Index

Additive rules of probability, 174–177, 181
  general, 177, 181
  mutually exclusive events, 176, 181
Alpha (level of significance), 367
Alpha (probability of type I error), 367
Alpha (population constant in least-squares line), 503
Alternate hypothesis H_1, 361
  for difference of two means (paired difference), 418
  for difference of two means (independent samples), 431
  for difference of two proportions, 450
  for left-tailed tests, 362, 379
  for right-tailed tests, 362, 379
  for test of correlation coefficient, 135–136, 503–504
  for test of goodness of fit, 484
  for test of independence, 471
  for test of mean, 363
  for test of proportion, 394
  for test of slope of least-squares line, 508
  for test of variance, 497
  for two-tailed tests, 362, 379
And (*A and B*), 169–170, 172, 175, 181. *See also* Probability
Arithmetic mean, 77, 96, 281. *See also* Mean
Averages, 74–80, 96
  mean, 77, 96, 210–211, 232–233, 259
  median, 75, 98
  mode, 74
  population mean μ, 77, 88, 210–211, 232–233, 249, 259, 281, 286, 289, 296, 316, 327, 333, 376, 378, 432, 436
  sample mean $\bar{x}$, 77, 96, 125, 210, 281, 316, 331
  trimmed mean, 78
  weighted, 79

b (slope of least-squares line), 137–138, 141, 503, 508
Back-to-back stem plot, 64
Bar graph, 48–49, 53
Benford's Law, 359–360
Bernoulli, 216
Bernoulli experiment, 216–217
Best-fitting line, 136–138. *See also* Least-squares line
Beta (probability of a type II error), 367
Beta (population coefficient of least-squares equation), 503, 508
Bias, 23
Bimodal histogram, 42
Binomial, 216–217, 230–233, 296
  approximated by normal, 296
  coefficients, 220
  distribution, 218–220, 230–233, 296
  experiment, 216–217
  formula for probabilities, 220
  histogram, 230
  mean of binomial distribution, 232–233
  standard deviation of binomial distribution, 232–233
  variable (*r*), 217, 296, 339, 394, 449
Bootstrap, 334
Boundaries, class, 37, 38
Box-and-whisker plot, 103

CV (coefficient of variation), 90
Categorical data, 5
Cause-and-effect relations, 130–131
Cells, 472
Census, 19
Central Limit Theorem, 288–291
Chebyshev's Theorem, 91–92
Chi-square (χ^2), 470–471, 476–477, 485, 494–495
  calculation of, 474, 477, 484, 487, 494, 497
  degrees of freedom for goodness-of-fit test, 486, 487
  degrees of freedom for independence test, 475, 477
  degrees of freedom for test of a variance, 494, 497
  distribution of, 470–471, 494–495
  test for goodness of fit, 484–487
  test for independence, 471–477
  test for a single variance or standard deviation, 494–497
Circle graph, 50–51, 53
Class, 34, 35
  boundaries, 37
  frequency, 36
  limits, 35
  mark, 36
  midpoint, 36
  width, 35
Cluster sample, 15, 16
Coefficient, binomial, 220
Coefficient of determination, r^2, 145
Coefficient of linear correlation, *r*, 123–126
  formula for, 125, 126
  testing rho (ρ), 135–136, 503–504
Coefficient of variation, *CV*, 90
Combinations rule, 193, 220
Complement of event *A*, 164–165, 181
Conclusions (for hypothesis testing), 371
  using critical regions, 387
  using *P*-values, 368–369
Conditional probability, 170
Confidence interval, 319, 321
  for difference of means, 430–432, 434, 435–436, 441
  for difference of proportions, 454
  for mean, 316–319, 320–321, 331, 333
  for predicted value of response variable, 512
  for proportion, 341
Confidence level, *c*, 316–318
Confidence prediction band, 514–515
Confounding variable, 22
Contingency tables, 179, 472
Continuity correction for normal approximation to binomial, 299
Continuous random variable, 207
Control group, 21, 22
Convenience sample, 16
Correlation
  Pearson product moment correlation coefficient *r*, 123–126
    formula for, 125, 126
    interpretation of, 127
    testing rho (ρ), 135–136, 503–504
Criterion for least-squares equation, 137
Critical regions, 383–384, 387
Critical values, 317, 329, 383
  for Chi-square distribution, 470
  for correlation coefficient, *r*, 135–136
  for normal distribution, 317–318, 383–384, 398, 432, 454
  for *t*, 329, 393, 436

Data
  continuous, 207
  discrete, 206
  paired (dependent samples), 416–417, 429–430
  population, 5
  sample, 5
  qualitative, 5
  quantitative, 5
Decision errors, types of, 366–368
Degrees of freedom (*d.f.*)
  for chi-square goodness-of-fit test, 486, 487
  for chi-square test of independence, 475, 477
  for chi-square test of variance, 494, 497
  for Student's *t* distribution, confidence interval for
    difference of means, 432, 436, 441–448
    mean, 331, 333
    prediction, 512
  for Student's *t* distribution, test of
    correlation coefficient, 503, 504
    difference of means, 432, 436, 441, 447, 448
    mean, 379
    paired difference, 418
    slope, 508
Deming, W.E., 49
de Moivre, 248
Density function, 250
Dependent events, 169–170
Descriptive statistics, 9
Deviation
  population standard, σ, 88, 210–211, 232–233, 249, 259, 281, 286, 289, 296, 319, 327, 376, 432, 494, 497
  sample standard, *s*, 83–85, 96, 281, 327, 331, 379, 436, 494, 497
  computation formula, 85, 96
Difference
  between two means, 430–432, 435–436, 441
  between two proportions, 449–450, 454
  paired difference test, 417–419
Discrete random variable, 206
Disjoint events. *See* Mutually exclusive events
Distribution
  bell-shaped, 249
  bimodal, 42
  binomial, 220, 230–233, 296
  chi-square, 470–471, 476–477, 494–495
  normal, 248–250
  probability, 208, 210–211
  sampling, 281, 284, 286, 288–291
  skewed, 42
  Student's *t*, 328–330, 379, 418, 436, 447, 448, 503–504, 508, 512
  symmetrical, 42, 249
  uniform, 42
Dotplot, 47
Double blind experiment, 22

E, maximal margin of error, 318, 323, 330, 340, 344, 346, 432, 436
  for least-squares prediction, 512
  for difference of proportions, 454
  for difference of means, independent samples, 432, 436
  for mean, 318, 330
  for proportion, 340, 344
EDA, 56–57, 103
Empirical rule, 250–251
Equally likely outcomes, 160–161
Equation of least-squares line, 137–138
Error of estimate. *See* Margin of error
Errors
  type I, 366–368
  type II, 366–368
Estimation, 316–317
  difference of means, 430–432, 434, 435–436, 441
  difference of proportions, 454
  mean, 316–321, 330–331, 333
  predicted value in linear regression, 512
  proportion, 340–341
Event, probability of, 160
Event, 162, 165
  complement of, 164–165, 181
  dependent, 169–170
  equally likely, 160–161
  failure *F*, binomial, 217
  independent, 169–170, 181
  mutually exclusive, 176, 181
  simple, 162, 165
  success *S*, binomial, 217
Expected frequency
  for contingency table, 472–473, 477
  for goodness of fit, 484–485, 487
Expected value, 211
  for binomial distribution, 232–233
  for general discrete probability distribution, 211
Experiment
  binomial, 216–217
  double-blind, 22
  randomized two-treatment, 21
  statistical, 20, 162, 165
Experimental design, 19–22
Explanatory variable in simple regression, 120, 131
Exploratory data analysis, 56–57, 103
Extrapolation, 141, 514

F, failure on a binomial trial, 217. *See also* Binomial
Factorial, 191
Fail to reject null hypothesis, 371
Five-number summary, 103
Frequency, 36
  expected, 472–473, 477, 484–485, 487
  relative, 37, 160
Frequency distribution, 38–42. *See also* Histogram
Frequency histogram, 38–42. *See also* Histogram
Frequency table, 34–37

Gauss, C.F., 248
Gaussian distribution. *See* Normal distribution
General probability rule
  for addition, 177, 181
  for multiplication, 170, 181
Goodness-of-fit test, 484–487
Gosset, W.S., 328
Graphs
  bar, 48–49, 53
  circle, 50–51, 53
  dotplot, 47
  frequency histogram, 38–42
  histogram, 38–42
  Pareto chart, 49, 53
  relative frequency histogram, 38–41
  residual plot, 149–150
  scatter diagram, 120, 154
  stem-and-leaf display, 57–59
  time series graph, 52, 53

Hinge, 104. *See also* Quartile
Histogram, 38–42
  bimodal, 42
  frequency, 38–42
  how to construct, 38–42
  relative frequency, 38–42
  skewed, 42
  symmetric, 42
  uniform, 42
Hypothesis tests, in general, 360–363, 365–369
  alternate hypothesis, H_1, 361
  conclusion, 371
  conclusion based on critical regions, 387

conclusion based on *P*-value, 368–369
critical region, 383–384
critical value, 383
level of significance, 367
null hypothesis, H_0, 361
P-value, 365
Power of a test, 367
Hypothesis testing (types of tests)
of correlation coefficient, 135–136, 503–504, 509
of difference of means, 430–432, 435–436, 441
of difference of proportions, 449–450
of goodness of fit, 484–487
of independence, 471–477
of mean, 376–377, 378–379, 383–385
of paired differences, 417–419
of proportion, 394–395
of variance or standard deviation, 494–497

Independence test, 471–477
Independent events, 169–170, 181
Independent samples, 429
Independent trials, 217
Individual, 5
Inference, statistical, 9
Influential point, 154
Interpolation, 141
Interquartile range, 100–101
Interval, confidence, 316–321, 330–331, 333, 340–341, 430–432, 435–436, 441, 454, 512
Interval level of measurement, 7, 8
Inverse normal distribution, 272

Large samples, 289, 296, 339
Law of large numbers, 162
Leaf, 58, 59
Least-squares criterion, 137
Least-squares line
formula for, 137–138
predictions, 141–142, 512
slope, 137, 138, 141, 508, 509
Level of confidence, *c*, 316–318
Level of significance, 367
Levels of measurement, 6–8, 79
interval, 7, 8
nominal, 7
ordinal, 7, 8
ratio 7, 8
Likert scale, 23
Limits, class, 35, 38
Linear regression, 136–141
Lower class limit, 35, 38
Lurking variable, 22, 23, 130, 131

Margin of error, 316, 344, 355
Marginal change, 141
Maximal error of estimate. *See E*, maximal margin of error
Mean. *See also* Estimation and Hypothesis testing
for binomial distribution, 232–233
comparison with median, 78
defined, 77
discrete probability distribution, 210–211
formula for grouped data, 96
formula for ungrouped data, 77
population, 77, 88, 210–211, 232–233, 249, 259, 281, 286, 289, 296, 316, 327, 333, 376, 378, 432, 436
sample, 77, 96, 125, 210, 281, 316, 331
trimmed, 78
weighted, 79
Median, 75, 98
Midpoint, class, 36
Mode, 74
Mu (μ), population mean, 77, 88, 210–211, 232–233, 249, 259, 281, 286, 289, 296, 316, 327, 333, 376, 378, 432, 436
Multiplication rule of counting, 187
Multiplication rule of probability, 170, 172, 181
for dependent events, 170, 172, 181
for independent events, 170, 172, 181
Mutually exclusive events, 176, 181

N, population size, 77
Negative correlation, 124–125
Nightingale, Florence, 32, 214
Nonresponse, 23
Nominal level of measurement, 7, 8
Normal approximation to binomial, 296
Normal distribution, 248–250, 259, 286, 288–291, 296
areas under normal curve, 259–261, 269
normal curves, 248–250
standard normal, 259
Null hypothesis, H_0, 361, 362. *See also* Alternate hypothesis, H_1
Number of degrees of freedom. *See* Degrees of freedom (*d.f.*)

Observational study, 20
Observed frequency (*O*), 473, 484
Or (*A or B*), 174, 175
Ordinal level of measurement, 7–8
Outlier, 105, 108

p (probability of success in a binomial trial), 217, 281, 339, 394
$\hat{p}$, point estimate of p, 241, 281, 340, 394, 449, 454
$\overline{p}$, pooled estimate of a proportion, 450
P-value, 365, 369
Paired data, 416–417
Paired difference test, 417–419
Parameter, 280–281, 316, 363
Pareto chart, 49, 53
Pearson, Karl, 124
Pearson product moment correlation coefficient r, 123–126
Percentile, 98–100
Permutations rule, 192
Pie chart, 50–51, 53
Placebo, 21
Point estimate
for population mean, 316
for population proportion, 340
for population probability of success, 340
Pooled estimate
of a proportion, $\overline{p}$, 450
of a standard deviation, s, 441, 448
of a variance, s^2, 418
Population
defined, 5, 280
mean μ, 77, 88, 210–211, 232–233, 249, 259, 281, 286, 289, 296, 316, 327, 333, 376, 378, 432, 436
standard deviation σ, 88, 210–211, 232–233, 249, 259, 281, 286, 289, 296, 319, 327, 376, 432, 494, 497
Population parameter, 280–281, 503
Positive correlation, 124–125
Power of a test, 367
Prediction for y given x, 141–142
Probability
addition rule (general events), 177, 181
addition rule (mutually exclusive events), 176, 181
binomial, 220
of the complement of an event, 164–165, 181
conditional, 170
defined, 160, 165
of an event, 160
multiplication rule (general events), 170, 172, 181

multiplication rule (independent events), 170, 172, 181
Probability distribution, 41, 208–211
continuous, 206
discrete, 206, 230
mean, 210–211
standard deviation, 210–211
Proportion, estimate of $\hat{p}$, 281, 340
Proportion, pooled estimate $\bar{p}$, 450
Proportion, test of, 394–395

q, probability of failure in a binomial trial, 217
Qualitative variable, 5
Quantitative variable, 5
Quartile, 100

r, number of successes in a binomial experiment, 217, 296, 339, 394, 450, 454
r, Pearson product moment correlation coefficient, 123–126
r^2, coefficient of determination, 145–146
Random, 12–13
Random number generator, 15, 30–31
Random number table, 13
Random sample, 12–13
Random variable, 206
Randomized two-treatment experiment, 21
Range, 83
Ratio level of measurement, 7, 8
Raw score, 258
Region, rejection or critical, 383–384
Regression, linear, 136–138
Reject null hypothesis, 371
Rejection region. *See* Critical region
Relative frequency, 37, 160
Relative frequency table, 37, 38
Replication, 22
Residual, 149, 505
Residual plot, 149–150
Resistant measures, 78
Response variable in regression, 120, 131
Rho (ρ), 130, 135–136, 503, 504

s, pooled standard deviation, 441, 448
s, sample standard deviation, 83–85, 96, 281, 327, 331, 379, 494, 497
s^2, sample variance, 83, 84, 85, 494, 497
S, success on a binomial trial, 217
Sample, 5
cluster, 15, 16
convenience, 15, 16
large, 289, 296, 339
mean, 77, 96, 232–233, 281, 316, 331
simple random, 12
standard deviation s, 83–85, 96, 281, 327, 331, 379, 494, 497
stratified, 15, 16
systematic, 15, 16
variance s^2, 83, 84, 85, 494, 497
voluntary response, 23
Sample size, determination of,
for estimating a mean, 322–323
for estimating a proportion, 346–347
Samples
independent, 429
repeated with replacement, 15, 30, 217
repeated without replacement, 15, 30, 217
Sample space, 162, 165
Sample test statistic. *See* Test statistic
Sampling, 11–16
cluster, 15, 16
convenience, 15, 16
simple random, 12
stratified, 15, 16
systematic, 15, 16
with replacement, 15, 30, 217
Sampling distribution. *See also* Central Limit Theorem
for proportion, 340
for mean, 280–289
Satterthwaite's formula for degrees of freedom, 436, 448
Scatter diagram, 120
Sigma
σ, 88, 210–211, 232–233, 249, 259, 281, 286, 289, 296, 319, 327, 376, 432, 494, 497
Σ, 77
Significance level, 367
Simple event, 162, 165
Simple random sample, 12
Simulation, 14, 20
Skewed distribution, 42
Slope of least-squares line, 137–138, 141, 503, 508
Standard deviation
for binomial distribution, 233
pooled, 450
for population standard σ, 88, 210–211, 232–233, 249, 259, 281, 286, 289, 296, 319, 327, 376, 432, 494, 497
for sample standard s, 83–85, 96, 281, 327, 331, 379, 436, 494, 497
for distribution of sample proportion, 340
for distribution of sample mean, 286, 288, 288–291
for testing a variance, 494–497
Standard error
of mean, 288
of proportion, 340
of slope, 508
Standard error of estimate, S_e, 506
Standard normal distribution, 259
Standard score, z, 256–257
Standard unit, z, 256–257
Statistic, 280–281, 503
Statistical experiment, 20, 162, 165
Statistical significance, 369
Statistics
definition, 5
descriptive, 9
inferential, 9
Stem, 58, 59, 62
Stem-and-leaf display, 57–59
back-to-back, 64
split stem, 62
Strata, 15
Stratified sampling, 15, 16
Student's t distribution, 328–330, 379, 418, 435, 436, 447, 448, 503–504, 508, 512
Sum of squares, SS, 84
Summation notation, Σ, 77
Survey, 23–24
Symmetrical distribution, 42
Systematic sampling, 15, 16

t (Student's t distribution), 328–330, 379, 418, 435, 436, 447, 448, 503–504, 508, 512
Tally, 36
Tally survey, 179
Test for independence, 471–477
Test of hypotheses. *See* Hypothesis testing
Test statistic, 363
for chi-square goodness-of-fit test, 484, 487
for chi-square test of independence, 474, 477
for chi-square test of variance, 494, 497
for correlation coefficient, rho (ρ), 503, 504, 509
for difference of means
dependent samples, 418, 419
independent samples, 432, 436, 441
for difference of proportions, 450
for mean, 363, 377, 379

for proportion, 395
for slope of least-squares line, 508, 509
Time series, 53
Time series graph, 52, 53
Tree, 188
Trial, binomial, 216–217
Trimmed mean, 78
Two-tailed test, 362–363
Type I error, 366–368
Type II error, 366–368

Uniform distribution, 42
Upper class limit, 35

Variable, 5
continuous, 207
discrete, 206
explanatory, 120, 131
qualitative, 5
quantitative, 5
random, 206
response, 120, 131
standard normal, 259. *See also* z value
Variance, 83
estimate of pooled, 440–441, 448
for ungrouped sample data, s^2, 83, 84, 85
population, 88, 494, 497
sample, s^2, 83, 84, 85, 494, 497
testing, 494–497
Variation
explained, 145
unexplained, 145
Voluntary response sample, 23

Welch approximation, 435
Weighted average, 79
Whisker, 103

x bar ($\bar{x}$), 77, 96, 125, 210, 281, 331, 316. *See also* Mean
z score, 256–257, 286, 291
z value, 256–257, 286, 291