INDEX

Numbers preceded by A refer to Appendix pages.

A
Accuracy, 14
Addition rule, 72–73
Analysis of variance, 306–308
  assumptions, 307
  completely randomized design, 308–334
  one-way, 309–334
  procedure, 307–308
  randomized complete block design, 334–346
  repeated measures design, 346–356
  two-way, 336–346
Arithmetic mean, 38
Average hazard rate, 760

B
Backward elimination, 564
Bayes’s theorem, 68, 80–83
Bernoulli process, 99–101
  \( \beta_1 \), confidence interval, 438
    hypothesis test, 432–434
Binomial distribution, 99–108
  parameters, 105–107
    table, A-3–A-31
      Use of table, 104–105
Biostatistics, 3
Birth rate, crude, 15–10
Bivariate normal distribution, 445
Bonferroni’s method, 324, 327
Box-and-whisker plot, 50–52

C
Case-fatality ratio, 15–14
Cause-of-death ratio, 15–8
Censored data, 752
Central limit theorem, 139–140
Central tendency, measures, 38–43
Chi-square distribution, 195–197, 600–657
  mathematical properties, 601–604
    table, A-41
      use in goodness-of-fit tests, 604–619
        small expected frequencies, 604
      use in tests of homogeneity, 630–634
        small expected frequencies, 633
      use in tests of independence, 619–630
        small expected frequencies, 625
    2 × 2 table, 625–627
Class interval, 22
Coefficient of determination, 427–428
Coefficient of multiple determination, 501–503
Coefficient of variation, 45–46
Combination, 101
Completely randomized design, 308–334
  ANOVA table, 317
    assumptions, 311
Compound symmetry, 348
Computers:
  and analysis of variance, 308, 321–323, 326–327,
    341–343, 350–351, 355
  and biostatistical analysis, 15–16
  and chi-square, 615, 623
  and descriptive statistics, 21, 22–30, 47
  and hypothesis testing, 232–233, 243–244, 258–259
  and interval estimation, 169–170
  and logistic regression, 573
  and multiple correlation analysis, 512–519
  and multiple regression analysis, 494–496
  and random numbers, 16
  and simple linear regression and correlation
    analysis, 450–451
  and stepwise regression, 560–563
Confidence coefficient, 169
Confidence interval:
for β1, 438
    multiple regression, 506
for difference between two population means, 177–185
    nonnormal populations, 179–180
for difference between two population proportions, 187–188
for mean of Y, given X, 441–442
for μY|X, 508–509
for population mean, 165–178
    nonnormal populations, 168–171
for population proportion, 185–186
    practical interpretation, 167
for predicted Y, 441–442, 508–509
    probabilistic interpretation, 167
for ratio of two variances, 198–201
for ρ, 454
for variance, 194–198
Confusion matrix, 219
Contingency table, 619
Correction for continuity, 152
Correlation coefficient:
    multiple, 510–513
    simple, 446–450
Correlation model:
    multiple, 510–519
    simple, 445–446
Cox regression model, 768–772
    hazard function, 768–769
Critical region, 224
Critical value, 224
Cumulative frequencies, 25
Cumulative relative frequencies, 25

D
Data, 2
    grouped, 22–37
    raw, 20
    sources, 3
Death rate:
    crude, 15-3
    fetal, 15-7
    specific, 15-3
    standardized, 15-3
Death rates and ratios, 15-3 through 15-10
Death ratio, fetal, 15-7
Decision rule, 218
Degrees of freedom, 45
Density function, 115
Descriptive statistics, 2, 19–64
Dispersion, measures, 43–49
Distribution-free procedures, 671
Dummy variable, 544–559

E
Epidemiology, 779
Estimation, 161–210
    in simple linear regression analysis, 434, 441
Estimator, 165
    robust, 170
Events:
    complementary, 74
    independent, 73–74
    mutually exclusive, 68
EXCEL:
    and binomial distribution, 106
Exclusive or, 73
Experiments, 10
    designing, 14–15
Exploratory data analysis, 52
Extrapolation, 442, 459–460

F
Factorial, 101
Factorial experiment, 358–369
    ANOVA table, 364
    assumptions, 362
False negative, 79
False positive, 79
Family-wise error rates, 506
F distribution, 199
    table of, A-42–A-51
Fecundity, 15-10
Fertility, 15-10
    measures, 15-10 through 15-12
Fertility rate:
    age-specific, 15-11
    cumulative, 15-12
    general, 15-10
standardized, 15-12
total, 15-12
Finite population correction, 141
Fisher exact test, 636–640
   table for, A-55–A-85
Fisher’s z, 453–454
   table, A-54
Fixed effects model, 311
F-max test, 198
Forward selection, 563
Frequency distribution, 22–37
Frequency polygon, 27
Friedman test, 712–716
   table for, A-102–A-103
F test, 316–317
G
Goodness-of-fit tests, 604–616
Grouped data, 22–37
H
Histogram, 25–28
Hypothesis, 215
   alternative, 216
   formulating, 14
   null, 216
   research, 215
   statistical, 216
Hypothesis tests, 215–303
   by means of confidence interval, 225–226
   difference between means, 236–249
   nonnormal populations, 242–243
   population variances known, 236–238
   population variances unknown, 238–243
   for \( \beta_i \), multiple regression, 504–506
   for \( \beta_1 \), simple linear regression, 427–432
   one-sided, 226–228
   purpose, 215, 220
   single population mean, 222–236
   non-normal population, 230–232
   population variance known, 222–228
   population variance unknown, 228–230
   single population proportion, 257–259
   single population variance, 264–266
   steps in, 216
   two population proportions, 261–264
   two population variances, 267–272
   two-sided, 226
I
Immaturity ratio, 15-14
Incidence rate, 15-13
Inclusive or, 73
Inferential statistics, 2, 162
Interaction, 359–360, 550
Interpolation, 442
Interquartile range, 48
Interval estimate, 165
Interval scale, 6
J
Joint distribution, 445
K
Kaplan-Meier procedure, 756–761
Kolmogorov-Smirnov test, 698–703
   advantages and disadvantages, 703
   and StatXact computer analysis, 702
   table for, A-99
Kruskal-Wallis test, 704–709
   table for, A-100–A-101
Kurtosis, 48–49
L
Least squares, method, 420
Least-squares line, 420–422
Levene’s test, 201, 270
Location parameters, 47
Log rank test, 763–765
Logistic regression, 569–581
Loss to followup, 751
M
Mann-Whitney test, 690–696
   table for, A-95–A-98
Mantel-Haenszel statistic, 650–653
Margin of error, 168
Mean, 38–40
   properties, 40
Measurement, 6
Measurement scales, 5–6
Median, 40
   properties, 41
Median test, 686–689

MINITAB:
and binomial distribution, 107
and box-and-whisker plots, 51–52
and chi-square, 615–616, 623, 632
and confidence intervals for a mean, 169–170
and descriptive measures, 47
and dummy variables, 546–547, 550, 555
and factorial experiment, 367–368
and frequency distributions, 27
and Friedman test, 716
and histograms, 26–27
and hypothesis testing, 253, 258–259
and Kruskal-Wallis test, 709
and Mann-Whitney test, 694–696
and median test, 689
and multiple correlation, 512, 515, 517
and multiple regression, 495, 508
and normal distribution, 126–127
and one-way ANOVA, 321–322
and ordered array, 20–21
and Poisson distribution, 111–112
and repeated measures ANOVA, 349–350
and sign test, 680
and simple linear regression, 421, 444
and Spearman rank correlation, 724
and stem-and-leaf displays, 29–30
and stepwise regression, 560–563
and two-way ANOVA, 341–342
and Wilcoxon test, 685

Mode, 41

Morbidity, 15-13
measures, 15-13 through 15-14

Mortality rate:
infant, 15-7
maternal, 15-6
neonatal, 15-7
perinatal, 15-7

Multicollinearity, 542
Multiple comparison, 322–326
Multiple correlation coefficient, 510–513
Multiple correlation model, 510–513
Multiplication rule, 71–72
Multivariate distribution, 510
Multivariate normal distribution, 510

N
Nominal scale, 6
Nonparametric statistics, 671–747
Nonrejection region, 218
Normal distribution, 118–127
applications, 122–127
characteristics, 118–119
standard, 118–122
table, A-38–A-39

O
Observation, 14
Observational study, 642–643
Odds ratio, 645–648
Ogive, 96
Operating characteristic curve, 277
Ordered array, 20–21
Ordinal scale, 6
Outliers, 52

P
Paired comparisons, 249–254
Parameter, 38
Partial correlation, 513–519
Partial regression coefficients, 492
Percentile, 47
Point estimate, 163
Poisson distribution, 108–113
table of, A-32–A-37
Poisson process, 109
Population, 5
finite, 5
infinite, 5
sampled, 164
target, 164
Power, 272–279
Precision, 14, 168
Predictive value negative, 80
Predictive value positive, 80
Prospective study, 642
Prediction interval
multiple regression, 508–509
simple linear regression, 441–442
Prevalence rate, 15-14
Probability, 65–85
posterior, 68
prior, 68
classical, 66–67
conditional, 70
joint, 71
marginal, 70, 75
objective, 66–67
personalistic, 67
properties, 68–69
relative frequency, 67
subjective, 67–68
Probability distributions, 92–132
of continuous variables, 113–128
of discrete variables, 93–113
  cumulative, 96–98
  properties, 95
Product-limit method, see Kaplan-Meier procedure
Proportional hazards model, see Cox regression model
Proportional mortality ratio, 15-8
p values, 225

Q
Qualitative variables, 4, 543–556
Quartile, 47–48

R
R
  and box-and-whisker-plots, 51
  and confidence interval between two means, 183
Random digits, table, A-2
  use, 9–10
Randomized complete block design, 334–346
  ANOVA table, 338
  assumptions, 337
Range, 43–44
Rank transformation, 672
Rate, 15-2
Ratio, 15-2
Ratio scale, 6
Regression:
  logistic, 569–581
  multiple, 489–510
    assumptions, 491
    equation, 491–492
    model, 490–492
  nonparametric, 727–730
  resistant line, 442–444
  simple linear, 413–446
    assumptions, 415–416
    equation, 417–423
    model, 414–416
    stepwise, 560–564
Rejection region, 218
Relative frequencies, 24–25
Relative risk, 643–645
Reliability, 14
Reliability coefficient, 167
Repeated measures design, 346–356
  assumptions, 347–356
  definition, 347
Research study, 10
Residual, 429
Resistant line, 442–444
Retrospective study, 643
Risk factor, 642

S
Sample, 5
  convenience, 165
  nonrandom, 164–165
  random, 164–165
  simple random, 7–10
  size for controlling Type II errors, 277–279
  size for estimating means, 189–191
  size for estimating proportions, 191–193
  stratified proportional to size, 13
  stratified random, 12
  stratified systematic, 12
  systematic, 11
Sampling distribution, 135–160
  characteristics, 136
  construction of, 135
  definition, 135
  of difference between sample means, 145–150
    nonnormal populations, 148
  of difference between sample proportions, 154–156
  of sample mean, 136–145
    nonnormal populations, 139–141
  of sample proportion, 150–153
SAS:
and chi-square analysis, 623–625
and descriptive measures, 47
and factorial experiment, 367, 368
and hypothesis testing, 233, 244–245
and logistic regression, 572–576
and multiple regression, 493, 496
and one-way ANOVA, 322
and repeated measures ANOVA, 350–351
and simple linear regression and correlation, 442–443, 450–451
and Tukey’s HSD test, 326
and two-way ANOVA, 341–342

Scatter diagram, 419–420
Scientific methods, 13–15
Secondary attack rate, 15-14
Sensitivity, 80
Significance level, 218–219
Sign test, 673–680
Simple random sampling, 7–10
without replacement, 7–8
with replacement, 7–8
Skewness, 41–42
Slope, 415
Spearman rank correlation coefficient, 718–724
  table for, A-104
Specificity, 80
Sphericity, 348
SPSS:
  and Fisher exact test, 640
  and frequency distribution, 25–26
  and kurtosis, 49
  and logistic regression, 577
  and Mann-Whitney test, 695–696
  and Mantel-Haenzcel test, 652–653
  and multiple regression, 493
  and odds ratio, 648
  and partial correlation, 516, 518–519
  and repeated measures ANOVA, 350–351
  and skewness, 43
  and survival analysis, 665–666
  and Tukey’s HSD test, 327
Standard deviation, 45
Standard error of mean, 139
Standard normal distribution, 118–122
  table of, A-38–A-39
Statistic, 38
Statistical inference, 7, 162
Statistics, 2
Step-and-leaf-display, 28–30
Stepwise regression, 560–569
Studentized range, 324
  table of, A-52–A-54
Student’s distribution, 172–177
  table of, A-40
Sturges’ rule, 22
Survival analysis, 750–776
  censored survival times, 752
    types, 752–753
    Cox regression, hazard function, 768–772
    cumulative distribution function, 754
    Kaplan–Meier procedure, 756–761
    nonparametric technique, 756
    probability of surviving, 756–757
    probability distribution function, 755
    statistical distribution functions, 753
    survival curves, comparing, 763–766
    time-to-event data, 751–756
T
  t distribution, 171–177
    and difference between means, 179–183
    population variances equal, 179–180
    population variances not equal, 180–183
    properties, 172
    table of, A-40
Time-to-event data, see Survival analysis
Test statistic, 217–218
Trimmed mean, 170
Tukey’s HSD test, 323–324
Tukey’s line, 443–444
Type I error, 219
Type II error, 219, 272–279
U
Unbiasedness, 163
Uniform distribution, 614–616
Unit of association, 459
V
Variable, 3
  continuous random, 4
dependent, 415
discrete random, 4
dummy, 544–556
explanatory, 490
extraneous, 307
independent, 415
predictor, 417, 490
qualitative, 4, 543–556
quantitative, 4
random, 4
response, 307, 417
treatment, 307
Variable selection
   procedures, 560–564

Variance, 44–45
   interval estimation, 194–197
Variance ratio, 316
Variance ratio test, 198, 268–272
Vital statistics, 778–796

W
Weibull distribution, 755–756
Wilcoxon test, 681–686
   table for, A-86–A-95

Y
Yates’ correction, 627
   y-intercept, 415