

Subject index

Symbols

β weights 224
 η^2 190
 ϕ 118

A

acquiring datasets 319
add, label define option 32
adjust, postestimation command . 202,
207
adjusted means 202
agreement, intraclass correlation .. 213
alpha reliability 291
tabulate command 87
analysis of covariance see ANCOVA
analysis of variance see ANOVA
ANCOVA 197–204
ANOVA assumptions 184
ANOVA,
degrees of freedom 189
equal-variance test 189
one-way 183–184
postestimation adjusted means ...
..... 202
repeated-measures 208
residual 199
two-way 204–208
anova command 206
ATS at UCLA 170

B

bar chart 96, 274
bar graph of means 192
Bartlett test of equal variances 189
beta weights 178, 224
limitation 255

block regression 242–250
Bonferroni multiple-comparison test ...
..... 174, 187
bookstore, Stata 317
bootstrap regression 230
Boston College Stata site 316
box plot 105, 196

C

casewise deletion 171
categorical predictors, regression .. 244
cause and effect 166
center command 253
chi-squared
table 115
test 112, 276
chitable command 115–116
clear command 27
codebook command 35–36, 45–46,
120
codebook example 23–27
coding system 22–27
coefficient of variation 104
collinearity 238
command structure 66
Command window 7, 8
comments in do-files 75
confidence interval
regression line 178
slope 178
constant 178
conventions used in the book 1
copying, HTML format 16, 78
copying results to word processor ... 16,
77
correlate command 171–172

- correlation,
 - interpreting 169
 - limitation 255
 - multiple comparison 174
- correlation ratio 190, 200
- count outcome variables 267
- covariate 198
- Cramér's V , measure of association ...
 - 118
- creating value labels 47–48
- criterion-related validity 298
- cross-tabulation 110

- D**
- data,
 - long format 141, 185
 - wide format 139, 184
- Data Editor 27–29, 47
- dataset contents 34
- dataset,
 - c10interaction 251, 257
 - c11barchart 274
 - cancer 9, 10
 - census 257
 - chores 151
 - create 19–21
 - descriptive_gss 90, 107, 108
 - divorce 261
 - environ 264
 - firstsurvey 35
 - firstsurvey_chapter4 68, 72, 81
 - gss2002 and 2006_chapter12 ...
 - 314
 - gss2002_chapter6 128
 - gss2002_chapter7 .. 136, 142, 143, 145, 152, 160
 - gss2002_chapter8 181
 - gss2002_chapter9 216
 - gss2002_chapter10 257
 - gss2002_chapter11 282
 - gss2006_chapter6 110, 128
 - gss2006_chapter8 162, 181
 - gss2006_chapter8_selected.. 175
 - gss2006_chapter9 191, 198
- dataset, *continued*
 - gss2006_chapter12 284, 291
 - gss2006_chapter12_selected....
 - 305
 - intraclass 213
 - kappa1 295
 - kuder-richardson 293
 - long 141
 - nlsy97_chapter7 156, 160
 - nlsy97_chapter11 267, 276
 - nlsy97_selected_variables.....
 - 216, 242
 - ops2004 220
 - partyid 186, 194, 217
 - positive 63
 - relate 42, 63
 - retest 289
 - severity 282
 - spearman 180, 181
 - states3 17
 - wide 139
 - wide9 208
- datasets, acquisition 319
- degrees of freedom, 114
 - ANOVA 189
 - one-sample t test 145
- dependent t test 151
- dependent variable 112
- describe command 9, 37–38, 43
- dfbeta command 237
- dialog box,
 - describe 37
 - drop 60–61
 - egen 57–59
 - generate 52
 - graph bar 126
 - graph pie 93
 - histogram 11–13, 96
 - keep 60–61
 - label define 30
 - label values 30
 - logistic 269
 - open 53
 - prtest 136–142
 - recode 49–50

- dialog box, *continued*
- `regress` 176, 220–221
 - `scatter` 163
 - Submit vs. OK 15
 - `summarize` 10
 - `tab1` 90
 - `tabi` 123–124
 - `table` 124
 - `tabstat` 104
 - `tabulate` 55, 110–114, 120
 - `ttest` 143–144, 147–149
- dictionary file 42
- difference of means test 145
- difference of proportions test 139
- do-file,
- comments 75
 - continuation line 146
 - introduction 4
 - version number 75
- Do-file Editor 72
- download datasets 319
- `drop` command 60–62
- dummy variables 244
- E**
- `edit` command 48
- effect size, 190, 201, 225
- η^2 190, 200
- `egen` command 51, 56–59, 285
- `egen count` command 285
- `egen rowmean` command 59, 285
- `egen rowmiss` command 57
- `egenmore` command 57
- entering data 27–29
- equal variance, Bartlett test 189
- `ereturn list` command 194
- `estat vif` postestimation command ..
..... 238–239
- `estimates store` command 276
- exit Stata 16, 34
- exponentiation 272
- external validity 171
- F**
- F* ratio 189
- F* test of unequal variances 150
- factor analysis, 300
- PCF 302
 - PF 301
 - commonality 303
 - eigenvalue 303, 308
 - exploratory factor analysis 301
 - extraction 303
 - factor score 303, 312
 - loading 303
 - oblique rotation 303, 310
 - orthogonal rotation 303, 309
 - postestimation 304
 - principal component analysis .. 302
 - principal-component factor analy-
sis 302
 - promax 310
 - rotation 303
 - scree plot 303, 308
 - simple structure 303
 - varimax 309
- factor variable 191
- `findit` command 115, 316
- fonts, fixed 77
- `foreach` command 33
- format,
- numeric 28
 - string 29
- `fre` command 91, 93
- frequency distributions 91
- `ftable` command 189
- G**
- gamma, measure of association 120
- `generate` command 51–56
- geometric mean 87
- Goodman and Kruskal’s gamma, mea-
sure of association 120
- GradPlan 317
- graph,
- bar chart 96, 127, 274
 - box plot 105
 - collinearity 238
 - hanging rootogram 226
 - heteroskedasticity 231

- graph, *continued*
- histogram 99, 226
 - medians 197
 - overlay two-way showing
 - interaction effects 251
 - pie chart 93
 - residual versus fitted 231
 - scattergram 162–168
- graph bar command 125–127, 192, 274
- graph box command 196
- graphics book 318
- GUI interface, Prefs 6
- H**
- harmonic mean 87
- help 4
- video 17
 - web-based 17
- help label command 34
- help, listcoef option 272–273
- heteroskedasticity 231
- hierarchical regression 242–250, 279–280
- histogram 99
- histogram command 11–15, 226
- HTML format 78
- I**
- increment in R^2 225
- independent variable 112
- indicator variables 244
- interaction term 251
- interaction, regression 251
- interactive table 122
- intercept 178
- interquartile range 104
- interval-level variables 86
- intraclass correlation 213, 214
- J**
- jitter(), scatter option 165
- K**
- kappa 294
- kappa with three raters 296
- kappa, weighted 296
- keep command 60–62
- Kendall's tau, measure of association ..
 120
- Kruskal–Wallis test, ANOVA alternative
 194
- Kuder–Richardson coefficient of
 reliability 293
- kurtosis 89, 101, 229–230
- kurtosis, test of significance 101
- L**
- label define command 30–33, 47
- label values command 30–33, 48
- label variable command 29
- labeling values 30–33
- labeling variables 20, 29
- likelihood-ratio chi-squared test... 276–
 277
- limitations of Stata 320
- list command 69–70, 186
- list, ereturn saved estimates 194
- list option, no label 186
- list, return saved statistics 194
- listcoef command 272–274
- listwise deletion 171
- log files 78
- log files and graphs 79
- log, .smcl extension 79
- logistic command 268–271
- logistic regression 119, 268–279
- bar chart 274
 - exponentiation 272
 - hypothesis tests 276
 - interpreting odds ratio 271
 - likelihood-ratio chi-squared test...
 276
 - logits 266
 - McFadden pseudo- R^2 270
 - nonlinear 262
 - odds ratio 264
 - percentage change 271
 - pseudo- R^2 270
 - S-curve 262

- logistic regression, *continued*
 vs. OLS regression 263
 Wald chi-squared test 276
- logit** command 262, 268–271
- logits 266
- lone**way command 214
- long format 141, 148, 185, 209–210
- lrdrop1** command 276–277
- lrtest** command 276
- M**
- manova** command 213
- maximum number of variables 320
- McFadden pseudo- R^2 270
- mean squares 189
- means, adjusted 202
- measure of association,
 η^2 190, 200
 odds ratio 119
 ϕ 118
 V 118
- median** command 158
- median, graph box plot 197
- menu, open 53
- missing values 25, 46, 54, 67
 types 45, 46
- missing, count 285
- modify**, **label define** option 32
- more** command 35, 59
- multicollinearity 238
- multiple comparison,
 Bonferroni 187
 Scheffé 187
 Šidák 187
- multiple comparison and correlation 174
- multiple correlation 222
- multiple regression command 220–225
- multiple regression diagnostics 234
- multiple regression with interaction
 term 251
- multiple regression,
 block 242–250
 categorical predictors 244
 dummy variables 244
- multiple regression, *continued*
 hierarchical 242–250
 indicator variables 244
 influential case 234
 nested 242–250
 outlier 234
 residual 231
 weighted data 240
- mvdecode** command 46, 54, 146
- N**
- naming variables 23
- nested regression 242–250, 279–280
- nestreg** command 248–250, 279–280
- NetCourses 319
- nolabel**, **edit** option 48
- nolabel**, **list** option 186
- nominal-level variables 86
- nonlinear 262
- nonparametric ANOVA alternatives 194
- nonparametric tests 156
 Mann–Whitney 156
 median 157
 rank sum 156
- normally distributed residuals 230
- numlabel** command 70, 93
- O**
- odds ratio 119, 264
 interpretation 271
 percentage change 271
- OLS regression vs. logistic regression 263
- omega2** command 201
- one-sample t test 143
- one-way ANOVA 183–184
- oneway** option
 bonferroni 187
 scheffe 187
 sidak 187
- open
 existing dataset 8–10
 Stata-installed dataset 8
- optifact** command 316

- option,
 add, label define 32
 help, listcoef 272–273
 jitter(), scatter 165
 modify, label define 32
 nolabel, edit 48
 percent, listcoef 273
ordinal-level variables 86
outlier 234
- P**
- paired *t* test 151
part correlation 225
pasting, 78
 reformatting 77
pasting results to word processor, ... 77
 formatting 77
pccorr2 command 226, 246
Pearson's chi-squared 120
percent, listcoef option 273
pie chart 93
plot a confidence interval 179
Poisson regression 267
postestimation command,
 adjust 202, 207
 estat vif 238–239
 predict 234–236, 251, 254
 test 247–248
power analysis 153
predict postestimation command
 234–236, 251, 254
predict postregression 234–236
predictive validity 298
probability tables 115
product term 251
project outline 42
prophecy formula 290
proportions,
 one-sample test 136
 two-sample test 139
prtest command 136–142
pseudo- R^2 270
pweight command 172–174
pweights 241
- Q**
- qualifier,
 if with missing values 67
 in 68
- R**
- R^2 222
 change 225
random sample 162
 how to draw 135
random sampling 133
randomization 133
 alternative to 198
 how to perform 135
ranksum command 156
recode command 49–51, 242
recoding 146
 ranges 49
regress command .. 176–179, 220–222,
 244–247, 251–253
regression,
 block 242–250
 bootstrap 230
 categorical predictors 244
 dummy variables 244
 hierarchical 242–250
 indicator variables 244
 influential case 234
 nested 242–250
 outlier 234
 residual 231
 robust 230
 weighted data 240
regression diagnostics 234
regression line, plotting 167
regression with interaction term ... 251
reliability,
 alpha 291
 equivalent forms 290
 kappa 294
 kappa with three raters 296
 Kuder–Richardson coefficient of re-
 liability 293
 prophecy formula 290
 split-half 290

- reliability, *continued*
 - test–retest 289
 - weighted kappa 296
- rename** command 209
- rename variable 209
- repeated-measures ANOVA 208
- repeated-measures *t* test 151
- replace** command 51, 146
- replace, save** option 62
- reshape** command 209–210
- reshape wide to long format 209–210
- residual 199, 230, 231
- response variable 191
- Results window 7
 - more** 59
 - scroll size 59
- return list** command 194
- reverse coding 285
- reverse-code variables 48
- Review window 7
- robust regression 230
- root mean squared error 177

- S**
- S*-curve 262
- sample** command 162
- sample, draw 162
- sample-size estimation 155
- sampsi** command 154–155
- save** command 33–34, 62
- saved estimates 194
- saved statistics 194
- scale construction 284
 - reverse coding 285
- scale creation 56
- scatter** command 163–168
- scattergram 162–168
- scattergram with confidence interval 179
- scattergram with **jitter()** option 165
- Scheffé, multiple comparison 187
- schemes 13
- scientific notation 224
- sdtest** command 150–151
- seed for starting a random sample 162
- semipartial correlation 225, 246
- set more off** command 59
- set seed** command 134, 162
- Šidák, multiple comparison 187
- significance,
 - statistical 170
 - substantive 170
- skewness 89, 101, 229–230
- sktest** command 229–230
- slope 178
- SMCL 79
- spearman** command 179–180
- Spearman’s rho 180
- split-half reliability 290
- ssc** command 316
- standardized beta coefficient 178
- standardized beta weights 224
- standardized regression coefficients 224
- Stat/Transfer 320
- Stata bookstore 317
- Stata code for textbook examples 316
- Stata Journal* 317
- Stata, limitations 320
- Stata listserver 316
- Stata Markup and Control Language *see* SMCL
- Stata NetCourses 319
- Stata Portal, UCLA 170, 316, 318
- Stata screen 5
- Stata tutorial 316, 319
- Stata/IC limitations 320
- Statalist 316
- statistical significance versus
 - substantive significance 170
- subcommand, **list** 194
- sum of squares 189
- summarize** command 10–11, 101, 175, 228–229
- summarize()**, **tabulate** option 204–206
- summary of data 25
- summary statistics, one/two-way table 204

sunflower command 165
sysuse command 8–9

T

t test,

 one-sample 143
 two-sample 145
tab1 command 91, 171
tab2 command 243
tabdisp command 210
tabi command 123–124
table 110
 means with repeated ANOVA . . 212
 one/two-way 204
 probability 115
 summary statistics 195, 203
table command 124–125, 203, 212
tabodds command 119
tabstat command 104, 195
tabulate command 55–56, 91,
 110–114, 120–122, 204–206
tau, measure of association 120
test of significance,
 kurtosis 101
 skewness 101
test postestimation command . . . 247–
 248
tests,
 Bonferroni 174
 chi-squared 112
 dependent *t* test 151
 likelihood-ratio chi-squared . . . 276–
 277
 likelihood-ratio chi-squared with
 logistic regression 276
 logistic regression 276
 long format for portions 141
 Mann–Whitney 156
 median 157
 multiple comparison with correla-
 tions 174
 nonparametric 156
 one-sample *t* test 143
 paired *t* test 151
 proportions 136, 139

tests, *continued*

 rank sum 156
 repeated-measures *t* test 151
 skewness and kurtosis 229–230
 two-sample *t* test 145
 unequal variances 150
 Wald chi-squared test 276
 wide format for proportions . . . 139
 z test for proportions 142
 z test with logistic regression . . 276
test–retest reliability 289
textbook examples using Stata
 programs 316
tolerance 238–239
toolbar, Stata 8
ttest command 143–149, 152
tutorial 319
 UNC, Population Center 316
two-by-two table 110
two-way ANOVA 204–208
twoway command 167–168, 251, 254
twoway connected command 207–
 208

U

UCLA Stata Portal 17, 170, 316
 reshaping data wide to long . . . 209
user-written commands 316

V

validity,

 criterion related 298
 external 171
 predictive 298
value labels 21, 30–33, 47–48, 70
variable labels 20, 29
variable name 20, 23
Variables window 7
variance inflation factor 238–239
version number in do-file 75

W

Wald chi-squared test 276
web search 316
weighted data 240–242

weighted kappa 296
weights, `pweights` 241
wide format 139, 149, 184, 209–210
working directory 8

X

`xtmixed` command 213

Z

z test,
 one-sample proportion 136
 two-sample proportion 139
zero-inflated models 267