE website 9
Unicode, strings 152–160
update merges 240–243
`use` command 14–15
`ustrlen()` function 158
`ustrlower()` function 153
`ustrltrim()` function 153–154
`ustrtitle()` function 153
`ustrupper()` function 153
`ustrword()` function 154–158
`ustrwordcount()` function 154–157
`usubstr()` function 154

V
validating data . . . . . *see* checking data
value labels 116–122
listing 122–127
multiple languages 127–132
problems 122–127
variable
labels 114–116
lists 487–490
with a Stata program . . 368–371
variables,
`1.`, `2.` . . . prefix 168–171
alphabetizing 201–207
categorical 168–171
checking . . . . . *see* checking data
commenting 132–135
computations across 186–188, 190–192
converting numeric to string . . 199
converting string to numeric 192–198
correcting . . . . . *see* correcting data
creating 146–150
date 172–179
date and time 179–186
deleting 480–483
dichotomizing 146–150
display formats 136–140
dropping 480–483
dummy 168–171
factor 168–171
`i.` prefix 168–171
indicator 168–171
keeping 480–483
labeling 114–116
modifying 146–150
recoding 160–165
reordering 140–144, 201–207
repeating commands across 344–351
standardizing 355–357
types of 459–469
Variables Manager 45–50
varlist 487–490
`version` command 318–321, 327

W
web resources 8–9
website for this book 3
`webuse` command 15–16
`week()` function 176, 185–186

wide datasets,
advantages 288–297
compared with long 288–297
compared with multilevel datasets 311–314
disadvantages 288–297
reshaping to long 303–305
problems 306–310
wrapper program 406, 409–410

Y

`year()` function 176, 185
years, two digit 172