

INDEX

- Page numbers followed by *f* and *t* refer to figures and tables, respectively.
- A**
- Abduction, 224–226, 241
 - "The Adventure of Silver Blaze" (Doyle), 32
 - Advertising, 211, 211*t*
 - Alpha (α), 123–124, 241
 - Ancient Rome, 14–15
 - Annualized volatility, 61, 62*t*
 - Apian, Peter, 2
 - Aristotle, 167, 221
 - Arithmetic mean (\bar{x}), 42, 56, 58, 241–242
 - AT&T, 7
 - Averages. *See also* Mean
 - defined, 242
 - frame of reference affecting, 51–52
 - and histograms, 49–51
 - types of, 40–49, 49*t*
- B**
- Bacon, Francis, 223
 - Bayeux Tapestry, 1
 - Bell curve. *See also* Normal distribution [N(x,s)]
 - about, 68–70
 - of normal distribution, 69*f*
 - and *p*-value, 186
 - Bell Labs, 7
 - Benford's law, 256
 - Bernoulli, Jacob, 141
 - Bernstein, Peter, 6
 - Best-fit line, 98
 - Biases:
 - confirmation, 170–172, 171*f*, 243
 - response, 60
 - sample, 29–31, 155, 250
 - survivorship, 32–33
 - Bierce, Ambrose, 222–223
 - Bins, 49, 50*f*, 136, 242
 - Birthday problem, 259–260
 - Black-Scholes option pricing model, 63, 170
 - Bookstaber, Richard, 93
 - Brackets, as bin notation 50
 - Brahe, Tycho, 226
 - Butterfly effect, 154, 167
 - Byrne, Donald, 93
- C**
- Calculus, 131–132, 132*t*, 134
 - Carson, Johnny, 234–235
 - Casablanca* (film), 210

- Causal explanation, 206–207
- Causation, 205–219
 defined, 242
 reverse, 216, 250
- Central limit theorem (CLT):
 about, 135–145, 228
 defined, 242
 examples, 137*f*, 139*f*, 142*f*, 144*f*
 and *p*-value, 186
- Chebyshev, Pafnuty, 80
- Chebyshev's theorem, 80–81, 242
- Chess, 93–94
- Cleveland Browns, 9–10
- Clockwork universe, 133
- CLT, *see* Central limit theorem
- Coca-Cola, 61–63, 62*t*
- Coefficient of determination (r^2),
 111–120, 242
 $1-r^2$, 116–117, 117*f*, 192
- Coefficient of variation, 59–61, 59*t*, 243
- Comparison, value of numbers based on,
 23–24, 23*t*
- Compensation rates, 20–21, 20*t*,
 31–32, 60*t*
- Confidence level:
 and critical value, 185, 185*f*
 defined, 123, 243
 and standard error, 154, 157, 158*t*, 189
- Confirmation biases, 170–172, 171*f*, 243
- Construct, defined, 164, 243
- Continuous variables, 17, 243
- Controls, 213, 218, 243
- Copernicus, 226
- Correlation:
 and causation, 208–209, 215*f*
 defined, 243
 fit measured through, 104*f*
 and linear relationships, 103, 103*f*
 negative, 99
 positive, 99
 spurious, *see* Spurious correlation
- Correlation coefficient (*r*), 97–109
 about, 97–100, 115
 calculating, 100–103, 100*t*
 defined, 244
- Count, of discrete variables, 16
- Covariance, 102
- Craft, 3–4, 244
- Critical value (CV), 185, 185*f*,
 197–198, 244
- Cromwell, Oliver, 176, 231
The Cult of Statistical Significance
 (Ziliak), 188
- CV (critical value), 185, 185*f*, 244
- D**
- Darwin, Charles, 166
- Data:
 aggregating, 235
 interval, *see* Interval data
 multi-modal sets of, 43–44
 nominal, *see* Nominal data
 ordinal, *see* Ordinal data
 quality of, 233
- Deduction, 221–222, 224, 225*f*, 227*f*, 244
- Descriptive statistics, 121, 244
- Dichotomous variables, 16, 244
- Dinosaurs, 206
- Discrete variables, 16–17, 244
- Distribution:
 leptokurtic, 78–79, 78*f*
 mesokurtic, 77, 78*f*
 platykurtic, 78, 78*f*
 power law, 91, 91*f*, 249
- Dow Jones Industrial Average, 92–93
- Doyle, Arthur Conan, 32
- E**
- Einstein, Albert, 226
Elements (Euclid), 221–222
- Elephant, parable of blind men and, 28
- Empirical rule, 71–72, 244
The End of Theory (Bookstaber), 93
- Endogeneity, 216, 244
- Enhrenberg, Andrew, 2
- Error:
 margin of, *see* Margin of error
 transposition, 255
- Euclid, 221, 223
- Euler, Leonhard, 132
- Excel:
 arithmetic mean, 42
 correlation coefficient, 103
 geometric mean, 44–45, 45*f*
 harmonic mean, 46*f*
 kurtosis, 77

- median, 42
- mode, 43
- normal distribution, 72, 79
- skew, 74
- standard deviation, 56, 63
- Explained variation, 115
- Explanation, causal, 206–207
- Explanations, 206, 238
- Extreme performance, 48, 85

- F**
- False negatives, 176, 238, 244–245
- False positives, 176, 237–238, 245
- Falsification:
 - defined, 111–112, 245
 - and null hypothesis, 164, 174–175, 175*f*
 - and r^2 , 192
- Fischer, Bobby, 93
- Fisher, Ronald, 133
- Fit:
 - correlation measuring, 104*f*
 - defined, 245
 - and slope, 200
- Fitzgerald, F. Scott, 128
- Frequency, normal distribution predicting, 73*f*
- Friedman, Milton, 121

- G**
- Gambling, 6
- Gates, Bill, 43
- Gauss, Johann Carl Friedrich, 68–69, 91–92, 258
- Generalists, numerate, 10–11
- Genesis (biblical book), 22
- Geometric mean, 44, 49*t*, 245
- German tank problem, 122, 123*t*
- Gettysburg Address, 22
- Gladwell, Malcolm, 9, 86
- Goal, defining your, 231–232
- Gossett, William, 156–157
- Gould, Stephen J., 64–65
- Green, George, 132–133
- Growth rate, 254–255

- H**
- H, *see* Hypothesis
- H_0 , *see* Null hypothesis

- Halley, Edmond, 1, 5, 7, 226, 239
- Halley's Comet, 2
- Harmonic mean, 45–46, 49*t*, 245
- Hawkins, Jeff, 106
- Heraclitus, 160
- Heterogeneity, 35
- Hippocratic Oath, 176
- Histogram, 49–51, 50*f*, 68, 136, 245–246
- Hypothesis (H), 227*f*, 228–229
 - about, 168–169
 - defined, 164, 246

- I**
- Iliad*, 3
- Induction, 223, 224, 225*f*, 227*f*, 246
- Inferences, 160–161, 227*f*, 228
 - about, 26
 - defined, 246
 - modified, 225
 - and sample size, 30*t*
 - and standard error, 148–149, 148*f*
- Inferential statistics, 121, 246
- Information:
 - defined, 246
 - as a surprise, 7–8
 - uncertainty reduced through, 8
- Intelligence, 22, 22*t*
- Interdependence, 33–35, 90, 234–235
- Interval data:
 - about, 18
 - defined, 246
 - example, 19*t*
- Inverse-square law, 2

- J**
- James, LeBron, 92
- Jenner, Edward, 219
- Jeter, Derek, 51–52
- Justice, David, 51–52

- K**
- Kepler, Johannes, 2, 226
- King, Martin Luther, Jr., 22–23
- Knight, Frank, 6–8
- Kurtosis, 77–79, 78*f*
 - defined, 246
 - negative, 78
 - positive, 78–79

L

Laboratory discoveries, 261–262
 Landon, Alf, 29
 Law of large numbers, 141, 246–247
 Leibniz, Gottfried, 132
 Leptokurtic distribution, 78–79, 78*f*
 Lewin, Kurt, 167
 Linear relationships:
 correlation measuring, 103–104, 103*f*
 and r^2 , 113*f*
 slope of, 192
The Literary Digest, 29–30, 155, 233
 Long-Term Capital Management (LTCM), 34
 Lorenz, Edward, 166–167
Love Story (film), 7
 LTCM (Long-Term Capital Management), 34
 Lurking variable, 216, 247

M

Machinery, 3–4
 Marbles, 27
 Margin of error, 127–130
 and alpha, 129*t*
 and capturing μ , 125*f*
 defined, 247
 in election poll, 256–257
 and inference, 154
 Marshall, Alfred, 166
 Marx, Karl, 207
Mathematical Principles of Natural Philosophy (Newton), 1
 “A Mathematical Theory of Communication” (Shannon), 7
 Math facts, 253–262
 Mean:
 arithmetic, *see* Arithmetic mean (\bar{x})
 geometric, 44, 49*t*, 245
 harmonic, 45–46, 49*t*, 245
 Measurement:
 of continuous variables, 17, 20–21
 outliers, 90, 94, 228, 249
 precision, 15, 252
 problems with populations, 25. 39–40. 148, 155
 problems with variables, 168

Median:

 about, 41*f*, 42–43, 49*t*
 defined, 247
 with five observations, 257–258
 without skew, 75
 Mediation, 216–217, 247
 Mesokurtic distribution, 77, 78*f*
 Mode, 41*f*, 43–44, 49*t*, 70, 247
 Model:
 defined, 164, 247–248
 theories shown in, 168*f*
 Moderation, 217, 248
 Modified inferences, 225
 Mueller, Robert, 189–190
 Multi-modal data sets, 43–44
 μ (population mean), 39, 121–134, 249

N

n , *see* Sample size
 N (population), 25, 26*f*, 249
 N(x,s), *see* Normal distribution
 NASA, 166
 Negative correlation, 99
 Negative kurtosis, 78
 Negatives, false, *see* False negatives
 Negative skew, 76
 Newton, Isaac, 1, 5, 132, 166
 New York Stock Exchange, 17
 Nominal data:
 about, 18
 defined, 248
 example, 19*t*
 Nonlinear relationships, 119*f*
 Normal distribution [N(x,s)], 67–81, 71*f*
 characteristics of, 70–71
 defined, 248
 example, 68*f*
 frequency predicted with, 73*f*
 and kurtosis, 77–79
 and p -value, 187
 problems with, 74
 and skew, 74–77
 Null hypothesis (H_0), 163–177, 229
 defined, 248
 and p -value, 181
 and slope, 196
 Numeracy, 227–229, 227*f*
 about, 1–12
 as craft, 5

- defined, 248
- information extracted through, 8, 8*f*
- and reasoning, 7
- Numerate generalists, 10–11
- O**
- Observations, 227*f*, 233–234
- Observer effect, 35
- Occam's razor, 218, 226, 248
- Olympic winners, 35, 36*t*, 37
- Omitted variables, 202
- Operationalize, defined, 248
- Ordinal data:
 - about, 18
 - defined, 249
 - example, 19*t*
- Outliers:
 - and correlation coefficient, 101
 - defined, 43, 249
 - and z-score, 86, 90
- P**
- P*, see *p*-value
- Palsgraf v. Long Island Railroad*, 217
- Parentheses, as bin notation, 50
- Pareto, Vilfredo, 90–91
- Pareto distributions, 91, 91*f*
- Pearson, Karl, 99
- Pearson's r , 99, 108
- Peirce, Charles Sanders, 224
- Pepsi, 61–63, 62*t*
- Placeholders, zero as, 13–15
- Plato, 167
- Platykurtic distribution, 78, 78*f*
- Popper, Karl, 111, 133, 169, 224
- Population (N), 25, 26*f*, 249
- Population mean (μ), 121–134, 249
- Positive correlation, 99
- Positive kurtosis, 78–79
- Positives, false, see False positives
- Positive skew, 76
- Possibilities, 260–261
- Post hoc, ergo propter hoc*, 209
- Power law distribution, 91, 91*f*, 249
- Pragmatism, 225
- Precedence, temporal, 209, 212
- Precision, zeros used for, 15–17
- Predictions, 11, 238
- Probability, 5, 6
- Propeller heads, 10–11
- Proposition, defined, 164, 249
- Ptolemy, 226
- p*-value (*p*), 175, 179–190, 229
 - and causation, 209
 - and critical value, 197–198
 - defined, 249
 - and false positives, 237
- Q**
- Questions, 231–239
- R**
- r*, see Correlation coefficient
- $r (s_y/s_x)$, see Slope
- r^2 (coefficient of determination), 111–120, 242
- Ratios:
 - about, 18, 253–254
 - correlation and, 108
 - defined, 250
 - example, 19*t*
 - zeros used for, 17–19
- Reasoning, 7–8, 58, 250
- Regression analysis, 194–195, 201, 212–213
- Reliability, 194, 236–237, 250
- Response biases, 60
- Reverse causation, 216, 250
- Risk, uncertainty vs., 6
- Robust sample, 28
- Roosevelt, Franklin, 29
- Royal Statistical Society, 2
- Rubin, Robert, 93
- Rubin, Vera, 226–227
- Russell, Bertrand, 224
- S**
- s*, see Standard deviation
- Sample, 25–26, 26*f*, 28, 250
- Sample biases, 29–31, 155, 250
- Sample mean (\bar{x}), 39–52, 41*f*, 49*t*, 70, 75
- Sample size (n), 25–38, 235
 - about, 58
 - count of observations determining, 17
 - defined, 250
 - and standard error, 155–159
 - varied, and CLT, 144*f*
- Samuelson, Paul, 227

Scales, 18–19

The School of Athens, 167

Science, 221–229

SEC (U.S. Securities and Exchange Commission), 20

Shannon, Claude, 7–8, 89, 232

Simpson's Paradox, 51, 52*t*, 251

Skew, 74–77, 75*f*, 76, 138, 140*f*, 250

Slope [$r(s_y/s_x)$], 105–106, 191–203, 251
 s/\sqrt{n} , *see* Standard error

Socrates, 221

Spurious correlation, 108, 216, 251

Standard deviation (s), 53–65

about, 54, 58

calculating, 54, 55*t*, 56

and Chebyshev's theorem, 80–81

and coefficient of variation,
59–61, 59*t*

and correlation coefficient, 102

defined, 251

and normal distribution, 70–71

uses of, 57–58

and volatility, 61–63, 62*t*

Standard error (s/\sqrt{n}), 147–162

and confidence level, 157, 158*t*, 189

defined, 251

Statistics:

about, 5–7

defined, 251

descriptive, 121, 244

importance of learning, 131

inferential, 121, 246

introduction of, 133

significance of, 181

uncertainty in, 7

Stevens, Stanley, 17–18

Success, 9

Sum, 258–259

Summation, 238–239

Survivorship biases, 32–33

Syllogism, 222

T

Tally, of dichotomous variables, 16

t distribution, 157

Temperature, 87–88, 88*f*

Temporal precedence, 209, 212

Test, 187, 227*f*, 229

Thales, 11

Theory, 164–165, 165*f*, 251

Thin sample, 28, 199, 214, 235, 250

Transposition error, 255

Trials, defined, 135, 251

Trinity College, 1

Trump campaign, 189–190

21 (film), 6

Two-tailed test, 187

U

Uncertainty, 6, 8

University of California, 51

Unpredictability, 93

U.S. Army, 18

U.S. Securities and Exchange Commission (SEC), 20

U.S. stock market, 92

V

Validity, 194, 222, 236, 251

Value, 193

Variables:

continuous, 17, 243

defined, 164, 252

dichotomous, 16, 244

discrete, 16–17, 244

lurking, 216, 247

omitted, 202

reliability of, 236–237

validity of, 236

Variance, 53–54

and arithmetic mean, 56

defined, 252

Variation, explained, 115

Volatility, 61–63, 62*t*, 252

W

Wald, Abraham, 32

Weather, 124

Whitehead, Alfred North, 167

Wiles, Andrew, 222

World War I, 133

World War II, 17, 32, 122, 123*t*

X

x , *see* Arithmetic mean; Sample mean

Z

z (z-score), 83–95, 252

Zero, 13–24

and bell curve, 70

benefits and purpose of, 13–14

and coefficient of determination, 112

and confidence interval, 198 f

correlation coefficient of, 197, 197 f

defined, 252

examples, 19–24

as placeholder, 13–15

for precision, 15–17

and ratios, 17–19

z-score (z), 83–95, 252

