

Subject index

Symbols

$\Delta\beta$ 274–275
 $\Delta\chi^2$ 275
 β see regression, standardized coefficient
* see do-files, comments
+ see operators
// see do-files, comments
#delimit 36–37
& see operators
_N 82–83, 95
_all 47–48
_b[] 185
_merge (variable) 310–312
_n 81–82, 95
| see operators
|| 104
^ see operators
~ see operators

A

Academic Technology Service .. see ATS
added-variable plot 206–209
additive index 79
adjusted count R^2 266–267
adjusted R^2 195–196
ado-directories 359
ado-files
 basics 335–337
 programming 337–351
aggregate see collapse
AIC 267
Akaike’s information criterion .. see AIC
Aldrich–Nelson’s p^2 326
alphanumeric variables ... see strings
analysis of variance see regression

angle() (axis-label suboption) ... 124–125
ANOVA see regression
Anscombe quartet 199, 201, 226
anycount() (**egen** function) 88
append 316–319
arithmetic mean see average
arithmetical expressions see expressions
ascategory (graph dot option) ... 149
ASCII files 292–301
ATS 354
augmented component-plus-residual plot 204
autocode() (function) 152
autocorrelation see regression, autocorrelation
average 15–16, 153, 156
avplots 207–208
aweight (weighting type) 68–69
axis
 labels 107–108, 122–125
 scales 117–119
 titles 107–108, 126–127
 transformations 118–119

B

balanced panel data 312–313
bands() (**twoway mband** option) .. 202–203
bar (graph type) ... 105, 108, 146–147, 159
bar charts 146–147, 159
batch jobs see do-files
Bayesian information criterion see BIC

Bernoulli's distribution ... see binomial distribution
 beta see regression, standardized coefficient
 bias 200
 BIC 267
 bin() (histogram option) 164–165
 binary variables..see variables, dummy
 binomial distribution 254–255
 BLUE..see Gauss–Markov assumptions
 book materials.....2–3
 bookstore.....353
 bootstrap 228–229
 box (graph type) 105, 161–163
 box plots.....22, 161–163
 Box–Cox transformation.....227
 bcskew0 227
 break see commands, break
 browse.....291–292
 by prefix..18–19, 60–61, 81–86, 95, 156
 by() (graph option) 131
 by() (tabstat option).....157
 bysort 60–61
 byte (storage type) 100

C
 calculator see pocket calculator
 caption() (graph option) 128–129
 capture.....38–39
 categories 138–139
 cd.....9–10
 center see variables, center
 chi-squared
 likelihood-ratio 143
 Pearson 143
 classification tables.....264–267
 clock position.....121
 cluster samples.....229–230
 cmdlog.....32–34
 CMYK 111
 CNEF 317
 coefficient of determination 191
 collapse 81
 comma-separated values.....see spreadsheet format

command + see commands, break
 command line..see windows, command
 commands
 abbreviations 14, 46–47
 access previous 14
 break 14
 e-class 71
 end of commands 36–37
 external.....45–46, 337, 354
 internal 45–46, 337, 354
 long see do-files, line breaks
 r-class.....71
 search.....23
 comments see do-files, comments
 component-plus-residual plot .. 203–204
 compound quotes 347–348
 compress 319
 compute see generate
 cond() (function) 347
 conditional-effects plot 222–223
 conditions.....see if qualifier
 confidence interval 187–188
 connect() (scatter option) .. 112–115
 connected (plotype) 112–115
 contingency table.....see frequency tables, two-way
 contract 68
 Cook's *D* 209–212
 cooksd (predict option) 209
 correlation
 coefficient 178–179
 negative 178
 positive 178
 weak 178
 count R^2 265–267
 covariate see variables, independent
 covariate pattern 267
 cplot 179
 cprplot 203–204
 Cramer's *V* 143
 cross-tabs see tables
 .csv see spreadsheet format
 Ctrl + Break see commands, break
 Ctrl + C see commands, break

cumulated probability function see
probit model

D

Data Editor 301–302
data matrix 291–292
data region 107
data types see storage type
datasets

ASCII files 294–301
combine 306, 308–319
describe 11–12
export 319–320
hierarchical 85, 313–316
import 293–294
load 10–11
nonmachine-readable 301–306
oversized 322–323
panel data 232–233
preserve 66
rectangular 292
reshape 232–236
restore 66
save 27–28, 319
sort 14–15
titanic 246

dates

combining datasets 305
elapsed dates 93–94
from strings 94

degrees of freedom see *df*

delete see *erase*

density 163, 165–167

describe 8–9, 11–12

destring 90

df 190

DFBETA 205–206

dfbeta 206

dictionary 298–301

dir 10, 27

directory

change 9–10

contents 10

working directory . . 3, 9–10, 58–59

discard 337

discrepancy 210–211, 272–273

discrete (histogram option) 144

display 55, 328–329

distributions

describe 137–174

grouped 150–152

do 26

do-files

analyzing 40

basics 25–27

comments 36

create 40

editors 25–26

error messages 26–27

execute 26

exit 39

from interactive work 29–34

line breaks 36–37

master 40–43

organization 39–43

set more off 37–38

version control 37

doedit 25, 30

dot (graph type) 105, 148–149,
159–161

dot charts 148–149, 159–161

double (storage type) 100

drop 12, 48

dummy variables see variables,
dummy

E

e() (saved results) 71–72

e(b) (saved result) 271

e-class see commands, *e-class*

edit 301–302

egen 87–88

EMF 135–136

Encapsulated PostScript see EPS

encode 90

endogenous variable see variables,
dependent

enhanced metafile see EMF

Epanechnikov kernel 167–168

EPS 135–136

erase.....43, 310
 ereturn list.....72
 error messages
 ignore.....38–39
 invalid syntax.....17
 error-components model.....240–242
 estat bootstrap.....228
 estat classification.....265, 290
 estat dwatson.....217
 estat effects.....230
 estat gof.....268, 290
 estat ic.....267, 277
 Excel files.....292–293
 exit (in Do-Files).....39
 exit Stata.....27–28
 exogenous variable.....see variables,
 independent
 expand.....68
 export.....see datasets, export
 expressions.....55–57
 extensions.....59
F
F test.....191–192
 FAQ.....23, 353
 fence.....162
 filenames.....58–59
 Fisher’s exact test.....143
 five-number summary.....156, 161
 fixed format.....298–301
 fixed-effects model.....236–240
 float() (function).....101
 float (storage type).....100
 foreach.....61–64, 313
 forvalues.....64–65
 free format.....296–298
 frequencies
 absolute.....139–140
 conditional.....141–142
 relative.....139–140
 frequency tables.....20–21
 one-way.....139–140
 two-way.....140–143
 frequency weights.....see weights
 function (plottype).....105

functions.....57
 fweight (weighting type).....66–68
 fysize() (graph option).....133–134
 fysize() (graph option).....133–134
G
 gamma coefficient.....143
 Gauss curve... see normal distribution
 Gauss distribution.....see normal
 distribution
 Gauss–Markov assumptions.....199
 GEE.....242
 generalized estimation equations... see
 GEE
 generate.....24, 75–86
 generate() (tabulate option)....147,
 219
 gladder (statistical graph).....107
 graph.....103–136
 graph region.....107, 116
 graphs
 3D.....107
 combining.....132–134
 connecting points.....113–114
 elements.....107–108
 export.....135–136
 multiple.....129–134
 overlay.....129–131
 print.....134–135
 titles.....128–129
 types.....104–107
 weights.....208
 grid lines.....119, 124–125
 grouping
 by quantiles.....151
 intervals with arbitrary width..152
 intervals with same width....151–
 152
 GSOEP.....4, 11–12, 96–97, 229, 232,
 306–308, 317
H
 help.....22–24
 help files.....351–352
 histogram.....144–146, 163–165

- histogram** (plotype)..... 105
 homoskedasticity.....see regression,
 homoskedasticity
 Hosmer–Lemeshow test..... 268
 Huber/White/sandwich estimator.. 216
- I**
- if qualifier 17, 52–55
imargin() (graph combine option)...
 133
 importing see data, import
 in qualifier..... 14–15, 51–52
infile..... 296–301
 influential cases..... 205–213, 272–275
input 302–304
 inputting data..... 301–306
insheet 295–296
inspect 138
 interaction terms 220–221, 280–281
invnormal() (function)..... 77
iscale() (graph combine option)....
 133
 iteration block..... 262–263
- K**
- kdensity** 165–170
 Kendall’s tau-b..... 143
 kernel density estimator..... 165–170
 key variable..... 309–311
- L**
- label data**..... 319
labels
 and values..... 99–100
 datasets..... 319
 display 99–100
 values..... 21, 98–99
 variables..... 21, 97–98
legend 107–108, 127–128
leverage 210, 272
lfit (plotype)..... 129–130
likelihood 255–256
 likelihood-ratio χ^2 264
 likelihood-ratio test .. 275–277, 279–280
limits 9
- line** (plotype) 108, 112–115
linear combination 180–181
linear probability model.....see
 regression, LPM
linear regression.....see regression
linearity assumption..... 202–204,
 268–272
list 12–14
local 73, 327–331
local macros.....see macros
local mean regression 269
loess see LOWESS
log() (function)..... 77
log (scale suboption) 118–119
log files
 finish recording..... 39
 interrupt recording 32
 log commands 31–34
 SMCL 38
 start recording..... 38–39
logarithm..... 77
logical expressions.....see expressions
logistic..... 261
logistic regression
 coefficients 259–262
 command..... 257–259
 dependent variable 249–254
 diagnostic 268–275
 estimation..... 254–257
 fit..... 263–268
 marginal effect 287
logit..... 257–259
logit model see logistic regression
logits 252–253
loops
 foreach 61–64
 forvalues..... 64–65
lower() (function) 90
LOWESS..... 204, 269–270
lowess (plotype)..... 269–270
lowess (statistical graph)..... 269–270
LPM see regression, LPM

M

macro
 extended macro functions 348–349
 local 73–74, 327–331
manuals 4
margin() (graph option) 116–117
marker
 colors 111–112
 labels 107, 120–122
 options 109–110
 sizes 112
 symbols 107, 110–111
master data 310
match see datasets, combine
matrix (command) 271
matrix (graph type) 105, 202, 205
maximum 16, 155–156
maximum likelihood
 principle 254–257
 search domain 263
mband (plottype) 202–203
mean see average
median 155–156
median regression 212–213, 231–232
median-trace 202–203
memory see RAM
merge 308–316
meta data 313
minimum 16, 155–156
missing
 encode 96–97
missing (tabulate option) 140
missing values see missings
missings
 coding 304–305
 definition 13
 in expressions 54–55
 set 17–18, 96
ML see maximum likelihood
mlabel() (scatter option) 120–122
mlabposition() (scatter option) 121
mlabsize() (scatter option) 121

mlabvposition() (scatter option) 121
MLE see maximum likelihood
mlogit 285–286
more off 37–38
MSS 189–190
multicollinearity 213–214, 218–219
multinomial logistic regression 284–288
mvdecode 18, 96
mvencode 96–97

N

net install 357
NetCourses 354
newlist 63
nonlinear relationships 224–225, 277–278
normal distribution
 density 281–282
 density function 282–283
note() (graph option) 128–129
notes 97
null model 263
numlabel 100
numlist 57–58, 63–64

O

observations
 definition 12
 list 12–14
odds 250–252
odds ratio 251, 260–261
odds-ratio interpretation 260–261
OLS 181–183
operators 55–56
options 19–20, 49–51
order 319
ordered logistic regression see proportional odds model
ordinal logit model see proportional odds model
ordinary least squares see OLS

P

package description 358
 panel data see datasets, panel data
 partial correlation see regression,
 standardized coefficient
 partial regression plot see
 added-variable plot
 partial residual plot see
 component-plus-residual plot
 PDF 135–136
 Pearson residual 267–268
 Pearson- χ^2 267–268
 percentiles see quantiles
 PICT 135–136
 pie (graph type) 105, 148, 159
 pie charts 148, 159
 plot region 107, 116–117
`plotregion()` (graph option) . . . 116–
 117
 PNG 135–136
 pocket calculator 55
 portable document format see PDF
 PostScript see PS
`ppfad.dta` 313
`predict` 186
 predicted values 185–186
 Pregibons $\delta\beta$ see $\delta\beta$
`preserve` 66
 probability interpretation 261, 262
`probit` 283–284
 probit model 281–284
`program define` 331–335
`program drop` 333
 programs
 and do-files 332
 debugging 333–334, 343
 define 331–332
 in do-files 334–335
 naming 333
 redefine 332–333
 syntax 340–343, 346–348
 syntax checks 348
 proportional odds model 289–290
 PS 135–136
 pseudo R^2 263–264

PSID 307, 317
`pwd` 27, 58–59
`pweight` (weighting type) 69–70

Q

quantile plot 170–173
 quantile regression 231
 quantiles 154–156
 quartiles 155–156
`quietly` 345
 Q–Q plots 173–174

R

`r` see correlation coefficient
`r()` (saved results) 71–72
`r(max)` (saved result) 71–72
`r(mean)` (saved result) 71–72
`r(mean)` (saved result) 64
`r(min)` (saved result) 71
`r(N)` (saved result) 71–72
`r(sd)` (saved result) 71–72
`r(sum)` (saved result) 71
`r(sum_w)` (saved result) 71
`r(Var)` (saved result) 71–72
`r-class` see commands, `r-class`
 R^2 191
 RAM 9, 320–322
 random numbers 63, 77–78
 random-effects model 241–242
`range()` (scale suboption) 117–118
 RAW 294
 raw data see RAW
 recode see variables, replace
`recode()` (function) 151–152
`recode` 86–87
 reference lines 107, 119–120
`regress` 25, 183–184
 regression
 ANOVA table 188–190
 autocorrelation 216–217
 coefficient 184–186, 194–195
 command 183–184, 193–194
 control 197–199
 diagnostics 199–217
 fit 190–192

- regression, *continued*
 - homoskedasticity 214–216
 - linear 24–25, 177–243
 - LPM 246–249
 - multiple 192–193
 - nonlinear relationships... 224–226
 - omitted variables 213–214
 - panel data 232–242
 - residuals 186
 - simple 180–183
 - standard error 188
 - standardized coefficient... 196–197
 - with heteroskedasticity ... 226–227
 - replace 24, 75–86
 - reshape 234–236
 - residual (predict option) 186
 - residual
 - definition 180
 - sum 182–183, 189
 - residual sum of squares *see* RSS
 - residual-versus-fitted plot 200–201, 215–216, 227
 - response variable *see* variables, dependent
 - restore 66
 - Results window ... *see* windows, Result window
 - return list 72
 - reverse (scale suboption) 118–119
 - Review window .. *see* windows, Review window
 - RGB 111
 - robust 216
 - root MSE 191
 - round() (function) 215
 - rowmiss() (egen function) 88
 - RSS 189
 - rstudent (predict option) 215
 - running counter 81–82
 - running sum 83–85
 - rvfplot 200–201
- S**
- sampling weights *see* weights
 - SAS files 292–294
 - save 27–28, 319
 - saved results 64, 71–74, 185, 271
 - scatter (plottype) 105, 108
 - scatterplot 177–178
 - scatterplot matrix 202, 205
 - scatterplot smoother 202–203
 - search 359
 - sensitivity 265–266
 - separate 173
 - sign interpretation 260
 - SJ 23, 353
 - SJ-ados 356–357
 - SMCL 38, 351–352
 - SOEP *see* GSOEP
 - sort 14–15
 - sort (scatter option) 114–115
 - specificity 265–266
 - spreadsheet format 294–296
 - SPSS files 292–294
 - SSC 358
 - ssc install 358
 - SSC-ados 358
 - standard deviation 16, 153
 - Stata Journal 353
 - Stata Press 353
 - Stata Technical Bulletin 353
 - stata.toc 358
 - Statalist 353
 - statistical inference 187
 - STB 23, 353
 - STB-ados 356–357
 - stereotype model 288–289
 - storage types 89–90, 100–101
 - strings 297–298, 305
 - display format 95
 - in expressions 90
 - replace substrings 91–92
 - storage type 89–90
 - to dates 94
 - to numeric 90
 - strpos() (function) 90–91
 - subinstr() (function) 92
 - subscripts 84–86
 - substr() (function) 91–92
 - subtitle() (graph option) ... 128–129

sum() (function) 83–85
summarize 15–16, 155–156
summarize() (**tabulate** option) ... 157
 summary graphs 159–161
 summary tables 157–159
 superposition 149, 160–161
 survey data 229–230
svmat 271
svy 229–230
 symmetry plot 214–215
symplot (statistical graph) ... 214–215
syntax 340–343, 346–348
 syntax diagram 45
 system files 292–293

T

tab-separated values ... see spreadsheet
 format
tab1 140
tab2 143
table 157–159
tabstat 156
tabulate 139–143
 tagged-image file format see TIFF
tau-b see Kendall's tau-b
tempvar 350–351
text() (**twoway** option) 121–122
 textbox options 127
 tick lines 107–108, 125–126
 TIFF 135–136
title() (**graph** option) 128–129
 total sum of squares see TSS
 total variance see TSS
trace 333–334
 TSS 189
 two-way table see frequency table,
 two-way
twoway (**graph** type) 105

U

U-shaped relationship 226
 unbalanced panel data 313
uniform() (function) 63, 77–78
update 354–355
 updating Stata 354–355

upper() (function) 90
use 10–11
using 58–59
 using data 310

V

V see Cramer's V
 value labels see labels, values
value label (**axis-label** suboption) ...
 124–125
 variable list see variables, varlist
 variables
 _all 47–48
 allowed names 76–77
 categorical 217–218
 center 32–33, 64, 72, 195, 221
 definition 12
 delete 12, 48
 dependent 177
 dummy 78–79, 147, 194,
 217–219, 278–280
 generate 24, 75–97
 group 150–152
 identifier 304
 independent 177
 multiple codings 305
 names 97
 ordinal 288
 replace 24, 75–97
 temporary 350–351
 transformations 107, 204, 213,
 216, 223–227
 varlist 14, 47–49
 Variables window see windows,
 Variables window
 variance see standard deviation
 variance of residuals see RSS
 variation 189
 varlist see variables, varlist
version 37
view 38

W

weights 65–70
whisker 162

wildcards 48
 windows
 change 26
 Command window 8
 font sizes 8
 Graph window 22
 preferences 8
 Results window 8
 Review window 8, 14
 scroll back 10
 Variables window 8
 windows metafile see WMF
 WMF 135–136
 working directory see directory,
 working directory
 wstata.exe 354

X

xi: 219
 xlabel() (twoway option) 122–125
 xline() (twoway option) 119–120
 xtick() (twoway option) 125–126
 xscale() (twoway option) 117–119
 xsize() (graph option) 116
 xt commands 232
 xtgee 242
 xtick() (twoway option) 125–126
 xtitle() (twoway option) 126–127
 xtreg 239–242

Y

ylabel() (twoway option) 122–125
 yline() (twoway option) 119–120
 ytick() (twoway option) 125–126
 yscale() (twoway option) 117–119
 ysize() (graph option) 116
 ytick() (twoway option) 125–126
 ytitle() (twoway option) 126–127

Z

zip archive 3