

An Introduction to Stata Programming

Second Edition

CHRISTOPHER F. BAUM

*Department of Economics and School of Social Work
Boston College*



A Stata Press Publication
StataCorp LP
College Station, Texas



Copyright © 2009, 2016 by StataCorp LP
All rights reserved. First edition 2009
Second edition 2016

Published by Stata Press, 4905 Lakeway Drive, College Station, Texas 77845

Typeset in L^AT_EX 2_ε

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN-10: 1-59718-150-1

ISBN-13: 978-1-59718-150-1

Library of Congress Control Number: 2015955595

No part of this book may be reproduced, stored in a retrieval system, or transcribed, in any form or by any means—electronic, mechanical, photocopy, recording, or otherwise—without the prior written permission of StataCorp LP.

Stata, **stata**, Stata Press, Mata, **mata**, and NetCourse are registered trademarks of StataCorp LP.

Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations.

L^AT_EX 2_ε is a trademark of the American Mathematical Society.

Subject index

A

added-variable plot 204
ado-path 7, 250, 341
adopath command 7
adoupdate command 4, 34
Akaike information criterion 98
append command 84, 110, 112, 187
args command 253, 263
asarray() Mata function 338
ASCII text file 19, 196
assert command 22, 87, 250
associative array 338
atanh() function 294
audit trail 84
autocorrelation 181
automated output 199
average marginal effects 140
avplot command 204

B

_b stored results 277
backslash 78
balanced panel 117
base levels 161
Bayesian information criterion 98
bcal command 12
bcal create command 132
binary file handling 197
biprobit command 294
Boolean condition 48
bootstrap prefix command .. 167, 270,
272
built-in commands 90
built-in function 353
business-daily data 129, 230
by prefix command 158
byable 52, 242

bysort prefix command 158
byte data type 96
bytecode 301

C

calendar variable 92
capture prefix command 17, 169
casewise deletion 41, 96, 193
cd command 6
ceil() function 50
center command 57, 320
certification script 250, 282, 392
char command 208
characteristics 207, 311
ckvar command 209
class programming 1
codebook command 24, 86
codebooks 24
collapse command 17, 57, 69, 105,
145, 193, 322
colon operator 304
column-join operator 192, 217, 303
column-range operator 303
comma-delimited file 21
comma-separated-values file 196
compiled function 340
complex arithmetic 301
compound double quotes 44
compress command 11, 29
cond() function 49, 295
conditional statements 308
confirm command 321, 343
constraint command 289
continuous price series 225
continuously updated generalized
method of moments
estimator 384

contract command 17, 69, 193
correlate command .. 41, 93, 168, 212, 379
correlation() Mata function 350
creturn command 62, 81
cross command 117
cross() Mata function 381
cscript command 282
 CSV file...see comma-separated-values file
 .csv file suffix 196
 current working directory 6
 CWD.....see current working directory

D

D. time-series operator 13
d0 evaluator 390
 daily data 129
 data dictionary 25
 data validation 84
decode command 43
 decrement operator 307
 delimiters 20
 delta method 265
 dereferencing 58
describe command 39, 91
destring command 42
 dialog programming 1
 dictionary file 196
 difference operator see D.
diparm() option 291
 directory separator 78
discard command 234
display command 35, 44
 Do-file Editor 8
dofm() function 142
 dot product 304
drop command 57
duplicates command 89, 116
dyex option 200
 dynamic panel data 294

E

e(b) matrix 96
e(sample) function 94, 96

e-class 90, 94
eform() option 243
egen anycount() function 67
egen bom() function 53
egen bomd() function 54
egen command 41, 51, 67, 280, 322
egen corr() function 54
egen count() function 52
egen eom() function 53
egen eomd() function 54
egen ewma() function 54
egen filter() function 54
egen functions 246
egen gmean() function 54
egen group() function 159
egen hmean() function 54
egen iqr() function 53
egen kurt() function 53
egen mad() function 53
egen max() function 52
egen mdev() function 53
egen mean() function 52, 228
egen median() function 53
egen min() function 52
egen mode() function 53
egen nvals() function 186
egen pc() function 53
egen pctlile() function 53, 280
egen rank() function 53
egen record() function 54
egen rndint() function 54
egen rowfirst() function 52
egen rowlast() function 52
egen rowmax() function 52
egen rowmean() function 52
egen rowmin() function 52
egen rowmiss() function 52, 67
egen rownonmiss() function 52, 67
egen rowsd() function 52
egen rowtotal() function 52
egen sd() function 53
egen semean() function 54
egen skew() function 53
egen std() function 53
egen tag() function 186

egen total() function.....47, 53, 70
 egen var() function.....54
 egenmore package.....53, 67, 186
 elasticity.....199, 214
 elementwise operations 304
 embedded spaces 21
 encode command.....42
 .eps format.....206
 ereturn display command.. 381, 385
 ereturn list command.....94, 97
 ereturn post command.....380
 estadd command.....102, 136
 estat ovtest command.....97
 estat vce command.....97
 estimates command.....97
 estimates for command.....100
 estimates notes command.....101
 estimates replay command.....97
 estimates save command....98, 101
 estimates stats command.....95
 estimates store command...97, 288,
 297
 estimates table command...97, 136
 estimates use command.....98
 estimation commands.....90
 estout command...102, 136, 199, 376
 eststo command.....199, 376
 esttab command.....136, 199
 Euclidian distance.....364
 exponential distribution 255
 export delimited command.....196
 export sasxport command.....196
 extended macro functions..60, 187, 190
 external declaration.....371, 389

F

F. time-series operator.....13
 factor variables 14, 161
 FAQs....see frequently asked questions
 fcast compute command 150
 file command.....196
 file handle.....197
 file open command.....197
 file read command.....198

file write command.....197
 fillin command.....117
 findexternal() Mata function...317
 finite-precision arithmetic.....286
 fixed effects.....297
 fixed format.....20, 24
 floor() function 50
 fopen() Mata function.....371
 foreach command.....59, 169, 190
 forvalues command.....169
 fracpoly prefix command.....169
 freduse command.....129
 free format 20
 frequently asked questions 33
 function 301
 function arguments.....316
 function library.....341
 function-evaluator program...263, 265

G

generalized method of moments....384
 generate command 44
 global macro 60, 78, 171, 288, 293
 gmm command 265
 GMM-CUE.....see continuously
 updated generalized method of
 moments estimator
 .gph file suffix 206, 220
 graph combine command 205
 graph display command 206
 Graph Editor.....207, 221
 graph export command.....206
 graph save command.....221
 graph twoway rarea command...279
 graphics.....203
 groupwise heteroskedasticity 270
 gsem command.....148
 gsort command.....55, 70, 212

H

hash table.....338
 help files.....244
 HTML.....102, 245

I

if qualifier 39, 40
 implied decimal 28
 import delimited command ... 13, 21, 23, 42
 import excel command 23, 186
 import sasxport command 19, 29
 impulse–response functions 152
 in qualifier 39
 include command 318
 increment operator 307
 indicator variable 47, 205
 inequality constraints 291
 infile command 20, 27, 121, 124
 infix command 25
 inlist() function 45, 88
 inner product 304
 inrange() function 45, 68, 87
 instrumental variables 384
 int() function 46, 50
 integer division 46
 interaction terms 162
 interval estimates 215
 invsym() Mata function 381
 irecode() function 51
 irf command 152
 irf create command 153
 ivreg2 command ... 251, 289, 384, 393
 ivregress command 388

J

J() Mata function 190, 305
 jackknife command 270
 joinby command 120

K

Kronecker product 304

L

L. time-series operator 13
 L2-norm 364, 365
 lag operator see L. time-series operator
 latent variable 138

L^AT_EX output .. 102, 191, 193, 199, 216, 370

lead operator see F.
 levelsof command 79, 190, 211
 likelihood function 95
 likelihood-evaluator program 253
 limits 353
 lincom command 98, 248, 277
 linear constraints 289
 linear filter see egen filter() function
 list subscripts 305
 liststruct() Mata function 336
 listtab command 155
 listwise deletion 193
 local macro 58, 171
 log price relative 226
 log2html command 245
 logit command 138
 long form 104
 longitudinal data 105
 loop constructs 169, 307

M

Mac OS X spreadsheet dates 23
 macro 57–60
 macro evaluation 58
 macro list functions 61, 79
 makematrix command 216
 Mann–Whitney test 218
 many-to-many merge 116
 many-to-one merge 114
 marginal effects 214
 margins command ... 14, 136, 199, 248
 marginsplot command 140, 215
 marksample command 238, 295
 mat2txt command 66, 191
 Mata 64, 301
 .mata file suffix 340
 mata mlib add command 342
 mata mlib create command 341
 mata mosave command 341
 match-merge 114
 matrix accum command 65

- matrix colnames** command....66, 95,
 190, 197, 344
matrix command.....64, 190
matrix language.....64
matrix list command....64, 96, 191
matrix programming.....301
matrix rownames command....66, 95,
 190, 197, 344
matrix score command.....267
matrix stripes.....344, 381
matsize command.....353
max() function.....71
maxbyte() function.....46
maximum likelihood.....289, 291, 384
maxindex() Mata function.....350
maxint() function.....46
maxlong() function.....46
mean() Mata function.....350
merge command.....84, 111, 114, 180,
 223
mf_neighbor() Mata function...364
mfp prefix command.....169
mi estimate prefix command....169
min() function.....71
minindex() Mata function.....365
missing() function.....40
missing string value.....41
missing values.....48
mkmat command.....65
ml command.....252, 384
mlexp command.....252
.mlib file suffix.....341
mlopts parsing tool.....261
mm_cond() Mata function.....308
mm_meancolvar() Mata function...365
.mo file suffix.....340
mod() function.....46, 142
month() function.....142
moremata package.....308, 342, 365
moving correlation.....181
moving-window statistics.....178
mrldif() function.....284
mvcorr command.....181
mvdecode command.....41
mvencode command.....41
mvsumm command.....178
N
naturally coded indicator.....161
nearest neighbor.....363
nested loop.....171, 173
nested macros.....58
nestreg prefix command.....169
net install command.....7
njc_stuff package.....207
nlcom command.....265, 289, 294
noisily prefix command.....169
normalden() function.....291
null model.....95
null pointer.....332
nullmat() function.....192, 218
numlist.....39
O
object file.....340
object-oriented programming.....1
ODBC.....30
odbc command.....30
one-liner.....301
one-to-one merge.....116
optimize() Mata function...384, 390
optimize_init() Mata function...389
optimize_init_evaluator() Mata
 function.....390
optimize_init_evaluator_type()
 Mata function.....390
optimize_init_params()
 Mata function.....390
optimize_init_which() Mata
 function.....390
optimize_result_value() Mata
 function.....390
optional arguments.....317
options.....237
order command.....39, 172, 196
outer join.....120
outer product.....304
outfile command.....196
outreg command.....97, 103
outtable command.....66, 191, 242

overlapping subsamples 162

P

panel data 74, 105, 294
 panel variable 92
`panelsetup()` Mata function 349
`panelsubmatrix()` Mata function 350
`panelsubview()` Mata function 353
 passing functions to functions 353
`pctile()` function 247
`pctrange` command 282
 .pdf format 206
`permute` command 167
 .png format 206
 point estimates 215
 pointer function 358
 pointer variables 331
 pointers 353, 381
 pointers to pointers 333
`post` command 195
`postfile` command 194
`predict` command 11, 248, 380
 prefix commands 157
`preserve` command 17, 193, 295
 price deflator 178
`probit` command 138
`profile.do` 6
`program` command 234
 program properties 243
 propensity score 363
`psmatch2` command 363
`putexcel` command 202
p-value 218
`pwcorr` command 41, 379

Q

`qofd()` function 144
`qreg` command 163, 376
 quantile regression 375
 quietly prefix command 169

R

range subscripts 305
`ranksum` command 218
 .raw file suffix 196

r-class 90, 93
`rcof` command 284
`real()` function 42
`recode` command 50, 172
`recode()` function 50
 record 70
 recursive estimation 162
`reg3` command 379
`regress` command 380
 relational operators 304
 relative difference 170
`reldif()` function 170, 285
`rename` command 112
`replace` command 44
 replay feature 261
 reserved words 317
`reshape` command 105, 150, 323
`restore` command 17, 193, 296
`return list` command 91, 136
 return types 316
 reverse recursive estimation 162
 rich text format 102, 103
`rnormal()` Mata function 306
`robvar` command 270
`rolling` prefix command 162, 178, 277
 rolling-window estimation 162
 root finders 342
`round()` function 50
 row-join operator 192, 217, 303
 row-range operator 303
 rowwise functions 52

S

S. time-series operator 13
 SAS XPORT Transport file 196
`save` command 6
 scalar 285
`scalar` command 62
 scheme programming 1
 Schwarz criterion 98
`_se` stored results 277
`search` command 4, 7, 33, 233
 seasonal difference operator see S.
 seemingly unrelated regression 379

semielasticity 199
 separate command 119, 205
 sersert command 221
 set rmsg command 395
 set seed command 165, 271
 set trace command 35
 similarity 364
 simulate prefix command 164, 270
 sort command 55
 sort order 91
 sortpreserve option 261
 space-delimited files 20
 spells 71, 294
 sphdist command 359
 spreadsheet data 23
 SQL databases 30
 SSC archive 33, 287
 ssc command 7, 52
 stack command 118
 st_addvar() Mata function 310
 standardized values see egen std()
 function
 stars, for significance 98
 Stat/Transfer 11, 29, 41, 196
 Stata blog 34
Stata Journal 7, 33, 34
 Stata Markup and Control Language ..
 102, 140, 244
 stata() Mata function 359
Stata Technical Bulletin 7, 33
 Stata tips 34
 Statalist ... 34, 125, 133, 179, 184, 185,
 211
 statsby prefix command 160
 statsmat command 66, 193
 st_data() Mata function 309
 stepwise prefix command 169
 st_global() Mata function 311
 st_local() Mata function 311
 st_matrix() Mata function ... 311, 344
 st_matrixcolstripe() Mata function
 344
 st_matrixrowstripe() Mata function
 344
 st_nobs() Mata function 311

st_numscalar() Mata function ... 311,
 319
 st_nvar() Mata function 311
 stock returns 211
 storage optimization 29
 storedresults command 287, 393
 strofreal() function 43
 string missing values 41
 strL 9
 strtoreal() Mata function 312
 structural vector autoregression 152
 structure 334
 st_sdata() Mata function 309, 311
 stset command 207
 st_sstore() Mata function ... 310, 311
 st_store() Mata function 310
 st_strscalar() Mata function ... 311
 st_subview() Mata function. 311, 356
 st_sview() Mata function 309, 311
 st_tsrevar() Mata function 386
 st_varindex() Mata function 312
 st_varname() Mata function 312
 st_view() Mata function 319, 359
 subview matrix 356
 suest command 147, 297
 sum() function 46, 182, 185
 summarize command 91, 212
 sureg command 379
 svmat command 212
 svy prefix command 169, 370
 svyset command 169, 208, 372
 syntax command 234, 247, 343
 sysdir command 6

T

tab-delimited file 21, 196
 tabstat command 49, 66, 69, 135,
 193, 217
 tabulate command 88, 370, 372
 tabulating results 98
 tempname command 194, 219
 temporary objects 331
 temporary variable 238
 tempvar command 238, 295
 test command 98, 147, 248

text editors.....19
 text files.....19
texteditors command.....19
 .tif format.....206
 time-series calendar.....141
 time-series operators.....13, 285
tokenize command.....174
tokens() Mata function.....321
tostring command.....43
 transmorphic matrix.....316
 transpose.....121
 transpose operator.....303
tsappend command.....152
tscollap command.....145, 322
tsfill command.....117, 179, 295
tsline command...130, 179, 183, 279
tsmktime command.....141, 325
tsset command...12-14, 54, 92, 129,
 141, 207, 295
tsspell command.....71, 295
ttest command.....93, 168, 218, 270

U

unbalanced panel.....93, 117, 294, 379
update command.....6, 33
use command.....6

V

var command.....150
 variable indices.....312
varlist.....39
vec command.....150
 vector autoregression.....149, 152
 vector error-correction model..149, 152
version command.....234
version prefix command.....169
 view matrices.....301, 310, 356
viewsource command.....90, 342
 virtual matrix.....301
 void function.....316

W

which command.....35, 90, 233
while command.....169
 wide form.....104, 379

wildcards.....39, 52, 107
 word processors.....19
 wrapper program.....277

X

xi prefix command.....161
xpose command.....121
xt commands.....379
xtabond command.....294
xtabond2 command.....294
xtile command.....51
xtline command.....150
xtreg command.....297
xtset command..14, 92, 106, 180, 199,
 207