

Index

• Symbols and Numerics •

{ } (braces), 94
! (factorial operator), 123
± (two values), 194
> (greater than), 141
≥ (greater than or equal to), 147
⊂ (subsets), 94
^ (estimated value), 291
| (conditional probabilities), 102, 104
∞ (infinity), 157
< (less than), 141
≤ (less than or equal to), 141
∉ (members not in set), 94
∈ (membership in set), 94
∩ (intersections), mathematical operation, 96–97
∪ (union), mathematical operation, 95–96
* (asterisk) character, 47
^ (caret) character, 47
1- (confidence coefficient), 195
2.71828 (e), 135
25th percentile, 66, 362
50th percentile, 66, 362
75th percentile, 66, 362

• A •

addition rule, probability theory
 formulas, 106–108, 363
 use of, 14
adjusted coefficient of determination
 (adjusted R^2), 313, 314–315
alpha (α) error, 206
alternative hypothesis (H_1)
 formula, 367
 F-test, 316
 hypothesis testing, population variance,
 257–259
 left-tailed test, 204
 overview, 235
 right-tailed test, 203–204
 t-test, 301, 321
 two-tailed test, 205

ANOVA (analysis of variance). *See also*
 F-distribution
 alternative hypothesis, 240
 critical value, finding, 247–248
 degrees of freedom, 247–248
 F-statistic, 246
 F-table, 247–248
 level of significance, 240–241
 null hypothesis, 240
 one-way, 239, 247, 248
 overview, 233, 239
 table, 318
 test statistic, computing, 241–246
 two-way, 239
 uses, 239
arithmetic mean, 40–42
arrangements, 124
asterisk (*) character, 47
autocorrelations, 308, 358, 370
average number of events in time period (λ),
 135, 265

• B •

b (intercept), 20, 38, 285, 286
 b (upper limit of interval), 145
base, rectangle, 144
bell-shaped curve, 151–153
beta (β) error, 206
bias errors, 360
BINOMDIST function, Microsoft Excel 2007, 127
binomial distribution
 expected value, 128
 factorial, 123–124
 formula, 123, 125–126, 364
 graphing on histogram, 129–131
 moments, 127–128
 overview, 16, 122–123
 process, 122
 standard deviation, 128
 uses, 122
 variance, 128
BINOM.IDIST function, Microsoft Excel 2010, 127
braces ({}), 94

• C •

- caret (^) character, 47
- causality errors, 357–358
- center of data set. *See* mean (average); median; mode
- centered moving averages, 343–344, 371
- central limit theorem (CLT), 18, 179–184, 365–366
- change in X (ΔX), 285
- change in Y (ΔY), 285
- chi-square distribution (χ^2). *See also*
 - goodness of fit test; hypothesis testing, population variance
 - applications, 252
 - chi-square table, 261, 264
 - features, 252–253
 - graphic illustration, 253–254
 - moments, 255–256
 - overview, 236, 252
 - for positive values, 252–253
 - positively skewed, 252–253
 - random variable, 255
- chi-square table, 261, 264
- class frequency distribution, 27
- class width, 26–27
- cluster samples, probability sampling, 171–172
- Cochrane-Orcutt procedure, 308
- coefficient of determination (R^2)
 - errors, 356
 - testing population regression equation, 298–299
- coefficient of variation (CV)
 - defined, 69, 72
 - formula, 69–70
- COMBIN function, Microsoft Excel, 125
- combinations
 - formula, 124
 - number of, 125
- complement, mathematical operation, 97–98
- complement rule, probability theory
 - formulas, 108–109, 363
 - as mathematical operation, 97–98
 - purpose, 14, 106
- conditional probabilities (\cap), 102, 104, 105
- confidence coefficient ($1 - \alpha$), 195
- confidence intervals
 - defined, 19, 187, 195, 366
 - errors, statistical analysis, 355
 - with known population standard deviation, 195–199, 366
 - statistical inference, 18–19
 - with unknown population standard deviation, 199–200, 366
- constant (e), 135
- constant (e^x), calculator key, 135
- continuous probability distributions, 16–17, 365. *See also* chi-square distribution; F-distribution; normal distribution; Student's t-distribution; uniform distribution
- continuously compounded interest, 135
- convenience samples, nonprobability sampling, 172
- correlation coefficient
 - versus covariance, 72, 82–85
 - defined, 13, 72
 - diversification, measuring, 88–90
 - errors, 357–358
 - interpreting, 85–86
 - measure of association, 71, 362
 - negative, 72
 - population, 80–82
 - positive, 72
 - sample, 73–77
 - scatter plots, 86–88
 - zero, 72
- covariance
 - versus correlation coefficient, 72, 82–85
 - defined, 13
 - measure of association, 71, 362
 - negative, 72
 - population, 77–80
 - positive, 72
 - sample, 73–77
 - zero, 72
- critical values, hypothesis testing
 - common, standard normal distribution, 213
 - defined, 208
 - degrees of freedom, 210
 - F-test, 319–320
 - large sample, 213
 - left-tailed test overview, 209
 - left-tailed test with small sample, 211–212
 - overview, 260
 - right-tailed test overview, 209
 - right-tailed test with large sample, 214–215
 - right-tailed test with small sample, 209–211
 - small sample, 209–210
 - standard normal table, positive values, 214
 - Student's t-distribution, 210
 - t-test, 304–305, 322–324

two-tailed test overview, 209
 two-tailed test with large sample, 215–216
 two-tailed test with small sample, 212–213
 cumulative frequency distributions, 30–31
 cumulative probabilities, 154
 CV (coefficient of variation)
 defined, 69, 72
 formula, 69–70
 cyclical effects, time series forecasting, 328

• D •

data
 class width, 26–27
 data set, 8, 11–12, 55
 defined, 8
 frequency distribution, cumulative, 30–31
 frequency distribution, qualitative, 29–30
 frequency distribution, quantitative, 25–29
 frequency distribution, relative, 27–29
 graphic analysis, 8–11
 graphing, 31–38
 measures of association, 13
 measures of dispersion, 12–13
 qualitative (non-numerical), 29–30, 24, 337
 quantitative (numerical), 24–27
 spread, 13
 decision rule, hypothesis testing
 left-tailed test and critical value, 216
 right-tailed test and critical value, 216
 two-tailed test and critical value, 216
 degrees of freedom (df)
 ANOVA, 247
 chi-square, 236–237
 denominator, 234, 236, 237, 370
 F-distribution, 234, 370
 numerator, 234, 236, 237, 370
 t-distribution, 189
 denominator degrees of freedom, 234, 236, 237, 370
 denominators, 45
 dependent variable (Y), 20, 283, 370
 discrete distributions. *See also* binomial
 distribution; geometric distribution;
 Poisson distribution
 defined, 121, 363
 expected value formula, 364
 moments, 363–364
 overview, 16
 standard deviation formula, 364

uses, 141
 variance formula, 364
 dispersion measures, 12
 distribution errors, 360
 diversification, measuring, 88–90
 double-counting, 106
 dummy variables, 337–338
 Durbin-Watson test, 308

• E •

e (constant), 135
 elements
 defined, 93
 membership in sets, 94
 empty sets, 97, 100–101
 equivalent standard normal probabilities, 161–164
 error term (ϵ), 22, 290
 errors
 alpha, 206
 beta, 206
 bias, 360
 causality, 357–358
 coefficient of determination, 356
 confidence intervals, 357–358
 correlation as causality, 357–358
 correlation coefficient, 357–358
 distribution, 360
 error sum of square, 242–244
 false positive (Type I), 205–207, 235, 301–302, 317
 false negative (Type II), 205–207, 235, 301–302, 317
 hypothesis test results misinterpretation, 356
 margin of, 18–19, 187, 194–195
 mean square, 245–246, 348–349
 misleading graphs, 353–355
 multicollinearity, 359
 normality assumption, 357
 null hypothesis, interpreting, 356
 overconfidence in forecasts, 359–360
 random, 360
 regression equation, interpreting, 358
 sampling, 242–244
 standard, 178, 180
 wrong conclusion in confidence interval, 355
 wrong distribution, 360
 ESS (explained sum of squares), 298–299
 estimated value, 291

estimates, 194
 estimators, 194
 events
 defined, 99, 112
 independent, 100, 102–103, 105
 intersection, 109–110
 mutually exclusive, 100, 108
 probability, computing, 101, 106–110
 sample space subset, 99
 e^x (constant), calculator key, 135
 $E(X)$ (expected value of X), 118, 145, 237
 Excel, Microsoft, functions in
 BINOMDIST, 127
 BINOM.IDIST, 127
 COMBIN, 125
 EXP, 135
 PERCENTILE, 66
 POISSON, 136
 POISSON.DIST, 136
 QUARTILE, 67
 RANDBETWEEN, 168, 169
 EXP function, Microsoft Excel, 135
 expected frequencies, 266
 expected value
 binomial distribution, 128
 chi-square distribution, 255
 defined, 117
 F random variable, 236–237
 formula, 117–118, 364
 geometric distribution, 133
 histogram, 119
 Poisson distribution, 136
 probability distribution, 117–119, 128
 t-distribution, 189–190
 uniform distribution, 145–146
 expected value of X ($E(X)$), 118, 145, 237
 explained sum of squares (ESS), 298–299
 exponential moving averages, forecasting,
 22, 371
 exponential smoothing, time series,
 345–347, 371
 exponents, 47

• F •

F random variable, 236
 factorial operator (!), 123–124, 268–269
 false positive (Type 1) error
 F-test, 317
 overview, 205–207, 235
 t-test, 301–302
 false negative (Type II) error
 F-test, 317
 overview, 205–207, 235
 t-test, 301–302
 fat tails, 357
 F-distribution
 as continuous probability distribution,
 233–234
 degrees of freedom, 234
 denominator degrees of freedom, 236, 237
 expected value, 237
 F-statistic, 246, 318
 F-table, 247–248
 level of significance, 234–235
 numerator degrees of freedom, 236, 237
 overview, 233–234
 positively skewed distribution, 234
 properties, 234
 rejection region, 248
 spreadsheet, 249–250
 versus t-distribution, 234
 fifth root, 44
 finite outcomes, 141
 finite population correction factor, 178
 first moment, probability distribution, 127
 first quartile (Q_1), 66, 362
 Fisher, Sir Ronald, 234
 forecasting. *See also* trends, forecasting
 models, 327
 overconfidence in, 359–360
 techniques, 21–22, 371
 formulas, statistical
 addition rule, 106–108, 363
 alternative hypothesis, 202–205, 367
 binomial distribution, 123, 125–126, 364
 complement rule, 108–109, 363
 confidence interval for population mean, 366
 critical values, population variances,
 369–370
 expected value, 117–118, 364
 geometric distribution, 131–132, 364
 left-tailed test, 204, 367
 multiplication rule, 109–110, 363
 normal distribution, 365
 null hypothesis, 202, 367
 Poisson distribution, 135, 364
 probability for standard mean, 365–366
 regression equation, 358, 371
 right-tailed test, 204, 367
 standard deviation, 364

- test statistic, equality of two population means, 277, 368–369
- test statistic, population mean, 207–208, 367–368
- test statistic, population variances, 266, 369–370
- two-tailed test, 205, 367
- uniform distribution, 147–150, 365
- variance, 119–120, 364
- fourth quartile (Q_4), 66
- frequencies, 24
- frequency distributions
 - cumulative, 30–31
 - histograms, 31–34
 - key points, 27
 - line graphs, 34–35
 - overview, 24
 - pie charts, 35–36
 - qualitative data, 29–30
 - quantitative data, 25–29
 - relative frequency distribution, 27–29
 - scatter plots, 36–38
- F-statistic
 - calculating, 246, 317–318
 - overview, 317–318
- F-table
 - finding critical values, 247–248, 278, 280
 - multiple regression analysis, 319
- F-test
 - alternative hypothesis, 316
 - comparing p-value with level of significance, 320
 - critical value, 319–320
 - decision, 320
 - level of significance, 317
 - null hypothesis, 316
 - overview, 316
 - test statistic, 317–318
- functions, 112
- statistical analysis, 153
- symmetry, 150
- values, 153
- generalized least squares (GLS) estimators, 294
- geometric distribution
 - expected value, 133
 - formula, 364
 - histogram, graphing, 134
 - mean, 42–43
 - moments, computing, 132–134
 - overview, 131
 - standard deviation, 134
 - variance, 133–134
- geometric probabilities
 - computing, 131–132
 - formula, 131–132
 - overview, 16
- GLS (generalized least squares) estimators, 294
- goodness of fit test
 - comparing population to normal distribution, 271–275
 - comparing population to Poisson distribution, 265–270
 - overview, 251, 264
 - as right-tailed test, 267
- GPA (grade point averages), 17
- graphing
 - histograms, 8–9, 24, 31–34
 - lines, 8–10, 24, 34–35
 - misleading, 353–355
 - pie charts, 8, 10, 24, 35–36
 - rectangle graph, 143–144, 146, 150
 - scatter plots, 8, 10–11, 24, 36–38
 - t-distribution, 190–192
 - types, 8, 38
 - uniform probabilities, 149–150
 - uses, 23
- greater than or equal to (\geq), 147
- greater than ($>$) symbol, 141
- gross return, 43

• G •

- Gauss, Johann Carl Friedrich, 150
- Gaussian distribution. *See also* standard normal distribution
 - formula, 365
 - goodness of fit test, 271–275
 - graphing on bell-shaped curve, 151–153
 - overview, 16–17, 140, 150
 - properties, 150–151

• H •

- H_0 (null hypothesis)
 - ANOVA, 240
 - dependent samples, 230
 - equal variances, 276
 - errors in interpreting, 356
 - F-distribution, 234–235
 - formula, 367

- H_0 (null hypothesis) (*continued*)
- F-test, 316
 - population variance, 256–257
 - single population mean, 202, 221, 235, 257
 - t-test, 301, 321
 - two population means, 221
- H_1 (alternative hypothesis)
- formula, 367
 - F-test, 316
 - hypothesis testing, population variance, 257–259
 - left-tailed test, 204
 - overview, 235
 - right-tailed test, 203–204
 - t-test, 301, 321
 - two-tailed test, 205
- height, rectangle, 144
- heteroscedasticity, 308, 358, 370
- histograms
- benefits, 175
 - binomial distribution, graphing, 129–131
 - defined, 8, 116
 - discrete distributions, 143
 - examples, 31–34
 - geometric distribution, 134
 - overview, 8–9, 24
 - Poisson distribution, 137–138
 - probability distribution, 116
 - sampling distribution, 176–177
- horizontal axis (x), 57, 86
- hypothesis testing, equality of two population variances
- alternative hypothesis, 276–277
 - critical values, 277–279
 - decision, 279–280
 - F-distribution, 275–276
 - level of significance, 277
 - null hypothesis, 276, 368
 - overview, 275–276, 367
 - test statistic, 277, 370
- hypothesis testing, population mean. *See also* ANOVA
- alternative hypothesis, 367
 - degrees of freedom, 210
 - hypothesized value of population, 202, 367
 - left-tailed test, 204, 367
 - level of significance, 367
 - null hypothesis, 367
 - right-tailed test, 203–204, 367
 - Student t-distribution, 210
 - two-tailed test, 205, 367
 - Type I error, 206–207
 - Type II error, 206–207
- hypothesis testing, population variance
- alternative hypothesis, 257–259
 - chi-square table, 261, 264
 - critical value, 260–264
 - formula, weighted average, 267
 - hypothesized value, 256–257
 - left-tailed test, 262
 - level of significance, 259
 - making decision, 263–264
 - null hypothesis, 256–257
 - right-tailed test, 260–262
 - test statistic, 259–260
 - test statistic formula, 266, 369
 - two-tailed test, 262–263
- hypothesis testing, steps
- alternative hypothesis, 202–205, 235, 367
 - critical value, comparing, 208–216
 - decision rule, 216–220
 - level of significance, 205–207, 235, 367
 - null hypothesis, 202, 235, 367
 - overview, 19, 201
 - test statistic, 207–208
- hypothesized value (σ_0^2), 256–257
- hypothesized value of population (μ_0), 202, 367
- 1 •
- i (index)
- expected value, 118
 - population covariance, 78
 - population regression equation, 290
 - sample arithmetic mean, 40
 - sample covariance, 73
 - sample variance, 56
- ln (natural logarithm), 287
- independent events, 100, 102–103, 105
- independent variables (X s), 20, 283, 370
- independent variables in regression equation (k), 323
- infinite outcomes, 141
- infinity (∞), 157
- integers, 65
- intercept (b), 20, 38, 285, 286
- intercept coefficient (β_0) of regression line, 289
- interquartile range (IQR)
- defined, 13, 64, 67
 - as measures of dispersion, 362
 - outliers, 68
- intersections (\cap), mathematical operation, 96–97, 109–110
- interval estimates, 194–195
- intervals, 24

IQR (interquartile range)
 defined, 13, 64, 67
 as measures of dispersion, 362
 outliers, 68
 irregular effects, time series forecasting, 328

• J •

Jarque-Bera test, 357
 Jevons, William Stanley, 357–358
 joint probabilities, 102, 103–104
 judgment samples, nonprobability
 sampling, 173

• K •

k (independent variables in regression
 equation), 323

• L •

Latin letters, 42, 174
 left-tailed test
 as alternative hypothesis, 203
 critical value, 209
 F-distribution, 279
 formula, 367
 interpreting, 203
 with large sample, 213
 population variance, 258, 262
 purpose, 204
 with small sample, 211–212
 t-test, 323
 two population means, 221
 less than ($<$), 141
 less than or equal to (\leq), 141
 level of significance (α)
 confidence intervals, 195
 equality of two population variances, 277
 F-distribution, 234–235
 F-test, 317
 hypothesis testing, population variance, 259
 t-test, 301–302, 321
 line graphs
 defined, 8
 examples, 34–35
 overview, 9–10, 24
 linear relationships
 multiple regression analysis, 310–311
 overview, 284–286
 scatter plots, 286–289
 linear trend, time series, 330–331, 371
 linearly related variables, 72

• M •

MA (moving averages), forecasting, 22,
 341–344, 371
 MAD (mean absolute deviation),
 348–349
 margin of error, 18–19, 187, 194–195
 marginal probabilities, 102–103
 mathematical operations, for sets
 complements, 97–98
 intersections, 96–97
 membership, 94
 overview, 93
 subsets, 94–95
 unions, 95–96
 mean (average)
 arithmetic, 40–42
 defined, 11, 39
 geometric, 42–44
 as measure of central tendency, 361
 relationship with median, 49–53
 weighted arithmetic, 44–46
 weighted geometric, 46–48
 mean, population (μ)
 arithmetic mean, 42
 normal distribution, 153
 as parameter, 174
 population variance, 61
 mean, sample (\bar{X})
 arithmetic mean, 40–41
 overview, 174
 sample correlation, 73
 sample covariance, 73
 sample variance, 56
 sampling distributions, 365–366
 mean absolute deviation (MAD),
 348–349
 mean square error (MSE), 245–246,
 348–349
 measures of association. *See also* correlation
 coefficient; covariance
 defined, 71
 overview, 362
 measures of central tendency. *See* mean
 (average); median; mode
 measures of dispersion, 362. *See also*
 interquartile range; percentiles;
 quartiles; standard deviation; variance
 measures of risk. *See* standard deviation;
 variance
 measures of uncertainty. *See* standard
 deviation; variance

- median
- defined, 11–12, 40, 48
 - as measure of central tendency, 361
 - population, calculating for, 48–49
 - relationship with mean, 49–53
 - sample, calculating for, 48–49
- members not in set (\notin), symbol, 94
- membership, mathematical operation, 94
- membership in set (\in), symbol, 94
- Microsoft Excel functions
- BINOMDIST, 127
 - BINOM.IDIST, 127
 - COMBIN, 125
 - EXP, 135
 - PERCENTILE, 66
 - POISSON, 136
 - POISSON.DIST, 136
 - QUARTILE, 67
 - RANDBETWEEN, 168, 169
- Microsoft website, 127
- midpoint, interval, 145
- mirror images, 152
- mode
- defined, 11–12, 40
 - determining, 53–54
 - features, 53
 - as measure of central tendency, 361
 - uses, 54
- moments
- binomial distribution, 127–128
 - chi-square distribution, 255–256
 - defined, 117, 127, 144, 251
 - discrete distribution, 363–364
 - expected value, 117–119
 - geometric distribution, 132–134
 - overview, 111, 121
 - probability distributions, 117–120
 - sampling distribution, 178
 - t-distribution, 189–190
 - uniform distribution, 144–147
 - variance, 119–120
- moving averages (MA), forecasting, 22, 341–344, 371
- MSE (mean square error), 245–246, 348–349
- MSTSR (treatment mean square), 246
- multicollinearity, 325, 338, 359, 370
- multiple regression analysis. *See also* regression analysis; simple regression analysis
- adjusted coefficient of determination, 313, 314–315
 - F-test, 316–320
 - linear relationship, 310–311
 - multicollinearity, 325, 338, 359, 370
 - overview, 21, 309
 - population regression equation, 21, 311–315
 - predicting value of Y, 313
 - sample regression equation, 21
 - t-test, 320–325
 - variance inflation factor, 325
 - variation, 314
- multiplication rule, probability theory
- formulas, 109–110, 363
 - purpose, 14
- mutually exclusive events, 100, 108
- mutually exclusive sets, 97

• N •

- n
- binomial probabilities, 123
 - expected value, 118
 - population covariance, 78
 - sample arithmetic mean, 40
 - sample covariance, 73
 - sample variance, 56
 - in sampling distribution, 178
- N , in sampling distribution, 178
- natural logarithm (\ln), 287, 310
- nCr function, combinations, 125
- negative correlation, 72
- negative covariance, 72
- negatively related variables, 36
- negatively skewed data set, 49, 51
- no trend, time series regression, 329–330, 371
- nonlinear least squares, 287
- nonprobability sampling
- convenience samples, 172
 - judgment samples, 173
 - overview, 172
 - purposive samples, 173
 - quota samples, 172–173
- normal distribution. *See also* standard normal distribution
- formula, 365
 - goodness of fit test, 271–275
 - graphing on bell-shaped curve, 151–153
 - overview, 16–17, 140, 150
 - properties, 150–151
 - statistical analysis, 153
 - symmetry, 150
 - values, 153
- normality errors, statistical analysis, 357
- not a subset ($\not\subseteq$), symbol, 94
- not in set (\notin), symbol, 94

n-period moving averages, time series, 341–342

null hypothesis (H_0)

- ANOVA, 240
- dependent samples, 230
- equal variances, 276
- errors in interpreting, 356
- F-distribution, 234–235
- formula, 367
- F-test, 316
- population variance, 256–257
- single population mean, 202, 221, 235, 257
- t-test, 301, 321
- two population means, 221

numerator degrees of freedom, 234, 236, 237, 370

numerators, 45

numerical (quantitative) data

- class width, 26–27
- classes, 24
- frequency distribution, 25–26
- overview, 24

numerical measure, 11

• O •

observed frequency (O), 266

one-way ANOVA hypothesis test, 239, 247, 248

operations, mathematical. *See* mathematical operations, for sets

ordinary least squares (OLS) estimators, 294

outliers

- defined, 12, 39
- interquartile range, 68

• p •

p (probability of success), 123, 132

paired samples, 369

parameters. *See also* population mean; population variance

- defined, 174
- Greek letters for, 174
- population standard deviation, 61–64

pdf (probability density function), 365

PERCENTILE function, Microsoft Excel, 66

percentiles

- computing, 65–66
- defined, 12, 64
- formula, 65
- as measures of dispersion, 362
- overview, 64–65

permutations, 125

pie charts

- defined, 8
- overview, 10, 24, 35–36

point estimates, 194–195

point estimators, 194

Poisson, Siméon Denis, 135

Poisson distribution

- expected value, 136
- formula, 364
- goodness of fit test, 265–270
- histogram, graphing, 137–138
- overview, 16, 134
- probabilities formula, 135
- standard deviation, 137
- table, 135–136
- variance, 137

POISSON function, Microsoft Excel 2007, 136

POISSON.DIST function, Microsoft Excel 2010, 136

population

- arithmetic mean, calculating, 42
- defined, 40, 165
- geometric mean, calculating, 42–44
- parameters, 174
- standard deviation, 64
- variance, determining, 61–64
- weighted arithmetic mean, calculating, 44–46
- weighted geometric mean, calculating, 46–48

population correlation coefficient, 78, 80–82

population covariance

- determining, 77–80
- formula, 77

population mean (μ)

- arithmetic mean, 42
- normal distribution, 153
- as parameter, 174
- population variance, 61

population regression equation

- multiple regression analysis, 311–315
- overview, 289–290
- testing, overview, 297–298
- testing, using coefficient of determination, 298–299

population standard deviation (σ), 61–64

population variance (σ^2), 61, 120

positive correlation, 72

positive covariance, 72

positive skewing

- chi-square distribution, 252–253
- data set, 49, 52
- distribution, 234
- F-distribution, 234

- positively related variables, 36
 - probabilities
 - addition rule, 106–108
 - basis, 98
 - complement rule, 108–109
 - distributions, 15–17
 - events, computing, 101
 - multiplication rule, 109–110
 - overview, 13–14
 - t-distribution, 193–194
 - theory, 93
 - types, 102–105
 - of union of two events, 106–108
 - probability density function (pdf), 365
 - probability distributions. *See also*
 - F-distribution; normal distribution; t-distribution
 - binomial, 16, 122–126, 128–131
 - chi-square, 236, 252–256, 261, 264
 - continuous, 141
 - defined, 115, 121
 - discrete, 141
 - expected value, calculating, 128
 - expected value moment, 117–119
 - first moment, 127
 - geometric, 16, 131–134, 365
 - histogram view, 116
 - moments, 117–120
 - Poisson, 16, 134–138, 265–270, 364
 - properties, 114
 - random variables, 114–116
 - sampling, 18, 166, 179–184, 365–366
 - second central moment, 128
 - standard deviation, 128
 - uniform, 16, 139–147, 365
 - variance, 128
 - variance moment, 119–120
 - p (probability of success), 123, 132
 - probability sampling
 - cluster samples, 171–172
 - defined, 167
 - simple random samples, 167–168
 - stratified samples, 170–171
 - systematic samples, 168–169
 - probability theory
 - addition rule, 14, 363
 - basic rules of, 14, 363
 - complement rule, 14, 363
 - computing probabilities, 106–110
 - continuous probability distributions, 16–17
 - discrete probability distributions, 16
 - elements, 93
 - multiplication rule, 14, 363
 - overview, 13–14
 - probability distributions, 15–16
 - probability types, 102–105
 - random experiments, 98–101, 111
 - random variables, 14–15, 111–116
 - sets, 93–98
 - probability values (p-values), 306, 320
 - product being computed, 46
 - products, geometric means, 42
 - purposive samples, nonprobability sampling, 173
 - p-values (probability values), 306, 320
- **Q** •
- Q_1 (first quartile), 66
 - Q_2 (second quartile), 66
 - Q_3 (third quartile), 66, 362
 - Q_4 (fourth quartile), 66
 - Q-Q plots, 357
 - quadratic trend, time series, 331, 371
 - qualitative (non-numerical) data
 - frequency distribution, 29–30
 - overview, 24
 - variable in regression equation, 337
 - quantile, 196
 - quantitative (numerical) data
 - class width, 26–27
 - classes, 24
 - frequency distribution, 25–26
 - overview, 24
 - QUARTILE function, Microsoft Excel, 67
 - quartiles
 - computing, 66–67
 - defined, 12–13, 64, 66
 - as measure of dispersion, 362
 - overview, 66
 - quota samples, nonprobability sampling, 172–173
- **R** •
- R^2 (coefficient of determination)
 - adjusted, 314–315
 - errors, 356
 - testing population regression equation, 298–299
 - RANDBETWEEN function, Microsoft Excel, 168, 169
 - random errors, 360

- random experiments
 events, 99–100
 overview, 14, 111
 probabilities of events, 101
 sample space, 99
- random number generator, 167–168
- random variables (X s)
 chi-square, 255
 defined, 111
 as function, 112
 independent, 20, 283, 370
 overview, 14–15
 probability distribution, assigning, 114–116
 role of, 111–114
- range, measure of dispersion, 56
- rectangle graph, 143–144, 146, 150
- regression analysis. *See also* multiple regression analysis; simple regression analysis
 multicollinearity, 325, 338, 359, 370
 overview, 20–21, 283–284
 regression equations, 358, 371
 spreadsheet programs for, 306–307
 SPSS software, IBM for, 306–307
- rejection region, F-distribution, 248
- relative frequency distribution, 27–29
- relative variation
 coefficient of variation, 69–70
 overview, 68
 relative risks, comparing, 69–70
- residual sum of squares (RSS), 298–300
- residuals, 292
- right-tailed test
 as alternative hypothesis, 203, 276
 critical value, 209
 defined, 203–204
 dependent samples, 230
 F-distribution, 277–278
 formula, 367
 goodness of fit test as, 265, 267
 with large sample, 213–215
 one-way ANOVA hypothesis test as, 248
 population variance, 258, 260–262
 with small sample, 210–211
 t-test, 323
 two means, 221
 two population means, 221
 two population means with equal variances, 224
 unequal variances, 276
- risk. *See also* standard deviation; variance
 reducing via diversification, 88–90
 relative, 69–70
- rounding down, 65
- rounding up, 65
- RSS (residual sum of squares), 298–300
- S •
- s (sample standard deviation), 57–60, 174, 365
- s^2 (sample variance), 56–57, 174, 365
- sample arithmetic mean, 40–41
- sample correlation coefficient, 73–74, 75
- sample covariance, 73–76
- sample mean (\bar{X})
 arithmetic mean, 40–41
 overview, 174
 sample correlation, 73
 sample covariance, 73
 sample variance, 56
 sampling distributions, 365–366
- sample mean for Y (\bar{y}), 73
- sample regression equation, 291
- sample space (S)
 complement, 97–98
 defined, 97, 99, 112
 events, 99–100
- sample standard deviation (s), 57–60, 174, 365
- sample statistics. *See also* sample mean;
 sample standard deviation; sample variance
 defined, 166, 174, 365
 most important, 365
- sample variance (s^2), 56–57, 174, 365
- samples
 arithmetic mean for, calculating, 40–41
 defined, 17, 40, 165
 geometric mean for, calculating, 42–44
 overview, 165–166
 weighted arithmetic mean for, calculating, 44–46
 weighted geometric mean for, calculating, 46–48
- sampling, nonprobability, 172–173
- sampling, probability
 cluster samples, 171–172
 defined, 167
 simple random samples, 167–168
 stratified samples, 170–171
 systematic samples, 168–169

- sampling distributions
 - central limit theorem, 18, 179–184, 365–366
 - computing moments, 180
 - defined, 166, 365
 - finding probability, 180–184
 - graphic illustration, 175–177
 - moments, 178
 - parameters, 174
 - standard error, 178, 180
 - statistics, 174
- sampling errors, 242–244
- sampling techniques, 17–18
- scatter plots
 - defined, 8
 - linear relationships, 286–289
 - overview, 10–11, 24, 36–38, 86
 - showing relationship between two variables, 71, 86
- seasonal variation, time series forecasting
 - dummy variables, 337–338
 - example of, 337–341
 - multicollinearity, 338
 - overview, 337
- second central moment, 128
- second quartile (Q_2), 66, 362
- sets
 - complement, 14, 97–98
 - defined, 14, 93
 - intersections, 14, 96–97
 - members not in, 94
 - membership in, 94
 - subsets, 94–95
 - unions, 95–96
- simple moving averages, 371
- simple random samples, probability sampling, 167–168
- simple regression analysis. *See also* multiple regression analysis; regression analysis
 - assumptions, 307–308, 370
 - autocorrelations, 308, 358, 370
 - Cochrane-Orcutt procedure, 308
 - coefficient of determination, 299–300
 - Durbin-Watson test, 308
 - errors in interpreting regression equation, 358
 - heteroscedasticity, 308, 358, 370
 - linear relationships, 284–286
 - population regression equation, estimating, 291–297
 - population regression equation, overview, 20, 289–290
 - regression equation formula, 293
 - regression equation intercept formula, 294
 - regression equation slope formula, 294
 - sample regression equation, 20, 291
 - software for, 306–307
 - White test, 308
- simulation studies, 143
- skewness of data distribution
 - negative, 49, 51
 - positive, 49, 52
- slope coefficient (β_1) of regression line, 289
- slope of line (m), 20, 38, 285, 286
- smoothing constant, time series, 345
- smoothing techniques, time series
 - centered moving average, 343–344
 - exponential smoothing, 345–347
 - moving averages, 341–343
 - overview, 341
- Solve My Math website, 141
- spreadsheets
 - adjusted coefficient of determination, 315
 - ANOVA hypothesis, 249–250
 - population regression equation, 312
 - for regression analysis, 306–307
- SPSS software, IBM, 306–307
- squared units, 120
- SSE (error sum of squares)
 - calculating, 242–244
 - constructing test statistic, 242
- SST (total sum of squares)
 - calculating, 245
 - constructing test statistic, 242
- SSTR (treatment sum of squares)
 - calculating, 244–245
 - constructing test statistic, 242
 - overview, 244
- standard deviation (σ)
 - binomial distribution, 128
 - chi-square distribution, 256
 - defined, 12, 120
 - formula, 364
 - geometric distribution, 134
 - as measures of dispersion, 362
 - overview, 56
 - Poisson distribution, 137
 - population, finding, 62–64
 - population, formula, 64
 - probability distribution, 120, 128
 - sample, formula, 57–61
 - of sampling distribution, 178
 - standard error, 178
 - t-distribution, 190
 - uniform distribution, 146–147
 - versus variance, 61

- standard error
 - defined, 178, 366
 - sampling distribution, 178
- standard error of the estimate (SEE), 302
- standard error of the regression (SER), 302–304
- standard normal distribution. *See also* normal distribution
 - certainty property, 157
 - overview, 154
 - properties, 157–158
 - symmetry property, 157–158
 - versus t-distribution, 188–189
- standard normal probabilities
 - Central Limit Theorem tables, 179
 - computing if Z is greater than or equal to specified value, 159
 - computing if Z is less than or equal to specified value, 155–158
 - computing in between, 160–161
 - computing overview, 154
 - properties, 157–158
- standard normal random variable (Z), 154, 272–273
- standard normal tables
 - computing greater than or equal to standard normal probabilities, 159
 - computing in between standard normal probabilities, 160
 - computing less than or equal to standard normal probabilities, 155–158
 - estimating confidence intervals, 196, 272, 273
 - finding probabilities, 184
 - negative Z values, 272–273
 - overview, 154
 - positive Z values, 273–274
- standard uniform probability distribution
 - defined, 143
 - formula, 149–150
 - uses, 143
- statistical analysis, 153
- statistical formulas
 - addition rule, 106–108, 363
 - alternative hypothesis, 202–205, 367
 - binomial distribution, 123, 125–126, 364
 - complement rule, 108–109, 363
 - confidence interval for population mean, 366
 - critical values, population variances, 369–370
 - expected value, 117–118, 364
 - geometric distribution, 131–132, 364
 - left-tailed test, 204, 367
 - multiplication rule, 109–110, 363
 - normal distribution, 365
 - null hypothesis, 202, 367
 - Poisson distribution, 135, 364
 - probability for standard mean, 365–366
 - regression equation, 358, 371
 - right-tailed test, 204, 367
 - standard deviation, 364
 - test statistic, equality of two population means, 277, 368–369
 - test statistic, population mean, 207–208, 367–368
 - test statistic, population variances, 266, 369–370
 - two-tailed test, 205, 367
 - uniform distribution, 147–150, 365
 - variance, 119–120, 364
- statistical inference
 - confidence intervals, 18–19
 - defined, 18, 166
 - hypothesis testing, 19
 - statistics, 18, 174
 - Statistics How To website, 127
 - strata, 170
 - stratified samples, probability sampling, 170–171
- Student's t-distribution
 - confidence interval for population mean, 195
 - degrees of freedom, 189
 - versus F-distribution, 234
 - graphing, 190–192
 - hypothesis testing, 210
 - interval estimates for known population
 - standard deviation, 195–199
 - interval estimates for unknown population
 - standard deviation, 199–200
 - median, 189
 - moments, 189–190
 - point estimates, 194–195
 - point estimators, 194
 - probabilities, 193–194
 - properties, 188
 - purpose, 188
 - standard deviation, 190
 - versus standard normal distribution, 188–189
 - t-table, 193–194
 - variance, 189–190
- Student's t-table, 193–194, 211, 323
- subsets (\subset), 94
- summary measures. *See also* moments
 - measures of association, 362
 - measures of central tendency, 361
 - measures of dispersion, 362
 - overview, 361
- summation operator (Σ), 40–42, 73, 117, 118

sums, arithmetic means, 42
 symmetrical data set, 49, 50–51, 150
 systematic samples, probability sampling,
 168–169

• T •

tables. *See also* standard normal tables

ANOVA, 318
 central limit theorem, 179
 chi-square, 261, 264
 coefficient of determination, 300
 critical values, standard normal
 distribution, 213
 F-table, 247–248, 278, 280, 319
 Poisson, 135–136
 standard error of β_1 , 303
 Student's t-distribution, 211, 305
 t-table, 193–194
 tails (distribution), 51, 152
 t-distribution
 confidence interval for population mean, 195
 degrees of freedom, 189
 versus F-distribution, 234
 graphing, 190–192
 hypothesis testing, 210
 interval estimates for known population
 standard deviation, 195–199
 interval estimates for unknown population
 standard deviation, 199–200
 median, 189
 moments, 189–190
 point estimates, 194–195
 point estimators, 194
 probabilities, 193–194
 properties, 188
 purpose, 188
 standard deviation, 190
 versus standard normal distribution,
 188–189
 t-table, 193–194, 211, 323
 variance, 189–190
 test statistic
 defined, 302
 formula, 207–208
 F-test, 317–318
 t-test, 302–304, 321
 tests. *See also* goodness of fit test; left-tailed
 test; right-tailed test; t-test; two-tailed
 test
 Durbin-Watson, 308
 F-test, 316–320
 Jarque-Bera, 357

one-way ANOVA, 239, 247, 248
 White, 308
 Texas Instruments calculators, 127
 third quartile (Q_3), 66, 362
 TI-83 calculator, Texas Instruments, 127
 TI-84 calculator, Texas Instruments, 127
 time series regression analysis
 defined, 328
 exponential smoothing, 345–347
 forecasting, 21–22, 336–337
 forecasts, comparing, 348–349
 possible effects, 328
 regression model, 328–329
 seasonal variation, 337–341
 smoothing techniques, 341–344
 trends, classifying, 329–332, 371
 trends, estimating, 332–336
 total sum of squares (TSS)
 calculating, 245
 computing, 299–300
 constructing test statistic, 242
 treatment mean square (MSTR), 246
 treatment sum of squares (SSTR)
 calculating, 244–245
 constructing test statistic, 242
 overview, 244
 trend effects, time series forecasting, 328
 trend lines, 37, 86, 287–288
 trend models forecasting, 22
 trends, forecasting
 estimating, 332–336
 higher-order, 371
 linear, 330–331, 335, 371
 no trend, 329–330, 371
 other possibilities, 332
 quadratic, 331, 336, 371
 TSS (total sum of squares)
 calculating, 245
 computing, 299–300
 constructing test statistic, 242
 t-statistic, 302, 307
 t-table, 193–194
 t-test
 comparing p-value with level of significance,
 324–325
 critical values, 304–305, 322–324
 decision, 324
 decision rule, 305–306
 level of significance, 301–302, 321
 null hypothesis, 301, 321
 overview, 301
 Student's t-distribution table, 305
 test statistic, 302–304, 321
 two values (\pm), 194

two-tailed test
 as alternative hypothesis, 203, 277
 dependent samples, 230
 F-distribution, 279
 formula, 367
 with large sample, 215–216
 overview, 203
 population variance, 257, 258, 259, 262–263
 with small sample, 212–213
 t-test, 301, 323
 two critical values, 209
 two means, 221
 two population means, 221
 two population means with equal variance, 224
 unequal variances, 276–277
 two-way ANOVA, 239
 Type I (false positive) error
 F-test, 317
 overview, 205–207, 235
 t-test, 301–302
 Type II (false negative error)
 F-test, 317
 overview, 205–207, 235
 t-test, 301–302

• U •

unconditional (marginal) probabilities, 102–103
 unexplained variation (RSS), 298–299
 uniform distribution
 defined, 142
 expected value, 145–146
 formula, 365
 moments, 143–144
 overview, 16, 139–140, 142–143
 probability density function, 365
 rectangle illustration, 143–144
 standard deviation, 146–147
 standard uniform distribution, 143
 variance, 146
 uniform probabilities
 computing probability that X is between two constants, 148
 computing probability that X is greater than or equal to x , 147
 computing probability that X is less than or equal to x , 147
 computing with formulas, 147–149
 computing with graphs, 149–150
 union (\cup), mathematical operation, 95–96
 union of two events, probability of, 106–108

unique values, 95
 universal set (S)
 complement, 97–98
 defined, 97, 99, 112
 events, 99–100
 unrelated variables, 36

• V •

variables. *See specific variables by name or type*
 variance
 binomial distribution, 128
 chi-square distribution, 255
 defined, 12, 120
 formula, 119–120, 364
 geometric distribution, 133–134
 as measures of dispersion, 362
 overview, 55–56
 Poisson distribution, 137
 population, determining for, 62–64
 population, formula, 61–62
 probability distribution, 119–120, 128
 sample, formula, 56–57
 sampling distribution, 178
 versus standard deviation, 61
 t-distribution, 189–190
 uniform distribution, 146
 variance analysis (ANOVA). *See also*
 F-distribution
 alternative hypothesis, 240
 critical value, finding, 247–248
 degrees of freedom, 247–248
 F-statistic, 246
 F-table, 247–248
 level of significance, 240–241
 null hypothesis, 240
 one-way, 239, 247, 248
 overview, 233, 239
 table, 318
 test statistic, computing, 241–246
 two-way, 239
 uses, 239
 variance inflation factor (VIF), 325, 359
 variation, 298–299
 Venn diagram, 94–95, 96, 97, 98
 vertical axis (y), 57, 86

• W •

weighted arithmetic mean, 44–46, 268–269
 weighted geometric mean, 46–48
 weighted least squares (WLS) estimators, 294

White test, 308
width, interval, 144
Wikipedia website, 143

• \bar{X} •

\bar{X} (sample mean)
arithmetic mean, 40–41
overview, 174
sample correlation, 73
sample covariance, 73
sample variance, 56
sampling distributions, 365–366
 χ^2 (chi-square distribution). *See also*
goodness of fit test; hypothesis testing,
population variance
applications, 252
chi-square table, 261, 264
features, 252–253
graphic illustration, 253–254
moments, 255–256
overview, 236, 252
for positive values, 252–253

positively skewed, 252–253
random variable, 255
x-axis, 57, 86
Xs (random variables)
chi-square, 255
defined, 111
as function, 112
independent, 20, 283, 370
overview, 14–15
probability distribution. assigning, 114–116
role of, 111–114

• ψ •

y-axis, 57, 86

• Z •

Z (standard normal random variable), 154
zero correlation, 72
zero covariance, 72