

Index

- abline** command, 152, 200, 552, 558
- abs** command, 122–123, 138
- absolute value, 87, 122–123, 195
- accumulating data, 42
- additive rules of probability, 224–227
 - disjoint events, 223–225, 229, 232
- addmargins** command, 226, 230, 233, 237
- adjusted *R*-squared, 498–499, 540, 584–585, 588
- aggregate** command, 439–440, 470
- AIC** command, 585, 588
- Akaike, Hirotugu, 585
- Akaike information criterion (AIC), 584–585, 588
- alternative hypothesis, 294–296, 300, 301, 327, 329, 330, 331, 332, 341–346, 353–356
 - defined, 294
 - directional, 300, 301, 332, 354–355
 - non-directional, 300, 308, 328, 344
- analysis of variance (ANOVA), 402, 403, 407, 414–415, 430–432, 571
 - Bonferroni adjustment, 427, 428–429, 461, 464–465
 - computation, 404–413
 - F*-distribution, 408
 - first-order effect variables, 574, 579
 - index of separation, 404
 - Kruskal–Wallis, 617–620
 - Levene’s test, 337–338, 342–344, 345–348, 413, 423–425
 - mean squares, 405
 - multiple *t*-tests, 399, 429
 - one-way, 399–402, 403, 427
 - assumptions of, 406–407
 - effect size, 414–416
 - F*-ratio, 406–414, 449–450, 453–454, 495
 - regression, 571
 - two-way, 437–441, 446, 447
 - assumptions, 446–447
 - decomposing score deviations, 455
 - effect size, 457–459
 - F*-ratio, 455
 - interaction effects, 441–446, 456–457
 - partitioning sum of squares, 448
 - within group variance (MSW), 405–406, 407–408, 417, 452
- anova** command, 494, 545–546, 571
- ANOVA. *See* analysis of variance (ANOVA)
- Anscombe dataset, 209, 499–501, 626
- aov** command, 412, 431–432, 465, 471–472
- arithmetic mean, 48, 76, 287
 - compared to mode and median, 76
- arithmetic operators, 97
- assumptions
 - binomial probability model, 242
 - independent samples *t*-test, 336–337, 340
 - paired samples *t*-test, 347–352
 - random sampling, 270
 - two-way ANOVA, 446–447
- asymmetric curves, 41–42
- asymmetric relationships, 484
- average. *See* arithmetic mean
- average squared deviation, 195
- average treatment effect, 386–387, 391–392
- b*-weight, 490, 515–517, 533–534, 540–542, 545–547, 570–573
 - multiple regression, 542–547
 - statistical significance, 493–497
- balanced design, 430, 448
- balanced factorial design, 448
- bar graphs, 27–28, 243–244
 - clustered, 172–175, 176, 179–181
 - histograms and, 35
 - percent distributions, 28
- barplot** command, 27, 28, 54, 172
- Basketball dataset, 109–110, 112, 552–553, 558, 626
- Bayes, Thomas, 234
- Bayes’ theorem, 234–235
- Bayesian information criterion (BIC), 584–585, 588
- behavioral sciences, 664, 666
- Bem Sex-Role Inventory*, 134, 664
- best-fitting linear equations, 193, 194, 196, 201
 - average squared deviation and, 194
 - least squares criterion, 196
- beta weights, 199, 541–542, 543
- between group variance, 403–405, 414, 449
- bias, 282–283, 384–385, 388–392, 554
 - reducing, 390–391
 - selection, 392, 554
- BIC** command, 585, 588
- bimodal distributions, 71–72
- binomial distribution, 244–245, 246, 259, 611
- binomial probability model, 242, 247, 248, 249
 - applicability, 242, 247
 - assumptions of, 242

INDEX

669

- distribution in, 242
- identical trials, 242, 247
- bivariate normal distribution, 481–483, 486, 489–490
- bivariate normality, 481, 483, 505, 593
- bivariate outliers, 507
- blocking design, 385, 394
- blocks, 385, 439, 571, 639
- Blood dataset, 321–322, 338–339, 342, 348–350, 360–362
- Bonferroni adjustment, 427, 428–429, 461, 464–465
 - multiple measures, 429
- Bonferroni and Scheffe post-hoc tests, 419, 426
- boot.mean** command, 276
- bootstrapping, 275–276, 666, 667
- box-and-whiskers plot, 48
- boxplot** command, 84, 113, 121, 123
- boxplots, 48–53, 82–84, 111–114, 176–179, 180
- causal studies, 380–381, 382, 385
- causality, 381, 387–388, 389, 391–392, 394–395
- centering, 116, 574, 578, 579
- central limit theorem (CLT), 277–282, 289–292, 294, 299–300, 305–307
- central tendency, 70, 111–113, 114–115
- charts. *See* graphs and charts
- chi-square distribution, 594–596, 598, 600, 617
- chi-square test of independence, 599, 601, 604–606
- classical test theory, 379
- clustered bar graphs, 172–175, 176, 179–181
- Cohen's *d*, 303–304, 308–309, 355–356, 378–379, 517–519
- column effects, 450, 456, 470
- column main effect test, 446, 457, 460
- column marginals, 174
- conf.int** command, 543
- confidence interval, 290–293, 318–324, 342–347, 356–360, 487–488
- confounders, 382, 384, 395, 553
- constants, 3
- contingency tables, 174, 176, 179, 606–608
- continuous variables, 8–9, 44, 46, 256
- control groups, 381–383, 385–391, 392–394
- Cook's influence values, 510–514, 520
- cor** command, 159, 171–172, 203
- cor.test** command, 182, 487–488, 538–539, 544, 557–558
- coronary heart disease, 381, 628–630, 666
- correlation analysis, 166, 205, 380
- counterfactuals, 387, 388
- Court TV, 71
- criterion variable. *See* independent variables
- Cronbach's alpha, 379
- crossed factorial design, 438–439, 450
- cross-sectional design, 347
- cross-tabulation, 173–176, 180, 182, 230
 - table, 173–174
- cumulative percent distribution, 43, 45
- Currency dataset, 148, 627
- cut** command, 32, 53
- δ (effect size), 306–308
- D* (standard error of estimate), 152–153, 160, 195–196, 201, 235–236
- DASL (Data and Story Library), 163, 320, 626–627, 639
- data analysis, 665, 667
- data management, 136–138
- data transformation
 - linear, 108–112, 115–117, 120, 125–127
 - additive, 115
 - defined, 109, 125, 204
 - multiplicative, 112, 138
 - reflection, 579
 - standard scores, 115
 - summary statistics and, 112–115
 - z*-scores, 115, 120
 - nonlinear
 - logarithms, 128–130
 - rank transformation, 133
 - square root transformations, 127–128, 130
- dataset, 637–639, 641
- datasets
 - Anscombe, 209, 499–501
 - Basketball, 109–110, 112, 138, 552–553, 558
 - Blood, 321–322, 338–339, 342, 348–350, 360–362
 - Brainsz, 148, 163, 627
 - Currency, 148, 627
 - Hamburger, 133, 156–159, 165–168, 198–200, 207–208
 - Ice Cream, 584, 630–632
 - Impeach, 631
 - ManDext, 412–413, 431–432, 439–440, 458–459, 470–471
 - ManDext2, 442–443
 - Marijuana, 39–41, 148–152, 181, 228–232, 242–246
 - NELS, 113–115, 119–121, 228–230, 247–249, 604–606
 - Politics, 637
 - States, 3, 51–52, 54–55, 147–151, 597–598
 - Stepping, 639
 - Temp, 82–85, 86–87, 89, 91–92, 97–98
 - UpperBodyStrength, 554–556, 640
 - Wages, 466, 472
- dbinom** command, 244–245, 248–249, 251, 259
- degrees of freedom, 318, 319, 336, 405, 406, 407–411, 594–597
 - ANOVA, 405
 - chi-square distribution, 594–597, 598, 600, 617
 - independent samples, 336, 341, 357
 - one-sample *t*-test, 318–319
 - t*-distribution, 318–320, 328, 329–330, 407
- density** command, 39
- density plots, 39, 54
- dependent, 278
- dependent variable, 414–415, 422–423, 437, 441–442, 454–455
- descriptive statistics, 2, 128–129, 221, 325, 339–340
- descriptive study, 379
- deviation, 87, 89–90, 194–195, 317–318, 320
- diagnostics, 499, 500, 511, 514
- dichotomous variables, 147, 166–168, 171–172, 176–177, 242–243
 - phi coefficient, 166, 171
 - point biserial correlation coefficient, 166
- discrepancy, 507–509, 512, 598

- discrete probability, 221, 223
- discrete variables, 9, 25, 44, 242
- disjoint events, 223–225, 229, 232
 - additive rules of probability, 224–227
- disordinal interaction, 442
- dispersion. *See* spread
- distributions, 41, 70–82, 84–96, 114–134, 273–282
- dnorm** command, 323
- effect size, 303–304, 307–309, 351–361, 414–416, 517–520
 - one-sample *t*-test, 329, 358–361
 - one-way analysis of variance, 431
 - partial eta squared, 457
 - two-way analysis of variance, 437
- endogeneity, 389
- equations, 87–88, 356, 458, 535, 545–547
- error, 304–305, 400, 417–418, 455–456, 499–500
- estimators, 268, 282, 336, 405
 - biased, 283
 - unbiased, 282–283, 317
 - variance, 336, 403, 405
- events, 221–230, 231–233, 234–236, 244–245, 247–249
- Excel, 16
- expected values, 208, 597–598, 601
- experimental design, 391, 554
 - blocking design, 385, 394
 - crossed factorial, 438–439, 450
 - factorial design, 438, 443, 448, 466
 - pretest–posttest design
 - one-group, 347, 386
 - randomized blocks designs, 439
 - regression discontinuity designs, 390
 - repeated measures, 619
 - single-sample, 316–318
 - true vs. quasi-experimental design, 391
 - unbalanced design, 448
- experimental mortality, 392–393
- experiments, 236, 242, 664, 666
 - agricultural, 401
 - probability, 242
- exponents, 1, 125
- external validity, 392, 393, 394
- factorial design, 438, 443, 448, 466
- factorial notation, 245
- family-wise error rate, 400–401, 427–429, 461, 466
- F*-distribution, 406–408, 426, 453
- firearms, 74–75, 76–77, 78–79
- first additive rule of probability, 224–225, 229, 232
- first-order effect variables, 574, 579
 - centering, 574
- Fisher, R.A., 298, 401, 606
- Fisher's exact test, 606–608, 610, 620
- Fisher's least significant difference (LSD), 417–420, 422, 425, 432
- fixed factors, 459
- fixed-effects models, 459
- following commands, 15, 25–26, 255–257, 576–577, 612
- Framingham Heart Study, 628
- F*-ratio, 406–408, 411–412, 449–450, 453–455, 495
- frequency distributions, 25–27, 29–30, 31–32, 70, 78
 - bar charts, 28
 - bimodal, 71
 - bivariate, 483
 - graphical representation, 25–26, 28
 - grouped and ungrouped, 33, 45
 - relative frequency, 26
 - table, 30–31, 53, 274–275
- F*-statistic, 491, 510, 513, 516
- F*-test, 309, 407, 416–417, 449
- Galton, Francis, 152, 198
- Gauss, Johann Carl Friedrich, 251
- global hypothesis, 427
- goodness-of-fit, 594, 597, 598–601, 619–620
- Gosset, William Sealy, 317
- grand mean, 405, 440, 450–451, 455–456
- graphs and charts, 1–2, 28, 37–39, 53–54
 - bar charts, 27–28, 243–244
 - boxplots, 48–53, 82–84, 111–114, 176–179, 180
 - histograms, 35–41, 49–51, 53–54, 84–85, 254–255
 - line charts, 37–40, 53–54, 441, 444
 - pie charts, 26, 29–30, 53–54
 - scatterplots, 148–152, 180–181, 198–201, 202–206, 500–507
 - stem-and-leaf plots, 33–34, 37–39, 52–53
- grouped frequency distribution, 32, 43, 53
- grouped frequency tables, 32
- grouping variable, 83, 414–415, 437–438
- Hamburger dataset, 133, 156–159, 165–168, 198–200, 207–208
- Handbook of Experimental Psychology*, 3, 667
- help** command, 15
- heterogeneity. *See* spread
- heteroscedasticity, 505, 664
- hierarchical analyses, 571
- hierarchical multiple regression, 571
- hist** command, 11–12, 35–37, 54, 84–85, 322–323
- histogram** command, 12
- histograms, 35–41, 49–51, 53–54, 84–85, 254–255
- homoscedasticity, 483–484, 486, 490, 501–505
- honestly significant difference (HSD), 416–418, 419–420, 422, 425–426
- hypergeometric probability distribution, 606–607
- hypothesis testing, 296–297, 298–299, 302, 303, 304–309, 330–331
 - alternative hypotheses, 294–296, 300–301, 327–332, 341–346, 353–356
 - effect size, 309
 - interval estimation and, 287, 302–303, 309
 - Neyman and Pearson null, 298
 - significance level, 297
 - Type I error, 304–305, 400–401, 407, 417–418, 427–429
- Ice Cream dataset, 584, 630–632
- identical trials, 242, 247
- ifelse** command, 134–135, 138
- Impeach dataset, 631
- inconsistency. *See* spread

INDEX

671

- independent groups designs, 352, 356, 399
- independent groups *t*-test, 380, 399, 401, 402
- independent samples, 334–336, 340–342, 355–359, 361–362, 515–517
 - assumptions of, 336, 347
 - degrees of freedom, 336, 341, 357
- independent samples design, 352
- independent samples *t*-test, 335, 340–342, 347, 357–359, 361–362
 - effect size, 344, 355, 361
- independent variables, 193, 437, 442
- index of separation, 404, 405–406
- inferences, 2, 201, 309
- inferential statistics, 2, 221, 268–270, 272–273
 - defined, 221
- infinite population, 272
- influential observations, 499, 507
- information criterion, 584
- instrumental variables, 389–390
- instrumentation, 392
- interact.plot** command, 575, 581–583, 586
- interaction, 441–446, 449–462, 465–468, 469–472, 566–584
 - disordinal, 442
 - ordinal, 442–443, 610
 - statistically significant two-way interaction, 575
- interaction models, 573
- interaction.plot** command, 442, 466, 470, 472
- internal validity, 391–392, 393–394
- interpretation, 1–2, 16, 177, 197–198, 578
- interquartile range (*IQR*), 48–49, 85–87, 92–95, 97–99, 113–116
- interval estimation, 287, 289, 293, 302
 - confidence interval and, 293
 - hypothesis testing and, 302–303
- interval level, 5–6, 8, 147
- interval-leveled variables, 147
- inverse normal probability function, 258
- IQ* scale, 10, 71–72, 118–119, 632
- IQR** command, 87
- joint distribution, 481
- Journal of Statistics Education*, 631, 666
- “Justice at work and reduced risk of coronary heart disease among employees” (article), 381
- kruskal.test** command, 619–620
- Kruskal–Wallis analysis of variance, 617–620
- Kruskal–Wallis rank sum test, 619
- law of large numbers, 236, 249–251, 416
- law of total probability, 236
- Learndis* dataset, 632
- least interval-level labeled variables, 147
- least significant difference, (LSD), 417–420, 422, 425, 432
- least squares criterion, 196
- least squares regression, 194, 196, 584, 664
- length** command, 158
- Levene’s test, 337–338, 342–344, 345–348, 413, 423–425
 - variances, 337–338, 344, 345
- levenes.test** command, 342, 346, 413, 423
- leverage, 507–509, 512, 519
- Likert dataset, 632–633
- Likert scale, 10, 134, 632
- line.graph** command, 37, 54
- line graphs, 37–40, 53–54, 441, 444
- linear equations, best-fitting, 193–194, 196, 201
- linear regression, 202, 203, 488–492, 499–505, 532–533
 - accuracy of prediction, 201–202
 - beta weights, 199, 541–542, 543
 - b*-weight for, 490, 533
 - diagnostics, 499, 500, 511, 514
 - equation, 196–197, 198–199, 201–202, 488–492, 531–533
 - best-fitting, 194, 196, 201
 - leverage, 507, 512
 - multiple, 489
 - R* as model of fit, 202
 - residual plots, 500–506
 - statistical significance, 490
- linear regression model, 201, 202, 209–211, 499–505
- linear relationship, 151–156, 159–161, 166, 209, 500–504
 - perfect, 193
 - perfect negative, 152, 158
 - perfect positive, 158
- linear relationships, 152–156, 161, 176
- linear transformation, 108–112, 113–117, 120, 125–127
 - defined, 109, 125, 204
 - standard scores, 115
 - z*-scores, 115, 120
- lines** command, 254, 323
- lm** command, 199–200, 519–520, 551–553, 557–558, 585
- log** command, 128
- log transformations, 128–130
- logarithms, 125–126, 128–130, 164
- logical operators, 97
- log-linear analysis, 610
- main effects, 446, 448, 456–457, 460, 469–472
- ManDext* dataset, 412–413, 431–432, 439–440, 458–459, 470–471
- Mann, Henry B., 610
- Mann–Whitney *U*-test, 614
- marginal effect, 536
- marginal sums, 226, 233, 237
- marginals, 174
- Marijuana* dataset, 39–41, 148–152, 181, 228–232, 242–246
- maturation, 392
- max** command, 170
- mean, 75–82, 273–283, 328–337, 415–430, 437–446
 - column marginal mean, 440, 446
 - grand mean, 405, 440, 450–451, 455–456
 - row marginal mean, 440–441, 446
 - two-way analysis of variance, 437, 446
 - weighted mean, 78–79, 384
- mean regression, 197
- mean squares, 405
- measurement
 - indirectly observable traits, 10
 - interval level, 5–6, 8, 593
 - level of, 3–4, 7–8, 10–11

- measurement (cont.)
 - nominal level of, 4, 7, 594
 - ratio level of, 6
 - scales, 6–7, 46, 51
 - units, 89
- measurement of variables, 3
- median, 74–76, 79–82, 92–95, 113–116, 134–135
- median splits, 134
- mixed-effects models, 459
- mode, 70–74, 75–76, 79–81, 92, 97
 - bimodal, 71
- the.mode** command, 74, 80, 97
- Moivre, Abraham de, 278
- monotonic transformations, 108
- multiple *R*-squared, 491, 495, 513, 516
- multiple regression, 519, 531–533, 534–536, 537–538, 557, 558
 - basic equation with two predictors, 532
 - hierarchical, 571
 - interaction, 566, 571, 586
 - introduction, 531, 664
 - model selection, 585
 - power analysis, 556
 - principles, 536
- multiplicative rule of probability, 227–229, 231–232
- mutually exclusive events, 225
- National Center for Education Statistics (NCES), 16, 633
- National Education Longitudinal Study. *See* NELS
- NELS (National Education Longitudinal Study), 93–95, 121, 578, 586–587, 633
- NELS dataset, 113–115, 119–121, 228–230, 247–249, 604–606, 633–634
- nominal variables, 4
- nonlinear transformations, 108, 125–126, 127
 - logarithms, 128–130
 - rank transformation, 133
 - square root transformations, 127–128, 130
- nonparametric methods, 593–594, 610
 - Wilcoxon sign test, 610
- non-randomized controlled study, 382, 383
- normal distribution, 252–255, 256–260, 277–279, 481–483, 489–490
 - bivariate, 481–486, 489–490, 497, 507
- normal probability model, 242, 251–252
- normality, 319–320, 322–325, 337, 412–413, 483–484
- null hypothesis, 294–302, 305, 308–309, 605–612, 614–619
- observational studies, 381–382, 389, 394–395, 666
- omnibus *F*-test, 417–418, 420, 425–427, 459
- one-sample *t*-test, 319–320, 325–327, 331–332, 335, 357–361
 - confidence intervals, 320
 - degrees of freedom, 318–319, 330, 332
 - effect size, 329–330, 361
 - equation, 317, 320, 329
- one-tailed tests, 300, 355
- operators, 97
- ordinal interaction, 442–443
- outcome variable, 515–517, 535–536, 537–539, 593–594, 610–611
- outliers, 51–52, 95–96, 121, 499–505, 507
 - bivariate, 507
 - influential, 499, 507
- paired samples design, 347, 349
- paired samples *t*-test, 346–348, 350–351, 355–359, 362, 611
- panel design, 386
- par** command, 84–85, 97
- parameters, 39–41, 268–269, 277–278, 407, 593
- parametric methods, 593
- parent populations, 336–337, 406, 412
- parsimony, 50
- partial regression coefficients, 536, 542
- partial slope, 546, 547–548
- partialled product, 570, 579
- pbinom** command, 250
- Pearson, 157–159, 160–163, 164–167, 180–182, 604–607
- Pearson, Egon S., 298
- Pearson, Karl, 152, 298, 597
- Pearson correlation coefficient value, 160–162
- Pearson product-moment correlation coefficient, 152–156, 157–158, 161–163, 164–167, 174–175
- percent.table** command, 26, 28, 32, 54
- percentiles, 45–48, 52–55, 322
 - ranks, 46, 400, 427
- pf** command, 411, 431, 453, 471
- phi coefficient, 166, 171, 174
- pie charts, 26, 29–30, 53–54
- pie** command, 54
- plot** command, 39–41, 54–55, 148–151, 181, 581–583
- pnorm** command, 260, 280, 295, 311
- point biserial correlation coefficient, 166–167, 168, 208
- point estimation, 287–288
- Politics dataset, 637
- polygon** command, 255, 257, 260, 310–311
- population, 268–283, 287–292, 316–318, 335–344, 614–619
 - in inferential statistics, 221, 268, 272, 273
 - infinite, 272, 277
 - population correlation coefficient, 485, 487
 - population standard deviation, 269, 287, 291
 - standard error of prediction, 489–490
 - target, 380, 393
- post-hoc tests, 416, 419, 459–460
- PostHocTest** command, 419, 432, 468, 472
- potential outcomes, 387
- power analysis, 353–355, 361–362, 472, 556
 - multiple regression, 556, 664
- power** command, 430, 469, 517–518, 557
- predict** command, 203, 212, 492, 515
- predictor variable, 207, 488, 536–537, 546–547, 558
- pretest–posttest design, 347, 386
- probability, 221–226, 227–236, 242–255, 272–274, 606–612
 - additive rules, 224–227
 - binomial model, 242, 244, 247, 248
 - complement rule, 223, 225, 231, 234–236, 256–257
 - conditional, 229–233, 234–235
 - discrete, 221, 242, 244, 254

discrete case, 221, 223	cut, 32, 53
disjoint events, 223–225, 229, 232	dbinom , 244–245, 249, 251, 259
experiments, 242	density , 39
law of large numbers, 236, 249–251	fisher.test , 607, 609, 620
law of total probability, 233–236	help , 15
models, 242	hist , 11–12, 35–37, 54, 84–85, 322–323
multiplicative rule, 227–229	ifelse , 134–135, 138
normal model, 242, 251–252	<i>install.packages</i> , 461, 465, 471–472
total, 222, 233–235, 236, 607	interact_plot , 575, 581–583, 586
probability distribution, 244, 256, 606	interaction.plot , 442, 466, 470, 472
probability models, 242	IQR , 87
binomial probability model, 242	kruskal.test , 619–620
normal probability model, 242	length , 158
theoretical, 242	levenes.test , 342, 346, 413, 423
theoretical probability model, 242	library, 13–15, 137, 353, 465, 471–472
propensity score matching, 389	line.graph , 37, 54
protected <i>t</i> -test, 417	lines , 254, 323
<i>p</i> -values, 295–298, 341–346, 407–411, 452–453, 604–608	lm , 199–200, 519–520, 551–553, 557–558, 585
chi-square test of independence, 606	log , 128
pwr.anova.test command, 430–432	max , 170
pwr.r.test command, 517–518, 520	the.mode , 74, 80, 97
	par , 84–85, 97
qqline, 322, 325–326, 360	pbinom , 250
qqline command, 322, 325, 360	percent.table , 26, 28, 32, 54
qqnorm command, 322, 325–326, 360	pf , 411, 431, 453, 471
qt command, 321, 341, 347, 360	pie , 54
quadratic relationships, 151, 159, 504	plot , 39–41, 54–55, 148–151, 181, 581–583
quantile command, 47, 55, 75, 97	pnorm , 255–260, 280, 295, 311
quantiles, 45, 47, 322	polygon , 255, 257, 260, 310–311
quartiles, 46–48, 55	PostHocTest , 419, 432, 468, 472
first, 47, 86	power , 430, 469, 517–518, 557
second, 47–48, 70, 74	predict , 203, 212, 492, 515
third, 47–48, 86	pwr.anova.test , 430–432
quasi-experimental design	pwr.r.test , 517–518, 520
external validity, 392, 393, 394	qnorm , 258, 260
internal validity, 391–392, 393–394	qqline , 322, 325, 360
quasi-experimental designs, 382, 391–392, 393	qqnorm , 322, 325–326, 360
	qt , 321, 341, 347, 360
<i>R</i> (measure of linear model fit). <i>See also</i> <i>R</i> -squared, 202–205, 209	quantile , 47, 55, 75, 97
<i>r</i> (Pearson product-moment correlation coefficient), 157–158, 159–162, 201–206, 208–209, 517–520	rank , 133, 138
<i>R</i> (software), 11–13	relational operators, 97, 123
command line, 12	scale , 120
commands	sd , 95, 98, 322–323, 325, 360–362
abline , 152, 200, 552, 558	se.skew , 94, 98, 325, 360–362, 412–413
abs , 122–123, 138	seq , 323
addmargins , 230, 233	set.seed , 276
aggregate , 439–440, 470	SIGN.test , 612, 620
AIC , 585, 588	sim slopes , 576, 581, 586–587
anova , 494, 545–546, 571	skew , 94, 98, 325, 360–362, 412–413
aov , 412, 431–432, 465, 471–472	skew.ratio , 94, 98, 325, 360–362, 412–413
barplot , 27, 28, 54, 172	stem , 33, 34, 54
BIC , 585, 588	summary , 136–137, 199, 557–558, 569, 585
boot.mean , 276	summary (Chapter 2), 53–55
boxplot , 84, 113, 121, 123	summary (Chapter 3), 97–99
chisq.test , 599–600, 603–606, 619–620	summary (Chapter 4), 138
choose , 248–250, 259	summary (Chapter 5), 181–182
cor , 159, 171–172, 203	summary (Chapter 6), 212
	summary (Chapter 7), 237
	summary (Chapter 8), 259–260

- R (software) (cont.)
 - summary (Chapter 9), 283
 - summary (Chapter 10), 310–311
 - summary (Chapter 11), 360–362
 - summary (Chapter 13), 431–432
 - summary (Chapter 14), 470–472
 - summary (Chapter 15), 519–520
 - summary (Chapter 16), 557–558
 - summary (Chapter 17), 586–588
 - summary (Chapter 18), 619–620
 - t.test**, 325–326, 342–344, 350, 353–356, 360–362
 - table**, 28–31, 54, 172–175, 224–226, 603–606
 - test**, 518, 556–557
 - var**, 89, 95, 98, 379
- console, 11
- data frames, 15
- data management, 136–138
- downloading, 2
- graphical parameters, 41
- interface, 11
- package installation, 13–15
- percentile estimation, 46
- R file, 136–138, 181, 321–323, 326
- .R scripts, 11–12, 136–138, 181, 294–295, 321–324
- sur package, 13–16, 25, 329–330, 632–634, 641
- r*-values, 517
- R*-squared, 202, 495, 544–545, 570–571, 579–580
 - adjusted, 498–499, 540, 584–585, 588
- random assignment, 382–383, 385, 388–389
- random factors, 333–334, 459
- random sampling, 269–272, 302, 380, 614
- randomized blocks designs, 439
- random-number generators, 271
- range and interquartile range, 85, 89
- rank** command, 133, 138
- recoding variables, 134
- rectangular distributions, 72
- reflection, 115, 117, 128, 579
- regression, 197–198, 388, 493–495, 533–534, 664, 667
- regression coefficients, 199, 536, 542
- regression diagnostics, 499, 500
- regression discontinuity designs, 390, 665
- regression equation, 198–200, 203–208, 493–496, 507–510, 550–553
 - standardized, 202, 541
- regression lines, 551, 556
- regression models, 198, 500, 584
- regression to the mean, 197–198
- relational operators, 97, 123
- relational studies, 380
- relationship, 151
- relationships
 - direction, 151
 - shape, 151
 - strength, 152
- relative frequency, 26–27, 172, 236
- reliability, 161, 164, 378–379
 - test–re-test, 379
- repeated measures designs, 352, 386
- residual plots, 500–506, 512
- residual values, 203, 212, 500, 509
- residuals, 195, 412–414, 493–496, 500–509, 512–513
- resistant statistics, 86, 90, 95–96, 121
- Richter scale, 126
- row marginal mean, 440–441, 446
- RStudio, 11–14, 15, 136, 641
- sample variance, 269, 282–283, 336
- sampling, with and without replacement, 271–273, 278
- sampling distribution of means, 273–274, 275–276, 277–282, 299, 305–307
 - central limit theorem, 277–278, 282, 289–292, 305–307
 - empirically, 273
 - theoretically, 277
- sampling with replacement, 228, 247, 272–273, 282
- sampling. *See also* random sampling independent and dependent samples, 228–229, 270, 271–282, 305–307, 426–427
- SAT, 638
- SAT Critical Reading score, 147, 205, 638
- scale** command, 120
- scatterplots, 148–152, 180–181, 198–201, 202–206, 500–507
 - direction, 152, 160
 - strength, 152–156
- school effectiveness in the United States, 16, 633
- Schwarz, Gideon E., 584–585
- sd** command, 95, 98, 322–323, 325, 360–362
- se.skew** command, 94, 98, 325, 360–362, 412–413
- selection bias, 392, 554
- seq** command, 323
- set.seed** command, 276
- sign test, 610–614, 620
- SIGN.test** command, 612, 620
- significance level, 297–298, 302, 307–308, 331
 - two-tailed test, 354–355, 360, 518
- sim slopes** command, 576, 581, 586–587
- simple effects, 460–463, 464, 466
- simple experiment, 221–222
- simple linear regression, 202–203, 207, 488–489, 491–492, 532–533
- simple slopes, 575–577, 581, 586
- Simpson's paradox, 384
- six-number summaries, 48
- skew.ratio** command, 94, 98, 325, 360–362, 412–413
- skewness, 90–92, 94, 113–115, 322, 323, 412, 413,
 - calculating, 126
 - defined, 92
 - ratio, 91–92, 94, 321–323, 412–413, 447–448
 - skewness statistic, 90–91, 325, 329
 - standard error of, 98, 412–413, 431
- Smith, Stanley, 3
- socioeconomic status, 30–32, 43, 147, 394–395
- Spearman rank correlation coefficient, 164, 611
- spread, 48–49, 82, 87–89, 111–113, 114–115
 - range and interquartile range, 85
 - selecting measures of, 92
 - standard deviation, 89, 95, 112, 489–490

INDEX

675

- variance, 82, 87–89, 95, 404
- sqrt** command, 92, 126, 130, 138
- square root transformations, 125, 126–131, 282, 283
- standard deviation, 89–90, 117–121, 252–253, 277–283, 541–542
- standard error of estimate, 195, 201, 490, 495
- standard error of prediction, 489–490, 494
- standard error of the mean, 303, 308, 342, 347
- standard normal distribution, 253, 258–260, 319, 320
- standard scores, 117–118. *See* z-scores
- standardized regression equation, 202
- Stanford Test of Manual Dexterity, 401, 437–438
- Stanford–Binet intelligence test, 160
- Stata, 2, 585
- States, 3, 51–52, 54–55, 147–151, 597–598
- statistical significance, 417, 497–499, 545, 571, 666
 - b*-weight for, 490
 - of effect size, 414
 - hierarchical multiple regression approach, 571
 - of *p*-values, 341
 - significance level, 297–298, 302, 307–308, 331
- statistics software, 2
- stem** command, 33, 34, 54
- stem-and-leaf plot, 33–34, 37–39, 52–53
- strength, 152–156, 160–161, 166, 554–556, 640
- studentized range distribution, 416–417
- sum of squares due to the regression (SSREG), 494
- summary** command, 136–137, 199, 557–558, 569, 585
- summary statistics, 95–96, 115, 129, 137–138, 209–212
 - effect of linear transformations, 112–115
 - six-number summaries, 48
 - of temperature, 97
- sur package, 13–16, 25, 329–330, 632–634, 641
- symmetric index, 483
- t.test** command, 325–326, 342–344, 350, 353–356, 360–362
- table** command, 28–31, 54, 172–175, 224–226, 603–606
- target populations, 379–380
- Temp dataset, 82–85, 86–87, 89, 91–92, 97–98
- test command, 355, 487–488, 517–518, 556–557
- testing by treatment interaction, 393
- theoretical probability model, 242
- total probability, 233–235
- transformations, 108–111, 114–115, 120, 127–131, 162–163
- translation, 115–116, 128, 131, 535
- treatment effect, 384–385, 388–389, 391, 392–393, 454
- true experiment, 382–383, 393
- t*-statistic, 317–319, 336–337, 338, 486–487
- t*-test, 324–330, 335–348, 350–351, 355–362, 515–517
 - independent samples, 334–335, 340–342, 355–359, 361–362, 515–517
 - one-sample, 319–320, 325–327, 358–361
 - paired samples, 346–348, 350–351, 355–359, 362, 611
 - protected, 417
 - two-sample, 344
 - Type I error, 401, 428, 437
 - Tukey, John, 52, 416, 466
 - Tukey test, 417, 418
 - two-sample designs, 332
 - two-way analysis of variance, 437, 438, 446, 469–470
 - assumptions, 446–447
 - effect size, 457–459
 - F*-ratio, 455
 - interaction effects, 441–446, 456–457
 - partitioning sum of squares, 448
 - sequence of tests in, 465
 - Type I error, 304–305, 400–401, 407, 417–418, 427–429
 - Type II error, 304–305, 418
 - unbalanced design, 448
 - ungrouped frequency distribution, 33, 45
 - univariate distributions, 25, 30, 70, 481
 - range of, 25, 70
 - unpredictability. *See* spread
 - unrelated samples. *See* independent samples
 - UpperBodyStrength dataset, 554–556, 640
 - validity, 378–379, 391–394, 407, 599–600
 - var** command, 89, 95, 98, 379
 - variability. *See* spread
 - variables, 128–136, 147–156, 161–165, 380–384, 531–545
 - absolute value, 87, 122–123, 195
 - categorical, 465, 601
 - centered, 574–575, 578
 - combining, 134
 - combining variables, 134, 135
 - confounding, 384
 - dependent, 414–415, 422–423, 437, 441–442, 454–455
 - dichotomous, 166–168, 171–172, 176–177, 207, 242–243
 - discrete, 9, 25, 44, 242
 - first-order effect, 574, 579
 - independent, 389, 437, 442
 - instrumental, 389–390
 - interaction, 442–446, 453–457, 460–462, 465–468, 567–584
 - local, 199, 203, 271, 276
 - measurement, 3, 4, 5–6, 161
 - moderator, 567, 576
 - nominal-leveled, 71, 77, 593–594
 - ordinal level, 4–5, 7–8, 177, 593–594
 - pairs of, 534
 - ratio-leveled, 6, 7, 27–28, 30
 - recoding, 134
 - redundant, 531–532, 536
 - standardized, 120, 138, 202
 - transformation, 108, 113, 115, 131
 - transformations
 - nonlinear, 125–131
 - ranking variables, 131
 - variance, 87–92, 400–406, 413–416, 437–438, 493–497
 - between group, 403–405, 414, 449
 - estimates, 283, 338, 403, 405

variance (cont.)	Yates' correction, 600
homogeneity assumption, 338–340, 342–344, 346–348, 416	zero-order correlations, 533
residual, 414, 416, 454	z-scores, 118–125, 156–158, 255–258, 296, 532
true, 379	outlier detection, 121–123
wages dataset, 466, 472	percentile ranks and, 124–125
Whitney, Ransom, 610	score comparisons across distributions, 123–124
Wilcoxon, Frank, 610–611	transformations, 202
Wilcoxon rank sum test, 614	z-statistic, 317
Wilcoxon sign test, 610–614, 620	z-test, 296, 316–317, 327, 331
within-group variance, 403–406, 414, 454	