CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-72583-5 — Statistics Using Stata
Sharon Lawner Weinberg , Sarah Knapp Abramowitz
Index
More Information

# Index

accumulating data 44–47
additive rules of probability 231–233
  disjoint (mutually exclusive) events 231–233
  Second Additive Rule 232–233
adjusted *R* squared values 523
Akaike, Hirotugu 612–613
Akaike information criterion (AIC) 612, 684–685
alternative hypothesis 297
analysis of variance (ANOVA) 297, 414–416
  assumptions of 419–422
  Bonferroni adjustment 444–446
  computations for 417–420, 469–474
  development of 414
  effect size 426–431
  F distribution 425
  F-ratios 427
  index of separation for 416, 419
  Kruskal–Wallis analysis of variance 639
  Levene's test 451
  mean squares 417–420
  multiple *t*-tests and 412–414
  one-tailed tests 303
  regression 518–520
  **robvar** command 425
  summary table 425
  *t*-tests 416
  two-tailed test 333
  in two-way analysis of variance 465
  variability between/within groups 416
  variance estimators 417
  *see also* two-way analysis of variance
**anova** command 429
  *t*-distribution and 420
*Anscombe.dta* 524–525, 673
arithmetic mean
  comparing with mode and median 84–85
  of dichotomous variable 82
  weighted 83–84
arithmetic operators 104
assumptions 622, 679–683
  of analysis of variance 419–420, 466–467
  of chi-square test of independence 633
  of independent samples 340–342
  of paired samples 354–359
asymmetric curves 95–99
average *see* arithmetic mean
average squared deviation 205
average treatment effect 398, 399, 404
  *b*-weight equation 556
  first-order effect variables 598–599
  in multiple regression 558
  as partial slope in multiple regression 568–571
  statistical significance 563–564
  testing for statistical significance 514–518

*b*, $\beta$ coefficients, with non-correlated predictors 555–556
*b*-weight 590, 594
balanced factorial designs 467–468
bar graphs 56
  clustered for bivariate relationships 184–187
  frequency and percent distributions of nominal data in Stata 28
Bartlett's test 429
*basketball.dta* 673
Bayes, Sir Thomas 239
Bayesian information criterion (BIC) 684–685
*Bem Sex-Role Inventory* 145
best-fitting linear equations
  average squared deviation in 205
  least squares criterion 205
  Pearson correlation coefficient and 202–209
  prediction lines 205
  regression lines 205
  residuals 204
  standard error of estimate in 205
between-group variance 416
bias
  populations and 285–286
  selection 405
bimodal distributions 75
bimodality 75
binomial probability model
  applicability 249–254
  approximation of using normal distribution 260
  distribution in 244–254
  exercises 261–268
  factorial notation 248
  identical trials 244, 249
  independent trials 244, 249
  law of large numbers 251
  normal probability model and 254–261



© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-72583-5 — Statistics Using Stata
Sharon Lawner Weinberg , Sarah Knapp Abramowitz
Index
More Information



binomial tail function 254
bivariate contingency table 630
bivariate normal distribution 509, 513
  homoscedasticity property 503–506
  linearity property 503–506
Blood 673
Bonferroni adjustment 444–446
  multiple measures 446–447
bootstrapping 360–368
*Brainsz.dta* 674

causal studies 391–395
centering 125
central tendency 74
chi-square test of independence, assumptions 633
Cohen's *d* 311
command summary for chapter 4–5, 614–616
continuous variables 8–9
correlation scale 171
Court TV 75
criterion variables 512
Cronbach's alpha 392
cross-sectional design
cross-tabulation 630, 632
crossed factorial design
**cumul** command 45
cumulative binomial probability value 252, 253
cumulative percent distributions 44–45
Currency 674

data analysis
  computers in 3
  context setting in 11–14
data management, Stata .do-file 147–150
data reliability 174
data sets
  *Anscombe.dta* 524–525, 673
  *Hamburg.dta* 159, 160, 676–677
  *Ice Cream.dta* 106, 506, 559, 560, 679–683
  *Impeach.dta* 679–683
  *Learndis.dta* 112, 154, 160, 375, 378–379, 678
  *Likert.dta* 125, 126, 144
  *ManDext.dta* 417
  *NELS.dta* 11, 13, 26, 43, 44, 74, 159, 236, 511
  *Statisticians.dta* 25
  *Stepping.dta* 499
  *TEMP.dta* 87
data transformation 54
  nonlinear 135–142
  recoding variables 144–147
data value
de Moivre, Abraham 282
degrees of freedom
  in ANOVA 420–422
  chi-square distribution 623–639
  independent samples 340
  *t*-distribution 320–321
  in univariate distributions 26
dependent samples 337–338
dependent variables 202, 217, 402, 427
**describe** command 108
descriptive statistics 2, 228, 269
  defined 2
descriptive studies 391–395
**detail** option 79, 424
diagnostics 524–529, 533–536
dichotomous independent variables 536
  interaction and 603–610
  multiple regression 571–576
dichotomous variables 4, 244
  defined 82
  mean values 82
  phi coefficient 184
  Point Biserial correlation coefficient 176–185
  simple linear regression and 214–217
discrepancy 531
discrete probability 228–230
discrete variables 8–9, 26, 47, 244
disjoint events 230
  additive rules of probability 231–233
disordinal interaction 461
dispersion *see* spread
**display** command 96, 168, 246–247, 253, 256, 302, 331, 423
distribution
  asymmetric curves 95–99
  shapes of 95–99
  *z*-scores 133–134
distribution location 74–86
  arithmetic mean as 80
  comparing mean, mode, and median 84–85
  mean 80–84
  median 78–80
  mode 74–78
  selecting measures of 99
  weighted mean 83–84
distribution-tied methods 622
division, in linear transformation 119
*Doctrine of Chances* (de Moivre) 282
**doedit** command 147, 148
**drop** command 108
Dynarski, Susan 410

**edit** command 19
effect size
  hypothesis testing and 305–306, 310
  independent sample *t*-test 349–353
  one sample *t*-test 333–336
  one-way analysis of variance 426–431
  paired sample *t*-test 359–360
  partial eta squared 476–477
  *R* 520–524
  simple linear regression and 217
  two-way analysis of variance 476–477
Efron, Bradley 360
**egen** command 34, 78, 129, 143, 166
**egen mode** command 77

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-72583-5 — Statistics Using Stata
Sharon Lawner Weinberg , Sarah Knapp Abramowitz
Index
More Information



endogeneity 402
equations
  for *b*-weights 556
  for hypothesis tests 369
  independent variables 210
  standardized regression equation 210
  for *t*-test values 369
  unstandardized regression equation 573, 594, 606
  *z*-values 298
estimators
  biased and unbiased 285
  variance 339, 340
**estpost** command 656
events 228
Excel
  correlation matrices 661–663
  regression output 663–668
  summary statistics tables 656
  working directory 655–656
expected values 400
experimental mortality 405, 406
experiments
  agricultural 414
  natural 410
  probability 244
  simple 228
*Exploratory Data Analysis* (Tukey) 55
external validity 404
  for quasi-experimental design 404, 405–406

*F*-distribution 420–422
  descriptive statistics 435
  Levene's test 427
  one-way analysis of variance 425, 427
  **robvar** command 427
  two-way ANOVA 469
factorial designs 467–468
factorial notation 248
family-wise error rate ($\alpha_{FW}$) 413, 431, 444
firearms 79
first additive rule of probability 231, 232, 236
first-order effects 594, 599–600
Fisher, Ronald A. 301, 414
Fisher's exact test 635, 637–638
  hypergeometric distribution 635
  Stata for 635
Fisher *r* to *Z* transformation 509
Fisher's least significant difference (LSD) 432
five-number summary 51–52
fixed-effects models 479
fixed factors 479
**forval** command 98
*Framingham* data set 17, 62, 113, 379
Framingham Heart Study 675
**freq** option 46
frequency distribution table 27, 32, 56
frequency distributions 26
  bar charts 28
  boxplots 51–56
  five-number summaries 51–52
  graphical selection summary 56
  grouped frequency distribution 34
  histograms 38–39
  line graphs 40–42
  mutually exclusive and exhaustive categories and 28
  ogive curve 46
  percentile ranks 46–51
  percentiles 46–51
  pie charts 31
  relative frequency 28
  skewness of 95–99
  stem-and-leaf displays and 35

Galton, Sir Francis 165, 206–207
**gen** command 120, 145, 146
generate **random** command 273
global hypothesis 444
global variables 98
goodness-of-fit
  chi-square and 625–632
  regression equations and 524
Gossett, William Sealy 320
grand mean 474–475
**graph bar** command 57
**graph box** command 53, 55, 88, 100–102
**graph hbar** command 28
**graph twoway** command 208
graphs
  graphical displays of variables 56
  modifying appearance of 56
grouped frequency distribution 34

*Hamburg.dta* 159, 160, 676–677
*Handbook of Experimental Psychology* 3
**help** command 30, 46, 51, 79, 107, 120
heterogeneity *see* spread
heteroscedasticity 527
hierarchical analyses 595
hierarchical multiple regression 595
**histogram** command 39, 276, 278, 325
histograms 38–39, 56, 57
homoscedasticity 506
honestly significant difference (HSD) 431
hypergeometric probability distribution 635, 637–638
hypothesis testing 296
  decreasing standard error of means 311
  effect size 306–307, 310
  equations for 369
  interval estimation and 305–306
  Neyman and Pearson null 301
  nondirectional hypothesis test 306
  one sample *t*-test 327
  pitfalls of 312
  power of the test 307–312
  significance level 310

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-72583-5 — Statistics Using Stata
Sharon Lawner Weinberg , Sarah Knapp Abramowitz
Index
More Information



two-way analysis of variance 466–467
type I and II errors and 307–312

*Ice Cream.dta* 106, 506, 560, 679–683
identical trials 244
*Impeach.dta* 679–683
inconsistency *see* spread
independent samples 338–340
assumptions of 340–342
degrees of freedom 340
independent samples *t*-test 373
ANOVA as extension 416
effect size 349–353
flowchart 371
independent trials 244
independent variables 558
dichotomous 536
relative importance of in multiple regression 565–566
simple linear regression 214–217
index of separation 416
indirect linear relationship 166
inferential statistics 228
defined 2
random sampling and 270–273
sampling and 269, 275
infinite populations 274
influential observations 529–530
information criterion 684–685
instrumental variables 402
instrumentation 405
interaction 460–465, 591
dichotomous and continuous independent variables 603–610
disordinal 462
ordinal 462
statistically significant two-way interaction 600–603
interaction models 592
internal validity 404
interpretation 14
interquartile range (IQR) 50, 52, 86, 89–91
interval estimation 312
central limit theorem and 293
confidence interval and 293
hypothesis testing and 305–306
interval estimation 293–296
interval-leveled variables 175, 437
interval measurement of variables 5–6
inverse normal probability function 259
**invnormal(*p*)** function 259, 261
IQ scale 10, 75

joint distribution 503
"Justice at work and reduced risk of coronary heart disease among employees" (article) 394

Kaplan, E. L. 312
**kdensity** command 41
Kruskal–Wallis analysis of variance 639, 647–648
**kwallis** command 648

**label define** command 126, 145
**label value** command 126, 145
law of large numbers 240, 251, 253
*Learndis.dta* 112, 154, 160, 375, 378–379, 678
least interval-leveled variables 159–166, 596, 597, 604
least significant difference (LSD) 432
least squares criterion 205
legend option 46, 575
Levene's test 439
*F*-distribution 438
one-way analysis of variance 438, 439
**robvar** command 439, 494
two-sample designs and 346, 351
two-way analysis of variance 494
variances 340, 341, 346
leverage 529–530
Stata for 530–531
Likert scale 10
*Likert.dta* 125, 126, 144
line graphs 40–42, 56, 57
Stata 601
linear composite variables 554
linear regression
*b*-weight for 514–518
equation 202
influential outliers 529–530
leverage 529–530
residual plots 524–529
statistical significance 516
linear regression equation 503
best-fitting 202–209
criterion variables 512
dichotomous independent variables and 214–217
prediction accuracy with 56–58, 512
linear regression model
*R* as measure of fit of 210–217
scatterplot importance for 217–218
linear relationships 503
linear transformation 119–121
defined 118
effect of summary statistics of a distribution 121–123, 151–153
effect on measures of central tendency 123
effect on measures of shape 123
effect on measures of spread 123
exercises 151–153, 154–156
Pearson correlation coefficient and 170
reflection as 126
shape of a distribution and 121
standard scores 126–127
translation as 125
*z*-scores 127–128
**list** command 78
median as 78
location characteristics
mode as 74
of a univariate distribution 74

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-72583-5 — Statistics Using Stata
Sharon Lawner Weinberg , Sarah Knapp Abramowitz
Index
More Information



log file, in Stata 13, 18–19
log-linear analysis 639
logarithms 135
logical operators 104
longitudinal designs
**lwidth** option 41, 42, 57

main effects 460
Mandex 679
*ManDext.dta* 417
Mann–Whitney *U*-test 642–647
marginal counts 183
marginal effect 558
**margins** command 481, 482, 600, 602, 606
**marginsplot** command 601, 602
*Marijuana.dta* 260, 676–677
marker option 53
math achievement of tenth-graders in southern US 3
**matlist** command 660
matrices 169
  in Excel 661–663
maturation, in quasi-experimental design 405
mean 80–84
  column marginal mean 460
  comparing with mode and median 84–85
  dichotomous variables 82
  grand mean 474–475
  one-way analysis of means (ANOVA) 414–416
  row marginal mean 460
  two-way analysis of variance
  weighted mean 83–84
mean squares 418
measurement of variables
  data transformation 54
  location in 99
  spread in 99
median 78–80
  comparing with mode and mean 84–85
median splits 145–146
metrics 102, 118
mixed-effects models 479
mode 74–78
  bimodal 74–78
  comparing with mean and median 84–85
  trimodal 76
moderator variables 591
monotonic transformations 118
multiple regression
  additive models 590
  *b*-weight as partial slope in multiple regression 568–571
  basic equation with two predictors 554–555
  correlated predictors 556–558
  dichotomous independent variables 571–576
    first-order effect variables 599–600
  hierarchical 595
  interaction 596
  introduction 553–554
  model selection 610–613
    *R*-squared significance 566–568
  summary of important principles of 556–558
multiple *t*-tests 412–414
multiplication, in linear transformations 119
multiplicative rule of probability 233–236
mutually exclusive events 230

*NELS.dta* 11–14, 19, 26, 43, 44, 74, 159, 178, 236, 511, 536, 634
**nestreg** command 595, 597
Neyman, Jerzy 296, 301
**nodots** option 362
noise components 337
nominal variables 4, 26–32
  chi-square goodness-of-fit test 623–639
  graphical display of 56
non-randomized controlled studies 395
  Simpson's paradox and 397
nondirectional hypothesis testing 306, 331
nonlinear transformation 118
  logarithm and square root transformations 135–142
  ranking variables 142–144
nonparametric methods 639–649
  ordinal dependent variables and 639–649
  Spearman rank correlation coefficient 640
  Wilcoxon sign test 3, 640
**norm** command 325
normal distribution 255
normal probability model 254–261
  approximation of a binomial probability 260
  complement rule 255
normality 322, 325
null hypothesis 297

observational studies 394
ogive curves 45–46
omnibus *F*-test 432
one sample *t*-test 320–321
  confidence intervals 323–333
  degrees of freedom 320–321
  effect size 333–336
  equation 321–323
  exercises 376–378
  flowchart 370
  hypothesis testing 330–333
  parent population normal deviation and 322–323
  Stata commands 372
one-tailed tests 303
one-way analysis of variance (ANOVA) 414–426
  assumptions 419–420
  computation 417–420
  effect size 426–431
    post-hoc multiple comparison tests 431–432
  summary table 425
**oneway** command 33, 423, 428, 430, 432, 433, 438
ordinal interaction 462
ordinal measurement 4–5

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-72583-5 — Statistics Using Stata
Sharon Lawner Weinberg , Sarah Knapp Abramowitz
Index
More Information



ordinal variables
  chi-square test of independence 623–639
  graphical display of 56
  nonparametric methods 639–649
  Spearman rank correlation coefficient and 174–176
orthogonal factorial designs 467–468
outcome variable 393
outliers 44, 53, 603
  bivariate
  influential 529–530
  *z*-score detection of 130–134
**over** commands 58–61
  Stata commands

*p*-values 299, 437, 472
  bootstrap technique 374
  chi-square test of independence 630
  Fisher's exact test 635
  in *t*-distribution 322, 333, 345
paired samples *t*-test 353–360
  flowchart 371
panel design 399
parameters 269
parametric methods 622
parent populations 275, 282
parsimony 52
partial eta squared effect size 477
partial regression coefficients 558
partial slope 568–571
partialed product 594, 605
partitioning total sum of squares, two-way analysis of variance 468–469
Pearson
  Egon 25, 301
  Karl 25, 165, 626
Pearson product moment correlation coefficient 165, 167, 185, 217, 503, 509–512
  best-fitting linear equation 202–209
  bivariate normal distribution and 506–509
  causation and 172
  correlation strength of linear relationship and 168
  data reliability and 174
  dichotomous variables and 176–185
  direct linear relationships 166
  indirect linear relationships 166
    linear regressions and 206, 209–210
  linear transformations effect and 172–173
  ordinal variables and 174–176
  perfect negative linear relationships 168
  perfect positive linear relationships 168
  phi coefficient 176, 181–185
  Point Biserial correlation coefficient 176–185
  restriction of range and 173–174
  shape of underlying distributions and 174
  Spearman rank correlation coefficient 174–176, 640
  Stata for 167, 184, 510, 511
  strength of relationship 168, 176
  zero value for 506–509
per-comparison error rate 413, 444
percent distribution table 27, 33
percentiles 46–51, 58
  ranks 46–51
percents 28
perfect linear relationships 168
phi coefficient 176, 181–185
  contingency tables 183
  marginal counts 183
pie charts 31–32
planned comparison tests, Bonferroni adjustment 444–446
Point Biserial correlation coefficient 176–185, 216, 349
point estimation 291
population
  bias and 285
  estimators for 285, 291
  infinite 274
    parent populations 282
  population correlation coefficient 507, 509–512
  population standard deviation 275, 291, 293, 412
  sampling of 269–270
  standard error of prediction 513
  target 393, 405–406
population mean estimation 291
post-estimation statistics 168
post-hoc multiple comparison tests, one-way analysis of variance 435
post-hoc multiple means test, two-way analysis of variance 479–481
post option 482
power, in hypothesis testing 307–312
power analysis 538–540, 613–614
  multiple regression 580–582
**power** command 448
pre-post design 353
**predict** command 211
prediction model
predictions
  accuracy of with linear regressions 209–210
  standard error of estimate 513
  standard error of prediction 513
predictor variable 512
**prefix** command/s 361
**preserve** command 108
pretest to posttest design 354, 399, 405, 406
probability
  additive rules 231–233
  complement rules 230
  conditional 236–239
  discrete 228–230
  experiments 244–254
  law of large numbers 240, 251, 253
  total 239
probability models
  binomial probability model 244–254, 260
  binomial probability model applicability 249–254
  normal probability model 254–261
  theoretical probability model 244
"Probable error of the mean" (article, Gossett) 320

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-72583-5 — Statistics Using Stata
Sharon Lawner Weinberg , Sarah Knapp Abramowitz
Index
More Information



propensity score matching 402
**putexcel** command 656
**pwcompare** command
**pwcorr** command 179
**pwmean** command 433

quasi-experimental designs 395, 404
  experimental mortality 406
  external validity threats to 406–408
  instrumentation in 405
  internal validity threats to 404
  maturation 405
  selection bias in 405

*R* 520–524, 558
  effect size 520
  effect size for simple regression 217
  linear regression model and 210–217
  values 511, 512
*R* squared
  adjusted 523
  change
    *r* to *Z* relationship 509
    statistical change 566–568
  simple regression 520–522
  statistical significance 522–523
  true population 523
random assignments, in causal studies 395–396
random factors 479
random number generator 271–272
  obtaining a simple 273
  with and without replacement 273–275
random sampling 270–273
range and interquartile range 89–91
rank function 143
ratio measurement of variables 6
raw total SS 475
RDD *see* regression discontinuity designs (RDD)
**recode** command 126
recoding variables 144–147
rectangular distributions 76
redundancy 553, 558
reflection 126
**regress** command 208, 219, 509, 515, 560, 563, 567, 569, 571
regression coefficients 514, 515, 531
regression diagnostics 524–529, 533–536
regression discontinuity designs (RDD) 403
regression lines 203
regression models
  Excel output 663–668
  *see also* linear regression; multiple regression
regression to the mean 206
**regressn** command 606
regressors
relational operators 104
relational studies 391–395
  construct measures 391–395
relative frequency 28
relative frequency distributions 28, 182, 240, 254, 505
  residual option 211
  studentized 531
reliability 391
**replace** command 145
residuals
  best-fitting linear equations 204
  sum of squares due to 518, 519
  variances for 418–419
resistant statistics 89, 94
**restore** command 108
**return list** command 168, 429
**r(max)** return value 98
**r(min)** return value 98
**robvar** command 340, 346, 425, 438, 479, 494
row marginal mean 460
runiform function 273

*S* curves 45
sampling, random 270–273
  independent and dependent samples 337–338
  inferential statistics and 269–270
  of populations 269–270
  with and without replacement 273–275
sampling distribution of means 275–280
  central limit theorem 281–285
  empirically 275–280
  exercises 287–290, 319
  theoretically 280–282
sampling with replacement 234
SAT Critical Reading score 159
scatterplot varying color and pattern 222
scatterplots 185, 222, 526
  between two variables 160–171
  direction 163
  linear regression models and 217–218
  regression line superimposed 508
  residual 526
  shape and 160–171
  strength 163
school effectiveness in the United States 11–14
Schwartz, Gideon E. 613
**seed** option 362
selection bias 405
**set seed** command 279, 283
shortened form of commands 120
sign test 640
signal components 337
significance level 300
**signtest recordwt** command 641
simple effects 482
simple experiment 228–230
  defined 228
simple linear regression 512–513
  dichotomous independent variables and 214–217, 536
  effect size and 217
  exercises for interval or ratio variables 219–227
simple slopes 600

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-72583-5 — Statistics Using Stata
Sharon Lawner Weinberg , Sarah Knapp Abramowitz
Index
More Information



Simpson's paradox 397
skewness 95–99
  calculating 96–97
  defined 43
  skewness statistic 95
  Stata for 96–97, 98
  **summskew** 98, 99, 102, 344
  **summskewcell** 325, 466, 467
  *t*-distribution 333
Skulls data set 683
socioeconomic status 77
**sort cum** command 46
Spearman rank correlation coefficient 174–176, 640
spread 86–95, 133
  range and interquartile range 89–91
  selecting measures of 99
  standard deviation 93–95
  variance 91–93
square root transformations 135
standard deviation 93–95
  see also *z*-scores
standard error of estimate 514
standard error of the mean 281
standard error of mean differences 339
standard error of prediction 514
standard normal distribution 255
standard scores see *z*-scores
standardized regression equation 210
Stata
  **anova** command 429, 486
  arithmetic operators 104
  as balanced option
  bar charts for frequency and percent distributions of nominal data in 28
  bar graph height sorting 30
  binomial tail function 254
  **bootstrap** command 276, 278, 361, 429, 597
  bootstrap estimation 517
  boxplots 100–102
  **browse** command 19
  by option 101, 102, 343
  **capture** command 278
  **cd** (change directory) command 278
  **centile** command 50
  chi-square goodness-of-fit test
    raw data 628
    tabulated data 627
  chi-square independence test 630, 635
  **chi2tail** function 625, 627
  **clear** command 108, 277
  **clustered bar graph** commands 184
  **codebook** command 108, 627
  command line input 13
  command line syntax 106
  command options 107
  command summary for Chapter 2 56–58
  command summary for Chapter 3 104–105
  command summary for Chapter 4 150
  command summary for Chapter 5 189
  command summary for Chapter 6 219
  command summary for Chapter 8 260
  command summary for Chapter 9 286–287
  command summary for Chapter 10 313–314
  command summary for Chapter 11 372–374
  command summary for Chapter 13 450
  command summary for Chapter 14 540–541
  command summary for Chapter 16 582
  command summary for Chapter 17 614–616
  command summary for Chapter 18 649
  Cook's distance 531
  **correlate** command 179, 211
  **csgof** command 627, 628
  **cumul** command 45
  cumulative binomial probability value 252, 253
  **describe** command 108
  **detail** option 79, 424
  discrepancy 531
  **display** command 96, 98, 168, 246–247, 252, 253, 256, 257, 259, 302, 331, 423
  .do-files 147–150
  **doedit** command 147, 148
  **drop** command 108
  **edit** command 19
  **egen** command 34, 78, 129, 143, 166
  **egen mode** command 77
  **erase** command 278
  **estat bootstrap** command 429, 430
  Excel and 656
  **exp** function for denoting exponential constant *e* 511
  **expand** command 628
  Explore procedure 100
  *F*-distribution tables and Stata functions 420–422
  finding *p*-value associated with a chi-squared goodness-of-fit test 627
  for Fisher's exact test 635
  **forval** command 98
  freq option 46
  frequency and percent distributions 27, 33
  **gen** or **generate** command 120, 146, 147, 355, 593
  **generate random** command 273
  **global k = r(max)** command 98
  **global m = r(max)** command 98
  **graph bar** command 30, 57
  **graph box** command 88, 101, 133
  **graph hbar** command 31
  **graph matrix** command
  **graph twoway** command 208
  **help** command 30, 41, 49, 52, 79, 107, 120
  helpful hints 105
  **histogram** command 38–39, 41, 276, 278, 325
  **if** in command line 106
  **immediate** command 633
  **in** in the command line 106
  independent samples *t*-test 373–374, 538
  interface structure 2–12
  jitter option 610
  **kdensity** command 41
  **kwallis** command 648

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-72583-5 — Statistics Using Stata
Sharon Lawner Weinberg , Sarah Knapp Abramowitz
Index
More Information



Stata (cont.)
**label define** command 126, 145, 195
**label value** command 126, 145–146, 195
lcolor option 41
legend option 46, 575
Levene's test 426–431
leverage 530–531
line graphs 40–42
**list** command 78
log files 18–19
logging out 25
logical operators 104
lwidth option 41, 42, 57
**margins** command 481, 482, 600, 602, 606
**marginsplot** command 601
marker option 53
**mysm(Oh)** option 610
**nestreg** command 595, 597
**nodots** option 361, 364
obtain influence 531–533
one sample *t*-test commands 372
**oneway** command 33, 423, 428, 429, 433, 438
online resources 106
opening data file 108
operators and expressions 104
**over** commands 58–61
paired samples *t*-test commands 374
**pcorr** command 565
Pearson product moment correlation 168, 184, 510, 511, 541
post-estimation statistics 168
post option 482
**power** command 447, 538–540
**predict** command 211, 532, 541
leverage 531, 541
**predict predicted** command 222
**predict residual** command 211, 219, 569
**predict stres rstudent** command 531, 541
**predict y1predict x1orthx2resid**, residuals 527, 541
**predict yhat** command 211, 493
**predicted** command 536
**predit yhat** command 516
**prefix** command 361
**preserve** command 108
product term 594
**pwcompare** command 486
**pwcorr** command 179, 511
**pwmean** command 433
**qnorm** command 325
rank function 143
**recode** command 126
**regress** command 208, 219, 509, 515, 560, 563, 567, 569, 571, 593
regression analysis 514–515
renaming variable in a bootstrap procedure 517
**replace** command 144–147
**restore** command 108
**return list** command 168, 429
**r(max)** 98
**r(min)** 98
**robvar** command 340, 346, 425, 479
runiform function 273
**save** command 19, 25, 147
**seed** option 362, 364, 517
**set seed** command 279, 283
significance testing 593–598
**signtest recordwt** command 641
skewness 96–97, 98
**sort** command 90, 95
**sort cum** command 46
**stem pertak** command 35
studentized residual values 531
**summarize** command 80, 82, 90, 93, 105, 106, 108, 207, 324, 342, 424
**summarize detail** command 98
**summarize prod** command 168
**summarize ses** command 49, 58, 79, 81
**summskew** command 98, 99, 102, 344
**summskewcell** command 324, 466, 467
**tab1** command 28
**tabi** command 633
**table** command 133, 338
**tabstat** command 49, 50, 80, 81, 91, 105, 424
**tabulate** command 28, 34, 77, 82, 131, 133, 184, 231, 233, 234, 236, 276, 428, 493, 632, 636
tabulated data 627
**test** command
testing interaction for statistical significance 593–598
**ttail** command 331
**ttest** command 326, 332, 346, 364
**twoway** command 41, 46, 160, 177, 179, 255, 541, 575
**twoway line** command 41, 46
**twoway scatter** command 507, 533
**use** command 19, 149
variable creation 569
**varname** command 99
**xlabel** option 177
statistical significance
*b*-weight for 563–564
of effect size 306–307
hierarchical multiple regression approach 595
of interactions 593–598
of *p*-values 345
of *R*-values 522–523
*Statisticians.dta* 25
descriptive 3, 26
post-estimation statistics 168
purpose of 1–2
summary statistics 86, 98
statistics software 2
status quo 391–395
stem and leaf displays 35–38, 57
**stem pertak** command 35
*Stepping.dta* 499, 684
strength
in Pearson product moment correlation coefficient 168
in scatterplots 163
studentized range statistic 427

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-72583-5 — Statistics Using Stata
Sharon Lawner Weinberg , Sarah Knapp Abramowitz
Index
More Information



studentized residual value 531
sum of squares due to residual 519
studentized range distribution 431
sum of squares due to the regression (SSREG) 519
**summarize** command 49, 58, 79, 80, 81, 82, 90, 93, 99, 106, 108, 168, 207, 324, 342, 656
summary statistics 86, 98, 106, 121–123
in Excel 656
of temperature 90
**summskew** command 98, 99, 102, 105, 344
**summskewcell** command 324, 466, 467
installing 324
symmetric distributions, 42
symmetric index 506

*t*-statistic 320
*t*-test
independent samples 340–342
paired samples 353–360
protected 432
Type I error 413
*see also* paired samples *t*-test; Stata, independent samples *t*-test
**tab1** command 28, 634
**table** command 133, 373
tables and Stata functions 420–422
**tabstat** command 49, 50, 80, 81, 91, 105, 424
**tabulate** command 28, 34, 77, 82, 131, 133, 184, 231, 233, 234, 236, 276, 428, 493, 633, 635
target 392–393
target populations 393
alternative hypotheses and 320
bootstrap 363
degrees of freedom 320–321
*t*-distribution 320, 340
*z*-statistics 319
*TEMP.dta* 87
**test** command
theoretical probability model 244
total probability 239
transformations 118
translation 125
trimodal distributions 76
true control groups 398–401
true experiment 395–396
**ttest** command 326, 332, 346, 364
Tukey, John 55, 431
two-factor designs 398, 457–460
two-sample designs 336–337
assumptions of 337–338
independent samples 353
Levene's test 346, 351
noise components in 337
repeated-measures design 353
signal components in 337
two-way analysis of variance
assumptions 466–467
balanced and unbalanced factorial designs 467–468
column main effect test 465
computation by hand 469–474
crossed factorial designs 458
degrees of freedom 469
effect size 477
*F*-test 469
fixed vs. random factors 479
grand mean 460, 475
hypothesis testing 465
interactions in 460–465
joint effects 475–476
line graph 461
modeling scores 475
partitioning total sum of squares 468–469
post-hoc multiple means test 479–481
randomized block designs 458
row main effect test 465
simple effects 482–484
two-factor design 457–460
twoway 609
**twoway** command 42, 46, 162, 177, 179, 255, 533, 575
**twoway line** command 42, 46
Type I error 413

unbalanced factorial designs 467–468
unbiased estimators 285
ungrouped frequency distribution 35, 47
uniform distributions 76
uniform random number generator 273
unimodal distributions 77
univariate distributions 26, 32
bimodal 75
data value occurrences 26
mean 80–84
median 78–80
mode in 74–78
range of 89–91
shape of 74
spread of 74, 86–95
unpredictability *see* spread
unrelated samples *see* independent samples
unstandardized regression equations 208
**use** command 19, 149

validity 391
external for quasi-experimental design 404, 405–406
internal for quasi-experimental design 404
values
cumulative binomial probability value 252, 253
expected 400
variability *see* spread
variables
choosing scale of measurement of 6–7
combining variables 147
interval measurement of 5
linear composites 554
measurement 3–5
nominal measurement of 4, 26–32
ordinal measurement of 4

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-72583-5 — Statistics Using Stata
Sharon Lawner Weinberg , Sarah Knapp Abramowitz
Index
More Information



variables (cont.)
ratio measurement of 6
recoding variables 144–147
redundant 553
variance 91–93
between group 416, 417
residual variance 426
test of means and 414–416
within group 416, 417
**varname** command 99

*Wages.dta* 613–614
weighted mean 83–84
Whitney, D. Ransom 639
Wilcoxon sign test 3, 640
Wilcoxon's Rank-sum test 642
within-group variance 416

xlabel option 177

Yates' correction 628

*z*-scores 127–128, 295, 300, 319
comparing different distributions and 133–134
outlier detection and 130–134
percentile ranks and 135
*z*-test 299
zero-order correlations 555

© in this web service Cambridge University Press & Assessment www.cambridge.org