

Index

Symbols and Numerics

- > (greater than), 146
- $\geq$ (greater than or equal to), 146
- 10th percentile (bottom 10%), 161
- 68-95-99.7% rule, 57, 85–87, 116
- 90th percentile (top 10%), 161

A

- ACT scores, comparing, 91–93
- alpha level, 231
- alternative hypothesis, 65, 225–226
- analyses
 - confidence interval, 18–19
 - correlation, 20–21
 - data from experiments, 289
 - data in surveys, 271–272
 - hypothesis tests, 19–20
 - margin of error, 18–19
 - overview, 17–18
 - regression, 20–21
 - two-way tables, 20–21
- anecdotes, 43, 346–347
- anonymous respondents, 269
- arithmetic errors, 36–37
- average (mean)
 - binomials, 146–147
 - calculating, 75–77
 - histograms, 113–114
 - median compared with, 78–80
 - normal distribution, 149
 - overview, 55–56
 - population, 55–56
 - random variables, 138–139
 - sample, 55–56
 - sampling distribution, 174

B

- bad sample, 53
- bar graphs
 - described, 101
 - evaluating, 105
 - frequency, 101
 - horizontal, 106
 - misleading, 336–337, 338
 - overview, 101
 - relative frequency, 101
 - scale of, 104–105
 - for state lottery profits, 103
 - tips for, 106
 - for tracking transportation expenses, 101–103
 - vertical, 106
 - for workplace pet peeves, 105–106
- bell-shaped curve, 58–59
- best-fitting line
 - calculating, 302–304
 - described, 302
 - example, 306
 - interpreting, 304–305
 - linear regression, 302–304
 - slope of, 302–305
 - y-intercept, 302, 304–305
- bias
 - in data, 2884
 - described, 53
 - in experiments, 14
 - overview, 55
 - sources of, 339–340
 - in surveys, 14
- big picture information determined with *t*-table, 170
- bimodal data set, 117

- binomial distribution
 - described, 17
 - mean, 146–147
 - standard deviation, 146–147
 - tables, 370–374
 - variance, 146–147
- binomials
 - conditions, checking step by step, 140
 - formula, finding binomial probabilities using a, 142–144
 - mean, 146–147
 - multiple outcomes, situations having, 141
 - normal approximation, 162–163
 - overview, 139
 - probability of success changes, 141
 - specific values of X , finding probabilities for, 145
 - standard deviation, 146–147
 - table, finding binomial probabilities using a binomial, 144–146
 - trials, no fixed number of, 140–141
 - trials, not independent, 141
 - variance, 146–147
 - X greater-than, less-than, or between two values, finding probabilities for, 146
- bivariate data, 295
- blind experiments, 61–62, 289
- borderline values in histograms, 110
- boxplots
 - creating, 120–121
 - described, 120
 - errors in interpreting, 126
 - horizontal, 121
 - interpreting, 121–126
 - median in, 124
 - for Old Faithful's eruptions, 124–125
 - outliers in, 125–126
 - skewed left, 121–122
 - skewed right, 121–122
 - symmetric, 121–122
 - variability in, 123–124
 - vertical, 121
- breaking down problems as exam strategy, 361–362
- business section of newspaper, statistics in, 28–29

C

- calculating
 - conditional distributions, 322–324
 - correlation, 297–298
 - joint distributions, 319–320
 - marginal distributions, 316–317
 - p -values, 230–231
 - regression line, 302–304
 - test statistic, 228
- Car Talk* (radio show), 259
- carrying out surveys, 268–271
- categorical data (qualitative data). *See also* graphing categorical data
 - calculating, 72–75
 - described, 95
 - summaries for, 14–16, 15
- categorical variables, 311
- causation, 67–68
- cause-and-effect relationship
 - correlation compared, 08–309
 - defined, 277
 - described, 21–22
 - determining, 329–330
 - as factor in misleading statistics, 43
 - questioning claims of, 21–22
- cells in two-way tables, setting up, 313
- census, 54
- center of data, 113
- Central Limit Theorem (CLT), 59–60, 178–181
- charts and graphs. *See also specific charts and graphs*
 - misleading statistics, 37–44
 - overview, 15–16
 - in summaries, 15–16
- children's healthcare, surveys about, 261
- clinical trials, 286
- collecting data for surveys, 268–269
- comparisons
 - defining experiment to make, 279–281
 - of two population averages in hypothesis tests, 245–247
 - of two population proportions in hypothesis tests, 251–253

- conclusions
 - avoiding jumping to, 332
 - cause and effect, questioning claims of, 21–22
 - drawing, 272–273
 - overstated results in, 21
 - overview, 21
- conditional distributions
 - calculating, 322–324
 - graphing, 322–324
 - overview, 321–322
- confidence intervals
 - choosing confidence level for, 206–207
 - estimating difference of two proportions, 219–221
 - interpreting, 204–205
 - misleading, 221
 - overview, 63–64, 203–204
 - for population mean, 210–212, 216–219
 - population proportion, 214–216
 - population variability, 209–210
 - sample size, 208–209, 213–214
 - t*-table, 169–170
 - width of, 205–206
- confidence levels
 - choosing, 206–207
 - described, 203–204
 - margin of error (MOE), 197
 - z*-value for, 207
- confidential respondents, 269
- confounding variables, 284–286, 344
- consumers of information, 36
- continuous data, 50
- continuous random variables, 136–137, 149.
 - See also* normal distribution
- control group, 61, 279
- correlation
 - calculating, 297–298
 - cause and effect relationship compared, 308–309
 - described, 20–21, 67–68, 297
 - detecting misinterpreted, 343
 - interpreting, 298–300
 - properties, 300

- correlation coefficient, 297
- countably infinite values, 50, 137
- counts, 107
- cover letter for surveys, 265
- crime rates, 37–39
- crime statistics, 37–39
- critical value, 207
- customer satisfaction statistics, 32–33
- cutoff points, 231
- cynical attitude about statistics, 22

D

- daily life statistics, 9–11
- data
 - bias in, 288
 - categorical, 50
 - continuous, 50
 - discrete, 50
 - numerical, 50
 - ordinal, 51
 - overview, 50–51
 - qualitative, 50
 - quantitative, 50
 - reliability of, 287–288
 - validity of, 28
 - yes/no, 50
- data collection
 - experiments, 287–289
 - hypothesis test, 226
 - overview, 12
 - random sample selection, 13
 - sample selection, 13
- data fishing, avoiding, 345–346
- data set, 51
- decisions based on results of experiments,
 - making informed, 292
- degrees of freedom (df), 166, 168
- dependent variables, 301, 324, 327–328
- descriptive statistics
 - categorical data, 72–75
 - empirical rule, 85–87
 - five-number summary, 93–94

- descriptive statistics (*continued*)
 - histograms, 78–80
 - interquartile range, 94
 - mean, 75–77
 - median, 77–78
 - overview, 14–15, 71–72
 - percentiles, 87–94
 - range, 84
 - standard deviation, 81–84
 - in summaries, 14–15
- designing studies to answer a research question
 - with experiments, 12
 - overview, 11
 - with surveys, 11–12
- detecting errors, 36–37
- discrete data, 50
- discrete random variables, 136–137, 149. *See also* binomials
- distribution. *See also specific distributions*
 - conditional, 322–324
 - described, 315
 - joint, 317–321
 - marginal, 315–317
 - overview, 16–17
 - two-way tables, 315–324
- double-blind experiments, 61–62, 288
- double-checking your work as exam strategy, 362–363

E

- empirical rule (68-95-99.7), 57, 85–87, 116
- erroneous data, removing, 44
- errors
 - in arithmetic, 36–37
 - described, 36, 174, 233–234
 - detecting, 36–37
 - erroneous data, removing, 44
 - in interpretation of survey results, 272
 - in interpreting boxplots, 126
 - missing data as factor in, 44
 - of omission, 36
 - reasonableness of projections, examining, 37

- type-1 errors (false alarms), 234
- type-2 errors (missed detection), 234–235
- estimates, 33, 202. *See also* confidence intervals
- ethical issues
 - experiments, 286–287
 - surveys, 265
- ethical review boards (ERBs), 265, 287
- evaluating
 - bar graphs, 105
 - pie charts, 100
 - time charts, 131
- exaggerations. *See* misleading statistics
- exam strategies
 - breaking down problems, 31–362
 - double-checking your work, 362–363
 - formulas, being comfortable using, 354–355
 - “if-then-how” chart, making an, 355–357
 - “know what you don’t know, and then do something about it,” 350
 - labeling everything in problem, 358–360
 - overview, 349–350
 - picture form, expressing your exam question in, 360–361
 - problems, understanding what the question is asking in test, 357–358
 - real exam conditions, practicing under, 353
 - sense, analyzing your answers to confirm they make, 363
 - “yeah-yeah” traps, avoiding, 351–354
 - on your own, trying problems, 353
- examples
 - media and statistics, 24–31
 - misleading graphs, 338
 - normal distribution, 151
 - regression line, 306
- excess data in time charts, 129–131
- experiments
 - analyzing data from, 289
 - bias in, minimizing, 14
 - blind, 61–62
 - comparisons, defining the experiment to make, 279–281

- confounding variables, 284–286
- control group, 61, 279
- data collection for, 287–289
- decisions based on results of, making informed, 292
- defined, 276
- double-blind, 61–62
- ethical issues, 286–287
- fake treatment, 280
- no treatment, 281
- overview, 12, 278
- placebo, 61
- placebo effect, 280
- random assignments of subjects to treatments, 283–284
- sample size, selecting, 281–283
- standard treatment, 281
- subjects, choosing, 283
- treatment group, 61, 252
- volunteers, finding, 284
- explanatory variable, 301
- extrapolation, 307

F

- factor, 276
- fake treatment, 280
- false alarms, 234
- fantasy sports, 28–29
- fatality rate, 25
- finite values, 50
- five-number summary, 93–94
- flat histograms, 114
- following up surveys, 270–271
- formulas
 - being comfortable using, 354–355
 - for finding binomial probabilities, 142–144
 - for Z-distribution (standard normal distribution), 153
- frequency
 - in bar graphs, 101
 - described, 15
 - in histograms, 108

G

- Gallup Organization, 190, 267
- “garbage in equals garbage out,” 13
- generalizing research results, 291
- grand total, 314
- graphing. *See also* misleading graphs
 - conditional distributions, 322–324
 - joint distributions, 320–321
 - marginal distributions, 317
- graphing categorical data
 - bar graphs, 101–106
 - pie charts, 96–101
- graphing numerical data
 - boxplots, 120–126
 - histograms, 108–119
 - time charts, 128–131
- groupings in histograms, 110, 117–118
- guidelines for p -values, 232–233

H

- high percentiles, 160–161
- histograms
 - borderline values in, 110
 - center of data, 113
 - creating, 108–111
 - described, 108
 - flat, 114
 - frequency, 108
 - groupings, 110, 117–118
 - interpreting, 111–115
 - mean (average), 113–114
 - median, 114
 - misleading, 117–119
 - misleading graphs, 339
 - outliers in, 114
 - relative frequency, 108
 - scale, 119
 - setting up, 108–111
 - shape of data, 111–113
 - showing ages of best actress Oscar winners, 108–111

- histograms (*continued*)
 - skewed left, 112, 116
 - skewed right, 111, 115
 - spread of data, 114–115
 - symmetric, 111, 116
 - time charts compared, 128
 - tips for, 115–116
 - variability in, 114–115
 - x-axis in, 110–111
 - y-axis in, 110–111, 119
- horizontal bar graphs, 106
- horizontal boxplots, 121
- horoscopes, statistics in, 31
- household incomes, comparing, 90–91
- hypotheses, 237–238
- hypothesis tests
 - alternative hypothesis, 225–226
 - comparing two population averages, 245–247
 - comparing two population proportions, 251–253
 - data, collecting, 226
 - described, 19–20, 64–65, 223–224
 - null hypothesis, 224–225
 - paired *t*-test, 247–250
 - population mean, 238–240
 - sample statistics, gathering, 227
 - setting up, 224–226
 - standard errors, measuring variability with, 227
 - standard scores, 228
 - test statistic, 228
 - testing for an average difference, 247–250
 - testing one population proportion, 243–244
 - t*-test, 240–243

I

- icons, explained, 6
- “if-then-how” chart as exam strategy, 355–357
- impact
 - of misleading statistics, 44–45
 - of surveys, 258–262
- implications of research results, 290–291
- incorrect numbers, looking for, 344–345

- independent ethics committees (IECs), 265, 287
- independent variables, 301, 324–327
- inflection point, 58, 150
- institutional review boards (IRBs), 265, 287
- Internet virus, statistics about, 24–25
- interpretation
 - boxplots, 121–126
 - cause-and-effect relationship, determining, 329–330
 - conclusions, avoiding jumping to, 332
 - confidence intervals, 204–205
 - correlation, 298–300
 - histograms, 111–115
 - predictions, making, 331
 - projecting from sample to population, 330–331
 - regression line, 304–305
 - scatterplot, 296
 - surveys, 271–273
 - test statistic, 228
 - time charts, 127–128
 - two-way tables, 315–324
- interquartile range (IQR), 84, 94, 123–124

J

- jargon, 49–68
- joint distributions
 - calculating, 319–320
 - graphing, 320–321
 - overview, 317–319

K

- “know what you don’t know, and then do something about it” as exam strategy, 350
- k*th percentile
 - calculating, 8–8985
 - described, 88

L

- labeling everything in problem as exam strategy, 358–360
- land development, statistics about, 26

leading questions, 265
level, 277
lies, 44
line graphs. *See also* time charts
 conditions for, 302
 explanatory variable, 301
 overview, 301
 regression line, 301, 302–304
 response variable, 301
lotto revenue, pie charts of, 97–98
low percentiles, 159–160

M

margin of error (MOE). *See also* confidence intervals
 checking for, 340–341
 confidence levels, 197
 finding, 193–194
 overview, 62–63, 189
 results, reporting, 194–195
 sample mean, 195–197
 sample size, 197–200
 sample variability, measuring, 191–193
 standard errors, 191–193
marginal column totals, 314
marginal distributions
 calculating, 316–317
 graphing, 317
 overview, 315–317
marginal row totals, 314
marginal totals, 314
mean (average)
 binomials, 146–147
 calculating, 75–77
 histograms, 113–114
 median compared with, 7–80
 normal distribution, 149
 overview, 55–56
 population, 55–56
 random variables, 138–139
 sample, 55–56
 sampling distribution, 174
 measurements, 107
media, statistics in
 business section, 28–29
 examples of, 24–31
 horoscopes, 31
 Internet virus, 24–25
 land development, 26
 medical malpractice insurance, 26
 microwave popcorn and cancer, 24
 motorcycle fatalities, 25–26
 movies, 30–31
 overview, 23–24
 school proficiency, 27–28
 sexual activity, 29–30
 sports section, 28
 surveys and polls, 27
 travel section, 29
 weather report, 30
media sources, 35
median
 boxplots, 124
 calculating, 77–78
 histograms, 114
 mean compared with, 78–80
 overview, 56
medical malpractice insurance, statistics about, 26
medical studies, 31–32
microwave popcorn and cancer, statistics about, 24
misleading graphs
 bar graphs, 336–337, 338
 confidence intervals, 221
 examples, 338
 histograms, 117–119, 339
 overview, 335–336
 pie charts, 336, 338
 time charts, 128–131, 337–339
misleading questions, 265
misleading statistics
 anecdotes as factor in, 43
 cause-and-effect relationships as factor in, 43
 crime statistics, 37–39
 impact of, 44–45

misleading statistics (*continued*)

- overview, 37
 - sample size as factor in, 42–43
 - sources, checking, 42
 - tornado statistics, 39
 - uncovering, 37–44
 - wording of questions as factor in, 43–44
- missed detection, 234–235
- missing data as factor in errors, 44
- mistake recognition
- anecdotes, avoiding influence of, 346–347
 - biased data, sources of, 339–340
 - confounding variables, revealing, 344
 - correlations, detecting misinterpreted, 343
 - data fishing, avoiding, 345–346
 - graphs, avoiding misleading, 335–339
 - incorrect numbers, looking for, 344–345
 - margin of error, checking for, 340–341
 - random samples, checking for, 341–342
 - sample size, checking, 342
- motorcycle fatalities, statistics about, 25–26
- movies, statistics about, 30–31

N

- negative t -values, 242
- 90th percentile (top 10%), 161
- no treatment in experiments, 281
- non-random sample, 53
- normal approximation, 162–163
- normal distribution
- examples of, 151
 - finding probabilities for, 156–158
 - mean, 149
 - overview, 58–59, 149–152
 - percentiles for, finding, 159–161
 - properties, 150
 - standard deviation, 149
 - Z-distribution (standard normal distribution), 152–155
- null hypothesis, 65, 224–225
- numerical data. *See also* graphing
- described, 75
 - summaries for, 15

O

- observational studies
- defined, 276
 - overview, 277–278
- observational study, 11
- Old Faithful's eruptions, boxplots for, 124–125
- omission errors, 36
- on your own, trying problems, 353
- ordering takeout, pie charts for, 9–101
- ordinal data, 51
- organizing data in surveys, 271–272
- original sources, 35
- outliers
- boxplots, 125–126
 - described, 56, 307
 - histograms, 114
- overstated results, 21, 290

P

- paired t -tests, 65, 247–250
- parameters
- described, 54, 201, 203
 - statistics linked to, 203
- percentiles
- ACT scores, comparing, 91–93
 - calculating, 88–89
 - described, 87–88
 - high, 160–161
 - household incomes, comparing, 90–91
 - interpreting, 89–93
 - k th percentile, 88–89
 - low, 159–160
 - 90th percentile (top 10%), 161
 - overview, 57
 - 10th percentile (bottom 10%), 161
 - t -table, 168–169
- percentiles for normal distribution
- finding, 159–161
 - high percentiles, 160–161
 - low percentiles, 159–160
 - steps for finding, 159
 - wording in percentile problems, translating, 161–162

- personal expenses, pie charts for, 96–97
- picture form, expressing your exam question in, 360–361
- pie charts
 - described, 16, 96
 - evaluating, 100
 - of lotto revenue, 97–98
 - misleading graphs, 336, 338
 - for ordering takeout, 98–99
 - for personal expenses, 96–97
 - for projecting age trends, 99–101
 - tips for, 101
- placebo, 61
- placebo effect, 61, 280
- planning and designing surveys, 263–266
- polls
 - analyzing data in, 271–272
 - bias in, minimizing, 14
 - carrying out, 268–271
 - collecting data for, 268–269
 - conclusions, drawing, 272–273
 - cover letter for, 265
 - errors in interpretation of survey results, 272
 - ethical issues, 265
 - following up, 270–271
 - impact of, 258–262
 - interpreting results of, 271–273
 - media and statistics, 27
 - organizing data in, 271–272
 - overview, 11–12, 257
 - planning and designing, 263–266
 - process for, 262–263
 - purpose of your survey, clarifying, 263
 - questions on, 265–266, formulating
 - response rate, 270
 - sample selection, 267
 - sources for, 258–260
 - target population, 264
 - timing for, 264–265
 - topics for, 260
 - type of, 264–265
- pollsters, 258
- population, 51–52
- population mean
 - confidence intervals for, 210–212
 - described, 55–56
 - hypothesis tests, 238–240
- population proportion, 183, 214–216
- population standard deviation
 - confidence intervals for population mean when you don't know, 212
 - confidence intervals for population mean when you know, 210–211
 - confidence intervals for the difference of two population means when you don't know, 218–219
 - confidence intervals for the difference of two population means when you know, 216–218
 - standard error, 176–177
- population variability, 209–210
- predictions
 - interpretation, 331
 - overview, 306–307
- probabilities
 - sample mean, 181–182
 - sample proportion, 185–186
 - t*-table, 168
- probability distributions, 137–138
- Probability For Dummies* (Rumsey), 104
- problems, understanding what the question is asking in test, 357–358
- process for surveys, 262–263
- projecting age trends, pie charts for, 99–101
- projecting from sample to population, 330–331
- properties
 - correlation, 300
 - normal distribution, 1505
- purpose of your survey, clarifying, 263
- p*-values
 - calculating, 230–231
 - defining, 220
 - guidelines for, 232–233
 - overview, 65, 229
 - test statistics and, 229–231

Q

qualitative data (categorical data). *See also*
graphing categorical data
calculating, 72–75
summaries for, 15
quantitative data, 50
questions on surveys, 265–266, formulating

R

random assignments of subjects to treatments in
experiments, 283–284
random digit dialing (RDD), 53
random samples
checking for, 341–342
described, 53–54, 267
selecting, 13
random variables. *See also* binomials
mean, 138–139
overview, 135–136, 172–173
standard deviation, 139
rates, 39
real exam conditions, practicing for exam under,
353
reasonableness of projections, examining, 37
regression, 20–21
regression line
calculating, 302–304
described, 301
example, 306
interpreting, 304–305
linear regression, 301, 302–304
slope of, 302–305
y-intercept, 302, 304–305
relative frequency
in bar graphs, 101
described, 15
in histograms, 108
relative standing, 57
reliability of data, 287–288
Remember icon, 6
research process, 48–49
research question design
with experiments, 12

overview, 11
with surveys, 11–12
residuals, 307
response, 276
response bias, 268
response rate, 270
response variable, 301
results
decisions based on results of experiments,
making informed, 292
generalizing research, 291
implications of research, 290–291
overstated, 21, 290
reporting, 194–195
surveys, interpreting results of, 271–273
right-tail probability, 168
Rumsey, Deborah
Probability For Dummies, 104
Statistics II For Dummies, 276, 302, 307, 324

S

sample mean
described, 55–56
margin of error (MOE), 195–197
probabilities for, finding, 181–182
sample proportion
margin of error (MOE), 193–194
sampling distribution, 183–185
sample selection
described, 13
surveys, 267
sample size
checking, 342
confidence intervals, 208–209, 213–214
defining, 282–283
determining needed, 214–216
as factor in misleading statistics, 42–43
margin of error (MOE), 197–200
selecting, 281–283
standard error, 175–176
t-distribution, 167
sample statistics, gathering, 227
sample variability, measuring, 191–193

- sample variance, 81
- samples
 - bad, 53
 - described, 52–54, 266
 - non-random, 53
 - random, 53–54, 267
 - self-selected, 53
 - size of, 267–268
 - target population, representing, 266–267
 - volunteer, 53
- sampling distribution
 - if X does not have normal distribution, 178–181
 - if X has normal distribution, 178
 - mean, 174
 - overview, 172–173
 - of sample proportion, 183–185
 - shape of, 178–181
- sampling error, 62, 340
- scale
 - bar graphs, 104–105
 - described, 39–42
 - histograms, 119
 - time charts, 128–129
- scatterplot
 - creating, 295
 - interpreting, 296
 - overview, 294–295
- school proficiency, statistics about, 27–28
- scientific method, 48–49
- self-selected sample, 53
- sense, analyzing your exam answers to confirm they make, 363
- setting up
 - cells in two-way tables, 313
 - histograms, 108–111
 - hypothesis test, 224–226
- sexual activity, statistics about, 29–30
- shape of data in histograms, 111–113
- shape of sampling distribution, 178–181
- significance level, 231
- 68-95-99.7% rule, 57, 85–87, 116
- size of sample, 267–268
- skepticism about statistics, 22
- skewed left
 - boxplots, 121–122
 - described, 79
 - histograms, 112, 116
- skewed right
 - boxplots, 121–122
 - described, 79
 - histograms, 111, 115
- slope of regression line, 302–305
- sources
 - of bias, 339–340
 - checking, 42
 - media, 35
 - original, 35
 - for surveys, 258–260
- specific values of X , finding probabilities for, 145
- sports section of newspaper, statistics in, 28
- spread of data in histograms, 114–115
- standard deviation
 - binomials, 146–147
 - calculating, 81–82
 - described, 56–57, 114
 - importance of, 83–84
 - interpreting, 82–83
 - normal distribution, 149
 - properties of, 83
- standard error
 - measuring variability with, 227
 - overview, 174
 - population standard deviation, 176–177
 - sample size and, 175–176
- standard normal distribution
 - formula for, 153
 - overview, 58–59, 152
 - relationship to t -distribution, 241–242
 - standardizing, 153–155
 - tables, 365–374
 - t -distribution compared, 165–167
 - Z -table used to find probabilities for Z , 155
- standard scores
 - hypothesis test, 228
 - overview, 57–58

- standard treatment in experiments, 281
- standardizing, 58
- start/end points, 128–129
- state lottery profits, bar graphs for, 103
- statistical analyses, 289
- statistical significance, 66–67
- statisticians
 - described, 47
 - in workplace, 33
- statistics. *See also* confidence intervals
 - in daily life, 9–11
 - described, 54, 203
 - linked to parameters, 203
- Statistics II For Dummies* (Rumsey), 276, 302, 307, 324
- subjects
 - choosing, 283
 - defined, 276
- sum of squares for error (SSE), 303
- summaries
 - for categorical data, 15
 - with charts and graphs, 15–16
 - with descriptive statistics (numbers), 14–15
 - for numerical data, 15
- surveys
 - analyzing data in, 271–272
 - bias in, minimizing, 14
 - carrying out, 268–271
 - collecting data for, 268–269
 - conclusions, drawing, 272–273
 - cover letter for, 265
 - errors in interpretation of survey results, 272
 - ethical issues, 265
 - following up, 270–271
 - impact of, 258–262
 - interpreting results of, 271–273
 - media and statistics, 27
 - organizing data in, 271–272
 - overview, 11–12, 257
 - planning and designing, 263–266
 - process for, 262–263
 - purpose of your survey, clarifying, 263
 - questions on, 265–266, formulating
 - response rate, 270

- sample selection, 267
- sources for, 258–260
- target population, 264
- timing for, 264–265
- topics for, 260
- type of, 264–265
- symmetry
 - in boxplots, 121–122
 - in data, 79
 - in histograms, 111, 116

T

- tables
 - binomial distribution, 370–374
 - t*-distribution, 368–369
 - z*-distribution, 365–367
- target population
 - defining, 264
 - described, 264
 - representing, 266–267
 - surveys, 264
- t*-distribution
 - hypothesis tests, 240–243
 - overview, 17, 165
 - sample size and, 167
 - tables, 368–369
 - z*-distribution, relationship to, 241–242
 - z*-distribution compared, 165–167
- Technical Stuff icon, 6
- teen drinking, surveys about, 261
- 10th percentile (bottom 10%), 161
- terminology
 - alternative hypothesis, 65
 - bad sample, 53
 - bell-shaped curve, 58–59
 - bias, 55
 - blind experiment, 61–62
 - categorical data, 50
 - causation, 67–68
 - census, 54
 - Central Limit Theorem, 59–60
 - confidence intervals, 59
 - continuous data, 50

- control group, 61
- correlation, 67–68
- data, 50–51
- data set, 51
- discrete data, 50
- distribution, 58–59
- double-blind experiment, 61–62
- empirical rule, 57
- estimates, 54
- experiments, 60–62
- hypothesis test, 59, 64–65
- inflection point, 58
- margin of error, 62–63
- mean (average), 55–56
- median, 56
- non-random sample, 53
- normal distribution, 57
- null hypothesis, 65
- numerical data, 50
- ordinal data, 51
- outliers, 56
- parameter, 54
- percentile, 57
- placebo, 61
- placebo effect, 61
- polls, 62
- population, 51–52
- population mean, 55–56
- p*-values, 65
- random digit dialing (RDD), 53
- random sample, 53–54
- relative standing, 57
- sample, 52–54
- sample mean, 55–56
- sampling error, 62
- self-selected sample, 53
- standard deviation, 56–57
- standard normal distribution, 58–59
- standard score, 57–58
- standardizing, 58
- statistic, 54
- statistical significance, 66–67
- surveys, 62
- treatment group, 61
- variable, 51
- volunteer sample, 53
- Z-distribution, 58–59
- z-values, 60
- test statistic
 - calculating, 228
 - described, 228
 - hypothesis test, 228
 - interpreting, 228
 - p*-values, 229–231
- testing for an average difference, 247–250
- testing one population proportion, 243–244
- time charts
 - evaluating, 131
 - excess data in, 129–131
 - histograms compared, 128
 - interpreting, 127–128
 - misleading, 128–131, 337–339
 - overview, 127
 - scale, 128–129
 - start/end points, 128–129
 - variability in, 128
- timing for surveys, 264–265
- Tip icon, 6
- tips
 - for bar graphs, 106
 - for histograms, 115–116
 - for pie charts, 101
- topics for surveys, 260
- tornado statistics, 39
- totals, calculating, 314
- tracking transportation expenses, bar graphs for, 101–103
- travel section, statistics in, 29
- treatment, 277
- treatment group, 61, 279
- t*-table
 - big picture information determined with, 170
 - confidence intervals, 169–170
 - overview, 167–168, 242–243
 - percentiles, 168–169
 - probabilities, 168

t-test, 65, 240–243

two-way tables

- cells, setting up, 313

- described, 20–21, 312–313

- distributions, 315–324

- interpreting, 315–324

- marginal totals, 314

- totals, calculating, 314

type-1 errors (false alarms), 234

type-2 errors (missed detection), 234–235

U

uncountable infinite, 137

uncovering misleading statistics, 37–44

unreported crimes, surveys about, 261

V

validity of data, 288

variability

- boxplots, 123–124

- histograms, 114–115

- time charts, 128

variable

- described, 16

- overview, 51

variance binomials, 146–147

variance of random variables, 138–139

variation

- overview, 80

- range, 84

- standard deviation, calculating, 81–82

- standard deviation, importance of, 83–84

- standard deviation, interpreting, 82–83

- standard deviation, properties of, 83

vertical bar graphs, 106

vertical boxplots, 121

volunteers

- finding, 284

- sample, 53

W

Warning icon, 6

weather report, statistics in, 30

width of confidence intervals, 205–206

wording

- in percentile problems, translating, 161–162

- of survey questions as factor in misleading statistics, 43–44

workplace pet peeves, 105–106

workplace statistics

- on customer satisfaction, 32–33

- for medical studies, 31–32

- overview, 31

- statisticians in workplace, 33

X

X greater-than, less-than, or between two values, finding probabilities for, 146

x-axis in histograms, 110–111

Y

y-axis in histograms, 110–111, 119

“yeah-yeah” traps, avoiding, 351–354

yes/no data, 50

y-intercept, 302, 304–305

Z

Z-distribution (standard normal distribution)

- formula for, 153

- overview, 58–59, 152–153

- relationship to *t*-distribution, 241–242

- standardizing, 153–155

- tables, 365–367

- t*-distribution compared, 165–167

- Z-table used to find probabilities for Z, 155

Z-table, 155, 365–367

z-values, 60