Subject index

Symbols

β weights 248
η^2 208
ϕ 126
`*` comment 79
`/*` and `*/` comment 79

A

acquiring datasets 376–377
`add`, `label define` option 2
agreement, intraclass correlation . . . 237
alpha reliability 325
`ameans` command 91
analysis of covariance . . . . . *see* ANCOVA
analysis of variance . . . . . *see* ANOVA
ANCOVA 215–226
ANOVA,
  degrees of freedom 207
  equal-variance test 207
  one-way 201–202
  repeated-measures 231
  two-way 226–231
ANOVA assumptions 202
ATS at UCLA 188

B

bar chart 101, 304
bar graph of means 210
Bartlett test of equal variances 207
beta weights 195, 248
  limitation 281
binary variables 267–268
block regression 266–274
Bonferroni multiple-comparison test 192, 205
bookstore, Stata 373
bootstrap regression 254
Boston College Stata site 372
box plot 111, 214

C

casewise deletion 189
categorical covariates 217, 221–222
categorical predictors, regression . . . 268
cause and effect 184
`center` command 279
chi-squared
  table 123
  test 121, 306
`chitable` command 123–124
`clear` command 29
`codebook` command 41–42, 51–52, 131
codebook example 25–28
coding system 24–28
coefficient of variation 110
Cohen's d 168
collinearity 262
command structure 72
Command window 8, 9
confidence interval
  regression line 195
  slope 195
constant 195
continuous covariates 217–219
conventions used in the book 1
copying, HTML format 18, 83
copying results to word processor . . . 18, 82
`correlate` command 189–190
correlation,
  interpreting 187
  limitation 281
  multiple comparison 192

correlation ratio 208, 224
count outcome variables 297
Cramér's V, measure of association 126
creating value labels 53–55
criterion-related validity 332
cross-tabulation 118

D

data,
  long format 153, 203
  wide format 152, 202
Data Editor 29–32, 38–39
dataset contents 40
dataset,
  acquisition 376–377
  `c10interaction` 276, 287
  `c11barchart` 304
  `cancer` 10, 11
  `census` 287
  `chapter6_aspirin` 129
  `chapter13_missing` 355
  `chores` 164
  create 21–23
  `descriptive_gss` 94, 113, 114
  `divorce` 291
  download . . . . . xxiii, 376–377
  `environ` 294
  `firstsurvey` 41
  `firstsurvey_chapter4` 75, 80, 86
  `gss2002 and 2006_chapter12` 348
  `gss2002_chapter6` 141
  `gss2002_chapter7` . . 149, 155, 157, 165, 178
  `gss2002_chapter8` 199
  `gss2002_chapter9` 240
  `gss2002_chapter10` 286, 287
  `gss2002_chapter11` 315
  `gss2006_chapter6` . . 118, 130, 141
  `gss2006_chapter6_10percent` 138
  `gss2006_chapter8` . . 180, 198, 199
  `gss2006_chapter8_selected` . . 192
dataset, *continued*
  `gss2006_chapter9` 209, 216
  `gss2006_chapter9_2way` 226
  `gss2006_chapter12` 318, 325
  `gss2006_chapter12_selected` 339
  `intraclass` 238
  `kappa1` 329
  `kuder-richardson` 327
  `long` 153
  `nlsy97_chapter7` 174, 177
  `nlsy97_chapter11` 297, 306
  `nlsy97_selected_variables` 240, 266
  `ops2004` 244
  `partyid` 204, 212, 241
  `positive` 70
  `regsmpl` 355
  `relate` . . . . . xxiii, 48, 70
  `relate_small` xxiii
  `retest` 323
  `severity` 315
  `spearman` 197, 199
  `states3` 19
  `wide` 151
  `wide9` 232
degrees of freedom, ... 123
  ANOVA 207
  one-sample t test 157
dependent t test 164
dependent variable 120
`describe` command 10, 43–44, 49
`dfbeta` command 261
dialog box,
  `describe` 43
  `egen` 65–67
  `generate` 60, 63
  `graph bar` 137
  `graph pie` 98
  `histogram` 13–15, 101
  `logistic` 299
  `margins` 222–223
  open 61
  `prtest` 149–155
  `recode` 56–58

dialog box, *continued*
`regress` 193, 245
`scatter` 181
Submit vs. OK 17
`summarize` 11
`tab1` 94
`tabi` 133–134
`table` 135
`tabstat` 110
`tabulate` 63–64, 118–122, 131
`ttest` 156, 159–161
dictionary file 48
difference of means test 157
difference of proportions test 151
do-file,
continuation line 159
introduction 6
Do-file Editor 77–81
download datasets 376–377
`drop` command 68
dummy variables 267–268

E
effect size, ... 208, 225, 249
η^2 . . . . . 208, 224
`egen` command 60, 65–67, 319
`egen count` command 319
`egen rowmean` command 67, 319
`egen rowmiss` command 66
`egenmore` command 66
entering data 29–32
equal variance, Bartlett test 207
`ereturn list` command 212
`estat vif` postestimation command 262–263
`estimates store` command 306
exit Stata 18, 41
exponentiation 302
external validity 189

F
F ratio 207
F test of unequal variances 163
factor analysis, ... 334
commonality 337
factor analysis, *continued* 334
eigenvalue 337, 342
exploratory factor analysis 335
extraction 337
factor score 337, 346
loading 337
oblique rotation 337, 344
orthogonal rotation 337, 343
PCF 336
PF 335
postestimation 338
principal component analysis . . 336
principal-component factor analysis 336
promax 344
rotation 337
scree plot 337, 342
simple structure 337
varimax 343
factor variables 209
`findit` command 123, 372
fonts, fixed 82
format,
numeric 31
string 31
`fre` command 96, 98
frequency distributions 95
`ftable` command 207

G
gamma, measure of association 131
`generate` command 60–65
geometric mean 91
Goodman and Kruskal's gamma, measure of association 131
GradPlan 373
graph,
bar chart 101, 137, 304
box plot 111
collinearity 262
hanging rootogram 251
heteroskedasticity 255
histogram 104, 251
medians 215

graph, *continued*
overlay two-way showing interaction effects 277
pie chart 98
residual versus fitted 255
scattergram 180–186
`graph bar` command 136–137, 210, 304
`graph box` command 214
graphics book 374
GUI interface, Edit ▷ Preferences 8

H

harmonic mean 91
help 6
video 19
web-based 19
`help`, `listcoef` option 302–303
`help label` command 41
heteroskedasticity 255
hierarchical regression 266–274, 309–310
histogram 104
`histogram` command 13–17, 251
HTML format 83

I

imputation . . . . *see* multiple imputation
increment in R^2 249
independent variable 120
indicator variables 267–268
interaction, regression 276
interaction term 276
interactive table 133
intercept 195
interquartile range 110
interval-level variables 90
intraclass correlation 237, 238

J

`jitter()`, `scatter` option 183

K

kappa 328
kappa, weighted 330
kappa with three raters 330
`keep` command 68
Kendall's tau, measure of association 131
Kruskal–Wallis test, ANOVA alternative 212
Kuder–Richardson coefficient of reliability 327
kurtosis 93, 106, 253–254

L

`label variable` command 55
labeling values 32–33
labeling variables 23
likelihood-ratio chi-squared test . . . 306–307
limitations of Stata 377
`list`, `ereturn` saved estimates 212
`list`, `return` saved statistics 212
`list` command 75–76, 204
`list` option, `nolabel` 204
`listcoef` command 302–304
listwise deletion 189
log, `.smcl` extension 84
log files 83
log files and graphs 84
`logistic` command 298–301
logistic regression 289–315
bar chart 304
exponentiation 302
hypothesis tests 306
interpreting odds ratio 301
likelihood-ratio chi-squared test 306
logits 296
McFadden pseudo-R^2 300
nested 309–310
nonlinear 292
odds ratio 294
percentage change 301
pseudo-R^2 300
S-curve 292
vs. OLS regression 293
Wald chi-squared test 306
`logit` command 292, 298–301

logits 296
long format 153, 160, 203, 233–234
`lrdrop1` command 306–307
`lrtest` command 306

M

MAR 351–353
`margins` command 220, 222–224
maximum number of variables 377
MCAR 351–352
McFadden pseudo-R^2 300
mean squares 207
measure of association,
    η^2 . . . . . 208, 224
    ϕ 126
    odds ratio 128
    V 126
`median` command 175
median, graph box plot 215
menu, open 61
`mi estimate` command 362–363
`mi ice` command 355
`mi impute` command 360–362
`mi register` command 360
`mi set` command 360–362
`mibeta` command 364–365
`misschk` command 355–359
missing, count 319
missing values . . 27, 52, 63, 73, 349–370
    types 51, 52
`misstable` command 355–357
`more` command 4, 41
multicollinearity 262
multiple comparison,
    Bonferroni 205
    Scheffé 205
    Šidák 205
multiple comparison and correlation 192
multiple correlation 246
multiple imputation 349–370
multiple regression command . . 245–249
multiple regression diagnostics 258
multiple regression with interaction term 276
multiple regression,
    block 266–274
    categorical predictors 268
    dummy variables 268
    hierarchical 266–274
    indicator variables 268
    influential case 258
    nested 266–274
    outlier 258
    residual 255
    weighted data 264
`mvdecode` command 52, 63, 159

N

naming variables 25
nested regression 266–274, 309–310
`nestreg` command . . . 272–274, 309–310
NetCourses 376
`nolabel`, `list` option 204
nominal-level variables 90
nonlinear 292
nonparametric ANOVA alternatives 212
nonparametric tests 173
    Mann–Whitney 174
    median 175
    rank sum 174
normally distributed residuals 254
`numlabel` command 76, 98

O

odds ratio 128, 294
    interpretation 301
    percentage change 301
OLS regression vs. logistic regression 293
`omega2` command 225
one-sample t test 155
one-way ANOVA 201–202
`oneway` option
    `bonferroni` 205
    `scheffe` 205
    `sidak` 205

open
existing dataset 9–11
Stata-installed dataset 9
`optifact` command 372
option,
`add`, `label define` 2
`help`, `listcoef` 302–303
`jitter()`, `scatter` 183
`percent`, `listcoef` 303
ordinal-level variables 90
outlier 258

P

paired t test 164
part correlation 249
pasting, ... 83
reformatting 82
pasting results to word processor, . . . 82
formatting 82
`pcorr` command 249, 270
Pearson's chi-squared 131
`percent`, `listcoef` option 303
pie chart 98
plot a confidence interval 196
Poisson regression 297
postestimation command,
`estat vif` 262–263
`predict` 258–260, 277, 280
`test` 271–272
power analysis 166
`powerreg` command 282–285
`predict` postestimation command 258–260, 277, 280
predict postregression 258–260
predictive validity 332
probability tables 123
product term 276
project outline 48
prophecy formula 324
proportions,
one-sample test 149
two-sample test 151
`prtest` command 149–155
pseudo-R^2 300
`pwcorr` command 190–192
`pweights` 265

Q

qualifier,
`if` with missing values 73
`in` 74

R

R^2 246
change 249
random sample 181
how to draw 147
random sampling 145
randomization 145
alternative to 216
how to perform 147
`ranksum` command 174
`recode` command 56–59, 266
recoding 159
ranges 57
`regress` command . . 193–196, 245–246, 268–271, 276–278
regression,
block 266–274
bootstrap 254
categorical predictors 268
dummy variables 268
hierarchical 266–274
indicator variables 268
influential case 258
nested 266–274
outlier 258
residual 255
robust 254
weighted data 264
regression diagnostics 258
regression line, plotting 185
regression with interaction term . . . 276
reliability,
alpha 325
equivalent forms 324
kappa 328
kappa with three raters 330

reliability, *continued*
Kuder–Richardson coefficient of reliability . . . 327
prophecy formula . . . 324
split-half . . . 324
test–retest . . . 323
weighted kappa . . . 330
`rename` command . . . 233
rename variable . . . 233
repeated-measures ANOVA . . . 231
repeated-measures t test . . . 164
`replace` command . . . 60, 159
`reshape` command . . . 233–234
reshape wide to long format . . . 233–234
residual . . . 254, 255
response variables . . . 209
Results window . . . 8
`more` . . . 4
scroll size . . . 4
`return list` command . . . 212
reverse coding . . . 319
reverse-code variables . . . 56
Review window . . . 8
robust regression . . . 254
root mean squared error . . . 195

S

sample, draw . . . 181
`sample` command . . . 138, 181
`sampsi` command . . . 170–173
`save` command . . . 39–40
saved estimates . . . 212
saved statistics . . . 212
scale construction . . . 318
reverse coding . . . 319
scale creation . . . 65
`scatter` command . . . 181–186
scattergram . . . 180–186
scattergram with confidence interval . . . 196
scattergram with `jitter()` option . . 183
Scheffé, multiple comparison . . . 205
schemes . . . 15
scientific notation . . . 248
scree plot . . . 341–342
S-curve . . . 292
`sdtest` command . . . 163–164
seed for starting a random sample . . 181
semipartial correlation . . . 249, 270
`set more off` command . . . 4
`set seed` command . . . 138, 146, 180
Šidák, multiple comparison . . . 205
significance,
statistical . . . 188
substantive . . . 188
skewness . . . 93, 106, 253–254
`sktest` command . . . 253–254
slope . . . 195
SMCL . . . 84
`spearman` command . . . 197
Spearman's rho . . . 197
split-half reliability . . . 324
`ssc` command . . . 372
standardized beta coefficient . . . 195
standardized beta weights . . . 248
standardized regression coefficients 248
Stat/Transfer . . . 377
Stata, limitations . . . 377
Stata bookstore . . . 373
Stata code for textbook examples . . 372
Stata/IC limitations . . . 377
Stata Journal . . . 373
Stata listserver . . . 373
Stata Markup and Control Language . . . *see* SMCL
Stata NetCourses . . . 376
Stata Portal, UCLA . . . 188, 372, 374
Stata screen . . . 7
Stata tutorial . . . 372, 376
Statalist . . . 373
statistical significance versus substantive significance . . . 188
subcommand, `list` . . . 212
sum of squares . . . 207
`summarize` command . . 11–13, 106, 193, 252–253
`summarize()`, `tabulate` option . . . 227–231
summary of data . . . 27

`sunflower` command 183
`sysuse` command 9–10

T

t test,
  one-sample 155
  two-sample 157
`tab` command 138
`tab1` command 95, 189
`tab2` command 267
`tabdisp` command 234
`tabi` command 133–134
table 118
  probability 123
  summary statistics 213
`table` command 135–136
`tabstat` command 110, 213
`tabulate` command 59, 63–65, 95, 118–122, 129, 131–133, 226–231
tau, measure of association 131
test of significance,
  kurtosis 106
  skewness 106
`test` postestimation command 271–272
test–retest reliability 323
tests,
  Bonferroni 192
  chi-squared 121
  dependent *t* test 164
  likelihood-ratio chi-squared . . . 306–307
  likelihood-ratio chi-squared with logistic regression 306
  logistic regression 306
  long format for portions 153
  Mann–Whitney 174
  median 175
  multiple comparison with correlations 192
  nonparametric 173
  one-sample *t* test 155
  paired *t* test 164
  proportions 149, 151
tests, *continued*
  rank sum 174
  repeated-measures *t* test 164
  skewness and kurtosis 253–254
  two-sample *t* test 157
  unequal variances 163
  Wald chi-squared test 306
  wide format for proportions . . . 152
  z test for proportions 155
  z test with logistic regression . . 306
textbook examples using Stata commands 372
tolerance 262–263
toolbar, Stata 9
`ttest` command 155–161, 165
tutorial 376
  UNC, Population Center 372
two-by-two table 118
two-way ANOVA 226–231
`twoway` command 185–186, 277, 280
`twoway connected` command 231

U

UCLA Stata Portal 19, 188, 372
  reshaping data wide to long . . . 233
user-written commands 372

V

validity,
  criterion related 332
  external 189
  predictive 332
value labels 23, 32–38, 53–55, 76
variable labels 23
variable name 22, 25
Variables Manager 33–38, 68
Variables window 8
variance inflation factor 262–263

W

Wald chi-squared test 306
web search 372
weighted data 264–266
weighted kappa 330
weights, `pweights` 265

wide format 152, 161, 202, 233–234
working directory 9

X

`xtreg` command 238

Z

z test,
    one-sample proportion 149
    two-sample proportion 151
zero-inflated models 297