

Index

A

Acceptance regions, 226f
 Acute lymphoblastic leukemia (ALL), 192–193
 Acute myeloblastic leukemia (AML), 192–193, 207
 Addition rule for probability, 73–75
 Adjusted rates, 54–55
 confidence intervals for, 185–188
 direct method for, 54
 indirect method for, 54
 variances of, 185–187
 Age groups, 23t, 28f
 death rates by, 49t, 55t, 93
 obesity and, 29
 Age-adjusted rates, 186
 AIDS, 80–81, 149
 Air pollution, 286, 291
 ALL. *See* Acute lymphoblastic leukemia
 α , 218–221
 versus β , 218–221
 reasonable values for, 218
 Alternative hypothesis, 215
 AML. *See* Acute myeloblastic leukemia
 Analysis of variance (ANOVA), 374–376, 378
 assumptions for use of, 323–324
 by general linear models, 344t
 linear model representation of, 339–342
 multiple comparisons, 329–333
 one-way, 324–329
 table, 327–329, 363–364, 369t
 table, for two-factor design with interaction, 338t
 for test score increase data, 338t
 two-way, 332–335
 two-way, with interaction, 335–339, 341
 with unequal observations in subgroups, 342–345
 Analysis strategies, 431–434
 contingency table analysis, 435–437
 descriptive, 434–435
 linear regression, 437–440

 logistic regression, 437–440
 preliminary, 432
 subpopulation, 433–434
 ANOVA. *See* Analysis of variance
 Antilogarithms, 44
 Approximations
 to binomial distributions, 126–131, 174
 normal, 256, 260
 normal, to binomial, 184–185
 to Poisson distributions, 126–131
 Arithmetic means, geometric v., 43
 Associations, 6
 hypothesis of no, 282–284
 odds ratio and, 277–278
 Assumptions
 for ANOVA, 323–324
 for linear regression, 357–358
 for logistic regression, 392–394
 for nonresponses, 146–147
 for stationary population, 97–98
 for survival length, 300
 for variance, 203, 237

B

Balanced repeated replication, 427–428
 Bar charts, 27–30, 64, 66, 110
 bar length in, 28
 clustered, 30f
 data represented by, 28
 showing binomial distribution, 106f
 Baseline hazards, 412
 Bayes' theorem, 84–87
 Bayes, Thomas, 85
 Berkson, Joseph, 387
 Bernoulli, Daniel, 91
 β , 218–221
 decreasing, 218–221
 reasonable values for, 218

- Biases, 135
 nonresponse, 145–146
 selection, 139
- Binomial confidence interval, 463t–464t
- Binomial distribution, 103–110
 approximations to, 126–131, 174
 bar charts, 106f
 in confidence intervals, 182–183
 mean and variance of, 107–110
 outcomes, 105t
 shapes of, 110
 variance of, 107–110, 199–200
- Binomial probabilities, 103–107, 111f, 454t–457t
- Binomial proportions, comparing two independent, 274
- Biofeedback, 265
- Biostatistics
 applications of, 1
 defined, 1
 government sources of data, 481–485
- Birthdays, 89–90
- Births, 73t, 77
 probabilities of, 76t
- Blind experiments, 154
- Blocks, 151, 155–156
 ranking observed values within, 263
- Blood pressure, 10, 12f, 25–27, 31t, 45f, 46, 199, 272t, 294, 350, 352t, 364, 382
 box plots, 47f, 48f
 dot plots for, 37f
 histograms of, 33–34, 35f
 scatter plots for, 38f
- BMI. *See* Body mass index
- Body mass index (BMI), 381–382, 391
- Box plots, 47
 of blood pressure, 47f, 48f
- Breast tumors, 85–86, 320
- Byssinosis, 335
- C**
- Caloric intake, 250–251, 253–254, 258t, 345
- Cancer, 294, 297–298, 300t
- Case-control design, 159–160, 281–282
- Cause-and-effect relationships, 157–158
- CDC. *See* Center for Disease Control
- Cell count, 279
 assuming no association, 274–277
- Cell frequencies, odds ratios and, 278
- Censored observations, 298, 300, 308
- Census, 19, 138–139, 481–482
- Center for Disease Control (CDC), 16, 252
- Central limit theorem, 124–126, 127–128, 189
- Central tendency, 39–45
 measures of, 42
- Chance, 135–137
- Change, measures of, 55–64
- CHD. *See* Coronary heart disease
- Chi-square distributions, 201f, 232
 critical values for, 465t
 goodness-of-fit, 271
 Yates' corrected, 275
- Chi-square test, 274–277
- Circular systematic sampling scheme, 142
- Cirrhosis, 345
- Clinical trial data, 22t
- Cluster sampling, 144–145
 multistage, 145
- CMF. *See* Compressed Mortality File
- CMH statistic, 290
- CMH test
 with small data sets, 314–316
 of survival distributions, 310–313
- Cochran, William, 288, 292
- Cochran-Mantel-Haenszel test, 288–290, 335
- Coefficient of determination, 355–357
- Coefficient of variation, 51
- Coefficients, estimation of, 351–353
- Cohort study design, 160
- Colorectal cancer, 319
- Combinations, 448
- Comparison groups, 149–150
- Completely randomized design, 339–340
- Complex sample designs, 421
- Compressed Mortality File (CMF), 483
- Computer packages, 343
- Conditional logistic regression, 407–409
- Conditional probabilities, 75–77, 78, 93, 95, 276
 of dying, 300
 estimators of, 301
- Confidence intervals, 169–170, 179t, 302, 365f
 for adjusted rates, 185–188
 binomial distribution and, 182–183
 for crude rates, 185–188
 for difference of two dependent means, 194–195
 for difference of two dependent proportions, 197–198
 for difference of two independent means, 188–194
 for difference of two independent proportions, 196–197
 for difference of two means and proportions, 188–198
 distribution-free, 171–175
 formation of, 187–188
 normal approximation to binomial and, 184–185
 normal distribution, 176–188
 one-sided, 173, 202
 for other measures, 200–205
 for Pearson correlation coefficient, 203–205
 for proportions, 182–185
 sample size and, 198–200
 for variance, 201–203
- Consistencies, 17
- Contingency table analysis, 435–437

- Continuity correction, 174, 289
 Continuous random variables, 71
 Conversion between bases, 445
 Coronary heart disease (CHD), 417
 Correction for continuity, 128
 Correlation coefficients, 60–64
 calculating, 61–62
 Pearson, 60–62
 Spearman rank, 63–64
 Cox regression model, 414
 CPS. *See* Current Population Survey
 Crossover design, 158
 Crosstabulation, 23t
 Crude rates
 confidence intervals for, 185–188
 variances of, 185–187
 Cumulative binomial distribution, 173t
 Cumulative distribution function, 107, 108f, 460t-461t
 of standard normal distribution, 119, 120f
 Current Population Survey (CPS), 483
 Cutoff points, 402
- D**
- Data, 2–4
 analyzing, 5
 in bar charts, 28
 clinical trial, 14
 collection in follow-up studies, 297–299
 common problems, 14
 continuous, 11
 design and, 4–5
 graphical representation, 22–39
 importance of, 2
 interpreting, 4
 negatively skewed, 36
 nominal, 11
 numerical representation of, 9–10
 ordinal, 11
 positively skewed, 36
 reliability and validity of, 11–13
 replication of, 6
 tabular representation, 22–39
 types of, 11
 Death rates
 age-specific, 49t, 55t, 93
 crude, 53, 54
 Decision rule, 216–218
 α and β and, 218–221
 Denominators, 52
 Descriptive analysis, 434–435
 Descriptive methods, 21–22
 Design, 135–137
 data and, 4–5
 effect, 424–425
 variations in, 158–160
 Design-based analysis
 components of, 422–426
 design effect, 424–425
 sample weights, 422–423
 Design-based inference, 421–422
 Design-weighted least squares (DWLS), 438
 Deviance residuals, 398
 by subject, 400f
 Diabetes, 30, 396–398
 Diagnosis Related Groups (DRGs), 67
 Diagnostics, 374–376
 Diets, 214–215
 DIG. *See* Digitalis Investigation Group
 DIG40, 38
 DIG200, 22, 31, 35f, 407–409, 410–415, 416
 basic patient characteristics in, 22t
 Digitalis Investigation Group (DIG), 21, 149, 153
 Digoxin, 21
 Discrete random variables, 71
 Disease status, 82t
 risk factors and, 280
 Distribution
 binomial, 103–110
 normal, 51, 116–124
 Poisson, 110–116, 125
 probability, 72
 Distribution-free intervals, 170–176
 confidence, 171–175
 prediction, 170–171
 tolerance, 175–176
 Doolittle, M.H., 274
 Dot plots, 37–39
 for blood pressure, 37f
 Double-blind experiments, 154
 Double-dummy procedure, 159
 Draft lottery, 136
 DRGs. *See* Diagnosis Related Groups
 Drug trials, 154
 Dunnett's method, 331–332
 DWLS. *See* Design-weighted least squares
 DXA. *See* Full body scan
- E**
- Education, measuring, 18
 Empirical distribution function, 122–123
 Epidemiology, 80–84
 predicted value negative in, 81–84
 predicted value positive in, 81–84
 rates and probabilities, 80–81
 sensitivity in, 81–84
 specificity in, 81–84
 Equal size intervals, in histograms, 35
 Error rates
 family, 329–330
 individual, 329–330

Error sum of squares (SSE), 337

Estimators

of conditional probabilities, 301

least square, 351

odds ratio, 277–278

pooled, 190

Expected values, in life table, 96–99

Experimental design, 5, 137–138, 148–158

blocking in, 155–156

comparison groups and randomization, 149–150

conditions of, 157

double-blind, 154

extraneous variables in, 155–156

limitations of, 156–158

sample size, 152–155

single-blind, 154

Exponential function, 445–446

Exponential growth, 58–59

population, 58–59

F

F distribution, 473t–475t

F statistic, 326–327, 328

Factorials, 448

design, 156

Failure-to-reject regions, 225f

Finite population correction (FPC), 425

Fisher, Ronald A., 270

Fisher's exact test, 276, 279–280

basis of, 279

Fisher's least significant difference method, 330–331

Follow-up studies, data collection in, 297–299

FPC. *See* Finite population correction

Frauds, 17

Frequency tables, 23–24, 114f

one-way, 23

rank sum, 259t

two-way, 23–24

Friedman test, 262–264

Full body scan (DXA), 484

G

Gauss, Carl, 116

Gaussian distribution. *See* Normal distribution

GDP. *See* Gross domestic products

Gender, 77, 78–79

survival distribution by, 314

General Linear Model (GLM), 379

Geometric growth, 57–58

Geometric means, 42–45

arithmetic means v., 43

GLM. *See* General Linear Model

Goodness-of-fit test, 269–273

chi-square statistic, 271

Gosset, W.S., 114, 181, 292

Graphical representation, 22–39

Graunt, John, 91

Gross domestic products (GDP), 24

Growth

exponential, 58–59

geometric, 57–58

linear, 55–57

H

Haenszel, William, 288

Halley, Edmund, 91

Hazard rates, 305

plots, 413f

standard errors and, 305t

Hazard ratios, 412

HDFP. *See* Hypertension Detection and Follow-up Program

Hinges, 47

Histograms, 30–35

blood pressure, 33–34, 35f

equal size intervals in, 35

intervals in, 31, 32t

Homicide, 417

Hypertension Detection and Follow-up Program (HDFP),

148, 152, 157, 311, 317

Hypotheses

alternative, 215

changing alternative, 219

conducting tests, 221–222

decision rule and, 216–218

equivalence of confidence intervals and, 224–225

of no linear trend, 284–286

null, 215, 254

one-side alternative, 224–225

one-sided tests, 216–217

preliminaries in tests of, 213–222

statistical and practical significance of, 243

test statistics, 215–216

testing, about difference of two means, 234–238

testing, about difference of two proportions, 238–240

testing, about mean, 223–229

testing, about Pearson correlation coefficient, 232–238

testing, about proportion and rates, 229–230

testing, about variance, 231–232

testing, of no association, 282–284

testing, of no trend, 284–286

tests of, and sample size, 240–243

two-sided tests, 216–217

type I errors, 215

type II errors, 215

I

ICC. *See* Intraclass correlation coefficient

Incidence, 80

- Inconsistencies, 15
 Independent events, 77–80
 Infant mortality rate, 52, 233t, 383–384
 Inference, design-based, 421–422
 Inference about coefficients, 357–364
 ANOVA table summary, 363–364
 assumptions for, 357–358
 regression diagnostics, 358–360
 slope coefficient, 361–362
 Y-intercept, 361–362
 Infiltration, 292
 Informed consent, 158
 Insecticides, 263
 Instrument calibration, 252–253
 Interaction, two-way ANOVA with, 335–339
 Interlaboratory testing programs, 252–253
 Interquartile range, 46
 Intersections, 85, 94
 Interval estimation, 169
 for $\mu_{Y|X}$, 364–368
 for $Y|X$, 364–368
 Interval scale, in histograms, 31, 32t
 Interval scales, 10–11
 Intervals, 178. *See also* Confidence intervals;
 Distribution-free intervals; Prediction
 intervals
 in histograms, 31, 32t
 one-sided, 180
 probability calculation for, 121–122
 two-sided, 180
 Intervention procedures, 347
 Intraclass correlation coefficient (ICC), 425
 Inverse transformation, 204
 Iron levels, 275
 Irregular patterns, 15
- J**
- Jackknife repeated replication, 428–430
 Jittering, 39
 Joint confidence coefficients, 478t–479t
Journal of the Royal Statistical Society, 21
- K**
- Kaplan-Meier estimates, 307t, 412, 414
 Kennedy, Galvin, 21
 Kruskal-Wallis (KW) test, 261–262
 KW test. *See* Kruskal-Wallis
- L**
- Lead concentrations, 267
 Least squares estimators, 351
 Legionnaires' disease, 159
 Leverage, 359t
 Life table, 484–485
 abridged, 92t
 annual, 485
 decennial, 485
 expected values in, 96–99
 first four columns in, 93–96
 preliminary annual, 485
 probability and, 91–99
 survival studies and, 299–306
 uses of, 92
 Likelihood ratio test (LRT), 392, 393
 Line graphs, 24–27, 65
 scales of, 25
 Linear growth, 55–57
 Linear model representation of ANOVA, 339–342
 Linear regression. *See also* Regression
 analysis of, 437–440
 assumptions for inference in, 357–358
 multiple, 368–380
 simple, 349–357
 Linearization method, 430–431
 Log odds, 388
 Logarithmic scale, 43
 Logarithmic transformation, 44
 Logarithms, 31
 conversion between bases, 445
 logic of, 445
 natural, 204, 445
 properties of, 445
 Logistic regression
 analysis of, 437–440
 conditional, 407–409
 multiple, 394–403
 ordered, 403–407
 simple, 387–394
 Log-rank test, comparison of survival distributions,
 313–314
 Lost-to-follow up, 298
 LRT. *See* Likelihood ratio test
 Lymphoma, 318–319, 419
- M**
- m* replicates, randomized block design with,
 341
 Mammography, 283
 Mantel, Nathan, 288, 292
 Mantel-Haenszel common odds ratio, 290–291
 Mass function
 binomial, 127f
 Poisson probability, 112
 probability, 105, 107t
 Matched-pairs studies, 280–282
 Maximum likelihood procedure, 390
 Maximum values, 32, 39
 McGill Pain Questionnaire, 265

- Mean, 40–42, 98
 arithmetic, 43
 binomial distribution, 107–110
 difference of two dependent, 194–195, 237–238
 difference of two independent, 188–194, 234–237
 extreme values and, 42
 geometric, 42–45
 Poisson distribution, 113–114
 sample, 40, 62, 303
 square error, 373
 squares, 325–326
 survival times, 302, 308–309
 testing hypothesis about, 223–229
 testing single, 241
 testing two, 242–243
- Measurements
 monitoring, 16
 post-test, 250
 pre-test, 250
 reliability and validity of, 12
- Median, 40–42, 98
 extreme values and, 42
 sample, 304
 survival curves and, 305f
 survival times, 302, 304
- Mental health, 441
- Minimum values, 32, 39
- MINITAB, 137
- Missing values, 14
- Mode, 40–42
 calculating, 42
- Multicollinearity problems, 376–378
- Multiple comparisons procedures, 329–333
 Dunnet's method, 331–332
 Fisher's LSD method, 330–331
 Turkey-Kramer method, 330, 334
- Multiple linear regression, 368–380
 dummy variables and, 378–380
 goodness-of-fit statistics and, 399–401
 model, 368–369
 multicollinearity problems, 376–378
 parameter estimates, ANOVA, and diagnostics in, 374–376
 ROC curve and, 403
 specification of, 369–374
- Multiple logistic regression, 394–403
 model and assumptions, 394–398
 residuals, 398–399
- N**
- National Cancer Institute, 301
- National Center for Health Statistics, 19, 482
- National Health and Nutrition Examination Survey (NHANES), 144, 152, 370, 431, 433, 437, 441–442, 483, 484
- National Health Interview Survey (NHIS), 483
- National Institute of Occupational Safety and Health, 245
- National Safe Workplace Institute (NSWI), 207, 264
- Negatively skewed data, 36
- NHANES. *See* National Health and Nutrition Examination Survey
- NHIS. *See* National Health Interview Survey
- No parameter estimation, 270–271
- Nominal scales, 10
- Nonparametric tests, 249
- Nonresponse, 145–146
 item, 147
 unit, 147
- Normal distribution, 51, 116–124
 comparison of survival distributions, 313
 confidence intervals based on, 176–188
 prediction intervals based on, 205–206
 standard, 118–119, 120, 121
 tolerance intervals based on, 206
 tolerance limits, 466t–467t
- Normal probabilities, 116–124
 calculation of, 119–122
 plot, 122–124, 376f
- NSWI. *See* National Safe Workplace Institute
- Null hypothesis, 215, 285
- Numerical representation, of data, 9–10
- O**
- Obesity, age and, 29
- Observations, 10
- Odds ratio, 277–278, 287
 cell frequencies and, 278
 Mantel-Haenszel common, 290–291
 sample estimator of, 277–278
- OLS. *See* Ordinary least squares
- One-sided tests, 216–217
- Ordered logistic regression, 403–407
- Ordinal scales, 10
- Ordinary least squares (OLS), 437
- P**
- Parameters, 41, 72, 139
 estimates, 374–376, 390–392
- Passive smoke variable, 287
- Patient characteristics, 24
- PDF. *See* Probability density function
- Pearson correlation coefficient, 60–62, 68, 200, 356
 confidence interval for, 203–205
- Pearson residuals, 398
 standardized, 398
 by subject, 400f
- Percentiles, 45–46
- Permutation, 448

- Pertussis, 115, 131
 PFT. *See* Pulmonary function test
 Physical activity, 345
 Placebos, 154
 Playfair, William, 21
 Point estimation, 169
 Poisson distributions, 110–116, 125
 approximations to, 126–131
 mass, 130f
 mean and variance of, 113–114
 Poisson probabilities, 111–113, 458t-459t
 finding, 114–116
 mass function, 112
 Poissonless plot, 115
 Pooled estimators, 190
 Populations
 exponential growth of, 58–59
 growth, 90–91
 stationary, 97, 98
 Positively skewed data, 36
 Poststratification, 423–424
 Power curve, 220f
 Predicted value negative, in epidemiology, 81–84
 Predicted value positive, in epidemiology, 81–84
 Prediction intervals, 169–170
 distribution-free, 170–171
 normal distribution, 205–206
 for $Y|X$, 366–368
 Prevalence, 80
 Probability, 99, 171
 addition rule for, 73–75
 binomial, 103–107, 111f
 of births by trimester, 76t
 calculating, 73–80
 calculating for intervals, 121–122
 conditional, 75–77, 78, 93, 95
 cumulative distribution function, 107
 defined, 71–72
 distribution, 72
 of dying, 95–96
 epidemiology and, 80–81
 life table and, 91–99
 mass function, 105, 107t
 plots, 126f, 375
 Poisson, 111–113
 randomized response technique and, 79–80
 sampling, 140
 in sampling, 87–89
 simulation and, 89–91
 survival, 94
 of type I errors, 217–218
 of type II errors, 217–218
 unconditional cell, 284
 Probability density function (PDF), 116–117, 117
 normal, 117, 118f
 plot of, 327f
 Problems, common data, 14
 Product-limit method, 306–310
 survival distribution estimated by, 309f
 Proportional hazard regression, 409–415
 Proportions
 confidence intervals for, 182–185
 difference of two dependent, 239–243
 difference of two independent, 196–197
 in simple logistic regression, 389–390
 testing hypotheses about, 229–230
 testing hypotheses about difference of two,
 238–240
 testing single, 241
 variance of, 109
 Public Citizen, 109
 Pulmonary function test (PFT), 387, 402
p-value, 222–223, 250, 373
- Q**
- Quality of care, 319
 Questionnaires, reliability and validity of, 12
- R**
- r* by *c* contingency table, 282–286
 testing hypothesis of no association, 282–284
 Race, 23t, 72t
 Radio estimates, 430
 Ramipril, 194, 195, 234, 266
 Random assignment, in experiments, 150–151
 Random block size method, 151
 Random digit dialing (RDD), 140
 Random digits, 452t-453t
 random numbers, generation of, 136–137
 Randomization, 149–150
 Randomized block design, 156, 336
 with *m* replicates, 341
 two-way ANOVA for, 332–335
 Randomized response technique, 13–14
 probability and, 79–80
 Range, 45–46
 Rank sums, frequency and relative frequency of,
 259t
 Rao, C.R., 2
 Rates, 51–55, 186. *See also* Adjusted rates; Crude rates;
 Death rates; Error rates; Hazard rates; Infant
 mortality rate
 defined, 52
 epidemiology and, 80–81
 specific, 53, 81
 stable, 187–188
 testing hypotheses about, 229–230
 vital, 52
 Ratio scales, 10, 11
 Ratios, 51–55

- RDD. *See* Random digit dialing
- Receiver Operating Characteristic (ROC), 84
curve, 401–403
- Regression, 351f, 354f. *See also* Linear regression
diagnostics, 358–360
dummy variables and, 378–380
stepwise, 372
sum of squares about, 354
sum of squares due to, 354
- Regular Care group, 158
- Rejection regions, 225f, 226f, 232, 258
for two-sided alternatives, 239
- Relative frequency, of rank sums, 259f
- Reliability, 11–13
defined, 11
of measurements, 12
of questionnaires, 12
- Replacement
sampling with, 87–88
sampling without, 88–89
- Replication, 6
- Residual sum of squares, 354
- Residuals, 358, 359f. *See also* Standardized residuals
deviance, 398
multiple logistic regression, 398–399
Pearson, 398
plots, 375, 376
- Restenosis, 380
- Reward patterns, 267
- Risk factors, disease status and, 280
- ROC. *See* Receiver Operating Characteristic
- Run-in period, 158
- S**
- Sample designs, 138–148
- Sample size, 46, 220, 254f, 260
confidence intervals and, 198–200
for experiments, 152–155
hypotheses tests and, 240–243
- Sample surveys, 137–138, 483–484
- Sample weights, 422–423
- Sampling, 138–148. *See also* Simple random sampling;
Systematic sampling
frame, 139–140
probability, 87–89, 140
with replacement, 87–88
without replacement, 88–89
replicated, 426–427
systematic, 142–144
unintended, 145–148
- Scatter plots, 38–39, 63f
for blood pressure, 38f
matrix, 39f
- Scatterplot matrix, 371f
- Selection bias, 139
- Selective Service, 136
- Sensitive questions, 13
- Sensitivity, in epidemiology, 81–84
- Sex, 23f
- Sign test, 249–253
uses of, 250
- Simple linear regression, 349–357
coefficient of determination, 355–357
estimation of coefficients in, 351–353
variance of $Y|X$ in, 353–355
- Simple logistic regression, 387–394
estimation of parameters, 390–392
proportion, odds, and logit in, 389–390
statistical interference and, 392–394
- Simple random sampling (SRS), 89, 141–142, 150
- Simulation, probability estimation and, 89–91
- Single-blind experiments, 154
- Slope coefficient, 361–362
- Smokers, 104, 346, 381, 394–396
- SMR. *See* Standardized mortality ratio
- Spearman rank correlation coefficient, 63–64
- Specificity, in epidemiology, 81–84
- Squares
mean, 325–326
sums of, 325–326, 337
- SSC. *See* Sum of squares for column factor C
- SSE. *See* Error sum of squares
- SSR. *See* Sum of squares for row factor R
- SSRC. *See* Sum of squares for interaction between factor R
and factor C
- SST. *See* Total sum of squares
- Stable rates, 187–188
- Standard deviation, 48–51, 119, 177
- Standard error, 177
estimated, 301–302
hazard rates and, 305f
- Standard normal curve, 122f
- Standard normal distribution
cumulative distribution function of, 119, 120f
percentiles of, 121
transforming to, 118–119
- Standardized mortality ratio (SMR), 55
- Standardized residuals, 358, 359f
plot of, 360f, 376f, 377f
- Stationary population, 97, 98
- Statistical interference, 392–394
- Statistical methods, 7
- Statistical tables, 451–479
- Statistics, 41, 72
- Stem-and-leaf plots, 35–36
to compare SBP, 36
- Stepwise regression, 372
- Strata, 144
- Stratification, 144
- Stratified random sampling, 144
- Student's *t*. *See* *t* distribution

- Subgroups, ANOVA, 342–345
- Sum of squares, 325–326, 337
- due to regression, 354
 - about regression, 354
 - residual, 354
 - total, 354
 - type I, 343
 - type III, 343
- Sum of squares for column factor C (SSC), 337
- Sum of squares for interaction between factor R and factor C (SSRC), 337
- Sum of squares for row factor R (SSR), 337
- Summation symbol, 74
- Surveys, telephone, 140
- Survival curve, 303f
- median and, 305f
- Survival distribution
- CMH test, 313, 314–316
 - comparisons of, 310–316
 - estimated by product-limit method, 309f
 - estimated by serum creatinine concentration, 312f
 - by gender, 314f
 - log-rank test, 313–314
 - normal distribution approach to, 313
- Survival studies
- follow-up, 297–299
 - product-limit method, 306–310
- Survival time, 298
- mean, 302, 308–309
 - median, 302, 304
- Symmetric distribution, 36
- Systematic sampling, 142–144
- circular, 142
 - repeated, 143
- T**
- t* distribution, 181, 192
- critical values for, 462t
- Tabular representation, 22–39
- Taylor series expansion, 448–450
- Test statistics, 215–216, 254, 289
- Thomas, Lewis, 157
- Tippett, 136–137
- Tolerance intervals, 169–170
- distribution-free, 175–176
 - normal distribution, 206
- Total sum of squares (SST), 337
- Turkey-Kramer method, 330, 334
- Two parameter estimation, 271–273
- Two-by-two contingency table, 273–282
- analyzing separately, 287–288
 - comparing independent binomial proportions, 274
 - expected cell counts assuming no association, 274–277
 - Fisher's exact test and, 279–280
 - matched-pairs studies and, 280–282
 - multiple, 286–291
 - odds ratio and, 277–278
- Two-sided tests, 216–217
- Type I errors, 215
- probabilities of, 217–218
- Type II errors, 215
- probabilities of, 217–218
- U**
- Unconditional cell probabilities, 284
- Upper respiratory infections (URI), 286, 287
- URI. *See* Upper respiratory infections
- V**
- Validity, 11–13
- defined, 13
 - of measurements, 12
 - of questionnaires, 12
- Variability, 45–51
- Variables, 10
- confounded, 149–150
 - continuous, 273
 - continuous random, 71
 - dependent, 138
 - discrete independent, 415
 - discrete random, 71
 - dummy, 378–380
 - extraneous, 155–156
 - independent, 78, 138, 371
 - indicator, 379
 - large differences in values, 16
 - predictor, 371, 439
 - scales used with, 10–11
- Variance, 48–51, 234
- of adjusted rates, 185–187
 - binomial distribution, 107–110, 199–200
 - changing, 50
 - coefficients of, 51
 - confidence intervals for, 201–203
 - of crude rates, 185–187
 - known, 177, 223–228
 - Poisson distribution, 113–114
 - population, 49, 192, 201–202
 - of proportions, 109
 - sample, 61
 - unknown, 180–182, 228–229
 - unknown but equal population, 190–194, 235
 - unknown but unequal population variance, 236–237
 - of $Y|X$, 353–355
- Variance estimation, 426–431
- jackknife repeated replication, 428–430
 - linearization method, 430–431
 - replicated sampling, 426–427

Variance inflation factors (VIF), 377
Veterans Administration Cooperative Duodenal Ulcer Study Group, 291
VIF. *See* Variance inflation factors
Vital statistics, 482–483

W

Washout period, 158
Weighted least squares (WLS), 438
Wilcoxon Rank Sum (WRS) test, 257–262
 critical values for, 469t–472t
Wilcoxon Signed Rank (WSR) test, 253–257, 267, 468t
 ties in, 255
Withdrawn alive, 298
Worksheets, 307

WRS test. *See* Wilcoxon Rank Sum test
WSR test. *See* Wilcoxon Signed Rank test

Y

Yates' correction, 275, 276
Y-intercept, 362–363
 $Y|X$
 interval estimation for, 364–368
 prediction intervals for, 366–368
 simple linear regression and, 353–355
 variance of, 353–355

Z

z value, 117, 227
 versus *t* value, 181–182