Subject index

Symbols

β weights 254
η^2 212
ϕ 129
`*` comment 81
`/*` and `*/` comment 81

A

acquiring datasets 383–384
`add`, `label define` option 2
agreement, intraclass correlation . . . 244
alpha reliability 333
`ameans` command 93
analysis of covariance . . . . . *see* ANCOVA
analysis of variance . . . . . *see* ANOVA
ANCOVA 221–232
ANOVA,
    degrees of freedom 211
    equal-variance test 211
    one-way 205–206
    repeated-measures 238
    two-way 232–238
ANOVA assumptions 206
ATS at UCLA 192

B

bar chart 103, 312
bar graph of means 217
Bartlett test of equal variances 211
beta weights 199, 254
    limitation 288
binary variables 273–274
block regression 272–280
blog . . . . . *see* Stata Blog
Bonferroni multiple-comparison test 196, 209
bookstore, Stata 380
bootstrap regression 260
Boston College Stata site 378
box plot 113, 220

C

casewise deletion 193
categorical covariates 223, 227–228
categorical predictors, regression . . . 274
cause and effect 188
`center` command 285
chi-squared
    table 125
    test 123, 314
`chitable` command 125–127
`clear` command 29
`codebook` command 42–43, 51–52, 134
codebook example 25–28
coding system 24–28
coefficient of variation 112
Cohen's d 172
collinearity 268
command structure 74
Command window 8, 9
confidence interval
    regression line 199
    slope 199
constant 199
continuous covariates 223–225
conventions used in the book 1
copying, HTML format 18, 85
copying results to word processor . . . 18, 84
`correlate` command 193–194
correlation,
    interpreting 191
    limitation 288

correlation, *continued*
multiple comparison 196
correlation ratio 212, 230
count outcome variables 305
Cramér's V, measure of association 129
creating value labels 53–56
criterion-related validity 340
cross-tabulation 120

D

data,
long format 157, 207
wide format 156, 207
Data Editor 29–32, 39–40
dataset contents 41
dataset,
acquisition 383–384
`c10interaction` 282, 294
`c11barchart` 312
`cancer` 10, 11
`census` 294
`censusfv` 19
`chapter6_aspirin` 132
`chapter13_missing` 363
`chores` 168
create 21–23
`descriptive_gss` 96, 115, 116
`divorce` 299
download . . . . . xxiii, 383–384
`environ` 302
`firstsurvey` 42
`firstsurvey_chapter4` 76, 82, 88
`gss2002 and 2006_chapter12` 356
`gss2002_chapter6` 144
`gss2002_chapter7` . . 153, 159, 161, 169, 182
`gss2002_chapter8` 203
`gss2002_chapter9` 247, 248
`gss2002_chapter10` 293, 294
`gss2002_chapter11` 323
`gss2006_chapter6` . . 120, 133, 144
dataset, *continued*
`gss2006_chapter6_10percent` 141
`gss2006_chapter8` . . 184, 202, 203
`gss2006_chapter8_selected` . . 196
`gss2006_chapter9` 215, 222
`gss2006_chapter9_2way` 232
`gss2006_chapter12` 326, 333
`gss2006_chapter12_selected` 347
`intraclass` 245
`kappa1` 337
`kuder-richardson` 335
`long` 157
`nlsy97_chapter7` 178, 181
`nlsy97_chapter11` 305, 314
`nlsy97_selected_variables` 248, 272
`ops2004` 250
`partyid` 208, 218, 248
`positive` 71
`regsmpl` 363
`relate` . . . . . xxiii, 48, 71
`relate_small` xxiii
`retest` 331
`severity` 323
`spearman` 201, 203
`wide` 155
`wide9` 239
degrees of freedom, ... 125
ANOVA 211
one-sample t test 161
dependent t test 168
dependent variable 122
`describe` command 10, 44–45, 49
`dfbeta` command 267
dialog box,
`describe` 44
`egen` 66–68
`generate` 61, 63–64
`graph bar` 140
`graph pie` 100
`histogram` 14–15, 103
`logistic` 307
`margins` 228–229

dialog box, *continued*
open 62
`prtest` 153–159
`recode` 57–58
`regress` 197, 251
`scatter` 185
Submit vs. OK 17
`summarize` 11
`tab1` 96
`tabi` 136–137
`table` 138
`tabstat` 112
`tabulate` 64–65, 120–124, 134
`ttest` 160, 163–165
dictionary file 48
difference of means test 161
difference of proportions test 155
do-file,
continuation line 163
introduction 6
Do-file Editor 79–83
download datasets 383–384
`drop` command 69
dummy variables 273–274

E

effect size, ... 212, 231, 255
η^2 . . . . . 212, 230
`egen` command 61, 66–68, 327
`egen count` command 327
`egen rowmean` command 68, 327
`egen rowmiss` command 67
`egenmore` command 67
entering data 29–32
equal variance, Bartlett test 211
`ereturn list` command 218
`estat vif` postestimation command 268–269
`estimates store` command 314
exit Stata 18, 42
exponentiation 310
external validity 193

F

F ratio 211
F test of unequal variances 167
Facebook . . . . . *see* Stata on Facebook
factor analysis, ... 342
commonality 345
eigenvalue 345, 350
exploratory factor analysis 343
extraction 345
factor score 345, 354
loading 345
oblique rotation 345, 352
orthogonal rotation 345, 351
PCF 344
PF 343
postestimation 346
principal component analysis . . 344
principal-component factor analysis 344
promax 352
rotation 345
scree plot 345, 350
simple structure 345
varimax 351
factor variables 215
`findit` command 125, 378
fonts, fixed 85
format,
numeric 31
string 31
`fre` command 98, 100
frequency distributions 97
`ftable` command 211

G

gamma, measure of association 134
`generate` command 61–66
geometric mean 93
Goodman and Kruskal's gamma, measure of association 134
GradPlan 380
graph,
bar chart 103, 140, 312
box plot 113
collinearity 268
hanging rootogram 257
heteroskedasticity 261

graph, *continued*
histogram 106, 257
medians 221
overlay two-way showing interaction effects 284
pie chart 100
residual versus fitted 261
scattergram 184–190
`graph bar` command 139–140, 217, 312
`graph box` command 220
graphics book 381
GUI interface, Edit ▷ Preferences 8

H
harmonic mean 93
help 6
video 19
web-based 19
`help`, `listcoef` option 310–311
`help label` command 42
heteroskedasticity 261
hierarchical regression 272–280, 317–319
histogram 106
`histogram` command 13–17, 257
HTML format 85

I
`ice` command 363
imputation . . . . *see* multiple imputation
increment in R^2 255
independent variable 122
indicator variables 273–274
interaction, regression 282
interaction term 282
interactive table 136
intercept 199
interquartile range 112
interval-level variables 92
intraclass correlation 244, 246

J
`jitter()`, `scatter` option 187

K
kappa 336
kappa, weighted 338
kappa with three raters 338
`keep` command 69
Kendall's tau, measure of association 134
Kruskal–Wallis test, ANOVA alternative 218
Kuder–Richardson coefficient of reliability 335
kurtosis 95, 108, 259–260

L
`label variable` command 55–56
labeling values 32
labeling variables 23
likelihood-ratio chi-squared test . . . 314–315
limitations of Stata 384
`list`, `ereturn` saved estimates 218
`list`, `return` saved statistics 218
`list` command 77–78, 208
`list` option, `nolabel` 208
`listcoef` command 310–312
listwise deletion 193
log, `.smcl` extension 86
log files 86
log files and graphs 87
`logistic` command 306–309
logistic regression 297–324
bar chart 312
exponentiation 310
hypothesis tests 314
interpreting odds ratio 309
likelihood-ratio chi-squared test 314
logits 304
McFadden pseudo-R^2 308
nested 317–319
nonlinear 300
odds ratio 302
percentage change 309
pseudo-R^2 308
S-curve 300

logistic regression, *continued*
vs. OLS regression 301
Wald chi-squared test 314
`logit` command 300, 306–309
logits 304
long format 157, 164, 207, 240–241
`lrdrop1` command 314–315
`lrtest` command 314

M

MAR 359–361
`margins` command 226, 228–229
`marginsplot` command 236–238
maximum number of variables 384
MCAR 359–360
McFadden pseudo-R^2 308
mean squares 211
measure of association,
η^2 . . . . . 212, 230
ϕ 129
odds ratio 131
V 129
`median` command 179
median, graph box plot 221
menu, open 62
`mi estimate` command . . 363, 369–370
`mi impute chained` command 363
`mi impute` command 367–368
`mi impute mvn` command 363, 368
`mi register` command 367
`mi set` command 367–368
`mibeta` command 370–372
`mim` command 363
missing, count 327
missing values . . 27, 52, 63, 75, 357–376
types 51, 52
`misstable` command 364–365
`more` command 4, 42
multicollinearity 268
multiple comparison,
Bonferroni 209
Scheffé 209
Šidák 209
multiple comparison and correlation 196
multiple correlation 252
multiple imputation 357–376
multiple regression command . . 251–255
multiple regression diagnostics 264
multiple regression with interaction term 282
multiple regression,
block 272–280
categorical predictors 274
dummy variables 274
hierarchical 272–280
indicator variables 274
influential case 264
nested 272–280
outlier 264
residual 261
weighted data 270
`mvdecode` command 52, 63, 163

N

naming variables 25
nested regression 272–280, 317–319
`nestreg` command . . . 278–280, 317–319
NetCourses 383
`nolabel`, `list` option 208
nominal-level variables 92
nonlinear 300
nonparametric ANOVA alternatives 218
nonparametric tests 177
Mann–Whitney 178
median 179
rank sum 178
normally distributed residuals 260
`numlabel` command 78, 100

O

odds ratio 131, 302
interpretation 309
percentage change 309
OLS regression vs. logistic regression 301
`omega2` command 231
one-sample t test 159
one-way ANOVA 205–206

`oneway` option
`bonferroni` . . . 209
`scheffe` . . . 209
`sidak` . . . 209
open
existing dataset . . . 9–11
Stata-installed dataset . . . 9
`optifact` command . . . 378
option,
`add`, `label define` . . . 2
`help`, `listcoef` . . . 310–311
`jitter()`, `scatter` . . . 187
`percent`, `listcoef` . . . 311
ordinal-level variables . . . 92
outlier . . . 264

P

paired t test . . . 168
part correlation . . . 255
pasting, . . . 85
reformatting . . . 85
pasting results to word processor, . . . 85
formatting . . . 85
`pcorr` command . . . 255, 276
Pearson's chi-squared . . . 134
`percent`, `listcoef` option . . . 311
pie chart . . . 100
plot a confidence interval . . . 200
Poisson regression . . . 305
postestimation command,
`estat vif` . . . 268–269
`predict` . . . 264–266, 283, 287
`test` . . . 277–278
power analysis . . . 170
`powerreg` command . . . 289–292
`predict` postestimation command . . . 264–266, 283, 287
predict postregression . . . 264–266
predictive validity . . . 340
probability tables . . . 125
product term . . . 282
project outline . . . 48
prophecy formula . . . 332
proportions,
one-sample test . . . 153
two-sample test . . . 155
`prtest` command . . . 153–159
pseudo-R^2 . . . 308
`pwcorr` command . . . 194–196
`pweights` . . . 271
`pwmean` command . . . 213–215

Q

qualifier,
`if` with missing values . . . 75
`in` . . . 76

R

R^2 . . . 252
change . . . 255
random sample . . . 185
how to draw . . . 151
random sampling . . . 149
randomization . . . 149
alternative to . . . 222
how to perform . . . 151
`ranksum` command . . . 178
`recode` command . . . 57–59, 272
recoding . . . 163
ranges . . . 57
`regress` command . . 197–200, 251–252, 274–277, 282–285
regression,
block . . . 272–280
bootstrap . . . 260
categorical predictors . . . 274
dummy variables . . . 274
hierarchical . . . 272–280
indicator variables . . . 274
influential case . . . 264
nested . . . 272–280
outlier . . . 264
residual . . . 261
robust . . . 260
weighted data . . . 270
regression diagnostics . . . 264
regression line, plotting . . . 189
regression with interaction term . . . 282

reliability,
alpha 333
equivalent forms 332
kappa 336
kappa with three raters 338
Kuder–Richardson coefficient of reliability 335
prophecy formula 332
split-half 332
test–retest 331
weighted kappa 338
`rename` command 53, 240
rename variable 240
repeated-measures ANOVA 238
repeated-measures t test 168
`replace` command 61, 163
`reshape` command 240–241
reshape wide to long format . . . 240–241
residual 260, 261
response variables 215
Results window 8
`more` 4
scroll size 4
`return list` command 218
reverse coding 327
reverse-code variables 56
Review window 8
robust regression 260
root mean squared error 199

S

sample, draw 185
`sample` command 141, 185
`sampsi` command 174–177
`save` command 40–41
saved estimates 218
saved statistics 218
scale construction 326
reverse coding 327
scale creation 66
`scatter` command 185–190
scattergram 184–190
scattergram with confidence interval 200
scattergram with `jitter()` option . . 187
Scheffé, multiple comparison 209
schemes 15
scientific notation 254
scree plot 349–350
S-curve 300
`sdtest` command 167–168
seed for starting a random sample . . 185
semipartial correlation 255, 276
`set more off` command 4
`set seed` command 141, 150, 184
Šidák, multiple comparison 209
significance,
statistical 192
substantive 192
skewness 95, 108, 259–260
`sktest` command 259–260
slope 199
SMCL 86
`spearman` command 201
Spearman's rho 201
split-half reliability 332
`ssc` command 378
standardized beta coefficient 199
standardized beta weights 254
standardized regression coefficients 254
Stat/Transfer 384
Stata Blog 379
Stata on Facebook 379
Stata on Twitter 379
Stata bookstore 380
Stata code for textbook examples . . 378
Stata Journal 380
Stata listserver 379
Stata Markup and Control Language . . . . . *see* SMCL
Stata NetCourses 383
Stata Portal, UCLA 192, 378, 381
Stata screen 7
Stata tutorial 378, 383
Stata, limitations 384
Stata/IC limitations 384
Statalist 379
statistical significance versus substantive significance . . . 192

subcommand, `list` 218
sum of squares 211
`summarize` command . . 11–13, 108, 197, 258–259
`summarize()`, `tabulate` option . . . 233–238
summary of data 27
`sunflower` command 187
`sysuse` command 9–10

T

t test,
one-sample 159
two-sample 161
`tab` command 141
`tab1` command 97, 193
`tab2` command 273
`tabdisp` command 241
`tabi` command 136–137
table 120
probability 125
summary statistics 219
`table` command 138–139
`tabstat` command 112, 219
`tabulate` command 60, 64–66, 97, 120–124, 132, 134–136, 232–238
tau, measure of association 134
test of significance,
kurtosis 108
skewness 108
`test` postestimation command 277–278
test–retest reliability 331
tests,
Bonferroni 196
chi-squared 123
dependent *t* test 168
likelihood-ratio chi-squared . . . 314–315
likelihood-ratio chi-squared with logistic regression 314
logistic regression 314
long format for portions 157
Mann–Whitney 178
tests, *continued*
median 179
multiple comparison with correlations 196
nonparametric 177
one-sample *t* test 159
paired *t* test 168
proportions 153, 155
rank sum 178
repeated-measures *t* test 168
skewness and kurtosis 259–260
two-sample *t* test 161
unequal variances 167
Wald chi-squared test 314
wide format for proportions . . . 156
z test for proportions 159
z test with logistic regression . . 314
textbook examples using Stata commands 378
tolerance 268–269
toolbar, Stata 9
`ttest` command 159–165, 169
tutorial 383
UNC, Population Center 378
Twitter . . . . . *see* Stata on Twitter
two-by-two table 120
two-way ANOVA 232–238
`twoway` command 189–190, 284, 287

U

UCLA Stata Portal 19, 192, 378
reshaping data wide to long . . . 240
user-written commands 378

V

validity,
criterion related 340
external 193
predictive 340
value labels 23, 32–39, 53–56, 78
variable labels 23
variable name 22, 25
Variables Manager 33–39, 69
Variables window 8
variance inflation factor 268–269

W

Wald chi-squared test 314
web search 378
weighted data 270–272
weighted kappa 338
weights, `pweights` 271
wide format 156, 165, 207, 240–241
working directory 9

X

`xtreg` command 246
`xtset` command 242

Z

z test,
    one-sample proportion 153
    two-sample proportion 155
zero-inflated models 305