

Subject index

Symbols

$\Delta\beta$ 278–279
 $\Delta\chi^2$ 279
 β see regression, standardized coefficient
* see do-files, comments
+ see operators
// see do-files, comments
& see operators
^ see operators
| see operators
|| 104
~ see operators

A

Academic Technology Service .. see ATS
added-variable plot 211–214
additive index 77
adjusted count R^2 270–271
adjusted R^2 199–200
ado-directories 369
ado-files
 basics 343–345
 programming 345–359
aggregate see collapse
AIC 271
Akaike’s information criterion .. see AIC
Aldrich–Nelson’s p^2 334
_all 43
alphanumeric variables ... see strings
analysis of variance see regression
angle() (axis-label suboption) ... 126–127
ANOVA see regression
Anscombe quartet ... 203–204, 206, 231
anycount() (egen function) 85
append 323–326

arithmetic mean see average
arithmetical expressions see expressions
ascategory (graph dot option) ... 153
ASCII files 298–307
ATS 364
augmented component-plus-residual
 plot 209
autocode() (function) 157
autocorrelation see regression, autocorrelation
average 9–10, 157–158, 160–161
avplots 212–213
aweight (weight type) 63–64
axis
 labels 107–108, 124–127
 scales 118–119
 titles 107–108, 128–129
 transformations 119

B

_b[] 189
balanced panel data 318
bands() (twoway mband option) .. 207–208
bar (graph type) ... 105, 109, 150–151, 164
bar charts 150–151, 164
batch jobs see do-files
Bayesian information criterion see BIC
Bernoulli’s distribution ... see binomial distribution
beta see regression, standardized coefficient
bias 205
BIC 271

bin() (**histogram** option) . . . 169–170
 binary variables..see variables, dummy
 binomial distribution 258–259
 BLUE..see Gauss–Markov assumptions
 book materials xxii–xxiii
 bookstore 363
bootstrap 234
box (graph type) 105, 166–168
 box plots 16, 166–168
 Box–Cox transformation 232
bcskew0 232
break see commands, break
browse 297–298
 by prefix . 12–13, 55–56, 79–83, 93, 161
by() (**graph** option) 133
by() (**tabstat** option) 162
bysort 56
byte (storage type) 101

C

calculator see pocket calculator
caption() (**graph** option) . . . 130–131
capture 34
 categories 142–143
cd 3–4
 center see variables, center
 chi-squared
 likelihood-ratio 147
 Pearson 147
 classification tables 268–271
 clock position 121–122
 cluster samples 234–235
cmdlog 28–30
 CMYK 112
 CNEF 324
 coefficient of determination 195
collapse 79
 comma-separated values see
 spreadsheet format
 command + see commands, break
 command line see windows,
 Command window
 commands
 abbreviations 8, 42–43
 access previous 8

commands, *continued*

break 8
 e-class 67
 end of commands 32
 external 41–42, 345, 364
 internal 41–42, 345, 364
 long see do-files, line breaks
 r-class 67
 search 17
 comments see do-files, comments
 component-plus-residual plot . 208–209
 compound quotes 355–356
compress 326
 compute see **generate**
cond() (function) 355
 conditional-effects plot 227–228
 conditions see **if** qualifier
 confidence interval 192
connect() (**scatter** option) . 113–116
connected (plottype) 113–116
 contingency table see frequency
 tables, two-way
contract 63
 Cook's *D* 214–217
cooksd (**predict** option) 214
 correlation
 coefficient 182–183
 negative 182
 positive 182
 weak 182
 count R^2 269–271
 covariate see variables, independent
 covariate pattern 271
cplot 183
cprplot 208–209
 Cramér's *V* 147
.csv see spreadsheet format
Ctrl+Break see commands, break
Ctrl+C see commands, break
 cumulated probability function see
 probit model

D

Data Editor 307–308
 data matrix 297–298

- data region 107
 - data types *see* storage types
 - datasets 298–299
 - ASCII files 300–307
 - combine 312, 314–326
 - describe 5–6
 - export 326–327
 - hierarchical 82, 320–323
 - import 299–300
 - load 4–5
 - nonmachine-readable 307–312
 - oversized 329–330
 - panel data 238
 - preserve 62
 - rectangular 298
 - reshape 238–241
 - restore 62
 - save 21–22, 326
 - sort 8–9
 - titanic 250
 - dates from strings 92–93
 - dates,
 - combining datasets by 311
 - elapsed 91–92
 - degrees of freedom *see* **df**
 - delete *see* **erase**
 - #delimit** 32
 - density 168, 170–172
 - describe** 2–3, 5–6
 - destring** 87–88
 - df** 194–195
 - DFBETA** 210–211
 - dfbeta** 211
 - dictionary 304–307
 - dir** 4, 21
 - directory
 - change 3–4
 - contents 4
 - working directory xxiii, 3–4, 54
 - discard** 345
 - discrepancy 215–216, 276
 - discrete (histogram option) 148
 - display 50, 336–337
 - distributions
 - describe 141–178
 - grouped 154–157
 - do** 20
 - do-files
 - analyzing 35–36
 - basics 19–21
 - comments 31–32
 - create 35–36
 - editors 19–20
 - error messages 20–21
 - execute 20
 - exit 34–35
 - from interactive work 25–30
 - line breaks 32–33
 - master 36–39
 - organization 35–39
 - set more off** 33
 - version control 33
 - doedit** 19, 26
 - dot** (graph type) 105, 152–154, 164–166
 - dot charts 152–154, 164–166
 - double** (storage type) 95, 101
 - drop** 6, 43
 - dummy variables *see* variables, dummy
- E**
- e()** (saved results) 67–68
 - e(b)** (saved result) 275
 - e-class *see* commands, e-class
 - edit** 307–308
 - egen** 84–86
 - elapsed dates *see* dates, elapsed
 - elapsed time *see* time, elapsed
 - EMF 138
 - Encapsulated PostScript *see* EPS
 - encode** 87–88
 - endogenous variable *see* variables, dependent
 - enhanced metafile *see* EMF
 - Epanechnikov kernel 172–173

EPS.....138
 erase.....39, 316
 ereturn list.....68
 error messages
 ignore.....34
 invalid syntax.....11
 error-components model.....245–248
 estat bootstrap.....233
 estat classification.....269
 estat dwatson.....222
 estat effects.....235
 estat gof.....272
 estat ic.....271, 281
 Excel files.....298–299
 exit (in do-files).....34–35
 exit Stata.....21–22
 exogenous variable.....see variables,
 independent
 expand.....63
 expected value.....204–205
 export.....see datasets, export
 expressions.....50–53
 extensions.....54

F
F test.....196
 FAQ.....17, 363
 fence.....167
 filenames.....53–55
 Fisher's exact test.....147
 five-number summary.....160, 166
 fixed format.....304–307
 fixed-effects model.....242–245
 float() (function).....101
 float (storage type).....101
 foreach.....57–59, 319, 353
 levelsof.....353
 forvalues.....60
 free format.....302–303
 frequencies
 absolute.....143–144
 conditional.....145–146
 relative.....143–144

frequency tables.....14–15
 one-way.....143–144
 two-way.....144–147
 frequency weights.....see weights
 function (plottype).....105
 functions.....52–53
 fweight (weight type).....61–63
 fysize() (graph option).....135–136
 fysize() (graph option).....135–136

G

gamma coefficient.....147
 Gauss curve... see normal distribution
 Gauss distribution.....see normal
 distribution
 Gauss–Markov assumptions.....203
 GEE.....247–248
 generalized estimation equations... see
 GEE
 generate.....18, 73–83
 generate() (tabulate option)....151,
 224
 gladder (statistical graph).....107
 graph.....103–138
 graph region.....107, 117
 graphs
 3-D.....107
 combining.....134–136
 connecting points.....114–115
 editor..108–109, 122–124, 136–137
 elements.....107–109
 export.....137–138
 multiple.....131–136
 overlay.....131–133
 print.....136–137
 titles.....130–131
 types.....104–107
 weights.....213
 grid lines.....120, 126–127
 grouping
 by quantiles.....155
 intervals with arbitrary width..157
 intervals with same width....156–
 157

GSOEP.....xxiii, 5–6, 97–98, 234, 237,
312–314, 324

H

help.....16–18
help files.....359–360
histogram.....148–150, 168–170
histogram (plottype).....105
homoskedasticity.....see regression,
homoskedasticity
Hosmer–Lemeshow test.....272
Huber/White/sandwich estimator..221

I

if qualifier.....11, 48–50
imargin() (graph combine option)...
.....135
importing.....see data, import
in qualifier.....8–9, 47–48
infile.....302–307
influential cases.....209–218, 276–279
input.....308–310
inputting data.....307–312
insheet.....301–302
inspect.....142
interaction terms....225–226, 284–285
invnormal() (function).....75
iscale() (graph combine option)....
.....135
iteration block.....266–267

K

kdensity.....170–175
Kendall's tau-b.....147
kernel density estimator.....170–175
key variable.....315–317

L

label data.....326
labels
and values.....100
datasets.....326
display.....100
values.....15, 99
variables.....15, 98–99

legend.....107–108, 129–130
leverage.....215, 276
lfit (plottype).....131–132
likelihood.....259–260
likelihood-ratio χ^2268
likelihood-ratio test..279–281, 283–284
limits.....3
line (plottype).....109, 113–116
linear probability model.....see
regression, LPM
linear regression.....see regression
linearity assumption.....206–209,
272–276
list.....6–8
local.....69, 335–339
local macro.....see macro
local mean regression.....273
loess.....see LOWESS
log() (function).....75
log (scale suboption).....119
log files
finish recording.....34
interrupt recording.....28
log commands.....27–30
SMCL.....33–34
start recording.....33–34
logarithm.....75
logical expressions.....see expressions
logistic.....265
logistic regression
coefficients.....263–266
command.....262–263
dependent variable.....253–258
diagnostic.....272–279
estimation.....258–261
fit.....267–272
marginal effect.....291
logit.....262–263
logit model.....see logistic regression
logits.....256–258
loops
foreach.....57–59
forvalues.....60
lower() (function).....88
LOWESS.....209, 273–274

lowess (plottype).....273–274
lowess (statistical graph).....273–274
LPM..... see regression, LPM

M

macro

 extended macro functions.... 356–357
 local 69–70, 335–339
manuals xxiv
margin() (graph option) 117
marker
 colors 112
 labels 107, 120–122
 options 110
 sizes 113
 symbols 107, 111–112
master data 316
match see datasets, combine
matrix (command) 275
matrix (graph type) 105, 206–207, 210

maximum 10, 160–161
maximum likelihood

 principle 258–261
 search domain 267
mband (plottype) 207–208
mean see average
median 159–161
median regression.... 217–218, 236–237
median-trace 207–208
memory see RAM
merge 314–323
 _merge (variable) 316–318
meta data 319
minimum 10, 160–161
missing

 encode 97
missing (**tabulate** option) 144
missing values see missings
missings
 coding 310–311
 definition 7
 in expressions 49–50
 set 11–12, 96–97

ML..... see maximum likelihood
mlabel() (**scatter** option) ... 120–122
mlabposition() (**scatter** option)
 121–122
mlabsize() (**scatter** option) 121–122
mlabvposition() (**scatter** option) ...
 122
MLE see maximum likelihood
mlogit 289–290
more off 33
MSS 194
multicollinearity 218–219, 223–224
multinomial logistic regression 288–292
mvdecode 12, 97
mvencode 97

N

_N 80–81, 93
_n 79–80, 93
net install 366–367
NetCourses 364
newlist 58–59
nonlinear relationships 229–230, 281–282
normal distribution
 density 285–286
 density function 286–287
note() (**graph** option) 130–131
notes 98
null model 267
numlabel 100
numlist 53, 59

O

observations

 definition 6
 list 6–8
odds 254–256
odds ratio 255, 264–265
odds-ratio interpretation 264–265
OLS 186–188
operators 51–52
options 13–14, 45–46

order 326
 ordered logistic regression *see*
 proportional-odds model
 ordinal logit model *see*
 proportional-odds model
 ordinary least squares *see* OLS

P

package description 368
 panel data *see* datasets, panel data
 PanelWhiz 312
 partial correlation *see* regression,
 standardized coefficient
 partial regression plot *see*
 added-variable plot
 partial residual plot *see*
 component-plus-residual plot
 PDF 138
 Pearson residual 271–272
 Pearson- χ^2 271–272
 percentiles *see* quantiles
 PICT 138
 pie (graph type) 105, 152, 164
 pie charts 152, 164
 plot region 107, 117
 plotregion() (graph option) 117
 PNG 138
 pocket calculator 50
 portable document format *see* PDF
 PostScript *see* PS
 ppfad.dta 319
 predict 190
 predicted values 190
 Pregibon's $\Delta\beta$ *see* $\Delta\beta$
 preserve 62
 probability interpretation 265, 266
 probit 287–288
 probit model 285–288
 program define 339–343
 program drop 341
 programs
 and do-files 340
 debugging 341–342, 351
 defining 339–340
 in do-files 342–343

programs, *continued*

 naming 341
 redefining 340–341
 syntax 348–351, 354–356
 proportional-odds model 293–294
 PS 138
 pseudo R^2 268
 PSID 313, 324
 pwd 21, 54
 pweight (weight type) 64–65

Q

quantile plot 175–177
 quantile regression 236
 quantiles 159–161
 quartiles 159–161
 quietly 353
 Q–Q plots 178

R

r *see* correlation coefficient
 r() (saved results) 67–68
 r(max) (saved result) 67–68
 r(mean) (saved result) 67–68
 r(mean) (saved result) 59
 r(min) (saved result) 67
 r(N) (saved result) 67–68
 r(sd) (saved result) 67–68
 r(sum) (saved result) 67
 r(sum_w) (saved result) 67
 r(Var) (saved result) 67–68
 r-class *see* commands, r-class
 R^2 195
 RAM 3, 327–329
 random numbers 58–59, 75
 random-effects model 247
 range() (scale suboption) 118
 RAW 300
 raw data *see* RAW
 recode *see* variables, replace
 recode() (function) 156
 recode 84
 reference lines 107, 120
 regress 19, 188

regression

ANOVA table 192–195
autocorrelation 221–222
coefficient 188–190, 198–199
command 188, 197–198
control 201–203
diagnostics 203–222
fit 195–196
homoskedasticity 219–221
linear 19, 181–248
LPM 250–253
multiple 196–197
nonlinear relationships 229–231
omitted variables 218–219
panel data 237–248
residuals 190–191
simple 184–188
standard error 192
standardized coefficient ... 200–201
with heteroskedasticity ... 231–232
replace 18, 73–83
reshape 239–241
residual (predict option) ... 190–191
residual
 definition 185
 sum 187, 194
residual sum of squares see RSS
residual-versus-fitted plot 205–206,
 220–221, 232
response variable see variables,
 dependent
restore 62
Results window .. see windows, Results
 window
return list 68
reverse (scale suboption) 119
Review window .. see windows, Review
 window
RGB 112
root MSE 195–196
round() (function) 220
rowmiss() (*egen* function) 85
RSS 194
rstudent (predict option) 220
runiform() (function) 58–59, 75

running counter 79–80
running sum 81–83
rvfplot 205–206

S

sampling weights see weights
SAS files 298–300
save 21–22, 326
saved results ... 59, 67–70, 189–190, 275
scatter (plotype) 105, 109
scatterplot 181–182
scatterplot matrix 206–207, 210
scatterplot smoother 207–208
search 369
sensitivity 269–270
separate 178
sign interpretation 264
SJ 17, 363
SJ-ados 366–367
SMCL 33–34, 360
SOEP see GSOEP
sort 8–9
sort (**scatter** option) 115–116
specificity 269–270
spreadsheet format 300–302
SPSS files 298–300
SSC 367–368
ssc install 368
SSC-ados 367–368
standard deviation 10, 158
Stata Journal 363
Stata Press 363
Stata Technical Bulletin 363
stata.toc 368
Statalist 363
statistical inference 191, 232–235
STB 17, 363
STB-ados 366–367
stereotype model 292–293
storage types 86–87, 100–101
strings 303, 311
 display format 93
 in expressions 88
 replace substrings 89–90
 storage type 86–87

strings, *continued*

- to dates 92–93
- to numeric 87–88
- strpos()** (function) 88–89
- substr()** (function) 90
- subscripts 82–83
- substr()** (function) 89–90
- subtitle()** (graph option) ... 130–131
- sum()** (function) 81–83
- summarize** 9–10, 160
- summarize()** (**tabulate** option) .. 161–162
- summary graphs 164–166
- summary tables 161–164
- superposition 153, 165–166
- survey data 234–235
- svmat** 275
- svy** 235
- symmetry plot 219–220
- sympplot** (statistical graph) ... 219–220
- syntax** 348–351, 354–356
- syntax checks 349
- syntax diagram 41

T

- tab-separated values ... see spreadsheet format
- tab1** 144
- tab2** 147
- table** 162–164
- tabstat** 160–161
- tabulate** 143–147
- tagged-image file format see TIFF
- tau-b see Kendall's tau-b
- tempvar** 358–359
- text()** (**twoway** option) 122
- textbox options 129
- tick lines 107–108, 127–128
- TIFF 138
- time from strings 95
- time, elapsed 93–96
- title()** (graph option) 130–131
- total sum of squares see TSS
- total variance see TSS
- trace** 341–342

- TSS 193
- two-way table see frequency tables, two-way
- twoway** (graph type) 105

U

- U-shaped relationship 208, 231
- unbalanced panel data 319
- uniform()** (function) .. see **runiform()** (function)
- update** 364–365
- updating Stata 364–365
- upper()** (function) 88
- use** 4–5
- using** 53–55
- using data 316

V

- V** see Cramér's *V*
- value labels see labels, values
- valuelabel** (axis-label suboption) 126–127
- variable list see variables, varlist
- variables
 - _all** 43
 - allowed names 74–75
 - categorical 222–223
 - center 28–29, 59, 68, 199, 226
 - definition 6
 - delete 6, 43
 - dependent 181
 - dummy 76–77, 151, 198, 222–224, 282–284
 - generate 18, 73–97
 - group 154–157
 - identifier 310
 - independent 181
 - multiple codings 311
 - names 98
 - ordinal 292
 - replace 18, 73–97
 - temporary 358–359
 - transformations 107, 209, 218, 221, 228–232
 - varlist 8, 43–45

Variables window see windows,
 Variables window
 variance see standard deviation
 variance of residuals see RSS
 variation 193
 varlist see variables, varlist
 vce(robust) 221
 version 33
 view 33–34

W

webuse xxiv
 weights 60–65
 whisker 167
 wildcards 44
 windows
 change 20
 Command window 1
 font sizes 2
 preferences 2
 Results window 1
 Review window 1, 8
 scroll back 4
 Variables window 1
 windows metafile see WMF
 WMF 138
 working directory see directory,
 working directory
 wstata.exe 364

X

xi: 224
 xlabel() (twoway option) 124–127
 xline() (twoway option) 120
 xtick() (twoway option) 127–128
 xscale() (twoway option) 118–119
 xsize() (graph option) 117
 xt commands 237
 xtgee 247–248
 xtick() (twoway option) 127–128
 xtitle() (twoway option) 128–129
 xtreg 245, 247

Y

ylabel() (twoway option) 124–127

ylines() (twoway option) 120
 ytick() (twoway option) 127–128
 yscale() (twoway option) 118–119
 ysize() (graph option) 117
 ytick() (twoway option) 127–128
 ytitle() (twoway option) 128–129

Z

zip archive xxiii