19 Glossary

100% Sum rule (margin values)	100% Sum Rule: If a margin value is a 100% sum, then the group is whole, and each component is a part. P. 188
Absolute risk reduction	Absolute risk reduction (ARR) is the percentage point difference in risk between those treated (exposed) and those untreated (unexposed) expressed as a percentage. P. 319
Accuracy, confirmation	Confirmation accuracy in a medical test is the percentage of diseased cases that test positive (the percentage of disease-free cases that test negative). P. 281
Accuracy, prediction	Prediction accuracy in a medical test is the percentage of positive outcomes that involve the disease (fraction of negatives that are disease free). P. 281
Adjective	Adjectives (grammar) are leading modifiers of nouns, pronouns or other adjectives (completed passes). P. 50
Alternate explanation	An alternate explanation for an association can be provided by a confounder. Suppose city hospitals have higher patient death rates than rural hospitals. Maybe rural hospitals give better care. An alternative: city hospitals have sicker patients. P. 53
Ambiguous	Ambiguous means vague on essentials. P. 15. See also Equivocal.
Anecdotal evidence	Anecdotal evidence involves a study or story about an individual or a small group based on limited experience. P. 71
Apples+oranges comparison	Apples and oranges comparison (a mixed-fruit comparison): a metaphor for any comparison of things that have a major difference. P. 25. See also a crude comparison.
Arbitrary evidence	Arbitrary evidence is evidence that is purely hypothetical or is weakly related to the argument. P. 21
Argument	An argument is a peaceful way to persuade someone to believe or do something by giving reasons. P. 46 See also valid argument, strong argument, sound argument and weak argument.
Arithmetic comparison	Arithmetic comparisons compare two values or variables by showing the size and direction of a two-group comparison or a two factor covariation. P. 62
Arithmetic compare kinds	Arithmetic two-group comparisons are of four kinds: difference, times ratio, percent difference and times difference. P. 62
Artifical intelligence	Artificial intelligence (AI) is the ability of machines to perform tasks that normally require human intelligence. P. 351
Assault	Assault is a violent crime excluding murder, sexual assault and robbery. P. 372

2023-StatLit-KH-Edit-0802.docx

Assault, Aggravated	Aggravated assault is an assault that involves a dangerous weapon. P. 372
Assault, Simple	Simple assault is an assault that does not involve a dangerous weapon. P. 372
Assembly	Assembly (assumptions) involves all the choices that influence a statistic and are not covered elsewhere under CARE. P. 27
Assembly solutions	Assembly solutions involve maturity and critical thinking: close-reading, hypothetical thinking, estimating the impact, and asking questions. P. 85
Association (statistical)	Association (statistical) is an observable quantitative connection between groups or characteristics. P. 12
Association, crude	A crude association is one that fails to take into account plausible confounders. It is also known as a mixed-fruit association or an apples and oranges association. P. 87
Association: flimsy	A flimsy association is an association that is easily influenced. P. 50
Association: solid	A solid association is an association that is resistant to being influenced (not easily influenced). P. 50
Association words	Association words assert association explicitly or describe associations involving fixed conditions or unrepeatable events. P. 16 See also Causation words and Between words.
Association1, Raw	Raw associations provide some underlying data for an association without specifying the order or the amount. P. 105
Association2, Ordered	Ordered associations provide the direction or order to an association, but lack arithmetic precision. P. 105
Association3, Arithmetic	Arithmetic associations provide order and precision for an association. P. 105
Attributable excess	The attributable excess is the difference between the bigger (exposure) rate and the smaller (control) rate. P. 237
Availability bias	Availability bias: selection bias where one's judgement about an event or statistic is influenced by how readily one recalls examples. P. 99
Average	Simple average: Add up the count or measure and divide by the number. P. 142 See also Mean and Weighted Average.
Bar chart	A bar chart is a chart that uses bar lengths or heights for the amounts. The bars can be vertical (column chart) or horizontal (row charts). The bar lengths can be counts or measures. With categorical data, the bars are separated. With quantitative data, the bars can touch (see Histogram). P. 118

Base-rate fallacy	The base-rate fallacy involves focusing on a high rate in one group while failing to compare it with a base-rate: the same kind of rate in a related group. P. 279
Base part	A base-part in a comparison of ratios is the base in the comparison and the part in the part-whole ratio. P. 241
Base whole	A base-whole in a comparison of ratios is the base in the comparison and the whole in the part-whole ratio. P. 241
Basis point	A basis point is one-hundredth of a percentage point. P. 105
Bayes comparison	A Bayes' comparison is a common-part ratio comparison in a group versus the entire population. It allows an exchange of the part with the test whole with no change in the numerical strength of the ratio comparison. P. 266
Bayes' rule	Bayes' rule: P(A|B) / P(A) = P(B|A) / P(B) for a 2x2 table with rows A and A' (non-A), and columns B and B' (non-B). P. 267
Between words	Between words are words whose meanings are "between" association and causation. They describe an association but suggest causation. P. 16
Bias	Bias (statistical) is systematic error in obtaining data. See subject bias, measurement bias and sampling bias. P. 38
Bias solutions	Bias solutions: Single blind (with a placebo), double blind, and randomness. P. 99
Big data	Big data (statistically) is any data set that is so big that all the associations are statistically significant. P. 91
Birth rate	The birth rate is the annual number of per births per 1,000 women. P. 391. See also fertility rate.
Butterfly Effect	Butterfly Effect: the idea that small things can have large impacts in complex systems. A powerful image is a butterfly flapping its wings and causing a typhoon. P. 294
Capture-recapture	Capture-recapture is a sampling technique that uses what is expected from randomness to estimate the size of a population that is not readily countable. P. 336
Case	A case is an instance of a disease or problem, typically based on a positive medical test. P. 277
Cases attributable	Cases attributable to exposure are the events (births, deaths) in the exposed group that are attributable to being exposed. P. 238
Causation words	Causation words assert causation, state sufficiency, or state a contra-factual. P. 16

Cause	A cause is an event or condition whose level or presence makes (or can make) a difference in something else. P. 14. See also necessary or sufficient condition, or probabilistic causation.
Census	A census is a survey involving every member of a population. P. 94. See also sample and population.
Centers rule	Three centers rule: The three measures of center have a natural order in most skewed distributions. The mean is most influenced by skew. It is alphabetic in English: mean, median, and mode (ascending) if skewed left; mode, median, and mean (descending) if skewed right. P. 127
Chance	Chance is a possibility of something happening. P. 33
Chance grammar	Chance grammar describes part-whole ratios using the keywords *chance, risk, odds, probability and likelihood*. P. 227
Cherry picking	Cherry picking is deliberate selection bias: intentionally selecting only those subjects or data that support a given claim. P. 43.
Clause	Clauses (grammar) are a group of words with a subject and predicate (men run). P. 50
Clinical trial	A clinical trial is an experiment where subjects are randomly assigned to treatment and control groups (or where treatments and controls are randomly assigned to subjects). P. 71.
Close reading	Close reading involves seeing how the presence, choice or absence of a single word or phrase could have a big impact on a statistic or statistical association. P. 85
Cluster sample	A cluster sample is a sample obtained by sampling (random or census) within randomly selected geographic subgroups or clusters of a population. P. 334. See also sample
Coefficient of Variation	The coefficient of variation (CV) is a relative measure of variation for data. Technically, it is the dimensionless ratio of the standard deviation to the mean. P. 176
Coincidence	Coincidence is a noteworthy connection between unlikely events with no obvious connection. P. 33
Columns, rows	Columns run vertically like columns in a building. Rows run horizontally like rows of seats in a theater. P. 181
Common-part comparison	A common-part comparison compares two ratios having a common part of two different wholes. P. 240.
Common part	A common part exists when two part-whole ratios share the same part. P. 240. See distinct parts.

Comparison	A comparison (statistical) compares two values. It can be two-valued (same or different), ordered (same, more, or less), or arithmetic. P. 12
Comparison bias	Comparison bias: selection bias where prior differences between the control group and the treatment/exposure group influence the comparison after a treatment/exposure. P. 99. This type of bias may be due to confounding.
Comparison, mixed-fruit	Mixed-fruit (crude) comparison: a comparison of groups having different mixtures of a related variable. Also called an apples and oranges comparison. P. 139
Comparison1, raw	Raw comparisons present the values to be compared. The user must figure out the order and the size of the comparison. P. 60
Comparison2, ordered	Ordered comparisons are non-arithmetic comparison of two values that indicate which value is bigger but does not mention the size of the comparison. P. 60
Comparison3, arithmetic	Arithmetic comparisons compare two values or variables by showing the size and direction of a two-group comparison or a two factor covariation. P. 62
Completion bias	Completion bias: Selection bias where the outcome is influenced differently by those who do or don't complete. P. 98
Component (tables)	Components are cells in a table within a group: a row, column, or range. P. 188
Confidence interval	A confidence interval is an interval around the sample statistic that contains the associated population statistic 95% of the time. P. 95
Confirmation bias	Confirmation bias: selection bias where data is selected because it agrees with one's thinking or beliefs. P. 99
Confounded, confounding	Confounded or *to confound* (found with) means to confuse; *confounding* means confusing. P. 22
Confounder	A confounder is anything that provides an alternate explanation for an association. Technically a confounder is a third factor that causes the result in an association, and is related to—but not caused by—the predictor. P. 23
Confusion of the inverse	Confusion of the inverse is a grammatical error where the top and bottom in a ratio are mistakenly exchanged without changing the number. P. 278. The confusion of the inverse in a part-whole ratio mistakenly exchanges part with whole. P. 187
Consumer Price Index (CPI)	The Consumer Price Index (CPI) is the weighted-average price of a basket of consumer goods over time. P. 393

Conjunctions	Conjunctions (grammar) are words that connect phrases or clauses (or, and, but). P. 50
Context	Context: the situation, circumstances, or reality. P. 6
Control for	Control for is mental: to take into account or adjust for. P. 82
Control group	The control group is the group not exposed (or lacks a condition) in an observational study, or not given the primary treatment in an experiment (they may be given nothing, a placebo or alternate treatment). P. 68
Control of	Control of is physical. In statistics, this involves assigning subjects to treatment or control groups in experiments. P. 82
Controlled study	A controlled study is any study involving two groups: one is treated or exposed; the other is the control group. P. 68
Convenience sampling	Convenience sampling is non-random sampling using easily-selected subjects. P. 335
Cornfield condition #1	Cornfield Condition #1: Simpson's paradox occurs only if the association between the confounder and outcome is bigger than the association between the predictor and outcome. Proved by Gastwirth. P. 295
Cornfield condition #2	Cornfield condition #2: Simpson's Paradox occurs only if the predictor-confounder association is bigger than the original predictor-outcome association. Proved by Cornfield. For example, the association between the hospitals and their patients in poor condition is stronger than the association between the hospitals and their patients' death rates. P. 316. Derived by Cornfield.
Correlation	Correlation is a standardized version of co-variation: the slope (covariation) of the standardized change in Y versus the standardized change in X. P. 176
Count data	Count data counts how many things have a discrete quality or quantity. P. 117 See also measurement data.
Co-variation	Co-variation summarizes the slope—the linear relationship—between two variables. For example, each additional year of schooling adds [on average] 12% to income. P. 12
Co-variation1, Raw	Raw covariation provides some of the data for a covariation, but not the order or size (when X is x1, Y is y1; when X is x2; Y is y2. P. 105
Co-variation2, Ordered	Ordered covariation provides the order, but not the strength (the more X, the more/less Y. As X increases, Y increases or decreases). P. 105

Co-variation3, Arithmetic	Arithmetic covariation provides both the order and the strength of the covariation (as X increases by x, Y increases or decreases by y. P. 105
Crime, assault	Assault is a violent crime excluding murder, sexual assault and robbery. P. 372
Crime, Property	Property crime is a crime that just involves property such as motor-vehicle theft, burglary/trespassing, and larceny (white-collar crime). P. 372
Crime, Violent	Violent crime is a crime against a person's body such as murder, rape, robbery, or assault. P. 372
Criminal homicide	Criminal homicide is an illegal killing. P. 365
Critical thinking	Critical thinking evaluates the strength of evidence in an argument. P. 1. See deductive or inductive reasoning.
Cross-level inference	Cross-level inference involves specification: inferring an association between variables at one level from an association between groups and those variables at a different level. P. 19
Cross-sectional study	A cross-sectional study is an observational study involving a moment in time (unemployment rate) or a short time interval (death rate). P. 71
Crude comparison	Crude comparison: a comparison that conflates (fails to take into account) important related factors. It is also known as a mixed-fruit comparison or an apples and oranges comparison. P. 25
Cumulative	Cumulative (unqualified) typically means the count or fraction less than or equal to (up to and including) a given amount. P. 121
Cut point	A cut point is a point on a continuous scale that marks the edge of a category. P. 134
Dashed lines	Dashed lines indicate an association in triangle diagrams. P. 23
Decile	Quintiles are fifths of the group; quartiles are fourths and deciles are tenths. P. 124
Deductive reasoning	Deductive reasoning (formal reasoning): reasoning where the conclusion must be true if the argument is valid and the premises are true. For example, Aristotle is a man. All men are mortal. Therefore, Aristotle is mortal. P. 19
Determiners	**Determiners** (quantities) are conditions that select the items being counted or measured. Determiners commonly involve the date or place: things that cannot be counted. Determiners can appear almost anywhere in a description or comparison. P. 129

Determiners (ratios)	Determiners (ratios) are conditions that select, determine or delimit the whole or part in a part-whole ratio. Determiners might include date, place, race or sex. Determiners control for confounders by selection. P. 183
Diabolical denominator	Diabolical denominator (ratios): Changing the denominator can change the size and direction of an association between two ratios. P. 271
Discriminate	Discriminate (verb) is ambiguous: to discern a difference (historically, good) or to judge with prejudice (currently, bad). P. 53
Disparate impact	Disparate impact: A disparity between group outcomes resulting from treatment that seem "facially" neutral. P. 53
Disparate treatment	Disparate treatment is an intentional decision to treat people differently based on their race or other protected characteristics. P. 53
Disparity	Disparity**:** A difference (lack of equality or parity) between group outcomes. P. 53
Distinct-parts comparison	A distinct-parts comparison compares two ratios involving different parts of a common whole. P. 240
Distinct parts	Distinct-parts exist when two groups or two part-whole ratios share the same whole. P. 240
Distribution grammar	Distribution (grammar) indicates that something (a 100% total) is being distributed or allocated by something. For example, distribution of heights by sex. P. 118
Distribution grammar rule	Distribution grammar rule: The keyword "distribution" should not be included when describing a single value or when comparing two values. P. 118
Distribution (mathematical)	Distribution (frequency distribution) refers to the count (fraction) of subjects (amounts) per group (interval). P. 117
Doing	Doing is when someone (or nature) manipulates the subject or something related to the subject. P. 14
Double-blind study	A double-blind study blinds the subject and the researcher from knowing which group a subject is in. P. 99
Double-ratio comparisons	Double ratio comparisons are ratio comparisons of ratios. P. 254
Double-who ambiguity	Double-who ambiguity: when two relative pronouns follow *percentage* in percentage grammar. P. 198
Dropout (school)	School dropouts are school leavers who don't graduate. School dropouts can be event dropouts or status dropouts. P. 203

Effect size	Effect size (confounder-sensitive) is the size of the comparison (typically a ratio) of means or percentages in a two-group comparison. P. 68
Effectiveness (drugs)	Effectiveness (drugs) in preventing an adverse outcome is the ratio of the rate in the control group minus the rate in the treatment group to the rate in the control group. P. 266
Equivocal	Equivocal describes a word that has several distinct meanings. For example, the word pen is equivocal: it refers to a writing implement and to an animal container. P. 49. See also Ambiguous.
Error	Errors are systematic deviations from what is real or true. P. 38
Error, Type 1	Type 1 error is a false alarm. In criminal trials: innocent people are found guilty. In hypothesis tests, a false research hypothesis is statistically-significant. P. 329
Error: Type 2	Type 2 error is a silent alarm. In criminal trials, a guilty person is found not-guilty. In hypothesis tests, a true research hypothesis is not statistically-significant. P. 329
Euthanasia	Euthanasia is a legal killing to relieve pain or suffering (a mercy killing). P. 365. See homicide.
Evidence	Evidence involves less-disputable claims that support the point of the argument. P. 46
Event dropout (school)	School event dropouts are students who dropped out during a given year. P. 203
Exclusive index values	Exclusive index values limit each subject to only one cell. P. 207 Non-exclusive index values allow some subjects to be in more than one cell. P. 207
Exhaustive index values	Exhaustive index values cover all relevant values of the index variable. P. 209 Non-exhaustive index values omit some relevant values. P. 209
Expected	Expected (empirically) means *on average* in the long run.333
Experiment	An experiment is a statistical study involving a doing: someone or nature manipulates or controls the subject. P. 69
Explanations	Explanations: Two-groups, same time: it takes a difference to explain a difference. One-group, two times: it takes a change to explain a change. P. 67
Explanatory percentage	Explanatory percentage (prediction versus explanation in a 2x2 table): The explanatory percentage has a size that is larger than that of its inverse: the predictive percentage. P. 287. See also Predictive percentage.

Explanatory power	The more important variable is the variable that has the higher explanatory power: the stronger association with the variable of interest. P. 304
Explanatory power rule	The explanatory power rule: the more important of two binary predictors is typically the one having the greater effect size: the bigger difference, times ratio or percentage difference in the variable of interest. P. 304
Explanatory size rule	Explanatory Size Rule: Given two inverse rates in a 2x2 table, the rate that is smaller is the predictive rate; the rate that is larger is the explanatory rate. P. 286
Exposure	Exposure can involve an action (getting a suntan) or a condition (being the first-born child). P. 68
Exposure group	The exposure group is the group in an observational study that is exposed (or has a condition). P. 68
Family (housing)	A family (in housing) is a household whose members are related by blood or law. P. 132
Fertility rate	The fertility rate is the annual number of births per 1,000 women ages 15-44. P. 391. See also birth rate.
Flimsy association	A flimsy association is an association that is easily influenced. P. 50
Fraction (everyday)	Fraction (everyday use) describes part-whole ratios non-arithmetically using percent grammar: a small fraction of Whole are Part. P. 211
Fraction (math)	Fraction (math use): a number expressed as a proportion of two (whole) numbers where the first or top is no larger than the second or bottom using percent-of-whole grammar. For example, the fraction of X which are Y is 0.05. A zero-to-one fraction is a "proper" fraction. P. 211
Frequency: rate	Frequency or velocity (rate): events per unit time (heart rate or speed. P. 219
Frequency distribution	A frequency distribution gives the count of subjects in each category or in a range of values. P. 118
Generalization (inference)	Generalization infers a property of a group (population) based on a related property in sample from that group. P. 47
Grammar	Grammar in this text includes ordinary English grammar and named ratio grammar. P. 28.
Group measure	A group measure (statistical) is the measure for a group. P. 129
Group-measure comparison	**Error! Reference source not found.** P. **Error! Bookmark not defined.**

Group in a table	A group in a table is the body of a row, column or table whose subjects make up the subjects in a margin value. P. 188
Growth rate	Growth rate**:** percentage or count change in amounts per unit time. P. 219
Half-table	A half-table is half of a full 100% table with enough information to complete the table. P. 201. A multiple one-way half-table consists of a series of one-way half-tables placed side-by-side as columns or rows. P. 201
Halo effect	The halo effect is when the researcher's optimism influences the data to support that optimism. P. 98
Hard science	Hard science is science based on experiments. P. 70
Hate crimes	Hate crimes are crimes that are motivated by hatred of a protected social group. P. 50
Hate incidents	Hate incidents are non-criminal actions (including speech) that are hurtful and are motivated by the victim's behavior or their social group. P. 50
Hawthorne effect	Hawthorne effect: a systematic bias in subjects' response when the subjects know they are being observed. P. 97
Helpful title (% tables)	Helpful titles are titles of rate or percentage tables that clearly identify what is whole or part. P. 201
Histogram	A histogram is a bar chart where a bar spans an interval of measures (a range of values), so the bars can touch. P. 119
Homicide	Homicide is a killing of one person by another. P. 365, See murder.
Household	A household is a group of people (related or unrelated) who share a common space. P. 353. See also family.
Human intelligence	Human intelligence (HI) involves thinking conceptually about the natures and causes of things in reality: to distinguish real from imaginary, similar from different, true from false, good from bad, right from wrong and beautiful from ugly. P. 351
Hypothesis, research	Research hypothesis: a claim involving effectiveness (a difference or change) that the researcher thinks will be true in the larger population. P. 329
Hypothetical thinking	Hypothetical thinking is "maybe" thinking: speculative thinking about what is unseen or unknown. P. 67

Hypothetical thinking: (assembly)	Hypothetical thinking about assembly is the broadest speculative thinking: thinking about how a statistic could have been created, defined, counted, measured, grouped, compared, adjusted, and presented. P. 85
Hypothetical thinking: (confounding)	Hypothetical thinking about confounders involves speculative thinking: what factors may be related to the outcome and the predictor, which of these may be plausible, and which may be bigger. P. 67
Hypothetical thinking (error-bias)	Hypothetical thinking about error (bias) involves speculative thinking: thinking about possible bias in a subject's response, in measurement, and in sampling. It also involves speculative thinking about possible mistakes, half-truths, ambiguity, and lies. P. 97
Hypothetical thinking (randomness)	Hypothetical thinking about randomness involves speculative thinking about the possible nature and size of the (underlying) data, how it may have been obtained, and if sampled, what might have been the sampled population. P. 90
Ideal experiment	An ideal experiment is an experiment that can be replicated on the same or identical subjects in the same condition. P. 14
Incidence	Incidence **(**rate**):** a relative frequency such as number of events in a time interval per group size (2020 birth rate per 1,000 population). P. 219
Indexes (tables)	Indexes are the values (words or numbers) in a table that indicate the content of the rows or columns. See exclusive and exhaustive. P. 181
Index values exclusive	Index values can be exclusive (non-overlapping) or non-exclusive (overlapping). Exclusive index values limit each subject to only one cell. Non-exclusive index values allow some subjects to be in more than one cell. P. 207
Index values exhaustive	Index values can 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. 209
Inductive reasoning	Inductive reasoning with statements (informal reasoning): reasoning where the conclusion is likely—but not certain—to be true even if the premises are true. For example, all swans I have seen are white. Therefore, all swans are white. P. 20
Inverse ratio	Inverse ratio: a ratio in which the numerator (top) and denominator (bottom) are switched. P. 278

Inverse relationship	An inverse relationship (grammar) involves two ratios that exchange part and whole: "percentage of men who run" versus "percentage of runners who are men." P. 286
Judgmental sample	A judgmental (purposive) sample is any non-probabilistic sample other than convenience, quota or snowball. P. 336
Law of averages	The Law of Averages holds that as sample sizes increase, the sample averages will approach the population average. P. 341
Law of very large numbers	Law of Very Large Numbers (chance**)**. Qualitatively: The unlikely is almost certain given enough tries. Quantitatively: If an event has one chance in N and there are N tries, then one event is expected: the chance of at least one event is more likely than not. P. 91
Lift ratio	See the Over-involvement ratio.
Line: Dashed	Dashed lines indicate an association in triangle diagrams. P. 23
Line: Solid	Solid lines with arrows represent causation in triangle diagrams. P. 23
Longitudinal study	A longitudinal study is an observational study that involves repeated measures: measurement of the outcome at two or more different times on the same or similar subjects. P. 71
Majority	A majority is more than half (50%). P. 192. See plurality.
Manslaughter	Manslaughter is an illegal killing (criminal homicide) that is not intended (accidental.). P. 365. See homicide or murder
Margin cells (tables)	Margin cells (tables) are cells at the edge of a table (top, bottom, left or right) that include part or all of the table. P. 202
Margin of error (sampling) #1	Margin of error (sampling) is an expected error: the error that includes 95% of *the closest* differences between a randomly sampled statistic and the associated population statistic. P. 94
Margin of error (sampling) #2	Margin of error (sampling) A range either side of a randomly-sampled statistic that includes the population statistic 95% of the time. P. 95. See also confidence interval.
Margin of error: subgroup	The subgroup margin of error is the *maximum* error *expected* for a particular subgroup. P. 348
Margin of error, survey	The survey margin of error is the *maximum* error *expected* for statistics involving the entire survey. P. 348
Margin values, location (tables)	Margin values are values in margin cells at the edge of a table (indicated by “All” or “Total”) that include all the subjects in a column, row or the entire table. P. 188

Margin values, content (tables)	Margin values (percentage tables) are either sums or averages. A sum is always bigger than the biggest value it includes; an average is always smaller. P. 188
Margin values missing, rule	Missing Margin Rule: If margins are missing and the index values are exclusive, they are wholes (unless they add to 100%). P. 207
Margin-value rule, 100% sum	100% Sum Rule: If a margin value is a 100% sum, then the group is whole, and each component is a part. P. 188
Margin-value rule, general	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. 203
Matching	Matching (statistics) involves selecting subjects so that each subject in the treatment/exposure group has a matching subject with the same value of the confounder in the control group. P. 84
Mean	The mean or average is the sum of the values divided by their count. P. 125 See also Weighted Average.
Measurement	See measurement data
Measurement bias	Measurement (researcher) bias is systematic error in measurements. P. 38. See also the Halo effect.
Measurement data	Measurement data measures continuous numerical quantities: characteristics that can have any value within some range. P. 117
Mediator	A mediator is the means (mechanism) by which the predictor causes the result. Technically, a mediator is a third factor that is caused by the predictor and causes the result. P. 53
Median	The median of a distribution is the middle value in a sorted list if odd (the middle of the two center values if even). P. 125
Medical test	Medical tests are tests that involve subjects with one of two conditions (diseased or disease-free), that return one of two outcomes (positive or negative), and that are calibrated by how well the test confirms the known presence or absence of the condition. P. 280
Medical test	See Medical test: simple, sensitivity or PPV.
Medical test, simple	A simple medical test is one where the accuracy (the error) in confirming is the same for diseased and disease-free. P. 282
Missing denominator	The missing denominator occurs when a common whole is omitted in a description of a ratio or ratio-based comparison of two part-whole ratios. P. 274

Mix-Matching	Mix-matching involves giving each group in a weighted average the same mixture of a confounder: either the mixture of the entire group or the mixture of one of the two groups chosen as the standard. P. 142
Mixed-fruit comparison	Mixed-fruit (crude) comparison: a comparison of groups having different mixtures of a related variable. Also called an apples and oranges comparison. P. 139
Mode	The mode (distribution) is the value or category with the highest count or amount. P. 126 A unimodal distribution has just one peak. A bimodal distribution has two peaks. P. 126
Multiple 1-way half-tables	A multiple one-way half-table consists of a series of one-way half-tables placed side-by-side as columns or rows. P. 204
Murder	Murder is an illegal killing (criminal homicide) that is intended. P. 365 See manslaughter
Named Ratios (grammar)	Named ratios are words that indicate the presence of a ratio: nouns such as rate, percentage or chance, or adverbs such as likely, risky, or prevalent. P. 179
Necessary condition	A necessary condition (logic) is one that must be satisfied before a result can occur. (If the result occurs, the necessary condition must have been true.) P. 172
Negative test	Negative is what a medical test should return if the subject is disease-free. P. 280. See also Positive Test.
Negative predictive value	Negative predictive value (NPV): Percentage of negatives without disease.. P. 315. See also Positive Predictive Value.
Negativity bias	Negativity bias: selection bias where people focus primarily or excessively on those ideas involving danger or risk that create fear or terror. P. 99
Nocebo effect	Nocebo (negative placebo) effect: subjects feel or do worse after receiving a placebo. P. 97 See the placebo effect.
Non-response bias	Non-response bias: Selection bias where the outcome is influenced by the mix of subjects that do not participate. P. 98
Nullify	Nullify is where an association vanishes after controlling for a confounder. P. 138
Number Needed to Treat	Number needed to treat (NNT) is the number that must be treated to save one life. P. 319 See Equations: NNT.
Observational study	An observational study is a statistical study that involves just a seeing: subjects can be measured but not manipulated. P. 69

Odds	Odds measure probabilities: *Odds = p/(1-p)*. Probabilities range from zero to one. The associated odds range from zero to infinity. Odds are always bigger than the associated probability. P. 267
Odds ratio	Odds ratios measure the ratios of odds: *Odds2/Odds1 = [p2/(1-p2)] / [p1/(1-p1)]*. Odds ratios are always bigger than the ratio of the underlying probabilities if they increase (smaller if they decrease). P. 267
Ordered comparison	Ordered comparisons are non-arithmetic comparison of two values that indicate which value is bigger but does not mention the size of the comparison. P. 60
Ordinary English grammar	Ordinary English grammar includes clauses, phrases, and selected parts of speech such as adjective, relative clauses, conjunctions, and prepositions. P. 28
Outcome of an association	The outcome (result, response or dependent factor) in an association is the factor whose size or existence is being predicted, explained, or influenced. P. 12
Over-Involvement Ratio	Over-involvement ratio (RP) is a ratio of exposure percentages: the percentage of cases that were exposed divided by the percentage of non-cases that were exposed. If RP>1, then RR>1. P. 266
Paradox	A paradox is a seemingly absurd or self-contradictory statement or result that is in fact well founded or true. See Simpson's paradox and Scanlan's paradox. P. 140
Part (ratio)	Part, in part-whole ratios, designates the group (men) which if applied to the whole (soldiers) gives the part within that whole (male soldiers). P. 182
Part-whole percentage	A part-whole percentage is the size of the part as a percentage of the size of the whole. P. 183
Percent grammar	Percent grammar describes a part-whole ratio when there is no other named ratio keyword and the % symbol (or "percent") is followed by "of" or a verb (are). P. 182
Percentage (everyday)	Percentage (everyday use) describes part-whole ratios non-arithmetically using percent grammar: a small percentage of Whole are Part. P. 211
Percentage (math)	Percentage (math use) describes part-whole ratios arithmetically using percentage grammar ("the percentage of Whole which are Part" or "among Whole, the percentage of Part"). P. 211

Percentage attributable to exposure	The percentage of exposure cases that are attributable to the exposure is the exposure excess as a percentage of the exposure rate: (Big% - Small%) / Big%. P. 237
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. 170
Percentage grammar	Percentage grammar describes part-whole ratios using the keywords *percentage, proportion or fraction.* P. 193
Percentage points	Percentage points measure the difference between two percentages. P. 64
Percentile	A percentile is the percentage of subjects who have scores at or below a value. P. 122
Percentile points	Percentile points measure the difference between two percentiles. P. 123
Percentile score	A percentile score is the score having a stated percentage of subjects “at or below" that value. P. 122
Phrase	Phrases (grammar) are a group of words without a subject or a predicate (male runners). P. 50
Placebo	A placebo is typically a fake (inert, non-therapeutic) medication or procedure for those in the control group. P. 73
Placebo effect	Placebo effect: subjects feel or do better after receiving a placebo. P. 97
Plurality	A plurality is the largest or most numerous. P. 192. See majority.
Point (argument)	The point of an argument is a more disputable claim supported by more factual evidence. P. 46
Political poll	Political polls predict: they forecast the vote in a future political event (like weather forecasts). P. 331 See also Surveys
Population	A population is any group of interest (people, plants or stars). P. 94
Positive predictive value	Positive predictive value (PPV): Percentage of positives that have the disease. P. 315 See also Negative Predictive Value.
Positive test	Positive is what a medical test should return if the subject is diseased. P. 280. See also Negative Test.
Predictive percentage	Predictive percentage (prediction versus explanation in a 2x2 table): The predictive percentage has a smaller size than that of its' inverse: the explanatory percentage. P. 287

Predictor	The predictor (explanatory or independent factor) in an association is the factor that predicts, explains or influences the existence or size of the outcome. P. 12
Prepositions (grammar)	Prepositions (grammar) are words that identify relationships: spatial (in, out), temporal (before, after), logical (if, then) or quantitative (out of, per). P. 50
Prevalence (rate)	Prevalence (rate): the ratio of two counts: group count divided by population count at the same time (unemployment rate). P. 219
Prevarication (word)	Prevarication is telling a half-truth by leaving out (omitting) an important part. P. 41
Probabilistic causation	Probabilistic causation is causation where "Something makes a difference sometimes." P. 14
Property crime	Property crime is a crime that just involves property such as motor-vehicle theft, burglary/trespassing, and larceny (white-collar crime). P. 372
Proportion (everyday)	Proportion (everyday use): (1) describes a balance (pleasing proportions) or lack thereof (blown out of proportion), (2) describes part-whole ratios arithmetically since proportion is typically preceded by "the" and followed by "of"'(the proportion of the population who are elderly)." P. 211
Proportional reasoning	Proportional reasoning (math use) involves an equality of two ratios: a / b = c / d. P. 211
Prosecutor's fallacy	The prosecutor's fallacy (the base-rate fallacy) involves treating the confirmation accuracy of a test as though it measured the predictive accuracy thereby ignoring a base rate: the prevalence of the disease in the group. P. 281
Pyramid, population	A population pyramid is a graphical distribution of a group's population by age and sex at a given time. P. 392
Quartile	Quintiles are fifths of the group; quartiles are fourths and deciles are tenths. P. 124
Quasi-experiment	A quasi-experiment is an experiment that is not an ideal experiment or a clinical (randomized controlled) trial. Intervention may be done by humans or nature. P. 71
Question bias	Question bias: Measurement bias where questions or answer choices are chosen to encourage a particular answer. P. 98
Quintiles	Quintiles are fifths of the group; quartiles are fourths and deciles are tenths. P. 124

Quota sample	A quota sample is a non-random sample designed to be representative on some characteristic where subjects are selected conveniently until the quota (pre-specified quantity) of subjects is obtained for each characteristic. P. 335
Random assignment	Random assignment: randomly assigning subjects to treatment and control groups or randomly assigning treatment and control to subjects. P. 72.
Random samples	Random samples are samples in which random selection or assignment was involved. P. 333. See simple random, stratified, cluster or systematic.
Random sampling: benefit and cost	Random sampling is more likely to yield a representative sample on unknown factors than any non-random process. Random sampling allows one to make statistical inferences, but it typically requires more time and money. P. 333
Randomized controlled trial	Randomized Controlled Trial (RCT): a controlled experiment (trial) involving random assignment. P. 106. See *clinical trial*.
Randomness	Randomness—pure chance—is the absence of any pattern that will help in predicting the next outcome. P. 33
Randomness solutions	Randomness solutions: Being sensitized to extremes due to coincidence, the law of very large numbers in big data and random error in small samples (sampling error). P. 90
Rank	Ranks (1st, 2nd, 3rd) measure the order or place of a value in a group of values with 1st being the best. P. 121
Rate	A rate is a ratio that uses *per* to introduce the unit of measure. P. 219. See also frequency, prevalence, incidence and growth rate.
Rate grammar: clause-based	Clause rate grammar describes a rate using an entire clause: a verb separates the part and whole. P. 222
Rate grammar: phrase-based	Phrase rate grammar describes a rate using just phrases. P. 222
Ratio (comparison)	A ratio is a quantitative relationship between two things involving or implying multiplication or division. P. 230
Ratio grammar	Ratio grammar describes a ratio using prepositions (nine men for every woman), per (miles per hour, 5 per cent), or the keyword ratio (the male-female ratio was 9:1). P. 230
Raw comparisons	Raw comparisons present the values to be compared. The user must figure out the order and the size of the comparison. P. 60
Recidivism (crime)	Recidivism occurs when a person is arrested any time after their release from prison. P. 395

Relative clause (grammar)	Relative clauses (grammar) are trailing modifiers of nouns or pronouns that start with relative pronouns such as "who" *or "that"* (passes that are completed). P. 50
Rel. frequency distribution	A relative frequency distribution shows the percentage in each group by the column height or row length. P. 119
Relative pronouns (grammar)	Relative pronouns (*who, that* and *which* as well as *what*, *where* or *when*) refer to prior nouns or noun phrases and introduce relative clauses. P. 194
Relative risk	Relative risk (RR): the risk of an outcome in an exposure group (Re) divided by the same risk in the control group (Rc) written as "the relative risk of <outcome> for <exposed> is Re/Rc." P. 266
Replication	Replication occurs when the experiment is repeated on the same or identical subjects in the same condition. P. 14
Representative samples	Representative samples are samples that match the population on some measured characteristics. P. 333
Representative sampling: benefit	Representative samples may take less time and money than random samples, but the sample statistic is not necessarily representative on other characteristics or the best predictor of the population statistic. The expected error is unknown. P. 333
Research hypothesis	Research hypothesis: a claim involving effectiveness (a difference or change) that the researcher thinks will be true in the larger population. P. 329
Researcher bias	Researcher bias: Measurement bias where a researcher's knowledge of who is in which group causes a change in the outcome. P. 98 Measurement is broader.
Respondent bias	See Subject bias. Subject is broader. Respondents are typically people; subjects can include other animals as well as plants.
Reversal (association)	Reversal is where an association changes direction—a difference changes sign—after controlling for a confounder. P. 138
Reverse causation	Reverse causation is when the outcome in an association is actually the cause of the predictor. P. 25
Safety effect	Safety effect: an increase in risky behavior because the subject knows they have safer equipment. P. 97
Sample	A sample is any part of the population. P. 94. See also census, random sampling and non-random sampling.
Sampling bias	See selection bias. Selection is broader; sampling is one form of selection.

Sample error	Sample error is the actual difference (typically unknown) between a sample statistic and the associated population statistic for a particular sample. P. 94 See Margin of error.
Scanlan's paradox	Scanlan's paradox: As a desirable outcome becomes more prevalent over time for two groups, the smaller their difference and disparity ratio tend to become. As the undesirable complements become less prevalent over time, their difference tends to decrease. This is good. But their disparity ratio tends to increase. That is bad. P. 276
Scanlan's rule	The rarer an outcome [becomes over time for two groups], the greater tends to be the relative difference [the ratio] in experiencing it and the smaller tends to be the relative difference [the ratio] in avoiding it. P. 275
School dropout	School dropouts are school leavers who don't graduate. School dropouts can be event dropouts or status dropouts. P. 203
School event dropout	School event dropouts are students who dropped out during a given year. P. 203
School status dropout	School status dropouts are adults of a certain age range not in school and who had not completed school. P. 203
Science, hard	Hard science is science based on experiments. P. 70
Science, soft	Soft science is science based on observational studies. P. 70
Selection	Selection (statistics) involves choosing subjects so they all have the same value (or range of values) of a confounder. P. 83
Selection bias	Selection (sampling) bias is systematic error in the outcome due to a non-representative selection from a population. P. 39. See also non-response, dropout, availabilitys, and confirmation bias.
Sensitivity versus specificity	Sensitivity in medical tests is the percentage of diseased cases that test positive. Specificity is the percentage of disease-free cases that test negative. P. 315
Serious delinquency	Serious delinquency: a financial loan that is in default with no payments for 90 days. P. 394
Simple medical test	A simple medical test is one where the accuracy (the error) in confirming is the same for diseased and disease-free. P. 282
Simple random sample	Simple random sample: the result when sampling from a known population that can be listed or identified so that each member is equally likely to be chosen. P. 334
Simpson's Paradox	Simpson's paradox is when an association of ratios has one direction at the group level and the opposite direction in each subgroup. P. 140

Single-blind study	A single-blind study uses a placebo to blind the subjects as to whether they are in the treatment or control group. P. 99
Skewed distribution	A skewed distribution has one peak with one tail longer than the other. A right-skewed distribution is pulled to the right with a longer right tail. P. 120
Snowball sample	A snowball sample is a sample obtained by finding a few members of the population (the seed) who refer people like them, etc. P. 336
Social statistics	Social statistics are statistics that involve people. P. 1
Soft science	Soft science is science based on observational studies. P. 70
Solid association	A solid association is an association that is resistant to being influenced (not easily influenced). P. 50
Solid lines	Solid lines with arrows represent causation in triangle diagrams. P. 23
Sound argument	A sound argument (deductive) is one that is valid and has true premises. P. 49
Specification	Specification (inference) applies an association in a group to a specific member of that group. P. 47
Spurious association	A spurious association is one that vanishes (no difference, no correlation) after controlling for another factor. P. 137
Standard deviation #1	Standard deviation: the average variability of data around the mean ignoring the sign. P. 175
Standard deviation #2	Standard deviation, arithmetically, is the square root of the average of the squared differences from the mean. P. 176
Standardizing	Standardizing involves taking something relevant into account that "levels the playing field." It may involve adjusting for inflation or population age differences or changes. It may involve using percentages as in standardized financial accounting. P. 142.
Standardizing ratios	Standardizing (statistics) converts a "mixed-fruit comparison" (an “apples and oranges” comparison) of ratios into an “apples and apples” comparison on a given characteristic. Standardizing can change an association in four ways: reverse, nullify, decrease without reversal, or increase. It does not control for the influence of assembly, randomness, or bias. Standardizing may involve mix-matching. P. 142.
Statistical generalization	A statistical generalization describes what is typical or expected in a group. P. 48

Statistical literacy A	Statistical literacy is the ability to read and interpret statistics in the everyday media. Statistical literacy is critical thinking about everyday statistics as evidence in arguments. Statistical literacy studies how everyday statistics are constructed and manipulated. P. 1
Statistical literacy B	Statistical literacy studies statistics in everyday usage. Statistically literate people think critically about the statistics behind claims. P. 4
Statistical literacy 1	Statistical literacy (broad) is critical thinking about statistics as evidence in arguments. P. 11
Statistical literacy 2	Statistical literacy (narrow) studies how everyday statistics are constructed and manipulated. P. 11
Statistical significance: 50% rule	Statistical Significance 50% Rule: If there is at least a 50% prior chance that the research hypothesis is true, then a statistically significant result means there is now a 95% chance that the research hypothesis is true. P. 344.
Statistical Significance Conf. Interval Rule	Statistically-Significant Confidence-Interval Rule: If two sample statistics have confidence intervals that do not overlap or touch, their difference is *statistically significant.* Otherwise, their difference may be *statistically insignificant*. P. 95.
Statistical significance Unlikely Rule 1	Statistical Significance Unlikely #1: If prior to the test the research hypothesis is unlikely to be true, then the confirmation error must be equally unlikely so that a statistically-significant result gives a 50% chance the research hypothesis is true. P. 346.
Statistical significance Unlikely Rule 2	Statistical Significance Unlikely #2: If a statistically significant result of a hypothesis test is to have a 50% chance the research hypothesis is true, then the confirmation error (p-value) must be less than the prior chance the research hypothesis is true. P. 346.
Statistical significance Unlikely Rule 3	Statistical Significance Unlikely #3: If a statistically significant result of a hypothesis test is to have a 90% chance the research hypothesis is true, then the confirmation error (p-value) must be less than F/9 where F is the prior chance the research hypothesis is true. For a 95% chance, the confirmation error (p-value) must be less than F/19. Page 346.
Statistically significant	Statistically significant describes an outcome that is very unlikely if due just to chance. P. 94
Statistics	*Statistics are numbers in context*. P. 6
Status dropout (school)	School status dropouts are adults of a certain age range not in school and who had not completed school. P. 203

Stereotype	A stereotype is a judgment about all (any one of) those in a group based on a characteristic of some members in that group. P. 47
Strong argument	A strong argument (inductive) is one where the premises give strong support for the conclusion. P. 49
Stratified sample	A stratified sample is a series of random samples each taken inside a given homogenous subgroup (strata) of the population. P. 334
Studies	Statistical studies involve the collection and analysis of data. Henceforth studies. Studies are either experiments (a doing) or observational studies (just a seeing). P. 69
Subgroup margin of error	The subgroup margin of error is the *maximum* error *expected* for a particular subgroup. P. 348
Subject bias	Subject (respondent) bias is when people lie or misremember in a systematic way. P. 38. See also the Hawthorne effect.
Sufficient condition	A sufficient condition (logic) is one such that the result must occur if the sufficient condition is true. P. 172
Surveys	Surveys summarize past or present conditions (like weather reports). P. 331. See also Polls
Survey margin of error	The survey margin of error is the *maximum* error *expected* for statistics involving the entire survey. P. 348
Symmetric distribution	A symmetric distribution has a mirror reflection around its center. P. 120
Systematic sample	A systematic sample is a sample obtained systematically from a place (or time) ordered population where the starting point is randomly selected. P. 334
Systemic discrimination	Systemic (institutional) discrimination is unfair treatment of members of a social group that is produced by institutional policies or practices that appear to be non-discriminatory. P. 53
Systemic racism	Systemic racism is systemic discrimination of a racial or ethnic group. P. 53
Tables	Tables are organizations of data into cells that are arranged in rows and columns. P. 181
Take CARE	Take CARE is a good admonition in dealing with statistics. In this course, each letter stands for a kind of influence: C for Confounding, A for Assembly or Assumptions, R for Randomness and E for Error/Bias. P. 10
Take into account (control)	Take into account: (1) to consider, regard, notice, pay attention to, or allow for. (2) Quantitative: to adjust for differences or changes. P. 25

Test and base (comparison)	Test and base are keywords used to identify the two groups in a two-group comparison. The base is the basis of the comparison. The test whatever is being compared with the base. P. 62
Test part (comparison)	A test-part in a comparison of ratios is the test in the comparison and the part in a part-whole ratio. P. 241
Test whole (comparison)	A test-whole in a comparison of ratios is the test in the comparison and the whole in the part-whole ratio. P. 241
Texas sharpshooter fallacy	The Texas sharpshooter fallacy is "a logical fallacy based on the metaphor of a gunman shooting the side of a barn, then drawing a target around the bullet holes to make it look like he hit the target. It illustrates how people look for similarities, ignoring differences, and do not account for randomness."92
Three-centers rule	Three centers rule: The three measures of center have a natural order in most skewed distributions. The mean is most influenced by skew. It is alphabetic in English: mean, median, and mode (ascending) if skewed left; mode, median, and mean (descending) if skewed right. P. 127
Times-based comparisons	Times-based comparisons are comparisons that involve division: times ratio, percent difference and times difference. P. 65
Times-less comparison	The new times-less comparison, T is (B/T) times less than B, is unambiguous provided both Test and Base must always positive. P. 65
Treatment group	The treatment group is the group in an experiment that is treated. P. 68
Triangle diagrams	Triangle diagrams show the relationships between three related factors: a predictor, an outcome (result), and a related factor such as a confounder. P. 23
Two-way half- table	A two-way half-table displays common-part half rates and percentages classified by two index variables. This format economizes on space. P. 205
Type 1 error	Type 1 error is a false alarm. In criminal trials: innocent people are found guilty. In hypothesis tests, a false research hypothesis is statistically-significant. P. 329
Type 2 error	Type 2 error is a silent alarm. In criminal trials, a guilty person is found not-guilty. In hypothesis tests, a true research hypothesis is not statistically-significant. P. 329
Uncontrolled study	An uncontrolled study involves just a single group. P. 68
Valid argument	A valid argument (deductive) is one where the conclusion must be true given that the premises are true. P. 49

Violent crime	Violent crime is a crime against a person's body such as murder, rape, robbery, or assault. P. 372 See also property crime.
Voter expectation	Voter expectation: (poll): Who do you think will win the upcoming election? P. 332 See also voter intention.
Voter intention	Voter intention (poll): If the election were tomorrow, who would you vote for? P. 332
Weak argument	A weak argument (inductive) is one where the premises give weak support for the conclusion. P. 49
Weasel words	Weasel words are words or phrases that suck the meaning out of a claim (just like weasels supposedly suck the yolk out of eggs). P. 41
Weighted average	Weighted average: weights the subgroup averages by their size. Multiply each group average by its size, add these products, and divide this total by the combined (total) size. P. 142
Weighted average line	The weighted-average line shows the weighted average for any mixture of the confounder. P. 149
Z-score	The Z-score is the distance from the mean measured in units of standard deviation. This ratio is dimensionless. P. 176

20 Index

21 Review of Named Ratios

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

"Percent" Grammar (p. 183). Determiners and modifiers can be added.

"% of" is present	"Among" is absent	#% of <whole> are <part>.
"% of" is absent	"Among" is present	Among <whole>, #% are <part>.
"% of" is present	"Among" is present	Among <whole1>, #% of <whole2> are <part>.

Example: 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" Grammar (p. 194). Other modifiers can be added.

	"Of" is	Leading Preposition	Subject: Whole and Part	Predicate
1	whole		The percentage of __ who* are __ {whole} {part}	is ___% ##
2	absent	Among _____, {whole}	the percentage who* are ____ {part}	is ___ % ##
3	whole2	Among _____, {whole1}	the percentage of __ who* are __ {whole2} {part}	is ___ % ##
4	part	Among _____, {whole}	the percentage of _____ {part}	is ___ % ##

* Other relative pronouns beside *who* include *that, which, what, when* and *where*.

Percentage grammar without "percentage": The <##>% of <whole> who are <part> is <predicate>.

Rules for Decoding Tables of Ratios (Percentages or Rates)

Margin values (percentage tables) are either sums or averages. A sum is always bigger than the biggest value it includes; an average is always smaller. (p. 188)

100% Sum Rule: If a margin value is a 100% sum, then the group is whole, and each component is a part. (p. 188)

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. 203)

Missing Margin Rule: If margins are missing and the index values are exclusive, they are wholes (unless they add to 100%). (p. 207)

Questions: Percent versus Percentage Grammar (p. 223).

Percent grammar: What percentage of <whole> are <part>?
Percentage grammar: What is the percentage of <whole> who are <part>?