Subject index

Symbols

β weights 216
η 184, 194
η^2 184
ϕ 109

A

acquiring datasets 275
`add`, `label define` option 33
`adjust`, postestimation command . . 195, 200
adjusted means 195
agreement, intraclass correlation . . . 207
`alpha` command 169–172
alpha reliability 168
analysis of covariance . . . . . *see* ANCOVA
analysis of variance . . . . . *see* ANOVA
ANCOVA 191–198
`angle()`, `twoway` option 200–202
ANOVA,
  degrees of freedom 183
  equal-variance test 183
  one-way 177–178
  postestimation adjusted means 195
  repeated-measures 202
  residual 194
  two-way 198–202
ANOVA assumptions 178
ATS at UCLA 159

B

bar chart 86, 263
bar graph of means 187
Bartlett test of equal variances 183
beta weights 166, 216
  limitation 245
block regression 232–240
Bonferroni multiple-comparison test 162, 181
bookstore, Stata 273
Boston College Stata site 272
box plot 95, 190

C

casewise deletion 160
categorical predictors, regression . . . 234
cause and effect 155
chi-squared table 106
chi-squared test 104, 264
`chitable` command 106–108
`clear` command 28
`codebook` command . . 26, 35–36, 42–43, 111
codebook example 24–28
coding system 24–28
coefficient of variation 94
collinearity 227
command
  `alpha` 169–172
  `anova` 180, 200
  `center` 243
  `chitable` 106–108
  `clear` 28
  `codebook` . . . 26, 35–36, 42–43, 111
  `correlate` 160–161
  `describe` 8, 37–38, 41
  `dfbeta` 226–227
  `drop` 56–58
  `edit` 45
  `egen` 48, 53–55
  `egen rowmean` 55
  `egen rowmiss` 54
  `egenmore` 54

command, *continued*
`ereturn list` ... 188
`estimates store` ... 264
`findit` ... 106, 272
`ftable` ... 107
`generate` ... 48–53
`graph bar` ... 116–117, 187, 263
`graph box` ... 190
`help label` ... 35
`histogram` ... 10–16, 218
`kap` ... 174
`keep` ... 56–58
`label define` ... 30–33, 44
`label values` ... 30–33, 45
`label variable` ... 30
`list` ... 65–66, 180
`listcoef` ... 261–263
`logistic` ... 258–260
`logit` ... 252, 258–260
`loneway` ... 208
`lrdrop2` ... 264–265
`lrtest` ... 264
`manova` ... 206
`median` ... 145
`more` ... 35, 56
`mvdecode` ... 43, 51
`nestreg` ... 238–240, 267–268
`numlabel` ... 66
`omega2` ... 195
`optifact` ... 272
`pcorr2` ... 217, 236
`prtest` ... 124–130
`pwcorr` ... 161–163
`ranksum` ... 144
`recode` ... 46–48, 232
`regress` ... 164–167, 212–214, 234–237, 241–242
`rename` ... 203
`replace` ... 48, 134
`reshape` ... 203–205
`return list` ... 188
`sample` ... 151
`sampsi` ... 142–143
`save` ... 34, 58
`scatter` ... 151–157
command, *continued*
`sdtest` ... 138–139
`set more off` ... 56
`set seed` ... 151
`sktest` ... 219–220
`spearman` ... 167–168
`ssc` ... 272
`summarize` ... 9–10, 91, 163, 218–219
`sunflower` ... 153–154
`sysuse` ... 7–8
`tab1` ... 83, 160
`tab2` ... 173, 233
`tabdisp` ... 205
`tabi` ... 113–114
`table` ... 115–116, 197, 206
`tabodds` ... 110
`tabstat` ... 94, 189
`tabulate` ... 51–53, 83, 102–106, 111–113, 198–200
`ttable` ... 107
`ttest` ... 131–137, 140
`twoway` ... 156–157, 241, 244
`twoway connected` ... 200–202
`use` ... 9
`webuse` ... 9
`xtmixed` ... 206
command structure ... 62
Command window ... 6, 7
comments in do-files ... 70
confidence interval
regression line ... 166
slope ... 166
constant ... 166
conventions used in the book ... 16
copying, HTML format ... 73
copying results to word processor ... 67
`correlate` command ... 160–161
correlation,
interpreting ... 157
limitation ... 245
multiple comparison ... 162
correlation ratio ... 184, 194
count outcome variables ... 257
covariate ... 193

Cramér's V, measure of association ... 109
creating value labels ... 44–45
cross-tabulation ... 102

D
data,
  long format ... 129, 179
  wide format ... 127, 179
Data Editor ... 28–30, 44
dataset
  `c10interaction` ... 241, 247
  `c11barchart` ... 263
  `cancer` ... 8, 9
  `census` ... 246, 247
  `chores` ... 139
  contents ... 34
  create ... 21–23
  `descriptive_gss` ... 83, 98, 99
  `divorce` ... 250
  `environ` ... 254
  `firstsurvey` ... 35
  `firstsurvey_chapter4` ... 64, 67, 75
  `gss2002_chapter6` ... 102, 118
  `gss2002_chapter7` ... 124, 130, 131, 133, 140, 147, 148
  `gss2002_chapter8` ... 150, 163, 175
  `gss2002_chapter9` ... 185, 192, 209, 210
  `gss2002_chapter10` ... 246, 247
  `gss2002_chapter11` ... 270
  `hsb2` ... 9
  `intraclass` ... 207
  `kappa` ... 173
  `long` ... 129
  `nlsy97_chapter7` ... 144
  `nlsy97_fp` ... 58
  `nlsy97_selected_variables` ... 169, 176, 210, 232, 257
  `ops2004` ... 212
  `partyid` ... 180, 188, 210
  `positive` ... 59
  `relate` ... 40, 59
  `severity` ... 270
dataset, *continued*
  `spearman` ... 167, 176
  `states3` ... 18
  `survey` ... 16
  `wide` ... 127
  `wide9` ... 203
datasets, acquisition ... 275
degrees of freedom, ... 106
  ANOVA ... 183
  one-sample t test ... 133
dependent t test ... 139
dependent variable ... 104
`describe` command ... 8, 37–38, 41
`dfbeta` command ... 226–227
dialog box
  `describe` ... 37
  `drop` ... 56–58
  `egen` ... 53–55
  `generate` ... 49
  `graph bar` ... 116
  `graph box` ... 97
  `graph pie` ... 85
  `histogram` ... 11–14, 86
  `keep` ... 56–58
  `label define` ... 31
  `label values` ... 31
  `logistic` ... 258–259
  `prtest` ... 124–130
  `recode` ... 46–47
  `regress` ... 164, 212–213
  `scatter` ... 151
  submit vs. OK ... 15
  `summarize` ... 10
  `tab1` ... 83
  `tabi` ... 113–114
  `table` ... 115
  `tabstat` ... 94
  `tabulate` ... 51–52, 102–106, 111
  `ttest` ... 131–132, 135–137
dictionary file ... 40
difference of means test ... 133
difference of proportions test ... 127
do-file,
  comments ... 70
  continuation line ... 134

do-file, *continued*
introduction . . . 3
version number . . . 70
Do-file Editor . . . 67
download datasets . . . 275
`drop` command . . . 56–58
dummy variables . . . 234

E
`edit` command . . . 45
effect size, η . . . 184, 194
`egen` command . . . 48, 53–55
`egen rowmean` command . . . 55
`egen rowmiss` command . . . 54
`egenmore` command . . . 54
entering data . . . 28–30
equal variance, Bartlett test . . . 183
`ereturn list` command . . . 188
`estat vif` postestimation command . . . 228–229
`estimates store` command . . . 264
exit Stata . . . 16, 35
exponentiation . . . 261

F
F ratio . . . 183
F test of unequal variances . . . 138
factor variable . . . 185
`findit` command . . . 106, 272
fonts, fixed . . . 72
format
numeric . . . 29
string . . . 30
frequency distributions . . . 83
`ftable` command . . . 107

G
gamma, measure of association . . . 111
`generate` command . . . 48–53
Goodman and Kruskal's gamma, measure of association . . . 111
GradPlan . . . 273
graph,
bar chart . . . 86, 116, 263
box plot . . . 95
graph, *continued*
collinearity . . . 228
heteroskedasticity . . . 220
histogram . . . 89, 218
medians . . . 191
overlay two way for ANOVA . . . 200–202
overlay two way showing interaction effects . . . 241
pie chart . . . 85
residual versus predicted . . . 221
scattergram . . . 150–157
sunflower scattergram . . . 154
`graph bar` command . . . 116–117
`graph box` command . . . 190
graphics book . . . 274
GUI interface, Prefs . . . 4–5

H
help,
video . . . 18
web-based . . . 18
`help`, `listcoef` option . . . 261–262
heteroskedasticity . . . 220
hierarchical regression . . . 232–240, 267–268
histogram . . . 89
`histogram` command . . . 10–16, 218
HTML format . . . 73

I
increment in R^2 . . . 216
independent variable . . . 104
indicator variables . . . 234
interaction term . . . 241
interaction, regression . . . 240
interactive table . . . 113
intercept . . . 166
interquartile range . . . 94
interval-level variables . . . 80
intraclass correlation . . . 207, 208

J
`jitter()`, `scatter` option . . . 154

K

`kap` command 174
`keep` command 56–58
Kendall's tau, measure of association 111
Kruskal–Wallis test, ANOVA alternative 188
kurtosis 82, 91, 219–220
kurtosis, test of significance 91

L

`label define` command 30–33, 44
`label values` command 30–33, 45
`label variable` command 30
labeling values 30–33
labeling variables 22, 30
`legend()`, `twoway` option 200–202
likelihood-ratio chi-squared test . . . 264–265
limitations of Stata 275
`list` command 65–66
`list`, `ereturn` saved estimates 188
`list` option, `nolabel` 180
`list`, `return` saved statistics 188
`listcoef` command 261–263
listwise deletion 160
log files 73
log files and graphs 74
log, `.smcl` extension 73
`logistic` command 258–260
logistic regression, ... 110, 258–267
  bar chart 263
  exponentiation 261
  hypothesis tests 264
  interpreting odds ratio 260
  likelihood-ratio chi-squared test 264
  logits 256
  McFadden pseudo-R^2 260
  nonlinear 253
  odds ratio 255
  percent change 260
  pseudo-R^2 260
  S-curve 252
logistic regression vs. OLS regression 253
logistic regression Wald chi-squared test 264
`logit` command 258–260
logits 256
`loneway` command 208
long format 129, 136, 179, 203–205
`lpattern()`, `twoway` option . . . 200–202
`lrdrop2` command 264–265
`lrtest` command 264

M

`manova` command 206
maximum number of variables 275
McFadden pseudo-R^2 260
mean squares 183
means, adjusted 195
measure of association,
  η 184, 194
  odds ratio 110
  ϕ 109
  V 109
`median` command 145
median, graph box plot 191
missing values 26, 43, 51, 63
  types 42, 43
`modify`, `label define` option 33
`more` command 35, 56
`msize()`, `twoway` option 200–202
`msymbol()`, `twoway` option 200–202
multicollinearity 227
multiple comparison
  Bonferroni 181
  Scheffé 181
  Šidák 181
multiple comparison and correlation 162
multiple correlation 214
multiple regression command . . 212–216
multiple regression diagnostics 224
multiple regression with interaction term 240
multiple regression,
  block 232–240

multiple regression, *continued*
categorical predictors 234
dummy variables 234
hierarchical 232–240
indicator variables 234
influential case 224
nested 232–240
outlier 224
residual 221
weighted data 230
`mvdecode` command 43, 51, 134

N

naming variables 24
nested regression 232–240, 267–268
`nestreg` command . . . 238–240, 267–268
NetCourses 274
`nolabel`, `list` option 180
`nolabels`, `edit` option 45
nominal-level variables 80
nonlinear 253
nonparametric ANOVA alternatives 188
nonparametric tests, ... 144
Mann–Whitney 144
median 145
rank sum 144
normally distributed residuals 220
`numlabel` command 66

O

odds ratio, ... 110, 255
interpretation 260
percent change 260
OLS regression vs. logistic regression 253
`omega2` command 195
`on`, `legend()` suboption 200–202
one-sample t test 131
one-way ANOVA 177–178
`oneway` option
`bonferroni` 181
`scheffe` 181
`sidak` 181
open,
existing dataset 7–9
Stata dataset on web 9
Stata installed dataset 7
web dataset 9
`optifact` command 272
option
`add`, `label define` 33
`help`, `listcoef` 261–262
`jitter()`, `scatter` 154
`modify`, `label define` 33
`nolabels`, `edit` 45
`percent`, `listcoef` 262
`order()`, `legend()` suboption 200–202
ordinal-level variables 80
outlier 224

P

paired t test 139
part correlation 216
pasting, ... 73
reformatting 72
pasting results to word processor, . . . 67
formatting 72
`pcorr2` command 236
Pearson's chi-squared 111
`percent`, `listcoef` option 262
pie chart 85
plot a confidence interval 167
Poisson regression 257
postestimation command
`adjust` 195, 200
`estat vif` 228–229
`predict` 224–226, 241, 244
`test` 237–238
power analysis 141
`predict` postestimation command 224–226, 241, 244
predict postregression 224–226
preferences 6
probability tables 106
product term 241
project outline 40

proportions,
one-sample test 124
two-sample test 127
`prtest` command 124–130
pseudo-R^2 260
`pwcorr` command 161–163
`pweights` 230–231

Q

qualifier
`if` with missing values 63
`in` 64

R

R^2 214
R^2 change 216, 217
random sample, ... 151
how to draw 122
random sampling 121
randomization, ... 121
alternative to 192
how to perform 122
`ranksum` command 144
`recode` command 46–48, 232
recoding, ... 134
ranges 47
`regress` command . . 164–167, 212–214, 234–237, 241–242
regression,
block 232–240
categorical predictors 234
dummy variables 234
hierarchical 232–240
indicator variables 234
influential case 224
nested 232–240
outlier 224
residual 221
weighted data 230
regression diagnostics 224
regression line, plotting 156
regression with interaction term . . . 240
reliability, ... 168, 170
alpha 170
kappa 173
`rename` command 203
rename variable 203
repeated-measures ANOVA 202
repeated-measures t test 139
`replace` command 48, 134
`replace`, `save` option 58
`reshape` command 203–205
reshape wide to long format . . . 203–205
residual 194, 220, 221
response variable 185
Results window 6
`more` 56
scroll size 56
`return list` command 188
reverse-code variables 45
Review window 6
root mean squared error 165

S

S-curve 252
sample,
draw 151
`sample` command 151
sample-size estimation 143
`sampsi` command 142–143
`save` command 58
saved estimates 188
saved statistics 188
scale creation 53
`scatter` command 151–157
scattergram 150–157
scattergram with confidence interval 167
scattergram with `jitter()` option . . 154
Scheffé, multiple comparison 181
schemes
`s1monoschrome` 14
scientific notation 216
`sdtest` command 138–139
seed for starting a random sample . . 151
semipartial correlation 216, 236
`set more off` command 56
`set seed` command 151
Šidák, multiple comparison 181

significance,
statistical 159
substantive 159
skewness 82, 91, 219–220
`sktest` command 219–220
slope 166
SMCL 73
`sort`, `twoway` option 200–202
`spearman` command 167–168
Spearman's rho 167
`ssc` command 272
standardized beta coefficient 166
standardized beta weights 216
standardized regression coefficients 216
Stat/Transfer 275
Stata bookstore 273
Stata code for textbook examples . . 272
Stata Intercooled version limitations 275
Stata Journal 273
Stata, limitations 275
Stata listserver 272
Stata Markup and Control Language . . . . . *see* SMCL
Stata NetCourses 274
Stata Portal, UCLA 159, 272, 274
Stata screen 3
Stata tutorial 272, 274
Statalist 272
statistical significance vs. substantive significance 159
subcommand
`list` 188
sum of squares 183
`summarize` command 9–10, 91, 163, 218–219
`summarize()`, `tabulate` option . . . 198–200
summary of data 26
summary statistics, one/two-way table 198
`sunflower` command 153–154
sunflower scattergram 154
`sysuse` command 7–8

T

t test,
one-sample 131
two-sample 133
`tab1` command 83, 160
`tab2` command 173, 233
`tabdisp` command 205
`tabi` command 113–114
table, ... 102
means with repeated ANOVA . . 206
one/two-way 198
probability 106
summary statistics 189, 197
`table` command 115–116, 197, 206
`tabodds` command 110
`tabstat` command 94, 189
`tabulate` command 51–53, 83, 102–106, 111–113, 198–200
tau, measure of association 111
test of significance
kurtosis 91
skewness 91
`test` postestimation command 237–238
tests,
Bonferroni 162
chi-squared 104
dependent *t* test 139
likelihood-ratio chi-squared . . . 264–265
likelihood-ratio chi-squared with logistic regression 264
logistic regression 264
long format for portions 129
Mann–Whitney 144
median 145
multiple comparison with correlations 162
nonparametric 144
one-sample *t* test 131
paired *t* test 139
proportions 124, 127
rank sum 144
repeated-measures *t* test 139
skewness and kurtosis 219–220

tests, *continued*
two-sample t test 133
unequal variances 138
Wald chi-squared test 264
wide format for proportions . . . 127
z test for proportions 130
z test with logistic regression . . 264
textbook examples using Stata programs 272
`title()`, `legend()` suboption 200–202
`title()`, `twoway` option 200–202
tolerance 228–229
tool bar, Stata 7
`ttable` command 107
`ttest` command 131–137, 140
tutorial 274
tutorial, University of North Carolina, Population Center 272
two-by-two table 102
two-way ANOVA 198–202
`twoway` command 156–157, 241, 244
`twoway connected` command 200–202

U

UCLA Stata Portal 18, 159, 272
UCLA web link
`reshape` command 203
`use` command 9
user-written commands 272

V

value labels 22, 30–33, 44–45, 66
`valuelabel`, `twoway` option . . . 200–202
variable labels 22, 30
variable name 22, 24
Variables window 6
variance inflation factor 228–229
version number in do-file 70

W

Wald chi-squared test 264
web search 272
`webuse` command 9
weighted data 229–232
weights
`pweights` 230–231
wide format 127, 137, 179, 203–205
working directory 7

X

`xlabel()`, `twoway` option 200–202
`xtmixed` command 206

Z

z test,
one-sample proportion 124
two-sample proportion 127
zero-inflated models 257