

INDEX

Note: Pages beginning with “A-” locate entries in the Appendix; “C-” locate entries in the CD.

- ±, 405
- Addition Rule for Disjoint Events, 238–49
 - Benford’s Law and, 239–41
 - with contingency tables, 242–43
 - General, 241–43
- alpha. *See* Type I error
- Alternative hypothesis, 454–57, 463
 - definition of, 455
 - structuring, 456
- American Academy of Dermatology, 400
- American Community Survey, 375
- American Medical Association, 400
- American Time Use Survey, 388
- Analysis of variance (ANOVA), C-19–37
 - definition of, C-19
 - null hypothesis in, C-19–20
 - one-way, C-19–37
 - conceptual understanding of, C-24–25
 - decision rule in, C-22
 - equal population variances in, C-21
 - normal distribution in, C-20–21
 - requirements of, C-20–22
 - robustness of, C-21
 - testing hypotheses with, C-22–27
 - using technology, C-23–24, C-33
 - Type I error in, C-20
- Anecdotal claims, 3
- ANOVA table, C-27
- Approach to problem solving, 6
- Area
 - under normal curve, 323–25, 345–47
 - finding, 326
 - interpreting, 325
 - as probability, 324, 325, 331, 340–41
 - as proportion, 324, 325, 331
 - standard normal curve, 332–36
 - as probability, 320–21
- Arithmetic mean. *See* Mean(s)
- Ars Conjectandi* (Bernoulli), 224, 299
- Associated variables, 179, 181
- At-least probability, 253
- Average, 106, 107. *See also* Mean(s)

- Bar graph(s), 57–59, 61, 70, 569–70
 - of conditional distribution, 569–70
 - frequency and relative frequency, 57–58
 - side-by-side, 58–59
 - technology to draw, 58, 70
- Before-after (pretest-posttest) experiments, 43
- Behrens-Fisher problem, 521
- Bell-shaped distribution, 80, 130
 - Empirical Rule to describe, 131–32
- Benford, Frank, 239, 560
- Benford’s Law, 239–41, 560
- Bernoulli, Jacob, 224, 299
- Bernoulli trial, 299
- beta. *See* Type II error
- Biased statistic, 127
- Bias in nonsampling error, 33
- Bimodal data set, 112
- Binomial experiment
 - criteria for, 298, 362
 - identifying, 298–300
- Binomial probability distribution, 298–314, 374
 - binomial probability distribution function (pdf), 302–3
 - binomial table to compute binomial probabilities, 303–5
 - constructing, 300–302
 - histograms of, 306–9, 318
 - mean and standard deviation of
 - binomial random variable, 305–6, 308–9
 - normal approximation to, 362–67
 - notation used in, 298
 - using technology, 305, 313
- Binomial probability distribution table, A-2–9
- Binomial random variable, 298
 - normal approximation to, 364–65
- Binomial tables, 303–5
- Bivariate data, 176. *See also* Relation between two variables
- Blood pressure, 5
- Box, George, 527
- Boxplots, 161–64
 - comparing two distributions using, 163–64
 - constructing, 161–62
 - distribution shape based on, 162–63
 - technology to draw, 168

- Callbacks, 34
- Cardano, Fazio, 229
- Cardano, Girolamo, 229
- Categorical variable, 6–7
- Causation, correlation vs., 184
- Cavendish, Henry, 371
- Cdf (cumulative distribution function), 305, 333
 - inverse, 349
- Cells, 243, 563
- Census, 54
 - definition of, 13
- Census Bureau, 28
- Center. *See* Mean(s)
- Central Limit Theorem, 385, 387–88, 463
- Central tendency, measures of, 107–23
 - arithmetic mean. *See* Mean(s)

- from grouped data, 142–43, 149
- median, 78, 107, 110–11, 116, 159, 162
 - computing, 110–11, 113, 123
 - definition of, 110
 - IQR and, 162
 - shape of distribution identified using, 113–16
- mode, 107, 111–13, 116
 - bimodality, 112
 - computing, 112
 - definition of, 111
 - multimodality, 112
 - of qualitative data, 112–13
 - of quantitative data, 112
- trimmed mean, 122
- Certainty, 225
- Chart(s)
 - Pareto, 58
 - pie, 60–61, 70
- Chebyshev, Pafnuty, 132, 133
- Chebyshev's Inequality, 132–34
- Chicago Tribune*, 218
- Chi-square (χ^2) distribution table, A-13
- Chi-square distribution, 551
 - characteristics of, C-8, C-14
 - critical values for, C-7–9, C-13
 - definition of, C-8
 - hypothesis testing about population standard deviation and, C-14
- Chi-square test, 180
 - for homogeneity of proportions, 570–74
 - definition of, 570
 - steps in, 571–73
 - for independence, 563–70
 - definition of, 563
 - expected counts in, 563–65
 - expected frequencies in, 566
 - steps in, 566–67
 - using technology, 570, 580
- Claim, 41
- Class(es)
 - data, 71, 73
 - width of, 73
- Classical approach to hypothesis testing, 463–64
 - in chi-square test
 - for homogeneity of proportions, 572
 - for independence, 567, 568
 - of difference between two population proportions, 537, 538
 - of difference of two means using independent samples, 522, 524
 - in goodness-of-fit test, 554–55, 556, 557–58
 - in least-squares regression model, 586, 588
 - about population mean
 - known population standard deviation, 465–69
 - unknown population standard deviation, 482, 483, 485
 - about population proportion, 494–95, 496, 497
 - about population standard deviation, C-15, C-16
 - right-tailed, large sample, 467–68
 - to testing claims about matched-pairs data, 510
 - two-tailed, small sample, 468–69
- Classical method, 227–31, 284
 - empirical method compared to, 229–31
- Closed question, 35
- Cluster sampling, 26–27, 29
- Coefficient of determination, 209–15
 - computing, 211–12
 - using technology, 212, 215
 - definition of, 209
- Column variable, 243, 563
- Combinations, 274
 - counting problems solved using, 270–73
 - definition of, 271
 - listing, 271
 - of simple random samples, 272–73
 - technology to compute, 272, 278–79
- Complement of event, 244
- Complement Rule, 244–45, 334
 - at-least probabilities and, 253
- Computational formula, 126
- Conclusions, stating, 459–60
- Conditional distribution, 569
 - bar graph of, 569–70
 - constructing, 569–70
- Conditional probability, 255–65
 - definition of, 256–57
 - independence and, 261
 - using the General Multiplication Rule, 258–61
- Confidence, level of, 5, 406, 410
 - margin of error and, 412–13
- Confidence interval(s), 403–52
 - definition of, 406, 595
 - determining appropriate procedure for constructing, 443–45
 - for difference between two population proportions, 520, 539–41
 - about the difference of two means, 526–28, 534
 - hypothesis testing about population mean using, 473–74
 - margin of error of, 406
 - for mean response, 595–96, 597
 - 95%, 407–9, 410, 438
 - about a population mean, 514
 - about a population mean: known
 - population standard deviation, 404–22
 - constructing and interpreting, 405–12, 414–15
 - margin of error in, 412–13
 - outliers and, 415, 421
 - point estimate of population mean, 404–5
 - sample size and, 410, 413–15
 - using technology, 412, 422
 - about a population mean: unknown
 - population standard deviation, 423–35
 - constructing and interpreting, 426–30
 - outliers and, 427, 428–29
 - Student's *t*-distribution, 423–25

- t -values, 425–26
 - using technology, 428, 434–35
- about the population mean difference
 - of matched-pairs data, 514–15, 520
- about a population proportion, 435–43
 - constructing and interpreting, 436–38
 - point estimate for population proportion, 435–36
 - sample size determination, 439–40
 - using technology, 437–38, 443
- about a population standard deviation, C-7–13
 - constructing and interpreting, C-10–11
 - critical values for chi-square distribution, C-7–9, C-13
 - using technology, 443, C-11, C-13
- about the population variance, C-10–11
- for slope of regression line, 589–90
 - constructing, 590
 - definition, 589–90
- about standard deviation, C-10–11
- Constants, 6
- Consumer Reports*, 48, 69, 141, 265, 314, 353, 400, 434, 492, 533, 579
- Contingency (two-way) table(s), 242–43, 563. *See also* Chi-square test
 - Addition Rule for Disjoint Events with, 242–43
 - conditional distribution, 569–70
- Continuity, correction for, 363
- Continuous data, 8
 - histograms of, 75–76
 - quantitative, mean for, 143
 - in tables, 73–75
- Continuous distributions. *See* Normal probability distribution; Uniform probability distribution
- Continuous random variable, 285–86
 - probability density functions to find probabilities for, 319–20
- Control
 - experiment vs. observational studies and, 15
- Control group, 5, 15
- Convenience sampling, 27–28
- Correction factor, finite population, 399
- Correction for continuity, 363
- Correlation, causation vs., 184
- Correlation coefficient
 - linear, 179–81
 - computing and interpreting, 182–83
 - definition of, 180
 - determination of linear relation and, 185
 - properties of, 179–81
 - technology to determine, 194–95
 - table critical values for, A-14
- Correlation matrix, 183
- Counting problems, 265–79
 - combinations for, 270–73, 274–75, 278–79
 - Multiplication Rule for, 265–68
 - permutations for, 268–70, 273–75, 278–79
 - without repetition, 267
- Counts, expected, 552–53, 563–65
- Critical region (rejection region), 466
- Critical value(s), 409, 466
 - for chi-square distribution, C-7–9, C-13
 - for F -distribution, C-34–37
- Cryptosporidium, 214
- Cumulative distribution function (cdf), 305, 333
 - inverse, 349
- Current Population Survey, 28
- Current Population Survey: Design and Methodology, The*, 28
- Cutoff point, 225, 226
- Darwin, Charles, 199
- Data, 3–4. *See also* Qualitative data; Quantitative data
 - bivariate, 176. *See also* Relation between two variables
 - classes of, 71, 73
 - collection of, 13–14. *See also* Experiment(s); Observational studies; Sampling
 - continuous, 8
 - histograms of, 75–76
 - quantitative, mean for, 143
 - in tables, 73–75
 - discrete, 8, 72–73
 - grouped, 142–49
 - mean from, 142–43, 149
 - standard deviation from, 144–46, 149
 - variance from, 144–46
 - polling, 14
 - raw, 54, 73
 - right skewed, 433
 - time-series, 81
 - univariate, 176
 - variability in, 3–4
 - variables vs., 8–9
- Data organization and summary, 54–105.
 - See also* Numerically summarizing data
 - graphical misrepresentations, 93–98
 - by manipulating vertical scale, 94
 - qualitative data, 8, 55–70
 - bar graphs, 57–59, 61, 70
 - pie charts, 60–61, 70
 - tables, 55–56
 - quantitative data, 8, 71–93
 - dot plots, 80
 - histograms, 72–73, 75–76, 92–93
 - shape of distribution, 80–81
 - split stems, 79
 - stem-and-leaf plots, 76–79, 92–93
 - tables, 71–72, 73–75
 - time-series graphs, 81–82
- Data sets, 54
 - comparing, 59, 123–24
 - small, 354
- Decision rule, 466, 470
- Degrees of freedom, 128, 426
- de Moivre, Abraham, 322, 325, 362
- Density function(s)
 - exponential, 384

- probability, 319–20
 - normal, 347
 - uniform, 321
- Dependent events, 250
- Dependent (response) variable, 14, 40, 41, 177–78
- Dependent sampling, 508–10
 - about two means. *See* Matched-pairs (dependent) design
- Depreciation rate, 204
- Depression, self-treatment for, 48
- Descriptive statistics, 4
- Designed experiment. *See also* Experiment(s): design of
 - defined, 14, 40
 - observational study vs., 14
- Determination, coefficient of, 209–15
 - computing, 211–12
 - using technology, 212, 215
 - definition of, 209
- Deviation(s), 209–10
 - explained, 210
 - about the mean, 125
 - total, 210
 - unexplained, 210–11
- Dice problem, 231
- Die
 - fair, 224
 - loaded, 224
- Dietary supplements, 48
- Diophantus, 230
- Discrete data, 8
 - histograms of, 72–73
- Discrete probability distributions, 284–317
 - binomial, 298–314, 374
 - binomial probability distribution function (pdf), 302–3
 - binomial table to compute binomial probabilities, 303–5
 - constructing, 300–302
 - histograms of, 306–9, 318
 - identifying binomial experiment, 298–300
 - mean and standard deviation of
 - binomial random variable, 305–6, 308–9
 - normal approximation to, 362–67
 - notation used in, 298
 - using technology, 305, 313
 - definition of, 286
 - identifying, 286–87
 - probability histograms of, 287–88
 - rules for, 287
 - binomial probability distribution function (pdf), 302–3
 - binomial table to compute binomial probabilities, 303–5
 - constructing, 300–302
 - histograms of, 306–9, 318
 - identifying binomial experiment, 298–300
 - mean and standard deviation of
 - binomial random variable, 305–6, 308–9
 - normal approximation to, 362–67
 - notation used in, 298
 - using technology, 305, 313
 - definition of, 286
 - identifying, 286–87
 - probability histograms of, 287–88
 - rules for, 287
- Discrete random variable, 285–97
 - continuous random variables
 - distinguished from, 285–86
 - mean of, 288–92, 293, 297
 - computing, 289
 - defined, 289
 - as an expected value, 291–92
 - interpreting, 289–90
 - using technology, 293, 297–98
 - variance and standard deviation of, 292–93
- Disjoint events, 238–41
 - independent events vs., 250–51
- Dispersion, measures of, 123–49, 159
 - Chebyshev’s Inequality, 132–34
 - Empirical Rule, 131–32
 - from grouped data, 142–49
 - range, 124–25
 - computing, 125
 - definition of, 124
 - interquartile (IQR), 159, 162
 - technology to determine, 142
 - standard deviation, 125, 129–31, 142, 180
 - of binomial random variable, 305–6, 308–9
 - confidence interval about, C-10–11
 - of discrete random variables, 292–93
 - from grouped data, 144–46, 149
 - interpretations of, 130
 - outlier distortion of, 155
 - population, 129–30, 406, 582–83
 - sample, 129–30, 583
 - of sampling distribution of sample mean, 382
 - technology to approximate, 146
 - technology to determine, 142
 - of two data sets, 130–31
 - variance, 125–29, 130, 142
 - of discrete random variables, 292–93
 - from grouped data, 144–46
 - population, 125–28, 144
 - sample, 125, 127–29, 144
 - technology to determine, 130, 142
- Distribution. *See* Discrete probability distributions; Frequency distribution(s); Normal probability distribution
- Distribution function, cumulative (cdf), 305, 333
 - inverse, 349
- Doctrine of Chance, The* (Leibniz), 322
- Dot plots, 80
- Double-blind experiment, 5, 40
- Dow Jones Industrial Average (DJIA), 94–95
- EDA (exploratory data analysis), 159
- Eléments de Géométrie* (Legendre), 198
- Elements of Sampling Theory and Methods (Govindarajulu), 28
- Empirical Method, 226–27, 229–31
 - classical method compared to, 229–31
- Empirical Rule, 131–32, 332, 343
 - unusual results in binomial experiment and, 308–9
- Equally likely outcomes, 227, 229
- Error(s)
 - input, 35
 - margin of, 405, 438
 - of confidence interval, 406
 - definition of, 412
 - level of confidence and, 412–13
 - sample size and, 413–15, 439–40
 - mean square due to (MSE), C-25

- residual, 197–98
- rounding, 129, 183
- sampling, 33–39
 - data checks to prevent, 35
 - frame and, 34
 - interviewer error, 34
 - misrepresented answers, 35
 - nonresponse and, 34
 - nonsampling errors, 33
 - order of questions, words, and responses, 36
 - questionnaire design and, 35
 - wording of questions, 35–36
- standard, 423, 583–85
 - computing, 583–84
 - definition of, 583
 - of the mean, 382
 - for sampling distribution of population proportion, 494
- Type I, 457–59, 464, 526
 - in ANOVA, C-20
 - probability of, 458–59, 464
- Type II, 457–59
 - probability of, 458–59
- Estimator, biased, 127
- Event(s), 224
 - certain, 225
 - complement of, 244
 - dependent, 250
 - disjoint, 238–41, 250–51
 - impossible, 225
 - independent, 249–51, 261
 - Multiplication Rule for, 251–52, 261
 - not so unusual, 225–26
 - simple, 224
 - unusual, 225–26
- Excel, 18
 - area under the normal curve using, 354
 - scores corresponding, 354
 - bar graph using, 58, 70
 - binomial probabilities using, 305
 - boxplots using, 168
 - chi-square tests using, 580
 - coefficient of determination using, 212, 215
 - combinations using, 272, 279
 - confidence intervals using, 599
 - for population mean where
 - population standard deviation is known, 422
 - for population mean where
 - population standard deviation is unknown, 435
 - for population proportion, 443
 - for population standard deviation, 443, C-13
 - correlation coefficient using, 195
 - difference of two population proportions using, 545
 - factorials on, 279
 - hypothesis testing using
 - about population mean assuming
 - known population standard deviation, 480
 - about population mean with
 - unknown population standard deviation, 493
 - about population proportion, 500
 - least-squares regression line using, 202, 208
 - least-squares regression model using, 594
 - linear correlation coefficient using, 183
 - mean and median using, 123
 - normal probability plot using, 361
 - one-way ANOVA using, C-23–24, C-33
 - permutations using, 269–70, 279
 - pie charts using, 70
 - prediction intervals using, 599
 - scatter diagrams using, 183, 195
 - simulation using, 237
 - standard error using, 584–85
 - standard normal distribution using, 344
 - time-series graphs on, 82
 - two-sample *t*-tests, dependent sampling using, 520
 - two-sample *t*-tests, independent sampling using, 534
 - z*-score from a specified area to the left using, 337–38
- Expected counts, 552–53, 563–65
 - in chi-square test, 563–65
 - in goodness-of-fit test, 552–53
- Expected value, 291–92
- Experiment(s), 54, 224, 227, 284
 - design of, 40–47
 - completely randomized design, 42–43
 - matched-pairs design, 43–44
 - simple random sampling and, 44
 - steps in, 41
 - double-blind, 5, 40
- Experimental group, 5
- Experimental units (subjects), 14, 40, 41
 - in matched-pairs design, 43
- Explained deviation, 210
- Explanatory (predictor or independent) variable, 40, 177–78
- Exploratory data analysis (EDA), 159
 - Exploratory Data Analysis* (Tukey), 159
- Exponential density function, 384
- Exponential probability distribution, 420–21
- Ex post facto* studies. *See* Observational studies
- Factor(s), 40, 41
- Factorial notation, 269
- Factorials, technology to compute, 278–79
- Factorial symbol, 268
- Fair die, 224
- F*-distribution, critical values for, C-34–37
- Federal Trade Commission, 400
- Fermat, Pierre de, 230, 231, 292
- Fermat's Last Theorem, 230
- Fibonacci sequence, 561–62
- Finite population correction factor, 399
- Fisher, Sir Ronald A., 40, 180, 553, C-13, C-20
- Fisher's *F*-distribution, C-34–37
- Five-number summary, 159–60

- Frame, 16, 25, 34
- Frequency, relative, 56
- Frequency distribution(s). *See also*
 Relative frequency distribution
 based on boxplots, 162–63
 bell shaped, 80, 130
 Empirical Rule to describe, 131–32
 center of (average), 106. *See also* Mean(s)
 characteristics of, 106
 chi-square, 551
 characteristics of, C-8, C-14
 critical values for, C-7–9, C-13
 definition of, C-8
 hypothesis testing about population
 standard deviation and, C-14
 comparing, 163–64
 from continuous data, 73–74
 critical value for, 409
 of discrete data, 71–72
 frequency, 55–56
 mean of variable from, 142–43
 of qualitative data, 55–56
 relative, 56
 shape of, 80–81, 106
 spread of, 106
 symmetric, 130
 variance and standard deviation from, 145–46
- F*-test, 527–28
 computing, C-25
 idea behind, C-24
- Gallup Organization, 15, 36
- Galton, Sir Francis, 199
- Gauss, Carl, 325
- Gaussian distribution. *See* Normal probability distribution
- General Addition Rule, 241–43
- General Multiplication Rule, conditional probability using,
 258–61
- Golden ratio, 561
- Goodness-of-fit test, 551–62
 characteristics of chi-square distribution, 551
 definition of, 552
 expected counts, 552–53
 technology for, 558
 testing hypothesis using, 555–59
 test statistic for, 554
- Gosset, William Sealey, 423
- Graph(s), 106
 characteristics of good, 95
 dot plots, 80
 histograms
 binomial probability, 306–9, 318
 of continuous data, 75–76
 of discrete data, 72–73
 probability, 287–88
 technology to draw, 92–93
 of lines, C-2–3
 misleading or deceptive, 93–95
 polar area diagram, 74
 stem-and-leaf plots, 76–79
 constructing, 76–79
 split stems, 79
 using technology, 92–93
- tables, 106
 binomial, 303–5
 continuous data in, 73–75
 discrete data in, 71–72
 open-ended, 73
 qualitative data in, 55–56
 time-series, 81–82
- Greek letters, use of, 108
- Grouped data, 142–49
 mean from, 142–43, 149
 standard deviation from, 144–46, 149
 variance from, 144–46
- Herbs, 48
- Histogram(s)
 binomial probability, 306–9, 318
 of continuous data, 75–76
 of discrete data, 72–73
 of discrete probability distributions, 287–88
 technology to draw, 92–93
- Homogeneity of proportions, chi-square test for, 570–74
 definition of, 570
 steps in, 571–73
- Huygens, Christiaan, 292
- Hypothesis/hypotheses
 alternative, 454–57, 463
 definition of, 455
 structuring, 456
 definition of, 455
 forming, 456–57
 null, 454–57, 463, 552, 555
 acceptance of, 459
 in ANOVA, C-19–20
 assumption of trueness of, 464
 definition of, 455
 structuring, 456
- Hypothesis testing, 453–506. *See also*
 Inferences; One-tailed test(s);
 Two-tailed tests
 choosing method for, 500–502
 definition of, 455
 illustration of, 454
 large samples, 467–68, 470–71, 475
 logic of, 462–65
 classical approach, 463–64
P-value approach, 464–65
 outcomes from, 457
 for a population mean assuming
 known population standard deviation, 462–80
 classical approach to, 465–69
P-values approach to, 469–73
 technology in, 473, 479–80

- using confidence intervals, 473–74
- for a population mean in with
 - unknown population standard deviation, 480–93
 - classical approach to, 482, 483, 485
 - outliers and, 482, 485
 - P*-value approach, 482, 483, 485
 - technology in, 484, 493
 - using large sample, 482–83
 - using small sample, 484–86
- for a population proportion, 493–500
 - classical approach to, 494–95, 496, 497
 - P*-value approach to, 494–95, 496, 497
 - right-tailed test, 495–96
 - technology in, 497–98, 500
 - two-tailed test, 496–97
- for population standard deviation, C-13–19
 - chi-square distribution and, C-14
 - classical approach to, C-15, C-16
 - left-tailed test, C-15–16
 - P*-value approach to, C-15, C-16
 - probability of Type I error, 458–59, 464
 - probability of Type II error, 458–59
 - small samples, 468–69, 471–72, 484–86
 - stating conclusions, 459–60
 - steps in, 454
 - Type I and Type II errors, 457–59
- Impossible event, 225
- Incentives, nonresponse and, 34
- Independence
 - chi-square test for, 563–70
 - definition of, 563
 - expected counts in, 563–65
 - expected frequencies in, 566
 - steps in, 566–67
 - conditional probability and, 261
- Independent events, 249–51, 261
 - disjoint events vs., 250–51
 - Multiplication Rule for, 251–52, 261
- Independent (explanatory or predictor) variable, 40, 177–78
- Independent samples, 508–10
 - difference between two means from, 522–26
 - confidence intervals regarding, 526–28, 534
 - technology for, 525, 534
 - testing claims regarding, 522–26
- Independent trials, 299
- Individual, population vs., 4
- Inferences, 375, 507–604. *See also*
 - Least-squares regression model
 - reliability of, 375
 - about two means
 - dependent samples, 508–20
 - independent samples, 522–26
 - about two population proportions, 534–45
 - confidence intervals, 520, 539
 - sample size requirements for, 541–42
 - testing, 535–39
 - using technology, 539, 545
- Inferential statistics, 5, 41, 404, 453
- Inflection points, 321–22
- Input errors, 35
- Integer distribution, 231
- Intercept, 198, 201–2
- Internet surveys, 27–28
- Interquartile range (IQR), 159, 162
- Interviewer error, 34
- Inverse cumulative distribution function, 349
- Journal of American Medical Association*, 13–14
- Kolmogorov, Andrei Nikolaevich, 260
- Landon, Alfred M., 33
- Laplace, Pierre Simon, 387
- Law of Large Numbers, 223–24, 284, 381
- Least-squares regression line, 195–208
 - coefficient of determination, 209–15
 - definition of, 198
 - equation of, 198–99
 - finding, 196–202, 208
 - using technology, 202, 208
 - interpreting the slope and *y*-intercept of, 201–2
 - sum of squared residuals, 202–3
- Least-squares regression model, 580–94
 - confidence interval about slope of, 589–90
 - constructing, 590
 - definition of, 589–90
 - confidence intervals for mean response, 595–96, 599
 - definition of, 583
 - example of, 581
 - inference on the slope and intercept, 585–89
 - using technology, 594
 - normally distributed residuals in, 585
 - prediction intervals for an individual response, 596–98, 599
 - requirements of, 581–83
 - robustness of, 587
 - sampling distributions in, 581–82
 - significance of, 580–94
 - standard error of the estimate, 583–85
- Left-tailed hypothesis testing. *See* One-tailed test(s)
- Legendre, Adrien Marie, 198
- Leibniz, Gottfried Wilhelm, 322
- Level of confidence, 5, 406, 410
 - margin of error and, 412–13
- Life on the Mississippi* (Twain), 208
- Linear correlation coefficient, 179–84
 - computing and interpreting, 182–83
 - definition of, 180
 - determination of linear relation and, 185
 - properties of, 179–81
 - using technology, 194–95
- Linear equation, C-1, C-4–5

- point-slope form of, C-3–4
- slope and y -intercept of line from, C-5
- slope-intercept form of, C-5
- Linear relation, 185
- Linear transformations, 122
- Lines, C-1–6
 - graphing, C-2–3
 - horizontal, C-3–4
 - slope of, C-1–2
 - vertical, C-1
- Literary Digest*, 33
- Loaded die, 224
- Lower class limit, 73
- Lurking variable, 3, 15, 177, 184

- Margin of error, 405, 438
 - confidence interval and, 406
 - definition of, 412
 - level of confidence and, 412–13
 - sample size and, 413–15, 439–40
- Matched-pairs (dependent) design, 43–44, 508
 - confidence intervals about population mean difference of, 514–15, 520
 - testing claims about, 509–13
 - using technology, 513, 520
- Mathematics, statistics vs., 4
- Matrix, correlation, 183
- Mean(s), 107–10, 116. *See also* Population mean;
 - Sample mean of binomial random variable, 305–6, 308–9
- as center of gravity, 109–10
- comparing three or more. *See* Analysis of variance (ANOVA)
- computing, 108–9
 - using technology, 113, 123
- definition of, 107, 108
- deviation about the, 125
- of discrete random variable, 288–92, 293, 297
 - computing, 289
 - defined, 289
 - as an expected value, 291–92
 - interpreting, 289–90
 - using technology, 293, 297–98
- from grouped data, 142–43, 149
- least-squares regression model and, 582
- outlier distortion of, 155
- of sampling distribution of sample mean, 382
- shape of distribution from, 113–16
- standard error of the, 382
 - technology to approximate, 146
 - trimmed, 122
 - weighted, 144
- Mean response, confidence intervals for, 595–96
- Mean squares, C-25
 - due to error (MSE), C-25
 - due to treatment (MST), C-25
- Mécanique céleste* (Laplace), 387
- Median, 78, 107, 110–11, 116, 159, 162
 - computing, 110–11
 - with even number of observations, 111
 - with odd number of observations, 110–11
 - using technology, 113, 123
 - definition of, 110
 - IQR and, 162
 - shape of distribution from, 113–16
- Method of least squares. *See* Least-squares regression line
- Midrange, 123
- MINITAB, 18
 - area under standard normal curve using, 333
 - area under the normal curve using, 354
 - scores corresponding, 354
 - bar graph using, 70
 - binomial probabilities using, 313
 - boxplots using, 168
 - chi-square tests using, 570, 580
 - coefficient of determination using, 215
 - confidence intervals using, 598, 599
 - for population mean where
 - population standard deviation is known, 412, 422
 - for population mean where
 - population standard deviation is unknown, 435
 - for population proportion, 438, 443
 - for population standard deviation, 443, C-13
 - correlation coefficient using, 195
 - difference of two population proportions using, 539, 545
 - histogram for continuous data using, 76
 - hypothesis testing using
 - about population mean assuming
 - known population standard deviation, 473, 480
 - about population mean with
 - unknown population standard deviation, 493
 - about population proportion, 497–98, 500
 - least-squares regression line using, 202, 208
 - least-squares regression model using, 594
 - mean and median using, 123
 - normality assessed using, 357–58
 - normal probability plot using, 361
 - normal random variable using, 348–49
 - number summary using, 160
 - one-way ANOVA using, C-23–24, C-33
 - pie charts using, 70
 - prediction intervals using, 598, 599
 - scatter diagrams using, 195
 - simulation using, 237
 - standard normal distribution using, 344
 - stem-and-leaf plot using, 78
 - testing claims about matched-pairs data using, 513
 - time-series graphs on, 82
 - two-sample t -tests using
 - dependent sampling, 520
 - independent sampling, 534
- Mode, 107, 111–13, 116
 - bimodality, 112
 - computing, 112
 - definition of, 111
 - multimodality, 112

- of qualitative data, 112–13
- of quantitative data, 112
- Models, 324
- Mood modifiers, herbal, 48
- Multimodal data set, 112
- Multimodal instruction, 531
- Multiple linear regression model
 - correlation matrix, 183
- Multiplication Rule, 563
 - for counting, 265–68
 - for Independent Events, 251–52, 261
- Multistage sampling, 28
- Mutually exclusive (disjoint) events, 238–41
 - independent events vs., 250–51

- Necomb, Simon, 239
- Negatively associated variables, 179, 181
- Newcomb, Simon, 560
- Newton, Isaac, 322
- Neyman, Jerzy, 463
- Nielsen Media Research, 28
- Nightingale, Florence, 74
- 95% confidence interval, 407–9, 410, 438
- Nominal variable, 13
- Nonparametric procedures, 429
- Nonresponse, 34
- Nonsampling errors, 33
- Normal curve
 - area under, 323–25, 345–47
 - finding, 326
 - interpreting, 325
 - as probability, 324, 325, 331, 340–41
 - as proportion, 324, 325, 331
 - standard normal curve, 332–36
 - inflection points on, 321–22
- Normal populations, sampling distribution of sample mean
 - from, 377–83
- Normal probability density function, 324, 347
- Normal probability distribution, 318–71, 374
 - applications of, 345–54
 - area under a normal curve, 345–47
 - value of normal random variable, 348–49
 - area under, 323–25
 - finding, 326
 - interpreting, 325
 - as proportion or probability, 324, 325, 331
 - assessing normality, 354–61
 - normal probability plots for, 354–61
 - confidence interval construction and, 415
 - graph of, 321–22
 - least-squares regression model and, 585
 - normal approximation to the binomial probability
 - distribution, 362–67
 - in one-way ANOVA, C-20–21
 - properties of, 319–31
 - statement of, 322–23
 - relation between normal random
 - variable and standard normal random variable, 325–27
 - standard, 326, 331–44
 - area under the standard normal curve, 332–36, 340–41
 - notation for probability of standard normal random
 - variable, 340
 - properties of, 331–32
 - t*-distribution compared to, 424
 - technology to find, 344
 - Z-scores for given area, 336–40
 - technology to find, 353–54
 - uniform probability distribution, 318, 319–21
 - definition of, 319
- Normal probability plot, 354–61
 - drawing, 355–57
 - linearity of, 355–57
 - technology to assess, 357–58, 361
- Normal random variable, 323–27
 - standard, 325–27, 340–41
 - standardizing, 326–27
 - value of, 348–49
- Normal score, 355
- Not so unusual events, 225–26
- Nouvelles méthodes pour la détermination des orbites des comètes* (Legendre), 198
- Null hypothesis, 454–57, 463, 552, 555
 - acceptance of, 459
 - in ANOVA, C-19–20
 - assumption of trueness of, 464
 - definition of, 455
 - structuring, 456
- Numerically summarizing data, 106–75
 - boxplots, 161–64
 - comparing two distributions using, 163–64
 - constructing, 161–62
 - distribution shape based on, 162–63
 - technology to draw, 168
 - five-number summary, 159–60
 - measures of central tendency, 107–23
 - arithmetic mean. *See* Mean(s)
 - from grouped data, 142–43, 149
 - median, 78, 107, 110–11, 113–16, 123, 159, 162
 - midrange, 123
 - mode, 107, 111–13, 116
 - shape of distribution from mean and median, 113–16
 - trimmed mean, 122
 - measures of dispersion, 123–49, 159
 - Chebyshev's Inequality, 132–34
 - Empirical Rule, 131–32
 - from grouped data, 142–49
 - range, 124–25, 142, 159, 162
 - standard deviation. *See* Standard deviation variance, 125–29, 130, 142, 144–46, 292–93
 - measures of position, 149–59
 - outliers. *See* Outliers
 - percentiles, 151–53, 346–47
 - quartiles, 153–54, 155
 - z-scores, 149–50, 326, 336–40, 355, 425
- Observational studies, 13–16

- defined, 14
- experiment vs., 14–15
- One-tailed test(s), 456, 466, 467–68, 470–71, 481–84, 495–96
 - of difference between two means, 509, 510, 511
 - of difference between two means: independent samples, 522
 - of difference between two population proportions, 536, 537
 - in least-squares regression model, 586–87
 - about population proportion, 495–96
 - about population standard deviation, C-15–16
- One-way ANOVA, C-19–37
 - conceptual understanding of, C-24–25
 - decision rule in, C-22
 - equal population variances in, C-21
 - normal distribution in, C-20–21
 - requirements of, C-20–22
 - robustness of, C-21
 - testing hypotheses with, C-22–27
 - using technology, C-23–24, C-33
- Open-ended tables, 73
- Open question, 35
- Ordinal variable, 13
- Outcomes, 223
 - equally likely, 227, 229
- Outliers, 106, 155–56
 - confidence intervals and
 - for population mean where population standard deviation is known, 415, 421
 - for population mean where population standard deviation is unknown, 427, 428–29
 - hypothesis testing about population mean and, 482, 485
 - quartiles to check, 155
- Parameters, 107, 108
- Pareto chart, 58
- Pascal, Blaise, 230, 231, 292
- Pearson, Egon, 463
- Pearson, Karl, 180, 321, 423, 463, 553
- Pearson product moment correlation coefficient.
 - See* Linear correlation coefficient
- People Meter, 28
- Percentile(s), 151–53
 - with index as integer, 151–52
 - with index not an integer, 152–53
 - k*th, 151–52
 - quartiles, 153–54
 - ranks by, 346–47
 - of specific data value, 153
- Permutations, 268–70, 274–75
 - computing, 269–70
 - using technology, 269–70, 278–79
 - definition of, 268–69
 - of distinct items, 274
 - with nondistinct items, 273–75
- Phone-in polling, 27–28
- Pie charts, 60–61
 - constructing, 60–61
 - technology to draw, 70
- Placebo, 5, 40
- Point estimate
 - definition of, 404
 - of population mean, 404–5
 - of population proportion, 435–36
 - of two population means, 514
 - of two population proportions, 536
- Points, problem of, 231
- Point-slope form of line, 197, C-3–4
- Polar area diagram, 74
- Polling, phone-in, 27–28
- Polling data, 14
- Pooled estimate of *p*, 536
- Pooled *t*-statistic, 527–28
- Pooling, 527
- Population, 4
 - with distribution unknown, 354
 - mean of (μ), 107–9
 - proportion of, 324
 - underrepresentation of, 34
- Population mean, 107–9, 142
 - confidence interval about, 514
 - confidence interval about, where the population standard deviation is known, 404–22
 - constructing and interpreting, 405–12, 414–15
 - margin of error in, 412–13
 - outliers and, 415, 421
 - point estimate of population mean, 404–5
 - sample size and, 410, 413–15
 - technology for, 412, 422
 - confidence interval about, where the population standard deviation is unknown, 423–35
 - constructing and interpreting, 426–30
 - outliers and, 427, 428–29
 - Student's *t*-distribution, 423–25
 - technology for, 428, 434–35
 - t*-values, 425–26
 - hypothesis testing about, 480–93
 - known population standard deviation, 462–80
 - unknown population standard deviation, 480–93
 - point estimate of, 404–5
- Population proportion, 324
 - confidence intervals about, 435–43
 - constructing and interpreting, 436–38
 - point estimate for population proportion, 435–36
 - sample size determination, 439–40
 - technology for, 437–38, 443
 - difference between two, 534–45
 - confidence intervals, 520, 539–41
 - sample size requirements for, 541–42
 - testing, 535–39
 - using technology, 539, 545
 - hypothesis testing about, 493–500
 - classical approach to, 494–95, 496, 497
 - P*-value approach to, 494–95, 496, 497
 - right-tailed test, 495–96

- technology in, 497–98, 500
- two-tailed test, 496–97
- point estimate for, 435–36
- sampling distribution of, 436
- standard error for, 494
- Population size, sample size and, 395
- Population standard deviation, 129–30
 - confidence intervals about, C-7–13
 - constructing and interpreting, C-10–11
 - critical values for chi-square distribution, C-7–9, C-13
 - using technology, 443, C-11, C-13
 - hypothesis testing for, C-13–19
 - chi-square distribution and, C-13
 - classical approach to, C-15, C-16
 - left-tailed test, C-15–16
 - P*-value approach to, C-15, C-16
 - least-squares regression model and, 582–83
 - margin of error of confidence interval and, 406
- Population variance, 125–27, 144
 - confidence intervals about, C-10–11
 - hypothesis testing about, C-14–15
 - in one-way ANOVA, C-21
- Population *z*-score, 150
- Position, measures of, 149–59
 - outliers, 106, 155–56
 - percentiles, 151–53
 - quartiles, 153–54
 - z*-scores, 149–50
- Positively associated variables, 179, 181
- Practical significance, 538
 - definition of, 474
 - statistical significance vs., 474–75
- Prediction intervals
 - definition of, 595
 - for an individual response, 596–98, 599
- Predictor (independent or explanatory) variable, 40, 177–78
- Pretest-posttest (before-after) experiments, 43
- Probability(ies), 221–83
 - Addition Rule for Disjoint Events, 238–49
 - Benford's Law and, 239–41
 - with contingency tables, 242–43
 - General, 241–43
 - area as, 320–21
 - area under normal curve as, 324, 325, 331, 340–41
 - at-least, 253
 - classical, 227–31, 284
 - Complement Rule, 244–45, 253, 334
 - conditional, 255–65
 - definition of, 256–57
 - independence and, 261
 - using the General Multiplication Rule, 258–61
 - counting problems, 265–79
 - combinations for, 270, 274–75, 278–79
 - Multiplication Rule for, 265–68
 - permutations for, 268–70, 273–75, 278–79
 - without repetition, 267
 - defined, 221, 223
 - Empirical Method to approximate, 226–27, 229–31
 - events and the sample space of probability experiment, 224
 - Multiplication Rule for Independent Events, 251–52, 261
 - relative frequency to approximate, 226
 - rules of, 223–37, 253
 - understanding, 225–26
 - simulation to obtain, 231–32
 - technology in, 237
 - subjective, 232
 - value of normal random variable corresponding to, 348–49
- Probability density function (pdf), 319–20, 324
 - normal, 324
- Probability distribution, 374. *See also* Normal probability distribution
 - binomial, 374
 - exponential, 420–21
 - testing claims regarding. *See* Contingency (two-way) table(s); Goodness-of-fit test
- Probability experiment, 54, 224, 227, 284
 - design of, 40–47
 - completely randomized design, 42–43
 - matched-pairs design, 43–44
 - simple random sampling and, 44
 - steps in, 41
 - double-blind, 5, 40
- Probability histograms of discrete probability distributions, 287–88
 - binomial, 306–9
- Probability model, 225, 284
 - for random variables. *See* Discrete probability distributions
 - from survey data, 227
- Proportion(s). *See also* Population proportion; Sample proportion
 - area under normal curve as, 324, 325, 331
 - homogeneity of, 570–74
 - definition of, 570
 - steps in, 571–73
 - value of normal random variable corresponding to, 348–49
- P*-value, definition of, 469
- P*-value approach to hypothesis testing, 464–65
 - in chi-square test
 - for homogeneity of proportions, 572
 - for independence, 567, 568
 - of difference between two means: independent samples, 522, 524
 - of difference between two population proportions, 537, 538
 - goodness-of-fit, 554–55, 556, 557–58
 - in least-squares regression model, 586, 588
 - of matched-pairs data, 510
 - in one-way ANOVA, C-22
 - about population mean
 - known population standard deviation, 469–73
 - unknown population standard deviation, 482, 483, 485

- about population proportion, 494–95, 496, 497
 - about population standard deviation, C-15, C-16
 - right-tailed, large sample, 470–71
 - two-tailed, 471–72
- Qualitative data, 8, 55–70
- bar graphs of, 57–59, 61, 70
 - frequency distribution of, 55–56
 - relative, 56
 - mode of, 112–13
 - pie charts of, 60–61, 70
 - tables of, 55–56
- Qualitative variable, 6–7
- Quantitative data, 8, 71–93
- dot plots of, 80
 - histograms of, 72–73, 75–76, 92–93
 - mode of, 112
 - shape of distribution of, 80–81
 - stem-and-leaf plots of, 76–79, 92–93
 - split stems, 79
 - tables of, 71–72, 73–75
- Quantitative variable, 6–7
- Quartiles, 153–54
- checking for outliers using, 155
 - using technology, 154
- Questionnaire design, 35
- Questions
- order of, 36
 - wording of, 35–36
- Questions and Answers in Attitude Surveys*
(Schuman and Presser), 36
- Queuing theory, C-18
- Randomization, 41
- Randomized block design, 44
- Random number generator, 18, 81
- Random number table, A-1
- Random sampling, 16–23, 25
- combinations of samples, 272–73
 - confidence interval construction and, 414–15
 - definition of, 16
 - illustrating, 16
 - obtaining sample, 16–23
- Random variable(s), 374
- binomial, 298
 - normal approximation to, 364–65
 - continuous, 285–86
 - probability density functions to find probabilities for, 319–20
 - definition of, 285
 - discrete, 285–97
 - continuous random variables distinguished from, 285–86
 - mean of, 288–92, 293, 297
 - variance and standard deviation of, 292–93
 - normal, 323–27
 - standard, 325–27, 340–41
 - standardizing, 326–27
 - value of, 348–49
- probability models for. *See* Discrete probability distributions
- statistics as, 374
- Range, 124–25, 142
- computing, 125
 - definition of, 124–25
 - interquartile (IQR), 159, 162
 - technology to determine, 142
- Ratio, golden, 561
- Raw data, 54
- continuous, 73
- Regression. *See* Least-squares regression model
- Regression analysis, 199
- Rejection region (critical region), 466
- Relation between two variables, 176–219
- correlation versus causation, 184
 - least-squares regression line, 195–208
 - coefficient of determination, 209–15
 - definition of, 198
 - equation of, 198–99
 - finding, 196–202, 208
 - interpreting the slope and y-intercept of, 201–2
 - sum of squared residuals, 202–3
 - linear correlation coefficient, 179–84
 - computing and interpreting, 182–83
 - definition of, 180
 - properties of, 179–80
 - using technology, 194–95
 - scatter diagrams, 177–79
 - definition of, 177
 - drawing, 177–78, 183, 194–95
- Relative frequency(ies)
- probability using, 226
 - of qualitative data, 56
- Relative frequency distribution, 56
- from continuous data, 73–75
 - of discrete data, 71–72
- Reliability of inferences, 375
- Replication, 41
- Research objective, identification of, 4
- Residual(s), 197–98
- normally distributed, 585
 - sum of squared, 202–3
- Response (dependent) variable, 14, 40, 41, 177–78
- Rewards, nonresponse and, 34
- Right skewed data, 433
- Right-tailed hypothesis testing. *See* One-tailed test(s)
- Rise, C-1
- Robustness, 410, 467, 510
- of least-squares regression model, 587
 - of one-way ANOVA, C-21
- Roman letters, use of, 108
- Roosevelt, Franklin D., 33
- Rounding, 8
- Rounding error, 129, 183
- Rounding up vs. rounding off, 414

- Row variable, 242–43, 563
- Run, C-1
- Sample(s)
- correlation coefficient, 180
 - defined, 4
 - matched-pairs (dependent), 508
 - confidence intervals for, 514–15, 520
 - testing claims about, 509–13
 - mean of (\bar{x}) 107–9
 - self-selected, 27
- Sample mean, 107–9, 142
- sample size and, 381
 - sampling distribution of, 375–92, 403, 407
 - definition of, 375–76
 - describing, 383
 - mean and standard deviation of, 382
 - from normal populations, 377–83
 - from not-normal populations, 384–88
 - shape of, 383
- Sample proportion
- computing, 393
 - definition of, 392
 - sampling distribution of, 392–400
 - describing, 392–95
 - probabilities of, 396–97
 - simulation to describe distribution of, 393–95
- Sample size, 28
- confidence interval about population mean
 - where population standard deviation is known and, 410
 - confidence interval about population proportion and, 439–40
 - for difference of two population proportions, 541–42
 - distribution shape and, 385–87
 - hypothesis testing and, 467–72, 475, 482–86
 - margin of error and, 406, 413–15, 439–40
 - population size and, 395
 - sampling variability and, 381–82
- Sample space, 224
- Sample standard deviation, 129–30, 583
- Sample variance, 125, 127–29, 144
- Sample z -score, 150
- Sampling, 16–39
- cluster, 26–27, 29
 - convenience, 27–28
 - dependent, 508–10
 - errors in, 33–39
 - data checks to prevent, 35
 - frame and, 34
 - interviewer error, 34
 - misrepresented answers, 35
 - nonresponse and, 34
 - nonsampling errors, 33
 - order of questions, words, and responses, 36
 - questionnaire design and, 35
 - wording of questions, 35–36
 - goal in, 16
 - independent, 508–10
 - multistage, 28
 - without replacement, 17
 - sample size considerations, 28
 - simple random, 16–23, 25
 - combinations of, 272–73
 - confidence interval construction and, 414–15
 - definition of, 16
 - illustrating, 16
 - obtaining sample, 16–23
 - stratified, 23–24, 25, 29
 - survey, 14
 - systematic, 25–26, 27, 29
- Sampling distribution(s), 373–402
- concept of, 375–77
 - of difference between two proportions, 535
 - of difference of two means, 521–22
 - in least-squares regression model, 581
 - of population proportion, 436
 - standard error for, 494
 - of sample mean, 375–92, 403, 407
 - definition of, 375–76
 - describing, 383
 - mean and standard deviation of, 382
 - from normal populations, 377–83
 - from not-normal populations, 384–88
 - shape of, 383
 - of sample proportion, 392–400
 - describing, 392–95
 - probabilities of, 396–97
- Scatter diagrams, 177–79
- definition of, 177
 - drawing, 177–78, 183, 194–95
- Seed, 18
- Self-selected samples, 27
- Shape. *See* Normal probability distribution
- Side-by-side bar graph, 58–59
- Sigma (σ) 108
- Significance, 538
- definition of, 463
 - of least-squares regression model, 580–94
 - practical, 538
 - definition of, 474
 - statistical vs., 474–75
 - statistical, 538
 - practical vs., 474–75
 - Type I error and, 459
- Simple events, 224
- Simple random sample, 16–23, 25
- combinations of, 272–73
 - confidence interval construction and, 414–15
 - definition of, 16
 - illustrating, 16
 - obtaining, 16–23
- Simpson's paradox, 604
- Simulation, 231–32
- using technology, 237
- Skewed distributions, 80

- IQR for, 162
- Skewness, 162
- Slope, 198, 201–2
 - defined, C-1
- Slope-intercept form of equation of line, C-5
- SPF (sun-protection factor), 353
- Split stems, 79
- Spread. *See* Standard deviation; Variance
- Spreadsheets. *See* Statistical spreadsheets
- Standard deviation, 125, 129–31, 142, 180
 - of binomial random variable, 305–6, 308–9
 - confidence interval about, C-10–11
 - of discrete random variables, 292–93
 - from grouped data, 144–46, 149
 - interpretations of, 130
 - outlier distortion of, 155
 - population, 129–30
 - confidence intervals about, C-7–13
 - hypothesis testing for, C-13–19
 - least-squares regression model and, 582–83
 - margin of error of confidence interval and, 406
 - sample, 129–30, 583
 - of sampling distribution of sample mean, 382
 - of two data sets, 130–31
 - using technology, 142, 146
- Standard error, 423, 583–85
 - computing, 583–84
 - definition of, 583
 - of the mean, 382
 - for sampling distribution of population proportion, 494
- Standard normal distribution table, A-10–11
- Standard normal probability distribution, 326, 331–44
 - area under the standard normal curve, 332–36
 - to left of z -score, 332–33, 335, 336
 - as a probability, 340–41
 - to right of z -score, 334–35, 336
 - technology to find, 333–34, 337–38
 - between two z -scores, 335, 336
 - z -scores for given, 336–40
 - notation for probability of standard normal random variable, 340
 - properties of, 331–32
 - t -distribution compared to, 424–25
 - technology to find, 344
 - z -scores for given area, 336–40
- Standard normal random variables, 325–27, 340–41
- Statistic, 107
 - biased, 127
 - as random variable, 373
 - test, 466, 494
- Statistical Abstract of the United States*, 231
- Statistical inference, 375
- Statistical significance, 538
 - practical significance vs., 474–75
- Statistical spreadsheets. *See also* Excel
 - area under standard normal curve using, 334
 - pie charts on, 60
 - Tally command, 56
- Statistical thinking, 3–4
- Statistics, 3–13
 - definition of, 3–4
 - descriptive, 4
 - inferential, 5, 41
 - mathematics vs., 4
 - process of, 4–6
 - variables in, 6–9
 - data vs., 8–9
 - discrete vs. continuous, 7–9
 - nominal vs. ordinal, 13
 - qualitative (categorical) vs. quantitative, 6–7
- Status quo* statement, 455, 456
- Stem-and-leaf plots, 76–79
 - constructing, 76–79
 - split stems, 79
 - using technology, 92–93
- Strata, 23
- Stratified random sampling, 23–24, 25, 29
- Student's t . *See* t -distribution
- Subject (experimental unit), 14, 40, 41
 - in matched-pairs design, 43
- Subjective probability, 232
- Sum of squared residuals, 202–3
- Sum of squares, C-27
- Sunscreens, 353
- Survey data, probability model from, 227
- Surveys, 54, 438
 - Internet, 27–28
- Survey sampling, 14
- Suzuki, Ichiro, 295
- Symmetric distributions, 80, 130
- Systematic sampling, 25–26, 27, 29
- Tables, 106, A-1–14
 - binomial, 303–5
 - binomial probability distribution, A-2–9
 - chi-square (χ^2) distribution, A-13
 - continuous data in, 73–75
 - critical values for correlation coefficient, A-14
 - discrete data in, 71–72
 - open-ended, 73
 - qualitative data in, 55–56
 - random number, A-1
 - standard normal distribution, A-10–11
 - t -distribution, A-12
- t -distribution, 423–25
 - finding values, 425–26
 - hypothesis testing and, 480–81, 482
 - properties of, 425, 480–81
 - standard normal distribution compared to, 424–25
- t -distribution table, A-12
- Technology. *See also* Excel; MINITAB; TI-83/84 Plus graphing calculator
 - ANOVA using
 - one-way, C-23–24, C-33
 - area under normal curve using, 346–47
 - area under standard normal curve using, 333–34, 337–38
 - binomial probabilities using, 305, 313

- boxplots using, 168
- chi-square tests using, 570, 580
- coefficient of determination using, 215
- combinations using, 272, 278–79
- confidence intervals using, 598
 - for population mean where population standard deviation is known, 412, 422
 - for population mean where population standard deviation is unknown, 428, 434–35
 - for population proportion, 437–38, 443
 - for population standard deviation, 443, C-11, C-13
- difference between two means using, 525, 534
- difference between two population proportions using, 539, 545
- exact P -values using, 558
- factorials using, 278–79
- goodness-of-fit test using, 558
- in hypothesis testing
 - about population mean assuming known population standard deviation, 473, 479–80
 - about population mean with unknown population standard deviation, 484, 493
 - about population proportion, 497–98, 500
- least-squares regression model using, 588–89, 594
- linear correlation coefficient using, 183, 194–95
- mean and standard deviation of discrete random variable using, 293, 297
- mean using, 146
- normal probability distribution using, 353–54
 - standard, 344
- normal probability plot using, 269–70, 361
- normal random variable using, 348–49
- number summary using, 160
- permutations using, 269–70, 278–79
- prediction intervals using, 598
- scatter diagram using, 183
- simple random sample using, 18–23
- standard error using, 584
- testing claims about matched-pairs
 - data using, 513, 520
- two-sample t -tests, independent sampling, 534
- Test statistic, 466, 494
- TI-83/84 Plus graphing calculator, 18–19, 22–23
 - area under normal curve using, 346–47, 353–54
 - scores corresponding, 354
 - area under standard normal curve using, 333
- binomial probabilities using, 313
- boxplots using, 168
- chi-square tests using, 580
- coefficient of determination using, 215
- combinations using, 272, 278–79
- confidence intervals using, 599
 - for population mean, 428
 - for population mean where population standard deviation is known, 422
 - for population mean where population standard deviation is unknown, 434–35
 - for population proportion, 443
- for population standard deviation, 443, C-13
- correlation coefficient using, 195
- difference between two population proportions using, 545
- factorials using, 278–79
- in hypothesis testing
 - about population mean assuming known population standard deviation, 479–80
 - about population mean with unknown population standard deviation, 484, 493
 - about population proportion, 500
- least-squares regression line using, 202, 208
- least-squares regression model using, 588–89, 594
- mean and median on, 123
- mean and standard deviation using
 - approximation, 146
 - from grouped data, 149
- normal probability plot using, 361
- one-way ANOVA using, C-23–24, C-33
- permutations using, 269–70, 278–79
- prediction intervals using, 599
- quartiles using, 154
- scatter diagrams using, 194
- simulation using, 237
- standard deviation of discrete random variable using, 293, 297
- standard normal distribution using, 344
- time-series graphs on, 82
- two-sample t -tests using
 - dependent sampling, 520
 - independent sampling, 534
- variance using, 130
- z -score from a specified area to the left using, 337–38
- Time-series data, 81
- Time-series graphs, 81–82
- t -interval, 427
- Total deviation, 209
- Transformations, linear, 122
- Treatment, 40
- Tree diagram, 230
- Trials, 298, 362
- Trimmed mean, 122
- t -statistic, 425–26
 - interpretation of, 424
 - pooled, 527–28
 - two-sample, 526–28, 534
 - Welch's approximate, 521–22
- Tukey, John, 159, 163
- Twain, Mark, 208
- Two-tailed tests, 455, 466, 468–69, 470, 471–72, 481–82, 496–97
 - of difference between two means, 509, 510
 - of difference between two population proportions, 536, 537
 - of difference of two means: independent samples, 522
 - in least-squares regression model, 586–87
 - about population proportion, 496–97
- Two-way table. *See* Contingency (two-way) table(s)

- Type I error, 457–59, 464, 526
 - in ANOVA, C-20
 - probability of, 458–59, 464
- Type II error, 457–59
 - probability of, 458–59

- Underrepresentation of population, 34
- Unexplained deviation, 209–10
- Uniform density function, 321
- Uniform probability distribution, 80, 318, 319–21
 - definition of, 319
- Unimodal instruction, 531
- Univariate data, 176
- Unusual events, 225–26
- Upper class limit, 73

- Value, expected, 291–92
- Variable(s), 6–9. *See also* Random variable(s)
 - associated, 179, 181
 - column, 243, 563
 - data vs., 8–9
 - defined, 6
 - dependent (response), 14, 40, 41, 177–78
 - discrete vs. continuous, 7–9
 - independent (explanatory or predictor), 40, 177–78
 - lurking, 3, 15, 177, 184
 - nominal vs. ordinal, 13
 - relation between two. *See* Relation between two variables
 - response, 14, 40, 41
 - row, 242–43, 563
- Variance, 125–29, 130, 142
 - of discrete random variables, 292–93
 - from grouped data, 144–46
 - population, 125–27, 144
 - sample, 125, 127–29, 144
 - technology to determine, 130, 142
- Venn diagrams, 238
- Vertical line, C-1
- Voluntary response samples, 27

- Weighted mean, 144
- Welch, Bernard Lewis, 521
- Welch's approximate t , 521–22
- Wiles, Andrew, 230
- Wunderlich, Carl Reinhold August, 488

- Yakovlena, Vera, 260
- y -intercept, 201–2, C-5

- Z-interval, 410
- z -score, 149–50, 425
 - area under normal curve using, 326
 - comparing, 150
 - expected, 355
 - for given area under standard normal probability distribution, 336–40
 - population, 150
 - sample, 150