Table of Figures

Figure 1: Inferences from Observed to Unobserved....18
Figure 2: Indicating Internal Causation....20
Figure 3: Association Diagrams....23
Figure 4: Association....24
Figure 5: Prevalence of Suicide by German Provinces in the 1500s....26
Figure 6: Argument as a House....28
Figure 7: House Graphs for Three Arguments....29
Figure 8: Statistics as Evidence in an Argument....32
Figure 9: Confounding in a Triangle Diagram....35
Figure 10: Examples of Confounding in Triangle Diagrams....36
Figure 11: Common Cause for an Association....37
Figure 12: Triangle Diagram: Golf Courses & Divorces....37
Figure 13: Mechanism....38
Figure 14: Story Diagrams for Father and his Kids....39
Figure 15: Child Abuse in Minnesota....45
Figure 16: Orphans Due to AIDS....45
Figure 17: Global Temperatures since 1940....46
Figure 18: Global Temperatures since 1820....46
Figure 19: Take CARE....61
Figure 20: Three Kinds of Arithmetic Comparisons....64
Figure 21: Grammar for Arithmetic Comparisons....65
Figure 22: Constant Growth: Absolute and Relative....71
Figure 23: Ratios With Constant Growth: Absolute and Relative....71
Figure 24: Antarctic Temperature Data....72
Figure 25: Major League Wins vs. Total Payroll....81
Figure 26: Experiments vs. Observational Studies....84
Figure 27: Golf Scores over Time....87
Figure 28: Golf Scores versus Age....88
Figure 29: Random Assignment Nullifies Prior Confounding....90
Figure 30: Randomized Assignment....90
Figure 31: Longitudinal Graph: Before/After Brief Run....93
Figure 32: Take Kaplan; Score Higher....99
Figure 33: Column Charts of Qualitative Data....114
Figure 34: Column Charts of Quantitative Data....114
Figure 35: Dot plot of Family Incomes....115
Figure 36: Shapes of Distributions....116
Figure 37: Distribution of Student Grades and House Prices....116
Figure 38: Cumulative Distribution of Houses by Price....118
Figure 39: Percentiles on ACT and SAT Tests....119
Figure 40: Percentiles for Family Incomes....121
Figure 41 U.S. 2000 Percentiles of Those Surviving and Dying by age....122
Figure 42 Heights of Boys in Percentiles by Ages 2 to 5....123
Figure 43 Minimum U.S. Family Income by Quintile....124
Figure 44 U.S. Adjusted Gross Income (AGI)....125
Figure 45 Distribution of House Prices: Mean....128
Figure 46 Distribution of House Prices: Mode....128

Figure 47 Distribution of House Prices: Median....129
Figure 48 Centers for Symmetric Unimodal Distributions....132
Figure 49 Centers for Right-Skewed Distributions....132
Figure 50: 2000 U.S. Distribution of Deaths by Age and Sex....133
Figure 51 Distribution of SAT Scores and Wages....137
Figure 52 Heart Attack Prevalence vs. Cholesterol Level....138
Figure 53 Scatter plot: Weight vs. Height for College Students....138
Figure 54 State NAEP Scores Confounded by Encyclopedia....144
Figure 55 Weighted-Average Graph: Silverware for Jill....147
Figure 56 Weighted-Average Graph: Silverware for Both....149
Figure 57 Weighted-Average Graph: Silverware Standardized....150
Figure 58 Triangle for Average Cost of Silverware....151
Figure 59 LA vs. WV NAEP Scores (Raw)....152
Figure 60 LA vs. WV NAEP Scores (Standardized)....153
Figure 61 Standardizing Race-Based Incomes: Step 1....155
Figure 62 Standardizing Race-Based Incomes: Step 2....156
Figure 63 Standardizing Race-Based Incomes: Step 3....157
Figure 64 Summarizing the NAEP example....159
Figure 65 Summarizing the income example....159
Figure 66 Standardizing House Prices on Size of Land....160
Figure 67 Take Into Account Confounder Influence....161
Figure 68 U.S. Income Distribution: Before/After Adjust....163
Figure 69 Precipitation and Population Z-scores in selected U.S. Cities....167
Figure 70 2000n NAEP 4th grad State Math Scores....168
Figure 71 Standardizing Test Scores to IQ....169
Figure 72 Height Distribution: Effect Size = 2.5 and 1....170
Figure 73 Effect Size of 3....172
Figure 74 Standard Normal Distribution....174
Figure 75 Random Process Outcomes....175
Figure 76 Height Distributions for Women, Men and Combined....177
Figure 77 Male-female ratio and Percentage who are Men by Height....178
Figure 78 Distribution of IQs (left) and Boy-Girl Ratios by IQ (right)....178
Figure 79: Median More Extreme than Mean....180
Figure 80 Four X-Y Plots....181
Figure 81 Correlation and Slope for a Normalized X-Y Plot....182
Figure 82: Controlling for relevant confounders....188
Figure 83: Template for Percent Grammar....192
Figure 84: Generic Part-Whole Decoding Rules....193
Figure 85: Percent Grammar Decoding Rules....193
Figure 86: Answer Part-Whole Questions using Counts....194
Figure 87: Create Part-Whole Ratio Table from a Count Table....195
Figure 88: Describe a specific Percentage in a 100% Table....196
Figure 89 Part-Whole Pie Chart: Age of Drivers in Accidents....199
Figure 90. Template for Percentage Descriptions....202
Figure 91 Margin Value Rule....215
Figure 92: Requirements for One-Way Half Tables in Columns....215
Figure 93: Rules for Identifying Half-Tables....216

Figure 94: Missing Margin Rule 220
Figure 95: Measuring Happiness....221
Figure 96: Template for Rate Phrase-Based Descriptions....226
Figure 97: Pie Diagram: Rate of Accidental Deaths....226
Figure 98: Rate Grammar Decoding Rules....227
Figure 99: Template for Clause-Based Rate Descriptions....230
Figure 100: Template for Chance Phrase-Based Descriptions....232
Figure 101: Template for Chance Clause-Based Descriptions....232
Figure 102: Wheel of Inferences....234
Figure 103: Rules for Decoding Descriptions of Ratios....235
Figure 104: Cases Attributable to an Exposure....243
Figure 105: Common-part comparison of Part-Whole ratios....246
Figure 106: Distinct-parts comparison of Part-Whole Ratios....247
Figure 107: Two Pie Diagram....247
Figure 108: Pie Compare for Example 1: Common Part....249
Figure 109: Pie Compare for Example 2: Distinct-Part Comparison....249
Figure 110: Compare Ratios Using Phrase-based Descriptions....251
Figure 111: Common-Part Comparison of Ratios: Percentage Grammar....254
Figure 112: Distinct-Part Comparison of Ratios: Percentage grammar....255
Figure 113: Creating Short Form Comparisons from Long Forms....255
Figure 114: Rules for *Likely* Comparisons of Ratios....258
Figure 115: Common-part comparison template Using *Likely*....261
Figure 116: Distinct-parts comparison Template Using *Likely*....262
Figure 117 Minn. population, arrestees, convictions and prisoners by race 270
Figure 118 Rules for Decoding Relative Risk Descriptions....274
Figure 119 Two Pies Describing Relative Risk....274
Figure 120: Comparison of Two Relative Risks....280
Figure 121: Percentage Under 15 vs. Birth Rate by Country....282
Figure 122: Percentage Under 15 vs. Death Rate....282
Figure 123: Low Weight Babies vs. Unmarried Moms by State....283
Figure 124: Leading Causes of Death in 2001....288
Figure 125: Comparing two three-factor percentages....289
Figure 126: Compare Suicides for Widows vs. Widowers....292
Figure 127: Ronald fisher and Jerome Cornfield....309
Figure 128: Death Rates by Hospital and by Patient Condition....312
Figure 129: Prevalence by Hospital: Death vs. Poor-Health....313
Figure 130 Death Sentence Rates....314
Figure 131 Distribution by Race of Murder....316
Figure 132 Magazine Subscription Renewal Rates....317
Figure 133 Distributions by Month: Renewals vs. Directs....318
Figure 134 Standardizing Magazine Subscription Renewals....319
Figure 135 Basketball Completion Rates by Team and Type of Shot....320
Figure 136 Prevalence by Team: Completions vs. Two Pointers....321
Figure 137 Auto Deaths and Airbag Presence....322
Figure 138: Weighted Average Plot: Hospital Death Rates....324
Figure 139: Standardizing Reverses Difference....325
Figure 140 Standardizing Decreases Difference....326

Figure 141: Comparative Binary Separations....330
Figure 142 Biorhythm Chart....332
Figure 143: Confidence Interval and Margin of Error....349
Figure 144: Exact Margin of Error vs. Sample Proportion....351
Figure 145 Statistical Significance using Conservative Intervals....356
Figure 146 Patients' Death Rates by Hospital:....363
Figure 147 Patients' Death Rates by Hospital: Standardized....363
Figure 148 Average Weights by Smoking Status....364
Figure 149 Average Weights by Smoking Status: Standardized....365
Figure 150 Low Birth-Weight Babies: Unstandardized, Significance....366
Figure 151 Low Birth-Weight Babies: Standardizing....367
Figure 152 Low Birth-Weight Babies: Standardized, Significance....368

Table of Tables

Table 1: NAEP Math Scores 2000, grade 8: MD vs AZ 33
Table 2: U.S. persons living with AIDS....43
Table 3: U.S. Accidental Deaths by Type of Accident....62
Table 4: U.S. Deaths, Miscellaneous, by Cause 67
Table 5: Average Weight by Smoking Status and Gender 80
Table 6: U.S. Hourly & Annual Wage, Selected Percentiles 121
Table 7 Distribution of Heights for U.S. Twenty-year olds 122
Table 8 Distribution of Weights for U.S. Twenty-year olds 122
Table 9: Average U.S. Family Income by Quintile and Top 5% in 2003 124
Table 10: 1999 U.S. Family Incomes by Family Type....133
Table 11: 1999 U.S. Family Incomes by Highest Education Level 135
Table 12: 2000 U.S. Family Incomes by Number of Wage Earners....136
Table 13: State Prison Operating Expenses: CA vs. NY 142
Table 14: State Prison Operating Expenses: MD vs. KS....142
Table 15: State Prison Operating Expenses: MN vs. ME....143
Table 16: State Prison Operating Expenses: MN vs. IA....143
Table 17: NAEP 2000 8th Grade Math Scores: VA vs. TX 144
Table 18 SAT Verbal Scores by Race: 2002 vs. 1981....145
Table 19: 2000 U.S. Married Family Income by Wife Working....146
Table 20: NAEP Scores 4th Grade Math: LA vs. WV 151
Table 21: Mean Family Incomes by Race & Marital Status....154
Table 22: Relation between Effect Size and Median Overlap....173
Table 23: Average Heights for Tall and Short....176
Table 24: Average Incomes ($1,000) for Rich and Poor....176
Table 25: Statistics Based on and Control for....185
Table 26: Toy Table of College Students....188
Table 27: U.S. Accidental Deaths by Race and Sex 189
Table 28: U.S. Accidental Deaths: Percentage by Sex within Race....190
Table 29: 1990 U.S. Accidental Death Rate by Race and Sex 190
Table 30: Toy Table of Students: Row Percentages....195
Table 31: Sample Table of Row Percentages 195
Table 32: Sample Table of Column Percentages....195
Table 33: Sample Table of Total Percentages....196
Table 34: World Population by Religion and Continent (1996)....197
Table 35 U.S. Distribution of Marriages by Age and Sex (%)....198
Table 36 U.S. Women: Distribution by Contraceptive Status....198
Table 37 U.S. Age of Drivers Licensed and In Accidents....199
Table 38 U.S. Lifetime Mobility by Religious Preference....205
Table 39: U.S. Teacher Status: 2004-05 vs. 1988-89 212
Table 40: U.S. Current (last month) Drug Use Percentages....213
Table 41 Sample One-Way Half Table 214
Table 42 Percentage Uninsured with Helpful Titles....214
Table 43 Percentage Uninsured with Unhelpful Titles....215
Table 44 Cigarette Smoking....215
Table 45 Moral Acceptability of Activities by Age Group 216
Table 46 Full Three-variable Table: Row-Percentages 217

Table 47 Smokers Two-Way Half Table (Helpful title)....217
Table 48 Smokers Two-Way Half Table (Unhelpful title)....218
Table 49 Percentage overweight or obese....218
Table 50 U.S. 2004 School Status Dropouts of 18-24 year olds....219
Table 51 Percentage of Marijuana or Hashish Users by Age....220
Table 52 U.S. Percentage of Smokers by Age....221
Table 53 Abortion Percentages by Party....221
Table 54 U.S. 1990 Accidental Deaths by Race and Sex....222
Table 55 U.S. 1990 Population by Race and Sex 222
Table 56 U.S. 1990 Accidental Death Rates by Race and Sex 224
Table 57 Auto Traffic Death Rates in the U.S.....227
Table 58: U.S. Unemployment Rates by Highest Grade, Sex & Race....252
Table 59: Students by Sex and Smoking Status....253
Table 60 Percentage who are Runners by Sex and Smoker 254
Table 61: IQ and Crime: Toy table of Counts 268
Table 62: Buyers versus Ad-Viewers (fictitious) 271
Table 63: Colonists For Separation vs. Readers of *Common Sense*....272
Table 64 Percentage of Births which are Low Weight....275
Table 65 Registered Adults vs Registered Voters 294
Table 66 Death rate per 1,000 Births by Cause: 1920 vs. 1960 294
Table 67 Greenhouse Gas Concentrations....296
Table 68 Gave Birth in Last Year By Sex among Young Adults 298
Table 69 Toy Table: Police Calls, Count Table 1....300
Table 70 Toy Table: Police Calls, Count Table 2....300
Table 71 Low IQ (≤90) and Prison 300
Table 72: Suicide-Bulemia Connection 301
Table 73: Medical Test—Quality vs. Predictive Power 302
Table 74: Outcomes of a Medical Test 303
Table 75: Quality of a Medical Test (Confirmation)....303
Table 76 Predictive Values of a Medical Test 304
Table 77 Medical Test, Blank Form for Counts 305
Table 78 Medical Test for Detecting HIV, 0.1% Prevalence....306
Table 79 Medical Test for Detecting HIV, 1% Prevalence....306
Table 80 Medical Test for Detecting HIV, 10% Prevalence....307
Table 81: Mortality by Hospital 310
Table 82: Mortality by Hospital and Patient Condition....310
Table 83: Mortality by Patient Condition and Hospital....312
Table 84: Distribution of Patient Conditions by Hospital 312
Table 85: Death Sentences by Race of Murderer 314
Table 86: Death Sentences by Race of Victim 314
Table 87: Death Sentence Data by Race of Murderer and Victim 315
Table 88 Renewal Rate by Division and Month....316
Table 89 Renewal Rate by Division....317
Table 90 Subscriptions Due by Division and Month....318
Table 91 Basketball Field Goals by Team & Shot 320
Table 92 Basketball Field Goals by Shot....320
Table 93 Percentage who smoke by sex, weight and height 330

Table 94 U.S. Poverty Rates by Race and Age....330
Table 95: Sensitive Question Surveys....342
Table 96. Conservative Margin of Error vs. Sample Size....352

Table of Stories

Story 1: Fruits, Veggies Cut Risk of Breast cancer....14
Story 2: Suicide Rate and Prevalence of Protestants....26
Story 3: Two Hunters and a Bear....29
Story 4: Similarities between Presidents Lincoln and Kennedy....48
Story 5: The Numerology of 9/11.....49
Story 6: Traders puzzled by eery 911 closing of S & P futures....50
Story 7: N.Y. Lottery Draws 9-1-1 on 9/11....50
Story 8: Surveys: Convenience vs. Random....53
Story 9: Many Dads Unknowingly Raising Others' Kids....54
Story 10: Survey Bias in Convenience Sample: Accuracy of Bible....54
Story 11: Loud Music Can Cause Lung Collapse....55
Story 12: Patriarchy, Poverty and Public Policy....56
Story 13: Risk of Dying....57
Story 14: IQs of U.S. Presidents....58
Story 15: Unfair boss could shorten your life: study....59
Story 16: Mistaken Comparisons in Stories....68
Story 17: Mistaken Comparisons in the News....68
Story 18: U.S. Federal Reserve Increases Interest Rates....73
Story 19: Earth 3 Million Miles Closer to Sun....73
Story 20 Bush Cuts Deficit in Half....74
Story 21 New York City "Dropout Rate"....77
Story 22: Does Smoking Increase the Risk of Lung Cancer?....88
Story 23: Do Magnets Reduce Pain?....91
Story 24: Prontosil Reduces Mother's Risk of Death....92
Story 25: Placebo Usage Criticized....107
Story 26: Teens who eat breakfast weigh less....109
Story 27: Percentiles Involving Admission to Law School....123
Story 28: Home Schooled Kids: Percentile Points Higher....130
Story 29: Education Gains at Edison-Managed Public Schools....131
Story 30 Each Daily Soda Increases Obesity Risk 60%....139
Story 31: U.S. Distribution of Incomes....163
Story 32: U.S. Price of Housing per Bedroom....164
Story 33: Coefficient of Variation, Commodities and Stocks....173
Story 34: TV for toddlers interferes with brain growth, says study....183
Story 35: Passive Smoke Raises Impotence Risk....241
Story 36: Do You Worry More About Your Health?....266
Story 37: Fertility Aids and Multiple Births....269
Story 38: Race of Hate-Crime Offenders....269
Story 39: Intimate Partner Violence....270
Story 40: Single-Parent Children Living in Poverty....278
Story 41: Births Attributable to Marriage....278
Story 42: Poverty Attributed to Lack of High School Diploma....278
Story 43: TV Watching Attributable to Race....279
Story 44: Heart Disease Deaths Attributable to Smoking....279
Story 45: Suicides: More Widows than Widowers....291
Story 46: Increase in Health Costs....292

Story 47: What to do when you test positive for HIV 304
Story 48. July 2006: 655,000 Excess Iraqi War Deaths 348
Story 49. Prozac failed anorexic women in study (Extract) 354
Story 50. Duct tape no magical cure for warts, study finds.....355
Story 51: Merck diabetes drug effective (Extract).....356
Story 52: Black candidates paint new picture for GOP politics (Extract).....358
Story 53: Imbalance of Boy vs. Girl Births in Canada Tied to Toxics.....359
Story 54: Clinton Poll (Margin of Error) 359

Table of Equations

Eq. 1 *N times as much as = (N-1) times more than = (N-1)100% more than* 66
Eq. 2 Percentile = 100[(RankAscending – 1) / (#Subjects – 1)].....120
Eq. 3 RankAscending = 1 + (#Subjects – RankDescending).....120
Eq. 4 Percentile = 100{1 – [(RankDescending – 1) / (#Subjects – 1)]} 120
Eq. 5 Mean = Sum of value / # of subjects 127
Eq. 6 Value(X) = Value(Xo) + Slope*(X – Xo).....139
Eq. 7 Weighted Average = $\frac{(W1)(M1)+(W2)(M2)}{W1+W2} = F1\bullet M1 + F2\bullet M2$ 146
Eq. 8 Percentage difference explained: 100% (Before – After)/Before 157
Eq. 9 Std value = Actual Value – Slope*(Actual – Mean Confounder).....160
Eq. 10 Range = Maximum – Minimum 165
Eq. 11 Interquartile Range = 75th percentile score– 25th percentile score...165
Eq. 12 Z = (Individual Score – Mean) / (Standard deviation).....166
Eq. 13. Standardized Score = Target Mean + Z * Target Standard Dev.....169
Eq. 14 Effect Size = (LargerMean—SmallerMean) / PooledStdDev.....171
Eq. 15 Coefficient of Variation = Standard Deviation / Mean 173
Eq. 16 Skewness = (Mean—Median) / Standard Deviation.....174
Eq. 17 Correlation = Standardized Slope = Slope (SDx/SDy) 181
Eq. 18 Part-whole percentage = (Part/Whole) x 100%.....191
Eq. 19 % Attributed to Exp = 100%(ExpRate—CntrlRate)/ExpRate.....240
Eq. 20 Cases Attributable to Exposure = (% Attributable)(Exposed cases)242
Eq. 21 Cases Attributable = (Rexposed - Rcontrol)(Exposed population) .242
Eq. 22 % Attributable to Exposure = 100% (RR – 1) / RR 276
Eq. 23 Cases attributable to exposure = [(RR – 1)/RR](# cases exposed) .277
Eq. 24. P(BC|A) ≤ P(C|AB).....289
Eq. 25. P(success) = *K*/*N* for *K* successes in *N* equally-likely outcomes. ...334
Eq. 26 Expected Sample statistic = Population statistic 339
Eq. 27 Expected Number of Sample Successes = *P* • Sample size 339
Eq. 28 Percentage of *N* who did X = (# *Yes* – 50%**N*)/(50% * *N*)*100.....342
Eq. 29 Guessing Adjusted Score = [# Right – (*N* / *k*)] [*k* / (*k* – 1)] 343
Eq. 30 Guessing Adjusted Score = # Right – [#Wrong / (*k* – 1)].....343
Eq. 31 Ntotal = N1stCapture / Fraction recaptured 344
Eq. 32 Confidence interval = sample statistic ± margin of error.....349
Eq. 33 Exact margin of error for proportion = $2\sqrt{p(1-p)}/\sqrt{n}$ 350
Eq. 34 Exact margin of error for average = 2*s* / √*n* 350
Eq. 35 Conservative margin of error = Range / $\sqrt{n}$ 352
Eq. 36 Conservative confidence interval for proportion = $p \pm [1/\sqrt{n}]$.....352
Eq. 37 Subgroup ME = Group ME *√(Group size / Subgroup size).....353

Glossary

100% Sum Rule	**100% Sum Rule:** If a margin value is a 100% sum, then the group is whole and each component is a *part.* P. 197
Adjusted gross income	**Adjusted gross income** (AGI), the official name for taxable income, is total income from taxable sources and subtract off allowable adjustments (exemptions and deductions). P. 125
Alternate explanation	Common **causes and** *confounders* are alternate explanations for an association P. 40
Arbitrary evidence	**Arbitrary evidence** is evidence that may be true but whose relevance to the argument is weak. P. 30
Arithmetic comparison	An **arithmetic comparison** is a numerical comparison of two statistics called the test and the base. **Arithmetic comparisons** are of three kinds: simple difference, simple ratio and relative difference. P. 63
Assembly	**Assembly** is a choice of a definition or presentation that influences the size of a statistic or a statistical association. P. 42
Association	**Associations** exist when things happen together repeatedly in space or time; specifically, an association is "a statistical dependence between two or more events, characteristics or other variables." P. 21
Association words	**Association words** describe an association explicitly: P. 24 See also **Causation words** and **Between words**.
Average	Average: See **Mean**.
Bar chart	A **bar or** column chart involves distinct categories, so there must be gaps between bars. P. 114
Basis points	**A basis** point is a one hundredth of a percentage point. P. 70
Bayes comparison	A **Bayes comparison** is an exchange of the part with the test whole in a common-part comparison of ratios with no change in the numerical strength of the comparison. P. 269
Best-fit line	The **best-fit line** summarizes the association between two variables. It is a straight line that typically goes through their joint center (the intersection of their means) with a slope that best fits the data. P. 139
Between words	**Between words** describe an association but suggest causation. P. 25
Bias	**Bias** is systematic error: a "systematic deviation of results or inferences from the truth." See Respondent bias, sampling bias and measurement bias. P. 52
Capture-recapture	**Capture-recapture is a sampling technique** that uses what's expected from randomness to estimate the size of a population that is not readily countable. P. 345

Cause words	**Causation words** assert causation explicitly. P. 25
Cause	**A cause is an event** or condition whose level or presence makes (or would make) a material difference in something else. P. 19
Cases attributable	Cases **attributable to exposure are those cases** in the exposed group that are attributable to being in the exposure group. P. 242
Chance	**Chance** is a term that includes luck, coincidence, and randomness. P. 47
Chance grammar	**Chance grammar** includes the keywords *chance, risk, odds, probability* and *likelihood.* P. 231
Coefficient of variation	The **Coefficient of Variation,** the ratio of the standard deviation to the mean, standardizes the standard deviation in terms of the size of the mean. P. 173
Coincidence	**Coincidence** is a noteworthy connection between unlikely events or conditions with no obvious causal connection. P. 47
Columns, rows	**Columns** run vertically like columns in a building; **rows** run horizontally like rows of seats in a theater. P. 189
Common cause	A **common cause** is a third factor that, like a confounder, is an alternate explanation for an association. Unlike a confounder, a common cause affects the predictor as well as the result. P. 36.
Common-part comparison	A **common-part comparison** compares ratios having a common part but different wholes, as in *whites are more likely than blacks to commit suicide.* P. 245
Comparison bias	**Comparison bias** occurs when the treatment or exposure group and the control group differ in important ways other than receiving a treatment or exposure. P. 106
Completion bias	**Completion/dropout bias** occurs when the outcome is influenced differently by those who do and don't complete. P. 105
Component	**Components** are cells within a group. Components are named by rows within a column group (by columns within a row group). P. 197
Confidence Interval	A **confidence interval for** a sample statistic is the interval of values within the margin of error. P. 349
Confidence Level	A **confidence level** is the chance that the confidence intervals generated by this procedure will include what is being estimated. P. 349
Confounder	A **confounder** is a factor associated with the predictor and a result in an association—a third factor that may cause the result but is not caused by the predictor. Confounders provide alternate explanations for an association; they involve alternate causes for the result and must also be associated with the predictor. P. 34

Control of, control for	**Control of** is physical; to take control of something is to set, to assign or determine some conditions or values of the independent factors. P. 96 **Control for** is mental or contextual thinking. P. 96
Control group	The **control group** is the group that does not receive the treatment or exposure. P. 92
Controlled study	A **controlled study** involves at least two groups or subgroups at the same time. While some use "controlled" to describe the same subjects at two different points in time, we use "longitudinal. P. 84.
Convenience sample	A **convenience sample** (a grab sample) is a sample "selected by easily-employed non-random methods." P. 102
Correlation	**Correlation** measures the slope of an association between two variables that have been standardized using Z-scores; the sign indicates whether the slope is positive or negative. P. 181
Cross-section study	A **cross-sectional study** involves "the relationship … between variables of interest in a defined population at one particular time." This may be a single moment in time (unemployment) or a time interval (death rate). P. 84
Determiners	**Determiners** are conditions that determine or delimit the size of the whole or part. P. 192
Distinct-part comparison	A **distinct-parts comparison** compares ratios having different parts but a common whole, as in *NFL players are more likely to be black than white.* P. 246
Double-blind study	A **double-blind study** blinds the researcher as well as the subject from knowing which group a subject is in, thereby eliminating bias for both. P. 106
Double-ratio comparisons	**Double ratio comparisons** are ratio comparisons of ratios. P. 248
Doubly-controlled	A **doubly-controlled study** involves a before/after comparison in the treatment/exposure group with a similar comparison in the control group. P. 85
Effect size	**Effect size** is the difference in means for two groups divided by their pooled standard deviation. P. 171
Error	**Errors** are systematic deviations from what is real or true. P. 52
Experiment	An **experiment** is "a study in which the observer intentionally alters one or more factors under controlled conditions in order to study the effects of so doing." P. 83
Frequency distribution	A **frequency distribution** gives the count of subjects in each group (categorical data) or in some range of values (continuous data). P. 113

Grade	**Grade** describes a slope as a percentage when the variables have the same units (e.g., a road or a walkway). P. 139
Group	A **group** is a row, column or table whose subjects make up the subjects in a margin value. P. 197
Half table	A **half table** is half of a full 100% table with enough information to complete the table. A **one-way half table** is a half-table of percentages involving a single column or row. **A multiple one-way half table** consists of a series of one-way half tables placed side-by-side (if columns) or on top of each other (if rows). P. 214
Halo Effect	The **halo effect** is when the researcher's optimism influences the data to support that optimism. P. 105
Hawthorne effect	The **Hawthorne effect** is a systematic change in response when the subjects know they are the subject of attention. P. 104
Histogram	A **histogram** is a chart where each bar is over an interval of heights, and the bars can touch. P. 114
Independence	**Independence** in statistics is when the existence of one event or condition has no influence on the probability of the other. P. 337
Index values exclusive	Index values can be exclusive (non-overlapping) or non-exclusive. **Exclusive index values** limit each subject to only one cell. **Non-exclusive index values** allow overlapping categories. P. 213
Index values exhaustive	Index values can also be exhaustive or non-exhaustive. **Exhaustive** index values cover all relevant values of the index variable. **Non-exhaustive** index values omit some relevant values. P. 213
Incidence	***Incidence*** is a ratio involving a time interval (death rate, birth rate). P. 227
Indexes	**Index variables** classify things in a table. **Index values** are values of an index variable. P. 189
Law of Averages	The **Law of Averages** holds that as sample sizes increase, the sample averages will approach the population average. P. 340
Law of very large Numbers	**The unlikely is almost certain given** enough tries. P. 50
Longitudinal study	A **longitudinal study** involves repeated measures: measurement of the outcome at two different times (before and after a treatment or exposure) on the same or similar subjects. P. 84
Luck	**Luck** gives chance a semi-causal status.47
Margin cells	**Margin cells** are cells at the edge of a table. P. 196
Margin values	**Margin values** are the values at the edge of a table (indicated by "All" or "Total") that include all the subjects in a column or row. P. 196

Margin value values	**Margin values** are either sums or averages. A sum is always bigger than the biggest value it includes, and an average is always **smaller.** P. 196
Margin-Value Rule	**Margin Value Rule:** If a margin value is a *sum*, then each component of the group is a separate *part.* If not a sum (if an *average),* then each component of the group is a separate whole. P. 215
Margin of Error	The **margin of error,** often abbreviated **ME,** is the variability expected in sample average when random samples are taken from a population. P. 349
Mean	The **mean** or **average** is the sum of the values divided by their count. P. 127
Measurement bias	**Measurement bias** is "systematic error arising from inaccurate measurements or classifications." P. 53
Mechanism	A **mechanism** is a means by which the predictor causes the result. It is not alternate explanation for their association. P. 38
Median	The **median** of a distribution is the middle value, the value having an equal number on either side. P. 129.
Mode	The **mode** is the value or category with the highest frequency. P 128
More important variable	A **more important variable** is one that explains more – that provides a stronger association with the variable of interest – than the predictor. P. 329
Named Ratio	**Named ratios** are nouns that indicate the presence of a ratio such as ratio, rate, percentage or chance in descriptions and likely or prevalent in comparisons. P. 250
Necessary condition	A **necessary condition** is one that must be satisfied before a result can occur. (If the result occurs, the necessary condition must have been true.) P. 159
Non-compare bias	Non-compare bias: See comparison bias
Non-response bias	**Non-response bias** is an instance of selection bias in which the subjects select out by not participating in a survey. P. 105
Normal distribution	A **Normal distribution** is a bell-shaped distribution—a distribution that is symmetric with a single peak or mode. P. 174
Normalizing	**Normalizing** takes into account—controls for—both center and spread in describing a value. Normalized values are usually expressed in units of standard deviation as Z-scores. P. 166
Nullify	**Nullify** is where an association vanishes after controlling for a confounder. P. 143

Observational study	An **observational study** is a study where "the investigator is observing without intervention." P. 83
Outcome	The outcome (result, or dependent factor) is the quantity in an association whose size or existence is being predicted, explained or influenced. P.21
Outliers	**Outliers** are extreme values in a distribution (e.g., more than two or three standard deviations from the mean). P. 167
Part	Henceforth, **part** will designate the group (e.g., men) which if applied to the whole gives the part within that whole (e.g., male soldiers). P. 191
Part-whole percentage	A **part-whole percentage** gives the size of the part measured as a percentage of the whole. P. 191
Percent grammar	Percent grammar includes the keywords percent, share or fraction. P. 190
Percentage attributable to exposure	The **percentage of the exposure rate or cases that is** attributable to the exposure is the excess between the exposure and control group rates as a percentage of the exposure rate. P. 240
Percentage explained by confounder	The **percentage difference explained by a confounder** is the percent reduction in the original difference after taking into account the influence of the confounder. P.157
Percentage grammar	**Percentage grammar** involves the keywords percentage, fraction or proportion. P. 200
Percentage points	**Percentage points** are used to measure a simple difference between two percentages. P. 69
Percentile	A **value's** percentile is the percentage of subjects who have scores at or below that value. P. 118
Percentile pnts	See **Percentiles**
Percentile score	Percentiles **also eliminate** the difficulty we saw with ranking. P. 118
Percentiles	**Percentile** points measure the difference between two percentiles. P. 119
Placebo effect	The **Placebo Effect** is a systematic response in subjects receiving a placebo: "an inert medication or procedure." P. 104
Point	The **point** is a disputable claim that an argument is designed to support. P. 17
Population	A **population** is any group of interest. P. 101
Prediction	**Predictions** look forward to future events or unknown conditions. **Confirmation** is the percentage of cases in which a known condition is accurately confirmed by the test. **Explanations** tend to look backward from outcomes to current or past conditions or events. P. 299

Prediction Interval	A **prediction interval** identifies the range in which randomly selected subjects are most likely when sampled from a bell-shaped **distribution.** P. 174
Predictor	**The** predictor (or independent factor) is the quantity in an association that predicts, explains or influences the existence or size of the outcome. P. 21
Prevalence	***Prevalence*** is a ratio taken at a moment in time (unemployment rate). P. 227
Probabilistic causation	**Probabilistic causation** is causation where something makes a difference sometimes. P. 19
Probability	**Probability** measures the chance of outcome. P. 333
Power rule, Explanatory	The **explanatory power rule** for comparing two binary variables holds that the more important variable (the one having greater explanatory power) has the greater difference for the variable of interest. P. 329
Quintiles	**Quintiles** are fifths of the group; **quartiles** are fourths and **deciles** are tenths. P. 123
Random assignment	**Random assignment** means randomly assigning subjects to the treatment and control group. P. 90.
Random samples	**Random samples** are samples in which each member of a population is equally likely to be selected. P.101.
Random sample benefit	**Random samples** are samples in which each member of a population is equally likely to be selected. P. 335
Randomized trial	**Randomized trials** are experiments in which random assignment as been used. P. 90
Randomness	**Randomness—pure** chance—is the absence of any pattern that would help in predicting the next outcome. P. 47
Range	The **range** of a set of numbers is, perhaps surprisingly, a single number, found by subtracting the smallest from the largest. P. 165.
Rank	**Ranks** (1st 2nd, 3rd) measure the order or place of a value in a group of values with 1st being the best. P. 116
Rate	A **rate** is a ratio that uses per to introduce the *unit* of measure. P. 223
Rate grammar	***Rate* grammar** includes *rate, incidence* and *prevalence.* P. 227
Regression	Regression: See slope, grade, best-fit line and correlation.
Regression to the mean	**Regression to the mean** is a chance-influenced process whereby extremes tend toward the average. P. 346
Rel. frequency distribution	A **relative frequency** distribution shows the percentage in each group by the height. P. 115

Representative samples	**Representative samples** are samples in which the sample mix matches the population mix on the factors of interest. P. 101
Representative sample benefit	**Representative samples** are samples in which the sample mix matches the population mix on the factors of interest. P. 335
Researcher bias	**Researcher bias** is a change in outcome due to a researcher's knowledge of who is in which group. P. 105
Respondent bias	**Respondent bias** is when people lie or misremember in a systematic way. P. 52
Reversal	**Reversal** is where an association changes direction—a difference changes sign—after controlling for a confounder. P. 143
Safety effect	A **Safety Effect** is an increase in risky behavior because the subject knows they have safer equipment. P. 104
Sample	A **sample** is a group surveyed or tested to learn about a larger group **of** interest, **called** the population. P. 101
Sampling bias	**Sampling or selection bias** is "systematic error due to … a non-random sample of a population." P. 53
Simple average	**A simple average for two** groups is just the sum of their averages divided by two. P. 145
Simpson's Paradox	**Simpson's paradox** is when an association has one direction at the group level and the opposite direction in each subgroup. P. 144
Single-blind study	A **single-blind** study blinds the subjects as to whether they are in the treatment or control groups and eliminates the placebo effect. P. 106
Skewed distribution	A **skewed distribution** has one peak with one side higher than the other or one tail longer than the other. A right-skewed distribution has a long right tail, a high left side or both. P. 116
Slope	**The** slope of a line is its steepness, measured by imagining a point moving along the line from left to right. The slope is how much the point P. 139
Spread	The **spread** of a distribution is any measure of the variation in the data about a center. P. 165
Spurious association	A **spurious association** is one that vanishes after taking into account the influence of a third factor. P. 34
Standard deviation	The **standard deviation,** the most common measure of spread, is related to the average positive variability of the data around the mean. P. 166
Standardizing	**Standardizing** involves giving each group the same mixture of a confounder. P. 154
Stanine	**Stanine** (STAndard NINE) assigns test scores to a nine-point scale based on rank: 1 is lowest, 5 is average and 9 is highest. P. 126

Statistical generalization	A **statistical generalization** is about the group as a unit: subjects are considered collectively, not individually. P. 18
Stat. Lit.	**Statistical literacy** studies statistics in everyday usage. P. 14
Statistically significant	**Statistically significant** (statistical **significance)** means that something (a value, a difference, a ratio or an association) is very unlikely if due just to chance. P. 102
Stereotype	A **stereotype** is a judgment (usually negative) about *all* members of a group solely because of their membership in that group. P. 18 A **universal** or **categorical claim** is about everyone in a group. A **particular or existential claim** is about just some of the group.
Sufficient condition	A **sufficient condition** is one such that the result must occur if the sufficient condition is true. P. 159
Symmetric distribution	A **distribution** is symmetric if it's a mirror reflection around a center. P. 116
Tables	**Tables** are organizations of data into cells that are arranged in rows and columns. P. 189
Test and base	The **test** (T) is the number of interest; the **base** (B) is the basis of the comparison. P. 63
Treatment group	The **treatment or exposure group** is the group receiving the treatment or exposed to a condition. P. 92
Triangle diagram	A **triangle diagram** shows the relationships between three related factors: a predictor, an outcome (result) and a related factor such as a confounder. P. 35
Weighted average	A **weighted average** weights the average for each subgroup by the number in the subgroup and then divides the result by **the total** number in the group. P. 145
Weighted average line	The **weighted-average line** shows the weighted average for any mixture of the confounder. P. 147
Z-score	A **Z-score** is a relative score: the number of standard deviations above the mean. P. 166

Index

% of
  describe ... 192
A=Assembly ... 42
adjusted gross income ... 126
associated margin ... 220
association ... 21, 24
  best fit ... 139
  correlation ... 181
  explained by ... 158
  Simpson's Paradox ... 309
  slope ... 139
  spurious ... 314
attributable ... 240
base ... 63
base indicators ... 63
base rate fallacy ... 297, 303
Bayesian statistics ... 361
best fit ... 139
bias
  measurement ... 52
  respondent ... 52
  sampling ... 52
body cells ... 189, 190
C=Confounding ... 16
C=Context ... 33
capture-recapture ... 345
CARE ... 16, 31
cases attributable ... 242
causation ... 19, 28
  probabilistic ... 20
center and skewness ... 132
chance
  confidence ... 350
  evolution ... 347
  expected value ... 338
  probability ... 333
  regression to the mean ... 346
clause description
  rates ... 230
coefficient of variation ... 173
column table ... 195
columns ... 189
common cause ... 36
common-part ... 247
common-whole ... 247
comparing extremes ... 177
comparison
  base indicators ... 63
  benefit of ... 329
  complements ... 75
  counts ... 63
  likely ... 258
  percentages ... 69
comparison, kind of
  1. simple difference ... 63
  2. simple ratio ... 63
  3. relative difference ... 63
components ... 197
confidence
  chance ... 350
confidence interval
  interpretation ... 349
  mean ... 353
  measure ... 354
confidence intervals
  difference in proportions ... 356
  proportions, conservative ... 352
confidence level ... 349
confounder ... 34
confounding ... 34
  averages ... 154
  counts/totals ... 142
  proportions ... 311
  statistical significance ... 362
construction of statistics ... 77
context
  confounding ... 34
control of ... 96
controlling for
  decrease w/o reversal ... 143
  decreases ... 143
  increase ... 143
  increases ... 143
  nullifies ... 142
  reverses ... 142
  spurious ... 142
convenience sample ... 102
correlation ... 181
counts, compare ... 63
cross-sectional study ... 84

data distributions – shape
skewness ... 115
determiners ... 193, 226
distribution
percentile ... 119
range, interquartile ... 165
standard deviation ... 166
distribution – spread
range ... 165
distribution – spread of
coefficient of variation ... 173
distribution spread
Normal ... 174
double blind ... 106
E = Error/Bias ... 52
effect size ... 171
evaluation ... 28
evolution ... 347
exclusive ... 213
exhaustive ... 213
experiment ... 83
double blind ... 106
single blind ... 106
explained by
confounder ... 158
explanatory power ... 329
exposure ... 240
exposure cases ... 242
exposure population ... 242
fallacy
base rate ... 297, 303
prosecutors ... 303
formula
margin of error ... 350
frequency distribution ... 113
group ... 197
guessing
adjusting for ... 343
guessing adjusted score ... 343
half tables
rates ... 224
two-way ... 254
incidence ... 227
independence ... 337
index values ... 189
index variables ... 189
inter-quartile range ... 165
keywords
association ... 24
between ... 25
causation ... 24
law of averagees ... 340
law of averages
compensation ... 340
dilution ... 340
Law of very large Numbers ... 50
likely ... 258
likely compare
common part ... 261
distinct part ... 261
longitudinal study ... 84
margin cells ... 189, 213
margin of error ... 349
formula ... 350
subgroups ... 353
margin value rule ... 214
margin values ... 197, 213
mean ... 128
confidence interval ... 353
measure
confidence interval ... 354
measurement bias ... 105
mechanism ... 38
median ... 129
median overlap ... 131
missing margin rule ... 220
mode ... 128
more important variable ... 330
necessary condition ... 159
normalizing ... 167
nullify ... 143
observational study ... 83
ordered comparison of
ratios ... 235
Over-involvement ratio ... 271
particular claim ... 18
part-whole percentage ... 191
percentage
attributable ... 240
part-whole ... 191
syntax ... 201
percentage diff. explained by ... 157
percentages - compare
two-way half tables ... 254

percentile score ... 118
percentiles ... 117
calculating ... 120
phrase description
rates ... 227
placebo effect ... 106
population ... 101
population size
estimating ... 344
prediction versus confirmation ... 303
prediction vs. explanation ... 298
prevalence ... 227
prevalent ... 258
probabilistic causation ... 19
probability
analytical ... 334
classical ... 334
proportion
confidence interval ... 351
conservative confidence interval. ... 352
exact confidence interval. ... 350
prosecutor's fallacy ... 303
quantitative data
three key properties ... 113
quintile ... 123
R = Randomness ... 47
random
independence ... 336
random sample ... 101
benefit ... 335
randomized response
chance ... 341
sensitive issues ... 341
range ... 165
inter-quartile ... 165
rank ... 116
stanine ... 126
ranks
calculating ... 117
rates
clause description ... 230
common errors ... 228
creating ... 223
incidence ... 227
not part-whole ... 228
part-whole indicators ... 225, 226
phrase description ... 227
prevalence ... 227
regression ... 139
regression to the mean ... 346
reject the null ... 361
relationship
independence ... 337
relative difference ... 63
representative sample ... 101
benefit ... 335
researcher bias ... 105
respondent bias ... 104
reversal ... 143
reversal condition ... 158
row table ... 195
rows ... 189
sample ... 101
convenience ... 102
random ... 101
representative ... 101
selection bias ... 105
sensitive topics ... 341
simple difference, counts ... 63
simple ratio ... 63
Simpson's Paradox ... 144, 309
single blind ... 106
skewness
left vs. right ... 116
rules for ... 132
slope ... 139
spurious ... 142
spurious association ... 82, 314
standard deviation ... 166
standardizing
averages ... 154
proportions ... 325
statistical significance ... 366
Z-scores ... 168
stanine ... 126
statistical generalization ... 18
statistical literacy ... 14
statistical significance
confounding ... 362
difference ... 356
difference in proportions ... 356
importance ... 360
rules. ... 356

stereotype 18
study
  cross-sectional 84
  longitudinal 84
sufficient condition 159
syntax
  percentage 201
  rates 227
tables 189
  % of 192
  100% total 196
  100% by column 195
  100% by row 195
  associated margin 220
  missing margin 220
taking into account .... *See* control for
test 63
trend 139
triangle diagram 35
universal claim 18
variables
  more important 330
weighted average graph 147, 324
weighted average line 148
whole
  inappropriate 294
  omitted 292
Z-score 166

Describing Part-Whole Ratios

1. "Among" and "per" always introduce a whole.
2. Leading prepositions introduce, determine or delimit a common whole.
3. In statements, a relative clause after "percentage" always contains a part.
4. Modifiers (leading adjectives or other trailing relative clauses) typically take on the status of whatever they modify. (Note exception in the previous rule)

"Percent" Grammar (P. 192). Leading prepositions can be added.

___ %	of _____	is/are _____
##	{whole}	{part}

E.g., In the U.S., 25% of women are smokers (or "25% of women smoke").

Among _____	___%	is/are _____
{whole} .	## .	{part}

E.g., In the U.S., among women, 25% are smokers (or "25% smoke").

The main verb separates part and whole. The whole is on the same side of the main verb as the % symbol; the part is on the opposite side (or is the verb).

"Percentage" Statement Grammar (P. 202). Leading prepositions can be added.

"Percentage who" present	The percentage of__ {whole}	who/that are _____ {part}	is/are ___% ##
"Percentage who" absent	Among _____, {whole}	the percentage of ____ {part}	is/are ___ % ##

"Rate" Phrase Statement Grammar (P. 227). Leading prepositions can be added.

R1: The _____rate [of _____] is ___ per _____.
{part} {whole} #/part #/whole

R2: Among _____, the rate of _____ is ___ per _____.
{whole} {part} #/part #/whole

"Rate" Clause Statement Grammar (P. 227). Leading prepositions can be added.

3a: Subject (whole) verb (part) **at a rate** of N per M.
3b: Subject (part) verb (intransitive) **at a rate** of N per M (whole).
4a: **The rate at which** subject(whole) verb (part) is N per M.
4b: **The rate at which** subject (part) verb (intransitive) is N per M.

Rules for Decoding Tables of Ratios (Rates or Percentages)

Margin values are either sums or averages. A sum is always bigger than the biggest value it includes, and an average is always **smaller.** P. 196

100% Sum Rule: If a margin value is a 100% sum, then the group is whole and each component is a part. P 197

Margin Value Rule: If a margin value is a *sum,* then each component of the group is a separate *part.* If not a sum (if an *average),* then each component of the group is a separate *whole.* P. 215

Missing Margin Rule: If the associated index values are exclusive, they are wholes (unless they add to 100%). Otherwise they can be parts or wholes. P. 220

In comparing two numbers, one is the test (T), the other is the base (B). (Page 65)

Simple Difference: # = (T - B): _____ is _____ more/less+ than _____.
{test} # or # percentage pts {base}

Simple Ratio: # = (T/B): _____ is _____ times as much/many as _____.
{test} # {base}

Relative Difference:
Percent difference: **# = 100(T-B)/B** ____ is _____ % more/less+ than ____.
Times difference: **# = (T-B)/B** ____ is __ times more/less+ than ____.
{test} # {base}

+: Difference comparisons allow endings in **"er"** such as great**er**, small**er**, etc.

COMPARING RATIOS: Common Part (page 245)

To delimit a common whole, leading phrases can be added before these templates. These templates show *percent difference.* Use templates above for other compares

"Percentage" Grammar, Long-Form Compare (P. 254)

The percentage of _____	**that are** _____	is ___	% more than	the percentage of _____	**that are** _____
test whole	common part	##	compare	base whole	common part

E.g., The percentage of women who smoke is 25% more than the percentage of men who smoke.

The percentage **of** _____	among _____	is ___	% more than	the percentage of _____	among _____
common part	test whole	##	compare	common part	base whole

E.g., The percentage of smokers among women is 25% more than the percentage of smokers among men

"Likely" Grammar Rules (P. 258)

1. "among" always indicates a whole
2. "to" indicates a part. (Also, to be, to do, to have, etc.)
3. A part-whole compare must have at least 3 part-whole terms with at least one part and one whole.
4. "as X is" or "than X is" means X is *linked* to the subject. Two linked terms have the same part-whole status.
5. "is likely to" without an object (e.g., *is likely to occur* or *is likely to happen*) indicates the subject is the part.

"Likely/Prevalent Among" Short Form Compare: Part as subject (P. 261)

_____ is/are	___	% more/less likely	among/in _____	than	among/in _____
common part	##	compare	test whole	Indicate	base whole

E.g., Among men, smoking was 50% more likely/prevalent in 1970 than in 2000.

"Likely To" Short-Form Compare: Test whole as subject (P. 261)

_____ is/are	___	% more/less **likely**	**to** _____	than	is/are _____
test whole	##	compare	common part	indicate	base whole

E.g., In 2000, women are 25% more likely to smoke than are men.