

Subject index

Symbols

$\Delta\beta$ 375–376
 $\Delta\chi^2$ 376–377
 β see regression, standardized coefficient
* see do-files, comments
+ see operators
== see operators
// see do-files, comments
(factor-variable operator) 307–308
(factor-variable operator) 308
& see operators
^ see operators

A

Academic Technology Service.. see ATS
added-variable plot 289–290
additive index 81
adjusted count R^2 366–367
adjusted R^2 274–275
ado-directories 474–475
ado-files
 basics 447–449
 programming 449–465
aggregate see collapse
Akaike’s information criterion 367
Aldrich–Nelson’s p^2 438
_all 43
alphanumeric variables... see strings
analysis of variance see regression

angle() (axis-label suboption) ... 141–142
ANOVA see regression
Anscombe quartet 279–280, 282, 311–312
anycount() (egen function) 92–93
append 429–432
arithmetic mean see average
arithmetical expressions see expressions
ascategory (graph dot option) ... 170
ASCII files 397–398, 402–410
ATE see average treatment effect
ATS 470
augmented component-plus-residual
 plot 285
autocode() (function) 173
autocorrelation see regression, autocorrelation
average 9–10, 174, 177
average marginal effects .. 361–362, 383
average partial effects see average marginal effects
average treatment effect 245–246
avplots 289
aweight (weight type) 66–67
axis,
 labels 119–120, 139–142
 scales 131–133
 titles 119–120, 143–144
 transformations 132–133

B

_b[name] 75, 262
balanced panel data 424
balanced repeated replication 219

- bands()** (**twoway mband** option) .. 283–284
bar (graph type) 117–118, 121, 166–168, 184
 bar charts 166–168, 184
 base category 302
 batch jobs *see* do-files
 Bayesian information criterion 367
bcskew0 312
 Bernoulli's distribution ... *see* binomial distribution
 beta *see* regression, standardized coefficient
 beta distribution 204
 bias 209, 281
bin() (**histogram** option) 190–191
 binary variables .. *see* variables, dummy
 binomial distribution 351–352
 BLUE .. *see* Gauss–Markov assumptions
 book materials xxii–xxiii
 bookstore 469
 boolean expressions 50, 53–54
 bootstrap 219
box (graph type) 117–118, 187–189
 box plots 15–16, 187–189
 Box–Cox transformation 312
 break *see* commands, break
browse 396
 by prefix 12–13, 58–59, 86–91, 102, 178
by() (**graph** option) 149
by() (**tabstat** option) 179–180
bysort 59
 byte (storage type) 111
- C**
c. (factor-variable operator) 308
 calculator *see* pocket calculator
caption() (**graph** option) 146
capture 34
 categories 158–159
 causality 244–246
cd 3–4
 center *see* variables, center
- central limit theorem 212–213
 chi-squared
 likelihood-ratio 163
 Pearson 162
 classification tables 364–367
 clock position 135–136
cmdlog 28–30
 CMYK 125
 CNEF 430
 coefficient of determination 269
collapse 86
 comma-separated values *see* spreadsheet format
 command + *see* commands, break
 command line *see* windows, Command window
 commands
 abbreviations 8, 42
 access previous 8
 break 8
 e-class 71
 end of commands 32
 external 41–42, 449
 internal 41–42, 449
 long *see* do-files, line breaks
 r-class 71–72
 search 16
 comments *see* do-files, comments
 component-plus-residual plot .. 285–286
 compound quotes 460
compress 433
 compute *see* generate
cond() (function) 84, 460
 conditional-effects plot ... 273–274, 360, 380–381
 conditions *see* if qualifier
 confidence intervals 231–232
connect() (**scatter** option) .. 126–129
connected (plottype) 126–129
 contingency table *see* frequency tables, two-way
contract 66
 Cook's *D* 290–294
cooksdi (**predict** option) 293

- correlation
 - coefficient 254–255
 - negative 254
 - positive 253–254
 - weak 254
- count R^2 365–367
- Counterfactual concept of causality ...
 - see causality
- covariate ... see variables, independent
- covariate pattern 367
- cplot 255
- cprplot 285–286
- Cramér's V 163
- .csv see spreadsheet format
- Ctrl+Break* see commands, break
- Ctrl+C* see commands, break
- cumulated probability function see
 - probit model

- D**
- Data Editor 410–411
- data matrix 395–397
- data region 119
- data types see storage types
- data-generating process 242–243
- data1.dta 5
- datasets 397–398
 - ASCII files 402–410
 - combine 415, 417–432
 - describe 5–6
 - export 433
 - hierarchical 89–90, 425–429
 - import 398–402
 - load 4
 - nonmachine-readable 410–415
 - oversized 435
 - panel data 328–329
 - preserve 126
 - rectangular 396
 - reshape 328–332
 - restore 126
 - save 22, 433
 - sort 9
- date() (function) 100–101

- dates
 - combining datasets by 414
 - display format 99
 - elapsed 99–100
 - from strings 100–101
 - variables 98–102
- .dct (extension) 57
- decode 319
- DEFF see design effect
- DEFT see design effect
- degrees of freedom see df
- delete see erase
- #delimit 32
- density 189, 191–193
- describe 2, 5–6
- design effect 221–222
- destring 95
- df 268
- dfbeta 287–288
- diary 102–103
- dictionary 407–410
- dir 4, 21
- directory
 - change 3–4
 - contents 4, 21
 - working directory xxiii, 3–4, 57
- discard 449
- discrepancy 292, 373
- discrete (histogram option) 164
- display 51, 441
- distributions
 - describe 157–199
 - grouped 171–173
- .do (extension) 57
- do 21
- do-files
 - analyzing 35
 - basics 20–21
 - comments 31–32
 - create 35–36
 - editors 20–21
 - error messages 21
 - execute 21
 - exit 34–35
 - from interactive work 25–30

- do-files, *continued*
 - line breaks 32
 - master 36–39
 - organization 35–39
 - set more off 33
 - version control 33
 - doedit 20, 26
 - dot (graph type) 117–118, 169–170, 184–186
 - dot charts 169–170, 184–186
 - double (storage type) 103, 111
 - drop 6, 43
 - .dta (extension) 57
 - dummy variables *see variables, dummy*
- E**
- e() (saved results) 71–76
 - e(b) 75
 - e(V) 75
 - e-class *see commands, e-class*
 - edit 410–411
 - efficiency 209–210
 - egen 92–94
 - elapsed dates *see dates, elapsed*
 - elapsed time *see time, elapsed*
 - EMF 154
 - Encapsulated PostScript *see EPS*
 - encode 95
 - endogenous variable *see variables, dependent*
 - enhanced metafile *see EMF*
 - Epanechnikov kernel 194
 - EPS 154
 - erase 38–39
 - ereturn list 74
 - error messages
 - ignore 34
 - invalid syntax 11
 - error-components model 337–339
 - estat classification 365
 - estat dwatson 300
 - estat gof 368
 - estat ic 367, 379
 - estimates restore 75
 - estimates store 75
 - esttab 315
 - Excel files 397–401
 - exit (in do-files) 34–35
 - exit Stata 22
 - exogenous variable *see variables, independent*
 - exp() (function) 54
 - expand 66
 - expected value 280–281
 - export *see datasets, export*
 - expressions 51–55
 - extensions 57
- F**
- F test 270
 - factor variables .. 29, 303–304, 307–308, 380, 382, 383
 - FAQ 16, 469
 - filefilter 404
 - filenames 56–57
 - Fisher’s exact test 162
 - five-number summary 177, 187
 - fixed format 407–410
 - fixed-effects model 332–337
 - float() (function) 112
 - float (storage type) 111
 - foreach 59–62, 424–425, 458
 - levelsof 458
 - forvalues 62–63, 208–209
 - free format 405–406
 - frequencies
 - absolute 159–160
 - conditional 161–162
 - relative 159–160
 - frequency tables 14–15
 - one-way 159–160
 - two-way 160–163
 - frequency weights *see weights*
 - function (plotype) 117–118
 - functions 54–55, 82–85
 - fweight (weight type) 64–66
 - fxsize() (graph option) 151–152
 - fysize() (graph option) 151–152

G

gamma coefficient 162
 Gauss curve ... see normal distribution
 Gauss distribution.....see normal distribution
 distribution
 Gauss–Markov assumptions.....279
 GEE.....338–339
 generalized estimation equations ... see
 GEE
 generate 17–18, 77–91
 generate() (tabulate option)....168,
 302–303
 gladder (statistical graph) 119
 .gph (extension).....57
 graph.....115–154
 graph region.....119, 130
 graphs
 3-D.....119
 combining.....150–152
 connecting points.....127–128
 editor..120–121, 137–138, 152–153
 elements.....119–121
 export.....153–154
 multiple.....147–152
 overlay.....147–148
 print.....152–153
 titles.....146
 types.....116–119
 weights.....290
 grid lines.....133, 141–142
 grouping
 by quantiles.....172
 intervals with arbitrary width..173
 intervals with same width....172–
 173
 GSOEP.....xxiii, 5, 106, 108, 219–220,
 223, 327, 415–417, 430

H

help.....16–17
 help files.....465–467
 histogram.....164–166, 190–191
 histogram (plottype).....117–118

homoskedasticity.....see regression,
 homoskedasticity
 Hosmer–Lemeshow test.....368
 Huber/White/sandwich estimator..298

I

i. (factor-variable operator)..303–304,
 382
 if qualifier.....11, 48–51
 imargin() (graph combine option)...
 151
 import.....see datasets, import
 importing.....see data, import
 in qualifier.....9, 47–48
 inference
 causal.....201, 242–249
 descriptive.....201, 213–241
 statistical.....201–249
 infile.....405–410
 influential cases.....286–295, 372–377
 inlist() (function).....55, 83–84
 input.....411–413
 inputting data.....410–415
 inrange() (function).....55, 83–84
 insheet.....403–404
 inspect.....158
 int() (function).....202–203
 interaction terms 304–308, 381–384
 invnormal() (function) ... 80, 204–205
 irecode() (function).....84–85
 iscale() (graph combine option)....
 151
 item nonresponse see nonresponse
 iteration block.....362–363

J

jackknife.....219
 jitter() (scatter option).....343

K

kdensity.....194, 205, 210–211
 kdensity (plottype).....194–195
 Kendall's tau-b.....163
 kernel density estimator.....191–195
 key variable.....420–421

L

label data 433
labels,
 and values 110–111
 datasets 433
 display 110–111
 values 18–19, 108–109
 variables 18, 108
legend 119–120, 144–145
leverage 292, 373
lfit (plottype) 147
likelihood 352
likelihood-ratio χ^2 364
likelihood-ratio test 377–379
limits 3
line (plottype) 121, 126–129
linear probability model *see*
 regression, LPM
linear regression *see* regression
linearity assumption 283–286,
 369–372
list 7–8
ln() (function) 54
local 73, 439–443
local macro *see* macro
local mean regression 369–370
loess *see* LOWESS
log() (function) 79–80
log (scale suboption) 132–133
log files
 finish recording 34
 interrupt recording 28–29
 log commands 27–30
 SMCL 33–34
 start recording 33–34
logarithm 79–80
logical expressions *see* expressions
logistic 358
logistic regression,
 coefficients 356–360
 command 354–356
 dependent variable 346–350
 diagnostic 368–377
 estimation 351–354
 fit 363–368

logit 354–356
logit model *see* logistic regression
logits 349–350
loops
 foreach 59–62
 forvalues 62–63
lower() (function) 95
LOWESS 285, 370
lowess (plottype) 370
lowess (statistical graph) 370
LPM *see* regression, LPM

M

m:1 match *see* 1:m match
macro
 extended macro functions 461–
 462
 local 73–74, 439–443
manuals xxiv
margin() (graph option) 130–131
marginal effects 361–362
margins ... 264–266, 273–274, 318, 359,
 371–372, 380–381, 383
marginsplot ... 265–266, 273–274, 318,
 360, 371–372, 380–381
marker,
 colors 125
 labels 119, 134–136
 options 123
 sizes 126
 symbols 119, 123–125
master data 420
match *see* datasets, combine
match type 420
matrix (graph type) 117–118, 283,
 286–287
maximum 10, 176–177
maximum likelihood
 principle 351–354
 search domain 363
mband (plottype) 283–284
mdy() (function) 99
mdyhms() (function) 104
mean *see* average
mean 214, 220–221, 239

mean comparison test 238–241
 median 176–177
 median regression 295, 324–327
 median-trace 283–284
 memory *see* RAM
 merge 417–429
`_merge` (variable) 420–421, 423–424
 metadata 424
`mi` 228–230
`mi()` *see* `missing()`
 minimum 10, 176–177
`minutes()` (function) 104–105
`missing` (tabulate option) 160
 missing values *see* `missings`
`missing()` (function) 54–55, 305
`missings`
 `coding` 413–414
 definition 7–8
 `encode` 107
 in expressions 50–51
 `set` 11–12, 105–106
 ML *see* maximum likelihood
`mlabel()` (`scatter` option) ... 134–136
`mlabposition()` (`scatter` option)
 135–136
`mlabsize()` (`scatter` option) ... 135–
 136
`mlabvposition()` (`scatter` option) ...
 136
 MLE *see* maximum likelihood
`mlogit` 388–389
 more off 33, 209
 MSS 268
 multicollinearity 296, 302
 multinomial logistic regression ... 387–
 390
 multiple imputation 223–230
`mvdecode` 12, 106
`mvencode` 107

N
`_N` 87–88, 102
`_n` 87, 102
`net install` 472–473
 NetCourses 469–470

`newlist` 61
 nonlinear relationships 309–310,
 379–381
 nonresponse 222
 normal distribution 204–206, 212
 `density` 385
 `density function` 385–386
`normal()` (function) 235–236, 239
`note()` (`graph` option) 146
 notes 107
 null hypothesis 234, 239
 null model 363
`numlabel` 111
`numlist` 55–56, 61–62

O

observations
 definition 5–6
 `list` 7–8
 odds 347–349
 odds ratio 347–348, 357–358
 odds-ratio interpretation 357–358
 OLS 258–260
 operators 52–53
 options 13–14, 45–47
`order` 433
 ordered logistic regression *see*
 proportional-odds model
 ordinal logit model *see*
 proportional-odds model
 ordinary least squares *see* OLS
`outreg` 315
`outreg2` 315

P

package description 474
 panel data ... *see* datasets, panel data
 PanelWhiz 415
 partial correlation *see* regression,
 standardized coefficient
 partial regression plot *see*
 added-variable plot
 partial residual plot *see*
 component-plus-residual plot
 PDF 154

Pearson residual 367–368
 Pearson- χ^2 367–368
 percentiles see quantiles
 PICT 154
 pie (graph type) 117–118, 168–169,
 184
 pie charts 168–169, 184
 plot region 119, 130–131
 plotregion() (graph option) 130–
 131
 PNG 154
 pocket calculator 51
 portable document format see PDF
 post-stratification weights .. see weights
 PostScript see PS
 ppfad.dta 424
 PPS see sample, PPS
 predict 264
 predicted values 263–266
 Pregibon's $\Delta\beta$ see $\Delta\beta$
 preserve 126
 primary sampling unit 215, 220
 probability density function 231
 probability interpretation 359, 360
 probability proportional to size see
 sample, PPS
 probit 387
 probit model 385–387
 program define 443–447
 program drop 445
 programs
 and do-files 444
 debugging 445–446, 455–456
 defining 443–444
 in do-files 446–447
 naming 445
 redefining 445
 syntax 453–456, 459–461
 proportional-odds model 392–393
 PS 154
 pseudo R^2 363–364
 PSID 416, 430
 PSU see primary sampling unit
 pwd 21, 57
 pweight (weight type) 67–68

Q

quantile plot 195–198
 quantile regression 325
 quantiles 175–177
 quartiles 176–177
 quietly 458
 Q–Q plots 199

R

r see correlation coefficient
 r() (saved results) 71–72
 r(max) (saved result) 71–72
 r(mean) (saved result) 62, 71–72
 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^2 269
 RAM 3, 434–435
 random numbers 61, 80, 202–203
 random-effects model 338
 range() (scale suboption) 131–132
 RAW 402–403
 .raw (extension) 57
 raw data see RAW
 rbeta() (function) 204, 243
 recode see variables, replace
 recode() (function) 84–85, 173
 recode 91–92
 reference lines 119, 133–134
 regress 19, 260–261
 regression
 ANOVA table 266–268
 autocorrelation 299–300
 coefficient 261–266, 272–274
 command 260–261, 271–272
 control 277–278
 diagnostics 279–300
 fit 268–270
 homoskedasticity 296–298
 linear 19, 253–340
 LPM 342–346

- regression, *continued*
 - multiple 270–271
 - nonlinear relationships... 309–312
 - omitted variables 295–296
 - panel data..... 327–339
 - residuals 266
 - simple 256–260
 - standardized coefficient ... 276–277
 - with heteroskedasticity ... 312–313
- replace 17–18, 77–91
- reshape 330–332
- residual (predict option) 266
- residual
 - definition 257
 - sum 259–260, 267–268
- residual sum of squares *see* RSS
- residual-versus-fitted plot 281–282, 297–298, 312–313, 345
- response variable *see* variables, dependent
- restore 126
- Results window .. *see* windows, Results window
- resultsset 319
- return list 72
- reverse (scale suboption) 132–133
- Review window .. *see* windows, Review window
- RGB 125
- rnormal() (function) 205–206
- root MSE 269
- round() (function) 82–83
- rowmiss() (egen function) 93
- RSS 267–268
- rstudent (predict option) 297
- Rubin causal model *see* causality
- runiform() (function) 61, 80, 202–203, 207
- running counter 87
- running sum 88–90
- rvfplot 281–282

- S**
- sample
 - PPS 216–217
- sample, *continued*
 - SRS 208, 213–215
 - cluster 215
 - multistage 216
 - stratified 217
 - two-stage 215–216
- sample 207
- sampling distribution 208–213, 217–218
- sampling probability 208, 215–217
- sampling weights *see* weights
- SAS files 397, 402
- save 22, 433
- saved results 62, 71–74, 262–263
- scatter (plotype) 117–118, 121
- scatterplot 253–254
- scatterplot matrix 283, 286–287
- scatterplot smoother 283–284
- _se [name] 75
- search 475
- secondary sampling unit 215
- selection probability *see* sampling probability
- sensitivity 365–366
- separate 199
- set obs 203
- set seed 202
- sign interpretation 357
- significance test 233–241
- simple random sample *see* sample, SRS
- SJ 16, 469
- SJ-ados 471–473
- SMCL 33–34, 466–467
- .smcl (extension) 57
- SOEP *see* GSOEP
- soepuse 415
- sort 9
- sort (scatter option) 128–129
- specificity 365–366
- spreadsheet format 402–404
- SPSS files 397, 402
- sqrt() (function) 54
- SRS *see* sample, SRS
- SSC 473

ssc install..... 473
 SSC-ados 473
 SSU see secondary sampling unit
 standard deviation..... 10, 174
 standard error..... 210, 213–215, 239
 standard normal distribution..... see
 normal distribution
Stata Journal..... 469
 Stata Press..... 469
Stata Technical Bulletin..... 469
stata.toc 474
 Statalist..... 469
statsby..... 319
 STB..... 16, 469
 STB-ados..... 471–473
 stereotype model..... 391–392
 storage types..... 94, 111–112
 strings..... 406, 414
 in expressions..... 95
 replace substrings..... 96–98
 storage type..... 94–95
 to dates..... 100–101
 to numeric..... 95
 variables..... 94–98
strpos() (function)..... 95–96
 student's *t* distribution..... 239–240
subinstr() (function)..... 97–98
 subscripts..... 89–91
substr() (function)..... 96–97
subtitle() (graph option)..... 146
sum() (function)..... 88–90
summarize..... 9–10, 176–177
summarize() (tabulate option).. 178–
 179
 summary graphs..... 184–186
 summary tables..... 178–184
 superposition..... 170, 185–186
 svy (prefix)..... 220–222, 240–241
svyset..... 219–220
 symmetry plot..... 296–297
sympplot (statistical graph).... 296–297
syntax..... 453–456, 459–461
 syntax checks..... 453
 syntax diagram..... 41

T

tab-separated values... see spreadsheet
 format
tab1 160
tab2 163
table..... 180–184
tabstat..... 177
tabulate..... 159–163
 tagged-image file format..... see TIFF
 tau-b..... see Kendall's tau-b
 Taylor linearization..... 219
tempvar 463–465
test..... 240–241
 test statistic..... 235–236
text() (**twoway** option)..... 136
 textbox options..... 144
 tick lines..... 119–120, 142–143
 TIFF..... 154
 time
 display-format..... 104
 elapsed..... 102–105
 from strings..... 104
title() (**graph** option)..... 146
 total sum of squares..... see TSS
 total variance..... see TSS
trace..... 446
 TSS..... 267
ttest..... 240
 two-way table... see frequency tables,
 two-way
twoway (graph type)..... 117–118
 .txt (extension)..... 57

U

U-shaped relationship... 284, 311–312
uniform() (function).. see **runiform()**
 (function)
 uniform distribution..... 202–204
 unit nonresponse..... see nonresponse
update..... 470–471
 updating Stata..... 470–471
upper() (function)..... 95
use..... 4
using..... 56–57
 using data..... 420

V

V see Cramér's *V*
 value labels see labels, values
valueLabel (axis-label suboption)
 141–142
 variable list see variables, varlist
 variables
 _all 43
 allowed names 78–79
 categorical 301, 341
 center 29, 62, 72, 272–274, 305
 definition 5–6
 delete 6, 43
 dependent 253
 dummy 80–81, 167–168, 272,
 301–304
 generate 17–18, 77–107
 group 171–173
 identifier 413
 independent 253
 multiple codings 414
 names 107–108
 ordinal 391
 replace 17–18, 77–107
 temporary 463–465
 transformations 119, 286, 295,
 298, 308–313
 varlist 8, 43–45
 Variables window see windows,
 Variables window
 variance see standard deviation
 variance of residuals see RSS
 variation 267
 varlist see variables, varlist
vce(robust) 298
version 33
view 33–34

W

webuse xxiv
 weights 63–68, 221–223
 whisker 187
 wildcards 44
 windows
 change 21

windows, *continued*

 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 154
 working directory see directory,
 working directory

X

xlabel() (**twoway** option) 139–142
xline() (**twoway** option) 133–134
.xls see Excel files
xtick() (**twoway** option) 142–143
xscale() (**twoway** option) 131–133
xsize() (**graph** option) 130
 xt commands 328
xtgee 338–339
xtick() (**twoway** option) 142–143
xtitle() (**twoway** option) 143–144
xtreg 336–338

Y

ylabel() (**twoway** option) 139–142
yline() (**twoway** option) 133–134
ytick() (**twoway** option) 142–143
yscale() (**twoway** option) 131–133
ysize() (**graph** option) 130
ytick() (**twoway** option) 142–143
yttitle() (**twoway** option) 143–144

Z

z transformation 235
 zip archive xxiii