INDEX

A
Absolute zero, 17, 18, 20
Age, and home ownership rates, 490–491
Age of vehicle, and cars in use in U.S., 107
Alpha (type 1) error, defined, 230, 249, 255
Alpha level: most commonly used, 230; selecting an, 229–230, 255; and table of critical t-scores, 517–518
Alternative hypotheses (H_a): explanation of, 249, 251–254; for tests about one mean, 259; for tests about one proportion, 267–268
Analysis of variance (ANOVA), one-way: analyzing variance, 339, 342–346; comparing means, 356–358, 360; defined, 98, 314, 338; exercises, 362–368; F-test, 350–354; F-test example, 354–355; hypotheses and assumptions, 339–341; key terms, 361; quiz answers, 362; quizzes, 341, 345, 353, 359; summary on, 360–361
Analysis ToolPak, Excel’s: installing, 50–51; for summary tables, 51–55
Anti-instructor comments, and number of class factions, 456–461
Arithmetic mean: defined, 82, 95–96, 102; formula for, 96, 97; from grouped data, 101; and interval/ratio data, 96–98; of list of values, 100–101; two mathematical properties of, 97; working with the, 98–99
Association of interval/ratio variables: exercises, 446–452; form of linear interval/ratio association, 444–445; key terms, 446; quiz answers, 446; quizzes, 432–433, 438, 441; and regression line (intercept and slope), 429, 444–445; significance testing for, 434–445; strength of, 439–441; summary on, 445–446; visualizing, 426–434
Assumptions: in general hypothesis testing process, 254–255; and one-way ANOVA, 339–341; regression, 463–464; in testing about one mean, 259–260; in testing about one proportion, 268
Average, defined, 82, 96, 98, 102. See also Mean

B
Bar charts: defined, 30–31; examples of, 30, 83, 84; exercise in making, 71–72; histograms versus, 36; modified, 31–32; and nominal association, 376–377; stacked, 37, 56, 86, 89
Bell curve: defined, 188, 189; properties of, 189–202. See also Normal distribution
Beta error (type II error), defined, 249–250
Bias and efficiency, 220–221
Bimodal, defined, 181–182
Binary variable, 18
Bins (categories), defined, 39–40
Box-and-whiskers plot: description of, 152, 153–154, 177; drawing, 155–158; exercise using, 181; quiz on, 159–160

C
Cartesian coordinate way of representing and reading quantitative information, 36
Case, defined, 14, 15, 21
Cases or observations, describing scores of a sample of, 26
Categorical data, dispersion of: example of, 115–117; index of qualitative variation (IQV), 118–119, 144–145; and proportion in nonmodal categories, 117–118; segregation index, 119–121, 145–146
Categories, and interval/ratio data, 39–40
Census bureau, data collected by, 12
Central tendency: basic idea of, 82–83; Excel for finding measures of, 108–109; exercises, 103–107; key terms, 102; mean, 82, 95–101; median, 82, 88–95; mode, 82, 83–88; quiz answers, 103; quizzes, 87, 92, 100; SPSS for finding measures of, 109–112; summary on, 102
Central tendency estimates: bias and efficiency, 220–221; introduction to, 219; point estimates, 220, 229–240; standard error of a mean, 221–225, 229; standard error of a proportion, 225–227, 229
Charts. See Bar charts; Histograms; Pie charts
Chi-squared: calculating, 318–319; comparing proportions across several groups, 314–315; defined, 177, 314; Excel for solving chi-squared problems, 322–324; exercises, 333–336; key terms, 332; quiz answers, 333; quizzes, 320–321, 331; SPSS for, 325–327; summary on, 331–332; table of critical values of the chi-squared test, 505–506; testing for multiple group differences, 315–320
Class factions, and anti-instructor comments, 456–461
Cluster samples, defined, 215–216
Coefficient of determination (big $R^2$-squared): defined, 440, 455–456; example using, 458–460
Coefficient of kurtosis, formula for, 173
Coefficient of skewness, 167–168
Coefficient of variation (CV): description of, 134–135, 139; example of calculating, 136–137; Excel for working with, 147–148
Comparing two groups, 280
Comparing two groups’ means, 280–289
Comparing two groups’ proportions, 289–294
Concordant pairs, 395–396
Confidence intervals: boundaries of, 230–231; defined, 212; importance of, 213; for means, 231–235; for proportions, 236–241; setting, 229
Confirmatory application of statistics, 7
Confirmatory phase of research, 10
Conservatism and sentencing attitudes, relationship between, 401–402
Constant: in regression equation, 455; variable versus, 15
Continuous variables: defined, 16; exercises illustrating, 22–23
Cramér’s V: description of, 382, 383, 405; example of, 390; formula for, 382
Critical value, defined, 254, 257
Cross-tabulations, and nominal association, 375
Cumulative percentages: description of, 34, 35, 89–91; rounding, 91
Cumulative relative frequencies, defined, 35

D
Data, defined, 14, 21
Death penalty: formal education and attitude toward, 416; and gender and beliefs about life after death, 409
Decision-making, in hypothesis testing process, 256–257
Degrees of freedom, and Pearson chi-squared test of independence, 318, 319–320
Degrees of freedom between groups ($df_b$), 352, 353
Degrees of freedom within groups (dfw), 352, 353
Dependent variables: defined, 15–16; exercises for identifying, 23–24
Deprivation theory, 441–442
Descriptive statistics, 5
Discordant pairs, 395, 396
Discrete variables: defined, 16; exercises illustrating, 22–23
Discussion section, of research report, 13–14
Dispersion: basic idea of, 114–115, 139; of categorical data, 115–121; Excel for working with, 144–148; exercises related to, 140–144; of interval/ratio data, 121–139; key terms, 140; quiz answers, 140; quizzes on, 122, 138; summary on, 139
Dispersion of categorical data: example of, 115–117; index of qualitative variation (IQV), 118–119, 144–145; and proportion in nonmodal categories, 117–118; segregation index, 119–121, 145–146
Dispersion of interval/ratio data: coefficient of variation (CV), 134–137, 139, 147–148; interquartile range, 123–126; mean absolute deviation (MAD), 127–129, 147; quiz on, 138; the range, 122–123; standard deviation, 130–134; two approaches to, 121–122; variance, 130, 132, 134
Displaying one distribution: exercises related to, 57–60; frequency distributions for interval/ratio variables, 38–43; frequency distributions for nominal variables, 26–32; frequency distributions for ordinal variables, 32, 34–37; key terms, 56; quiz answers, 57; quizzes, 33, 38, 43; and SPSS (statistical software), 61–79; summarizing data with Excel, 43–56; summarizing variation in one variable, 26; summary on, 55
Distinctiveness, 17, 18, 20
Dummy variables: creating, 269–270; defined, 18, 270; and regression model, 473; two-sample test for proportions using, 290, 291–292, 307
E
Ecological fallacy, 14
Educational degree by race, from General Social Survey, 327. See also Formal education
Effect modification, 473–474
Effect sizes, 305
Efficient estimates, 221
Empirical generalizations, 10
Equal intervals, 17, 18, 20
Equal probabilities, 176
Eta-squared: defined, 344–346; example, 346–350
Euler’s constant (e), 493
Excel: Analysis ToolPak, 50–55; for chi-squared problems, 322–324; and confidence interval for a mean, 234–235; and confidence interval for a proportion, 238–240; functions for summarizing data, 43–50; and kurtosis, 175; for linear regression analysis, 461; and measures of central tendency, 108–109; and measures of variation, 144–148; and the normal distribution, 199, 206–207; and skewness, 175; and SPSS, 65, 68–70; for standard error of a mean, 224–225; for two-sample tests, 301–302
Excel’s Analysis ToolPak: installing, 50–51; for summary tables, 51–55
Expected cell counts or frequencies, 318, 319
Expected value of the statistic, 218, 240
Exploratory application of statistics, 7
Exploratory phase of research, 10
Exponential distributions, 176

**F**
50th percentile of a distribution, 89
Firearms, racial groups and their access to, 413–414. See also Gun ownership
Fisher’s exact test, 317
Form of group differences, 329–330
Form of the relationship, 376
Formal education: among adults in the U.S., 89, 90; and attitude toward death penalty, 416; and church attendance, 417, 418, 419
Formula Builder, in Excel, 45, 46, 54
Formula for: coefficient of kurtosis, 173; coefficient of skewness, 168; confidence interval of a mean, 231–234; confidence interval of a proportion, 236, 237; Cramér’s V, 382; gamma, 397; index of qualitative variation (IQV), 118; kurtosis (coefficient of kurtosis), 170; large-sample hypothesis test about a single sample proportion, 268–269; logistic regression model, 493; mean, 96–97; mean absolute deviation (MAD), 127; multiple regression, 470; Pearson zero-order product-moment correlation coefficient, 437, 442; phi coefficient, 382; proportional reduction in error (PRE), 383; relative frequencies (percentages and proportions), 28–29; segregation index, 119; skewness, 164; Somers’ d, 398; Spearman’s rho, 400; standard deviation, 130, 131; standard error of the mean, 221–222; standard error of a proportion, 225; tau-b, 398; tau-c, 398; test statistic, 256, 283, 284, 288, 289; testing a hypothesis about a single population mean, 260; variance, 130, 132; Z-score, 193
F-ratio, defined, 351. See also F-test
Frequency chart, defined, 26
Frequency chart rules: and interval/ratio variables, 40–41; and nominal variables, 27–28; and ordinal variables, 34–35
Frequency distributions: and Excel tools, 43–55; exercises using, 57–60; for interval/ratio variables, 38–43, 56; for ordinal variables, 26–32, 56
Frequency polygon, 41, 43, 56
Frequency tables (exercise 1 using SPSS), 65–70
Full rank-order ordinal variables, 20, 34, 85

**G**
Gamma: defined, 395, 406; in example, 402; formula for, 397
Gender and game preference, 374
Gender gap in computer use, 333
General Social Surveys (GSS), examples of data taken from, 59, 60, 61, 72, 103–105, 294, 315, 325–327, 334–335, 363, 364, 373, 379, 409, 410, 413, 414, 416, 417, 419, 480, 482, 501, 503
Global null hypothesis, 462, 496
Goodman and Kruskal tau, 384, 385, 391, 406
Goodness of fit: defined, 7, 455; in examples, 458, 484; in logistic regression, 496, 497–498
Graphical displays: and interval/ratio variables, 41–43, 56; and nominal variables, 29–32, 56; and ordinal variables, 36–38, 56. See also Bar charts; Histograms; Pie charts
Gross national product (GNP) per capita, frequency charts of, 40, 41
Group differences, describing, 327–330
Grouped ordinal data, finding the median in, 92–93
Grouped ordinal variable type, defined, 18, 34, 85
Gun ownership: and attitude toward death penalty, 501–502; General Social Survey data on, 334–335
H
Height, finding the median, 88, 89
Heteroscedasticity, defined, 431
Histograms: bar charts versus, 36–37; defined, 36, 56; examples of, 36, 52, 150, 151, 180, 244, 481; in Excel, 51, 52; of normal distribution, 188, 189, 190–191; for ordinal data, 36; for visualizations of interval/ratio variables, 41, 95, 96
Home ownership rates, and age, 490–491
Homogeneity of variances, Levene test for, 282, 285, 292, 307, 309, 310, 311, 312, 357
Homoscedasticity assumption, 436, 465
Hypothesis: defined, 10; and research problem, 12
Hypothesis testing: alternative hypotheses, 251–254; assumptions, 254–255; calculating test statistics, 255–256; defined, 7, 248; exercises, 275–278; key terms, 275; making a decision, 256–257; null hypothesis, 249, 250–251; quiz answers, 275; quizzes, 258, 263, 271; selecting an error level, 255; summary on, 273–275; and theory, 248–249; and type I error, 249; and type II error, 249–250
Hypothesis testing about one mean: alternative hypotheses, 259; assumptions, 259–260; brief explanation of, 258; example of, 264–265; Excel for, 266; making a decision, 262; null hypothesis, 259; selecting an error level, 260; the t-test, 260–262; usefulness of, 263
Hypothesis testing about one proportion: alternative hypotheses, 267–268; assumptions, 268; example of, 272–273; making a decision, 270; null hypothesis, 267; selecting an error level, 268; the test statistic, 268–270; two approaches for, 267
Hypothesis testing for two samples: comparing two groups, 280; comparing two groups’ means, 280–289; comparing two groups’ proportions, 289–294; Excel for, 301–302; exercises on, 308–312; interpreting group differences, 302–305; key terms, 308; Levene’s test for equality of variances, 282, 285, 292, 307; nonindependent samples, 296–300; quiz answers, 308; quizzes, 286, 293, 298, 306; significance versus importance, 304, 307; summary on, 307–308; testing the difference between two groups’ proportions, 294–296
Hypothesis testing in regression, 461–463
I
i (subscript i), 97
Income distribution for nonblack and black American families, histograms showing, 150–151
Independent variables, 15–16
Index of dissimilarity: defined, 119–121; Excel for computing, 145–146
Index of qualitative variation (IQV): defined, 118–119; Excel for computing, 144–145
Infant mortality and fertility data, example using, 403–404
Infant mortality and life expectancy: interval/ratio association of, 427; ordinal association of, 426
Infant mortality to education with linear regression line, relationship of, 430
Infant mortality to education with nonlinear regression line, relationship of, 431
Inference, defined, 212
Integrated Public Use Microdata Series (IPUMS) project, 12
Intercept (Y-intercept), 445, 455
Interquartile range, calculating the, 123, 124–126
Inter-University Consortium for Political and Social Research (ICPSR), 12
Interval variables, defined, 20
Interval/ratio association: exercises, 446–452; form of linear, 444–445; key terms, 446; quiz answers, 446; quizzes, 432–433, 438, 441; and regression line (intercept and slope), 429, 444–445; significance testing for, 434–445; strength of, 439–441; summary on, 445–446; visualizing, 426–434
Interval/ratio data: and the mean, 96–98, 102, 108–109; and the median, 88, 91, 108; and the mode, 87, 88, 108
Interval/ratio data, dispersion of: coefficient of variation (CV), 134–137, 139, 147–148; interquartile range, 123–126; mean absolute deviation (MAD), 127–129, 147; quiz on, 138; the range, 122–123; standard deviation, 130–134; two approaches to, 121–122; variance, 130, 132, 134
Interval/ratio scales, description of, 38–39
Interval/ratio variables: frequency distributions for, 38–43, 56; and SPSS software, 72–79
Intervening variable, controlling for, 477–478

J
Jointly exhaustive categories, 39

K
Kurtosis: coefficient of, 170–174; defined, 152, 169, 178; Excel for calculating, 175; negative, 169, 170, 175; positive, 169, 170; quiz on, 172; skewness and kurtosis together, 174–175; statistical measures of, 169–171

L
Lambda, 384, 385, 390
Legalization of marijuana, relationship between education and attitude toward, 410–413
Leptokurtic distribution, 169, 170
Levels of measurement, 16–18, 20
Levels or values of a variable, 15
Levene’s test for equality of variances, 282, 285, 292, 307, 309, 310, 311, 312, 357
Linearity assumption, 436, 464
Literacy and urbanism, link between, 422–423
Little r and coefficient of determination, 440; defined, 434, 439–440; example using, 441–444; formula for, 439; quiz on, 441
Log likelihood, 496, 501, 502
Log of the odds, 493, 495, 498–499, 500, 503
Logistic regression analysis: applying, 496–498; exercises, 501–504; extending, 499–500; goodness of fit in, 496, 497–498; interpreting effects in, 493–495; key terms, 501; logistic curve, 490–491; logistic regression model, 492–493; and maximum likelihood (ML), 495–496; and partial effects, 498–499; quiz answers, 501; quizzes, 496; summary, 500–501; when to use, 492
Logit, 494, 495
Lower bound, 230

M
Margin of error, 212
Marginal totals, 28
Marijuana, relationship between education and attitude toward legalization of, 410–413
Marital status and race, association between, 315, 328, 380, 384
Marriage, testing theories about prevalence of, 294–296
Maximum likelihood (ML), 495–496
Mean: defined, 82, 95–96, 102; Excel for finding mean of interval/ratio variable, 108–109; formula for, 96, 97; from grouped data, 101; and interval/ratio data, 96–98, 108–109; of list of values, 100–101; SPSS for finding mean of a distribution, 112; two mathematical properties of, 97; working with the, 98–99
Mean absolute deviation (MAD): calculating the, 127–129; Excel for working with, 147; formula for, 127
Median: defined, 82, 88, 102; Excel for finding, 108; in grouped ordinal data, 92–93; height,
88, 89; and interval/ratio data, 91; in list of values, 94–95; and ordinal data, 89–91, 102; quiz on, 92; SPSS for finding median of a distribution, 111–112

Mediating variable, 475

Methods section, of research report, 12–13

Mode: defined, 82, 83, 102; Excel for finding mode of nominal data, 108; and interval/ratio data, 87–88; and nominal data, 83–85, 102; and ordinal data, 85–86; quiz on, 87; SPSS for finding mode of variables, 109–111

Multicollinearity assumption, 464

Multimodal, defined, 160, 161

Multinomial logistic regression, 499, 500

Multiple regression: applying, 474–478; defined, 469–470; estimated regression equation, 471–472; extending, 472–474; formula for, 470; hypotheses in, 471; ordinary least squares (OLS) for estimating, 470; quiz, 474

Mutually exclusive categories, 39

N

N, sample size, 29, 96, 97

n = 1, 131

National Center for Health Statistics (NCHS), 12

National Health Interview Survey, 12

National Opinion Research Center (NORC), 60, 61

Negative kurtosis, 169, 170, 175

Negative skew, 163, 166, 168

Nominal and ordinal distributions, shape of, 152–153


Nominal data, mode of: Excel for finding, 108; explanation of, 83–85, 102

Nominal dichotomy variable type, 18

Nominal polytomy variable type, 18

Nominal variables: defined, 18; frequency distributions for, 26–32, 56

Nonindependent samples: description of, 296–297; quiz, 298; test for difference in mean differences, 297–299; testing two nonindependent means (example), 299–300

Nonprobability samples, 214

Normal distribution: defined, 151–152, 178, 188; Excel for working with, 206–207; exercises, 203–206; finding areas “under the curve”, 197–201; properties of, 189–191; quiz answers, 203; quizzes, 192, 194, 197; shape of curve of histogram of, 190–191; standard normal or Z distribution, 188, 192–194; summary on, 201–202; working with Z-scores, 194–197

Normal distribution function in Excel (NORMDIST), 199

Normality assumption, 435–436

Null hypothesis: defined, 249, 250–251; global, 462; for a test about one mean, 259; for a test about one proportion, 267

O

Observation, units of, 14, 21

Observed cell counts or frequencies, 318, 319

Observing a datum, 10

Odds ratio, 387

Odds that an event will occur, 493

One mean, tests about: alternative hypotheses, 259; assumptions, 259–260; description of, 258; example, 264–265; Excel for, 266–267; making a decision, 262; null hypothesis, 259; quiz, 263; selecting an error level, 260; six steps to, 259–262; t-test, 260–262

One sample, hypothesis testing for: alternative hypotheses, 251–254; assumptions, 254–255; calculating test statistics, 255–256; defined, 7, 248; exercises, 275–278; key terms, 275; making a decision, 256–257; null hypothesis, 249, 250–251; process of, 250–258; quiz answers, 275; quizzes, 258, 263, 271; selecting an error level, 255; summary on, 273–275; tests about
one mean, 258–267; tests about one proportion, 267–273
One-tailed alternative hypothesis: and chi-squared test, 317; defined, 252; examples of, 257, 259; and significant findings, 304; and table of critical t-scores, 517, 518; for tests comparing two-group means, 281–282
One-way ANOVA: analyzing variance, 339, 342–346; comparing means, 356–358, 360; defined, 338; exercises, 362–368; F-test, 350–354; F-test example, 354–355; hypotheses and assumptions, 339–341; key terms, 361; quiz answers, 362; quizzes, 341, 345, 353, 359; summary on, 360–361
Ordering in magnitude, 17, 18, 20
Ordinal association: significance testing for, 393, 395; strength of, 395; visualizing, 391
Ordinal data: histograms for, 36; and the median, 89–92, 102, 108; and the mode, 85–86, 108
Ordinal logistic regression, 499, 500
Ordinal variables: defined, 18; frequency distributions for, 32, 34–38, 56; two kinds of, 34, 85
Ordinary least squares (OLS), 455, 470–471
Overrepresented in survey research, 267
P
Parameter, defined, 6, 7
Partial effects, assessing, 498–499
Pearson chi-squared test of independence: and association between two nominal variables, 375; calculating, 318–319; comparing proportions across several groups, 314–315; defined, 314; Excel for solving chi-squared problems, 322–324; exercises, 333–336; key terms, 332; quiz answers, 333; quizzes, 320–321, 331; SPSS for, 325–327; summary on, 331–332; testing for multiple group differences, 315–320
Pearson zero-order product-moment correlation coefficient: and coefficient of determination, 440; defined, 434, 439–440; example using, 441–444; formula for, 439; quiz on, 441
Perfect negative rank-order association, 396
Perfect positive rank-order association, 395, 396
Phi coefficient, 382, 405
Pie charts: description of, 32; Excel for creating, 71; and nominal association, 376, 377–378; quiz question on, 33
Platykurtic distribution, 169, 170, 171, 175
p-level: defined, 230, 255; selecting the, 255, 260, 262; and table of critical t-scores, 517–518
Point estimates: assessing confidence in, 229–241; defined, 212, 220
Policy implications of findings, 13
Political conservatism and sentencing attitudes, relationship between, 401–402
Pooled estimator, defined, 283
Population, inferring to the, 5–7
Population parameter, 212
Population proportion (P_u), 236
Pornography, gender and opinions about, 409
Positive kurtosis, 169, 170
Positive skew, 163, 165, 166
Post hoc tests, 357–358, 360
Predicting outcomes with regression analysis: basic description of, 454; exercises, 480–488; key terms, 480; multiple regression, 469–479; quiz answers, 480; quizzes, 462, 464, 474, 478–479; simple linear regression, 454–469; summary on, 479
Predicting with nonlinear relationships, 490–491. See also Logistic regression analysis
Predictions, use of statistics for, 7–8
Probability sampling: kinds of, 214–215; methods, 6–7
Promotion, data on being passed over for, 335–336
Proportion: confidence interval for a, 238–240; standard error of a, 225–227; tests about one, 267–273
Proportional reduction in error (PRE) logic, 381, 383
Qualitative variables, 17–18
Qualitative variation, index of (IQV): defined, 118–119; Excel for computing, 144–145
Quantitative variables, 18
Quiz answers for chapter quizzes on: association of interval/ratio variables, 446; association with categorical variables, 407; central tendency, 103; chi-squared, 333; dispersion, 140; displaying one distribution, 57; distributional shape, 178; general introduction to statistics, 21–22; hypothesis testing for one sample, 275; hypothesis testing for two samples, 308; inferential statistics, 241; logistic regression analysis, 501; normal distribution, 203; one-way ANOVA, 362; regression analysis, 480
Quizzes on: association of interval/ratio variables, 432–433, 438, 441; association with categorical variables, 374, 379, 381, 386–387, 388–389, 394, 398–399, 400–401; central tendency, 87, 92, 100; chi-squared, 320–321, 331; dispersion, 122, 138; displaying one distribution, 33, 38, 43; distributional shape, 159–160, 161, 172; general introduction to statistics, 6, 8, 11, 15, 17, 19; hypothesis testing for one sample, 258, 263, 271; hypothesis testing for two samples, 286, 293, 298, 306; inferential statistics, 218, 228–229; logistic regression analysis, 496; normal distribution, 191, 194, 197; one-way ANOVA, 341, 345, 353, 359; regression analysis, 462, 464, 474, 478–479
Regression line (intercept and slope): description of, 429, 444–445; example of, 458, 459
Regression prediction equation, 444
Regression analysis: exercises, 480–488; key terms, 480; multiple regression, 469–479; predicting outcomes with, 454; quiz answers, 480; quizzes, 462, 464, 474, 478–479; simple linear regression, 454–469; summary on, 479
Research, basic elements of: dependent and independent variables, 15–16; discrete and continuous variables, 16; exercises, 22–24; key terms, 21; levels of measurement, 16–18, 20; quiz answers, 21–22; quizzes on, 15, 17, 19; summary on, 20–21; units of analysis, 14; variables, 15
Research hypothesis, defined, 10, 248–249. See also Hypothesis testing
Research report: common format for, 11; discussion, 13; methods, 12–13; peer review and publication of, 13–14; research problem, 12; results, 13
Research studies, identifying variables in, 23–24
Residuals: analyzing, 329–330; and multiple regression, 464–465; and simple regression, 458
Sales figures collected by the Recording Industry Association of America, 333–334
Sample, defined, 5, 14
Sample survey, 212
Sample variance ($s^2$), 130
Samples: cluster, 215–216; stratified, 215
Sampling distributions, defined, 217, 240
Sampling error, 216–217
Sampling variability, 217
Scatter plot: example of, 427; preparing a, 428–429; seeing association in a, 429–430
Scheffé test, interpreting the, 358, 360
Scientific theories, defined, 9
Secularization hypothesis in sociology of religion, 417–419
Segregation index: defined, 119–121; Excel for computing, 145–146
Sexual activity and religious service attendance, relationship between, 482–483
Sigma, 97, 221
"Sigma hat," 221
Significance, importance versus, 304, 307
Significance testing. See Hypothesis testing
Simple linear regression: applying, 465–469; defined, 454–455; example of, 456–461; hypothesis test in regression, 461–463; regression assumptions, 463–464; regression equation, 455–456
Skewed distribution, description of, 99
Skewness: basic idea of, 163–164; coefficient of, 167–168; defined, 152, 163; Excel for calculating, 175; formula for, 164; and kurtosis together, 174–175; negative skew, 163, 166, 168; and nominal variables, 152; positive skew, 163, 165, 166; quiz on, 166; statistical measures of, 164–166
Slope (b), 445, 455
Social class differences, in voting for George Bush, 502–504
Social research and theories, 9–11
Software, statistical. See Excel; SPSS
Somers’ d: defined, 395, 406; in example, 402; formula for, 398
Spearman’s rho (rank-order correlation coefficient): defined, 395, 406, example using, 403–404; exercises using, 421–423; formula for, 400; quiz on, 400–401
SPSS (software): Data View, 62; and distributional shape, 183–185; exercises using, 65–79; introduction to, 61–64; launching, 61; and measures of central tendency, 109–112; for Pearson independence chi-squared test, 325–327; Variable View, 63–64
Spuriousness, controlling for, 475–477
Squared deviations from the mean, 98
Stacked bar charts, 37, 56, 86, 89
Standard deviation: defined, 130–131; example of calculating variance and, 132–134; formula for, 130, 131
Standard error of a mean: calculating, 222–224; and confidence intervals, 229; Excel for calculating, 224–225; formula for, 221–222
Standard error of a proportion, 225–227
Standard error of the regression coefficient, 467
Standard errors, 217–219
Standard normal distribution: description of, 188, 192–194; working with Z-scores, 194–197
Standardized regression (beta) coefficient, 469
Statistic, defined, 6
Statistical independence, defined, 373
Statistics: descriptive, 5; exercises after introduction to, 22–24; four tasks for, 4–8; for hypothesis testing, 7; inferential, 5–7; key terms, 21; for prediction, 7–8; quiz answers, 21–22; quizzes on, 6, 11, 15, 17, 19; reasons for studying, 4; in research process, 9–14; summary on role of, 20–21

Stem-and-leaf plot, 185

Stratified samples, defined, 215

Strength of group differences, 328–329

Strength of the relationship, 375

Student’s t-distribution: defined, 177, 260; for hypothesis testing about one mean, 255, 260; table of critical t-scores, 517–518

Subscripts, in statistics, 29, 97

Sum of deviations from the mean, 97

Sum of squares between groups (SSB): description of, 343–344; example, 349–350

Sum of squares total (SST): description of, 342–343; example, 346–347

Sum of squares within groups (SSW): defined, 344; example, 348–349

Symmetry, of the normal distribution, 190

T

Tables: critical values of the chi-squared test, 505–506; critical values of the F-test, 507–509; table of critical t-scores, 517–518; Z-score table, 511–516

Tau-b: defined, 395, 406; in example, 402; formula for, 398

Tau-c: defined, 395, 406; formula for, 398

Television watching: predicting average daily hours of, 483–486; two theories on, 480–482

Test statistic: calculating, 255–256, 265; for small samples, 284

Testing, hypothesis: alternative hypotheses, 251–254; assumptions, 254–255; calculating test statistics, 255–256; making a decision, 256–257; null hypothesis, 250–251; quiz, 258; selecting an error level, 255; summary on, 273–275; and theory, 248–249; and type 1 error, 249; and type II error, 249–250

Testing about one mean: alternative hypotheses, 259; assumptions, 259–260; description of, 258; example, 264–265; Excel for, 266–267; making a decision, 262; null hypothesis, 259; quiz, 263; selecting an error level, 260; six steps to, 259–262; t-test, 260–262; usefulness of, 263

Testing about one proportion: alternative hypotheses, 267–268; assumptions, 268; example of, 272–273; making a decision, 270; null hypothesis, 267; selecting an error level, 268; the test statistic, 268–270; two approaches for, 267

Theories and social research, 9–11

Theory, observation, and hypothesis testing, 250–251. See also Hypothesis testing

t-tests: Excel for two-sample t-tests, 301–302; formula for, 260; paired sample, 302; SPSS one-sample t-test results, 261–262; table of critical t-scores, 517–518; unpaired or independent, 301

Tukey, J., 155, 185

Two-sample tests: comparing two groups, 280; comparing two groups’ means, 280–289; comparing two groups’ proportions, 289–294; Excel for, 301–302; exercises on, 308–312; interpreting group differences, 302–305; key terms, 308; nonindependent samples, 296–300; quiz answers, 308; quizzes, 286, 293, 298, 306; summary on, 307–308; testing the difference between two groups’ proportions, 294–296

Two-tailed alternative hypothesis: defined, 252; examples of, 257, 259; and table of critical t-scores, 517, 518; for tests comparing two-group means, 281–282

Type I (or alpha) error, defined, 230, 249, 255

Type I error level: most commonly used, 230; selecting a, 229–230, 255; and table of critical t-scores, 517–518

Type II (or beta) error, defined, 249–250
U
Unbiased point estimate, 220
Unemployment and property crimes, relationship between, 441–444
Unimodality: description of, 151, 158, 160; of a normal distribution, 190; in real-world applications, 160–163
United States Bureau of the Census, data collected by, 12
Units of analysis, 14, 21
Univariate descriptive statistics, 55
Upper bound, 230
Urbanism and literacy, examining link between, 422–423

V
Valid frequencies (Valid Percent), 35
Variables: defined, 15, 21; dependent and independent, 15–16; discrete and continuous, 16, 22–23; dummy variable, 269–270, 473; exercises for identifying, 22–24; levels of measurement of, 16–18, 20
Variance: analyzing, 339, 342–346; defined, 130; example of calculating the, 132–134; formula for, 130, 132
Voting for Bush, and social class differences, 502–504

W
Wald chi-squared test, 498, 500

X
X bar, 96, 97
x-axis of bar chart, 31

Y
y-axis of bar chart, 31
Y-intercept, 445, 455

Z
Z distribution: defined, 177, 188; properties of, 192–193; and Z-scores, 193–194
Z-scores: and confidence intervals for means, 231–234; and confidence intervals for proportions, 236, 237; defined, 192–194, 232; Excel for working with, 206–207; quiz on, 197; using the Z-score table, 511–516; working with, 194–197, 202