

Subject index

A

- adding commands/programs 33
- adjust for multiple comparisons . . . 126–129
 - experimentwise error rate 126
 - familywise error rate 126
 - Scheffé 126–127
 - Tukey’s HSD 126–129
- anova** command 119
 - factorial ANOVA 144–146, 159–160
 - one-way ANOVA 119

B

- benefits of Stata 4–5
- bootstrap** command 342
 - computing the confidence interval 343
- by processing 18
- bysort** prefix 268

C

- categorical predictors 77–86
 - dummy coding 78–86
 - changing the base category . . . 84
 - factor variables 83–84
 - generate** and **replace** commands 79
 - generate** command only 80
 - interpretation of coefficients 81–83
 - null hypothesis 83
 - rules 78
 - tabulate** command 80
 - using with **regress** command 81
 - incorrect implementation 78

- cd** command 12
- conditional variance 53
 - formula 53
 - root mean squared error 53
- confidence intervals 58
 - coverage rate 59–60
 - formula 58
 - interpretation 59–60
- confirmatory factor analysis . . . 377–396
 - χ^2 -difference test 346–349
 - causal models 321
 - computing reliability (ω) 340
 - connection to regression 323
 - correlated residuals 350
 - reliability 353
 - estimating in Stata *see sem* command
 - global fit 392–396
 - goodness of fit 328
 - χ^2 test 330–334
 - χ^2 392–393
 - CFI 394
 - printing in Stata *see estat* **gof** command
 - RMSEA 393
 - SRMR 394–395
 - TLI 394
 - identification 383, 401–402
 - constraints 384–386
 - just-identified 386
 - overidentified 386
 - underidentified 386
 - intercepts 400
 - introduction 320
 - latent variables 321
 - mean structure 381, 383

- confirmatory factor analysis, *continued*
 model-implied covariance 323,
 382
 nested models 346
 parallel items 349
 path model 321
 parts 321
 predicted covariance matrix ($\hat{\Sigma}$) ..
 326, 329, 334
 prediction equation 323–326
 R^2 for an item 335
 estimating in Stata ... see **estat**
eggof command
 formula 335
 reliability
 ω versus α 343
 impact of weak loadings 343
 internal consistency 343
 tau-equivalence 344
 sample versus asymptotic variance
 336
 saturated model 331
 shared variance across all items ...
 338
 standardized solution 389
 tau-equivalent models 346
 unstructured model 321
 variance of an item 382
 variance structure 381
 versus exploratory factor analysis ..
 377
- contrast** command
 factorial ANOVA 147–149,
 164–165
mcompare() option 148, 165
nowald option 165
pveffects option 165
 repeated measures 176, 196
small option 176, 196
 small samples 190
- corr** command 322, 359
covariance option 329
- cumulative normal distribution function
 (Φ) 211
- D**
 Data Editor 15
describe command 14
 do-files see reproducible analysis
 dummy coding see categorical
 predictors, dummy coding
- E**
 Early Childhood Longitudinal Program
 data 398
- effect size
 η^2 149–150
 ω^2 150–151
 f 223
 Cohen's d 116
 versus Hedges's g 116
 estimating in Stata see **esize**
 command
 factorial ANOVA 149–151, 165
 partial- η^2 149
 partial- ω^2 150
- egen** command 17
mean() option 252, 268
tag option 268
- egenmore** command
semean option 252
- esize** command 116
esize twosample command 116
- estat bootstrap** command 343
estat eggof command 335
estat esize command
 factorial ANOVA 150–151
 omega option 151
- estat framework** command
fitted option 329, 334, 383
- estat gof** command 330, 391, 392
- estat ic** command
 following **mixed** 179
- estat icc** command 250
- estat wcorrelation** command ... 178,
 182, 184, 187, 196
- estimates store** command 348
- exploratory factor analysis 358–377
 common factor model 359
 communality 364

exploratory factor analysis, *continued*
 data reduction 359
 eigenvalues 363
 equation 360
 estimating in Stata *see* **factor**
 command
 extracting factors 360
 choosing the number 366
 eigenvalues-greater-than-one
 rule 366
 scree plot 367
 extraction methods 360
 principal-component factor.. 360
 interpreting loadings 362
 notation 360
 orthogonal factors 362
 parallel analysis 368
 promax rotation 374
 rotation
estat common command 376
estat structure command
 376
rotate command 372, 374
 simple structure 370
 uniqueness 364
 varimax rotation 370
 versus confirmatory factor analysis
 358
 versus principal-component analy-
 sis 365
 extending syntax over multiple lines ...
 25

F

factor command 362
 parallel analysis *see* **paran**
 command
pcf option 362
 scree plot *see* **screeplot**
 command
 factor variables *see*
 categorical predictors, dummy
 coding, factor variables

factorial ANOVA

benefits 131
 degrees of freedom 142–143
 effect size *see* effect size
 estimating in Stata *see* **anova**
 command
 interactions 139–140, 156–158
 null hypothesis 140, 156
 main effects 138–139, 155–158
 null hypothesis 139
 versus first-order effects 139–
 140
 marginal means 138–139,
 154–155
 one-way marginal means 155
 two-way marginal means 154
 notation 133–134
 partitioning the variance .. 140–142
 SS between 141
 SS interaction 141–142
 SS main effects 141
 simple effects 146–149, 163–165
 estimating in Stata *see*
contrast command
 null hypothesis 147
 source table 142–143
 three-factor design 151–166
 three-way interactions 156,
 158–159
 null hypothesis 158–159
 two-factor design 134–151
 visualize data ... 134–138, 152–153
 file paths 12

G

General Social Survey 38
generate command 17
_n function 288
 handling missing data 18–19
 getting help 32–33
graph box command 24, 287
by() option 153
 factorial ANOVA 135–136, 153
over() option 153
 repeated measures 171, 192

graph dot command 261
 graphics introduction 22–27

H

histogram command 23, 281, 290

I

in qualifier 15, 16
 interactions 90–109
 categorical by continuous .. 91–107
 continuous by continuous 107–109
 interpretation of coefficients
 108
 margins command 108–109
 marginsplot command 108–109
 visualizing relationships 108–109
 dichotomous by continuous 91–101
 factor-variable notation 94
 first-order coefficients 96
 interpreting coefficients 95–101
 multiple intercept, multiple slope
 model 99–101
 factor notation 100
 lincom command 100
 polytomous by continuous 101–107
 interpretation of coefficients
 102–103
 margins command 104
 marginsplot command 104
 using factor notation to prevent
 errors 102
 visualizing the interaction .. 104–107
 probing meaning 96–101
 problems with dichotomization ...
 107–109
 product between variables 94

interactions, *continued*

 simple regression equation .. 96–101
 lincom command 98–99
 margins command 97–98
 marginsplot command ... 97–98
 plotting 97–98
 interocular trauma test 329

L

label values command 19
 label variable command 19
 labels 19
 value labels 19
 variable labels 19
 lincom command 73, 305
 between-clusters and within-cluster
 relationships 271
 contrast after the regress
 command 73–74
 planned comparisons 123–125
 linear combinations see lincom
 command
 lines 43
 intercept 43
 slope 43
 list command 15
 combined with if and in qualifiers
 17
 repeated measures 173
 lrtest command ... 279, 298, 348, 350,
 353, 389
 repeated measures 199

M

margins command 50–51, 299, 302,
 305, 309
 expected values 65
 following the anova command
 145–146, 160
 inferential uncertainty 65–70
 marginsplot command 66–70, 299,
 302, 305, 309
 by option 309
 factorial ANOVA 136–138

- marginsplot** command, *continued*
 - following the **anova** command 145–146, 160–163
 - methods for improving 67–70
 - noci** option 309
 - xdimension()** option 162
 - mata** command 363, 364
 - mean** 40
 - measurement invariance** 398–427
 - across groups 400–413
 - configural invariance 400–406
 - invariant factor means 412–413
 - invariant factor variances 412
 - metric invariance 407
 - residual invariance 408
 - scalar invariance 407–408
 - sem** command 402–405
 - structural invariance 412–413
 - across time 413–427
 - configural invariance 413–424
 - identification 414–424
 - metric invariance 424–425
 - residual invariance 426
 - scalar invariance 425–426
 - sem** command 417–424
 - structural invariance 427
 - steps 399–400
 - using the CFI to evaluate 409–411
- mixed** command
 - between-clusters and within-cluster relationships 270
 - cov(unstructured)** option 278
 - covariance(unstructured)** option 297
 - cross-level interaction 304
 - dfmethod()** option 175, 190
 - dfmethod(satterthwaite)** option 178, 190
 - display residual correlation matrix see **estat wcorrelation** command
 - fit indices see **estat ic** command
 - longitudinal data 297
 - noconstant** option 175, 306
- mixed** command, *continued*
 - nofetable** option 249
 - noheader** option 249
 - random slopes 278
 - random-effects specification 175
 - random-intercept model 248–252
 - reml** option 175
 - repeated measures 175, 195
 - residuals()** option
 - by()** option 198
 - residuals(ar 1)** option 181
 - residuals(exchangeable)** option 178
 - residuals(independent)** option 175, 176
 - residuals(toeplitz)** option 183
 - residuals(unstructured)** option 186, 195
 - separate intercepts, separate slopes model 306
 - stddeviations** option 178, 184, 249, 279
 - syntax basics 175–176
 - t()** option for **residuals()** 181, 183, 186
 - time-invariant covariate 301
 - time-varying covariate 308
- multilevel models
 - adding a predictor 262–264
 - atomistic fallacy 267
 - basic model 244
 - Bayes’s theorem 255
 - between-clusters variance 244, 245
 - between-clusters versus within-cluster effects 264–266
 - between-clusters versus within-cluster relationships 267–273
 - caterpillar plot 253–254
 - centering 292
 - complete pooling 254
 - conditional independence 246
 - contextual effect 273
 - correlated random effects 276

- multilevel models, *continued*
- definition of random effects... 246
 - ecological fallacy..... 267
 - fitting them in Stata... see **mixed** command
 - interpreting random slopes... 280–283
 - intraclass correlation..... 250–252, 280
 - fitting them in Stata... see **estat** **icc** command
 - issues with clustered data... 239–243
 - level-specific relationships..... 266
 - longitudinal data..... 296
 - need for growth model..... 290
 - random slope..... 296
 - longitudinal data introduction... .. 287
 - no pooling..... 254
 - nonindependence..... 239–242
 - sources..... 242–243
 - substantive benefits... 242–243
 - partial pooling..... 255–259
 - partitioning variance..... 244–246
 - predicting cluster means.. 252–261
 - random effects
 - compared to fixed effects... 247
 - examples..... 246
 - random intercepts..... 246–249
 - random slopes..... 273–283
 - separate intercepts, separate slopes model..... 306
 - time-varying covariate..... 307
 - total relationships..... 266
 - total residual..... 244
 - versus repeated-measures models.. .. 243
 - visualizing partitioning..... 244
 - within-cluster variance... 244, 245
- multiple comparisons..... 120–129
- adjust for multiple comparisons... .. see adjust for multiple comparisons
 - adjusting in Stata... see **pwcompare** command
 - α_{joint} 121–122
 - protecting against α inflation..... .. 122–129
- multiple regression..... 86–90
- centering predictors..... 87–88
 - interpreting coefficients..... 87–90
 - model fit..... 86
 - R^2 87
 - root mean squared error..... 87
 - partial slopes..... 88
 - regress** command..... 86
 - relationship among predictors.. 89–90
 - rescaling predictors..... 88–89
- N**
- nlcom** command..... 338, 340, 341, 353
- normal()** function..... 211
- null hypothesis significance testing.. 60–63, 202
- alternative hypothesis..... 60
 - criticisms..... 63–64
 - null hypothesis..... 60
 - p*-value..... 63
 - regress** output..... 63
 - steps..... 61–63
- O**
- one-way ANOVA..... 116–129
- alternative hypothesis..... 117
 - between versus within variance... .. 117–119
 - degrees of freedom..... 118
 - estimating in Stata... see **oneway** command or **anova** command
 - follow-up tests..... 119–129
 - mean squares..... 118–119
 - multiple comparisons..... see multiple comparisons

- one-way ANOVA, *continued*
 - null hypothesis 117
 - ratio of variances 119
 - sum of squares 117–118
- oneway** command 119
- P**
- paran** command 368
 - pcf** option 368
- partitioning variance
 - regression 51–53
- planned comparisons 122–125
 - contrast weights 122
 - degrees of freedom 123
 - F* test 123
 - linear combinations 122–125
 - Stata estimation
 - see **test** command and **lincom** command
 - sum of squares 123
- point-and-click 9
 - problems 9
 - versus command language 10
- population parameters versus sample-based estimates 44
- power see also **simulate** command
 - definition 202
 - estimating in Stata see **power** command
 - null and alternative distribution ..
 - 202–210
 - simulation 204–210
 - type M error 230, 233–235
 - type S error 230–233
 - z* test 210–214
 - z* test formula 211
- power** command 214–229
 - correlation 220–223
 - detectable correlation .. 222–223
 - sample-size calculation 221–222
 - diff()** option 218
 - factorial ANOVA 226–229
 - graph()** option 216
 - knownsds** option 215
- power** command, *continued*
 - one-way ANOVA 223–226
 - f* 223
 - sample-size estimates 225
 - onecorrelation** 221–223
 - oneway** 224–226
 - parallel** option 225–226
 - sample-size estimates 229
 - t* test 215–220
 - detectable difference ... 216–217
 - sample-size calculation 218
 - twomeans** 214–215
 - twoway** 227–229
 - parallel** option 229
 - varcolumn()** option 227
 - varrow()** option 227
 - varrowcolumn()** option 227
 - varying multiple parameters ... 219
 - z* test 214–215
- predict** command 49, 70–73
 - residual** option 49
 - fitted** option 299
 - inferential uncertainty 70–71
 - plotting the results 71–73
 - predictive uncertainty 71–73
 - stdf** option 72
 - stdp** option 70
- predictive uncertainty see **predict** command, predictive uncertainty
- program** command 204
 - bootstrapping 342
- program define** command 40
- Project Manager see reproducible analysis
- pwcompare** command 191
 - effects** option 191
 - mcompare()** option 191
 - Scheffé 126–127
 - small** option 191
 - small samples 190
 - Tukey's HSD 127–129

R

- R^2 52
 - formula 52
 - interpretation 53
 - Stata output 52
- regress** command 45
 - output 45
- regression
 - centering predictors 47
 - conceptual introduction 43
 - equation 43
 - expected values 44
 - general bivariate equation 44
 - intercept computation 45
 - intercept interpretation 47–48
 - margins** command ... *see* **margins** command
 - predict** command ... *see* **predict** command
 - predicted values 49–51
 - reason for error term 43
 - residuals 48–49
 - definition 48
 - slope computation 45
 - slope interpretation 46–47
 - Stata command *see* **regress** command
- reliability 315
 - definition 319
 - factor analysis...*see* factor analysis
 - partitioning variability in items 319
 - reasons for factor analysis 320
 - shared variance across items... 318–320
 - variability in item scores 318
- repeated measures
 - Σ 174
 - benefits of mixed models.. 167–168
 - block diagonal matrix 189
 - covariance among residuals 174
 - repeated measures, *continued*
 - covariance structures 176–189
 - autoregressive 180–182
 - compound symmetry ... 177–180
 - independent 176–177
 - Toeplitz 183–185
 - unstructured 186–189
 - degrees of freedom 189–190
 - denominator degrees of freedom 179
 - estimating in Stata *see* **mixed** command
 - heteroskedastic residuals.. 197–200
 - maximum likelihood versus least squares 167
 - model formulation 172–173
 - multiple factors 192–197
 - nonindependence 173
- replace** command 17
- replication 5–8
- reproducible analysis 5, 27–31
 - log files 30
 - text versus SMCL 30
 - Project Manager 30–31
 - role of do-files 28
 - annotating 29
 - versus point-and-click 29
 - suggestions for 28
 - workflow 31
 - recommendations 31
- reshape** command 135, 171–172
 - i()** option 172
 - j()** option 172
 - naming recommendations 172
- reshape long** command 172
- reshape wide** command 172
- residual variance *see* conditional variance
- residuals *see* regression residuals
- root mean squared error *see* conditional variance

S

sampling distribution 56, 202
 mean of 56
 shape of 56
 simulating *see* simulation
 standard deviation of 56
 standard error .. *see* standard error

screepplot command 367

sem command 380
 constraining latent correlation
 388
 constraining latent variables... 386
 constraints 328, 345
 equation goodness of fit *see*
estat **gof** command
 making sense of the output... 328
 model builder versus syntax... 326
 one-factor model 327
 options
 coeflegend 338, 340
 covariance() 351, 381
 covstructure 349
 ginvariant 402, 407, 408
 group(varname) 402
 nm1 336
 parallel items constraints 349
 standardized option 389
 syntax basics 327
 tau-equivalent constraints 347
 tau-equivalent model 344

simdemo command 40

simple effects *see* factorial ANOVA

simulate command 40–41
 arguments 41
 options 41
 power 205–209

simulation 40
 sampling distribution... 54–56, 58

standard errors 56
 regression 56–58

Stata interface 10
 Command window 10
 Properties window 11
 Results window 11

Stata interface, *continued*

 Review window 11
 Variable window 11

statsby command 275, 292–293

sum of squares 52
 model 52
 residual 52
 total 52

summarize command 20–21
 assisting with centering 47
 repeated measures 170
 returned values 47

T

t test 114–116
 effect size 115–116
 formula and procedures 114
 null hypothesis 114

table command 21–22
by() option 134, 152
contents() option 134
 factorial ANOVA 134–135, 152
format() option 134
 marginal means 138

tabstat command 113
stat option 336

tabulate command 21–22

test command
 planned comparisons 123–125

ttest command 114–115

twoway
function command 260
lfit command 269, 292, 294
rcap command 276, 293
scatter command 25–27, 260,
 269, 276, 291–294, 369

type I error rate 203
 type II error 203

U

use command 12
 file paths 12

V

- validity 315
- variable types
 - date and time 14
 - numeric 14
 - string 15
- violin plots 113
 - `vioplot` command 113

W

- wide versus long datasets 169–170
- working directory 12
 - benefits of setting 13

Y

- Yale–Brown Obsessive Compulsive Scale 398