

Glossary

absolute zero vs. arbitrary zero

When the number “0” means “none,” it is an *absolute zero*. On the other hand, when “0” doesn’t mean “none” — when it simply means the number halfway between “1” and “-1” (as in the Fahrenheit temperature scale), or the number one unit below “1” (as in the cumulative grade point scale used in most universities), it is an *arbitrary zero*. Absolute zero is a characteristic of *ratio scaling*, whereas an arbitrary zero is usually associated with *interval scaling*.

abstract concept

Concepts are abstract ideas. Examples include “truth,” “love,” “compassion,” “alienation,” “prejudiced,” “exciting.” They have no physical existence. They are like the names you put on the mental file folders in which you sort your experiences so you can make sense of them. Because your experiences are different from the experiences of other people, the contents of your mental file folders are different from the contents of theirs. Even though you may use the same label that other people use, because your mental file folder contains different experiences, your labels mean something different for you than the same labels mean for other people.

Accidental sample

Also known as *convenience sample*. A non-probability sample in which you select whoever happens to be available.

accuracy

When you estimate the amount of time it will take you to do something, your estimate may be more or less accurate. The closer your estimate is to the actual amount of time it takes, the more accurate it is. Accuracy is an issue in research when you use a sample statistic to estimate a population parameter.

alternate hypothesis

While the null hypothesis says that the pattern you see in your sample data is due to sampling variability and does *not* reflect a pattern that exists in the population your sample represents, the *alternate hypothesis* says that the pattern you see in your sample data is *not* due to sampling variability and that it *does* reflect a pattern that exists in the population. The alternate hypothesis is sometimes called the *research hypothesis*.

alternative explanation

If you are interested in the validity of a theory and you test a *hypothesis* that is logically implied by the theory, you will probably feel that your theory is supported by your data. It is usually possible, however, to come up with a different theory that also implies the same hypothesis. This different theory provides an *alternative explanation* for your data.

amount

When you are counting discrete objects (books, people, cars), you talk about the number of things you see. When you are measuring a continuous quantity (sand, fear, anxiety), you talk about the *amount*.

analysis of variance

Also known as ANOVA. An analytic procedure that tests the size of the differences among the means of two or more groups by means of comparing the between-groups variance (differences between groups) to the within-groups variance (differences between members of groups and other members of the same groups).

antecedent condition

A condition or event that happens *before* something else happens. Possibly, something that caused something else to happen.

applied research

Research that is intended to address a specific problem or issue where the result of the research has an immediate application. For example, a study of the relative efficiency of a company's internal communication systems. Applied research is usually contrasted with *basic* research.

arithmetical operations

Things you normally do with numbers, including addition, subtraction, multiplication, division, square root.

assumptions behind inferential statistics

Most inferential statistics make assumptions about the sample or the population the sample comes from. The most common one is that the sample is randomly selected from the population.

average

An informal somewhat vague word meaning common, not too good and not too bad, ordinary, or typical. If you mean the result you get when you add up all the numbers and divide by how many there are, you're talking about the *mean*.

average deviation

A useless measure of dispersion for interval or ratio data. The mean of the absolute values of the deviation scores for a set of data. If you don't take the absolute values, the sum of the deviation scores will be zero which would give a mean of zero, resulting in a misleading estimate of the spread of the data.

basic research

Also called *pure* research. Sometimes described as *curiosity-driven* research. Research that is not done in order to produce results that can directly address a specific problem or situation.

bell-shaped curve

This is a descriptive name for the *normal* curve, which has the shape of a cross-section of a bell. SEE ALSO *normal curve*.

between-groups degrees of freedom

ANOVA (*analysis of variance*) requires two *degrees of freedom*. The one for the numerator is

the number of groups minus one. Also known as *numerator* degrees of freedom.

between-groups sum of squares (SS_b)

Used in ANOVA (*analysis of variance*), this is a measure of the differences between group means and the grand mean. It is calculated as the sum of the squared differences between the means of each group and the grand mean, weighted according to the number of cases in each group.

between-groups mean square (MS_b)

Used in ANOVA (*analysis of variance*), this is also a measure of the extent to which group means differ from one another and from the grand mean. It is calculated as the *between-groups sum of squares* divided by the *between-groups degrees of freedom*.

between-groups variance

Same as *between-groups mean square*.

bias

Systematic distortion or error. When a measurement is biased, it is consistently distorted in a particular direction. CONTRAST WITH *random error*.

biased variance

Also known as *systematic* variance. The portion of variance in one variable that can be explained or accounted for by another variable is biased variance. This means that the value of the variable being explained is determined at least in part by the second variable rather than being random or *unbiased*. CONTRAST WITH *random variance*.

bins

When you make a frequency histogram for a variable for which there are many values, like Age, or a continuous variable like Height, you will probably want to somehow collapse the values into a smaller set of *bins* and have one vertical bar on the histogram for each bin. Ex: for Age you might use UNDER 20, 21-30, 31-45, 46-60, OVER 60.

boundary conditions

The circumstances or conditions in which the generative force in a theory is likely to explain its effects.

categorical variable

A variable that can take only a small number of values is *categorical* or *discrete*. Family size, profession, and gender are usually considered to be categorical. CONTRAST WITH *continuous variable*.

causal explanation

A theory that explains patterns by reference to uncontrollable antecedents employs *causal explanation* and is called a *law*.

causal relationship

A situation in which a change in one thing causes a subsequent change in another thing. SEE *cause*.

cause

An antecedent condition (something that happens first) that produces a consequent *effect* (something that happens later as a result of what happened earlier) over which the people or objects involved have no control. The effect happens whether or not the people or objects involved want it to; it is an unavoidable consequence of the cause. SEE ALSO *necessary cause*, *sufficient cause*, *probabilistic cause*.

cell's contribution to chi-squared

When you calculate chi-squared, in each cell of the table, you subtract the expected value from the observed value, square the difference, and divide the result by the expected value. The result of this is the cell's contribution to chi-squared. The sum of these contributions is chi-squared.

census

A sample that includes the entire population.

central tendency

This is an unfortunate name for a way of summarizing the data in your sample that tells you what the typical value, the value seen most often, or the central value is. Measures of central tendency include the mode, the median, and the mean.

Chi-squared — χ^2

Chi-squared is a measure of the extent to which the values in your data differ from the values you would expect to see if a certain assumption about your data is true. For bivariate data, the assumption is that the two variables are not related to one another. For a single variable, the assumption is

that the frequency distribution of your data matches a pre-defined set of probabilities.

circular definition

A definition that uses the term it is trying to define as part of the definition. EX: *negative advertising is advertising that contains negative terms*. Also, a definition so complicated and hard to understand that it makes your head spin.

closed question

A question for which the answer is selected from a short list of possibilities. Multiple-choice exam questions are closed questions.

column percents

In a *cross-tabulation*, these are the numbers that tell you what percent of the cases in a column fall in a given row. The sum of the column percents in a column should be 100%. SEE ALSO *row percents*.

computational formula

A formula that you can use to calculate a statistic that differs from the *definitional* formula in three ways: it looks more complicated; it is a lot less work to use; and the result is likely to be more accurate because there is less rounding error.

concept

A word that expresses an *abstraction* formed by *generalization from particulars*. For example, "prejudice." We group all the things we consider to be examples of prejudice together because we are interested in what it is that they have in common. So we can talk about it, we give the grouping a name, "prejudice." The name we give the grouping is thus an abstraction formed by generalization from all the particular examples.

conceptual definition

A definition that clearly explicates the meaning of a concept. It should include a specification of all essential qualities of the concept. It should be clear, complete, precise, and not circular.

conceptualizing

The process of developing conceptual definitions, of specifying the meaning of your concepts and the relationships between #or among them, and of relating your conceptual definitions to those of other researchers.

confidence estimate

How confident you can be that your estimate of a population parameter is accurate to within a specified range. SEE ALSO *confidence interval*, *confidence level*.

confidence level

The estimated probability that a population parameter lies within a given confidence interval. When a newspaper article says that the results show that 23% of the population supports the President and that these results are accurate within 3 percent 19 times out of 20, the confidence level is 95% (19 times out of 20). SEE ALSO *confidence estimate*, *confidence interval*.

confidence interval

The range of values within which a population parameter is estimated to lie at a specified level of confidence. The confidence interval in the example above is from 20% to 26%. Also, the minimal distance you want between yourself and an angry Doberman. SEE ALSO *confidence estimate*, *confidence level*.

confounded

Confused, mixed up, misunderstood.

consistency

The extent to which a measure's results do not change when the measurement is repeated. You get almost the same result every time.

construct

A concept created explicitly for a specific scientific purpose.

construct validity

The extent to which your construct or concept is a single unidimensional one, and your measuring instrument is able to measure the intended construct and nothing more. If the construct is solid and coherent and doesn't have a range of aspects that interact with your measurement approach, and if various operational definitions of the construct result in identical or similar measurements, then the construct and measurement approach together have construct validity.

contingency table

Another name for *crosstabulation table*.

continuous variable

A variable that can take any value in the range from its highest and lowest value. Age, water, income, and freedom are continuous. CONTRAST WITH *categorical variable*.

control group

In an experiment, the group to which the treatment (stimulus) is not administered. CONTRAST WITH *experimental group*.

convenience sample

Also known as *accidental sample*. A non-probability sample in which you select whoever happens to be conveniently available.

correlation

A measure of the strength of association between a pair of *continuous* variables, based on the sum of the *cross products* of the z-scores of the variables. Also known as *Pearson's r* or *Pearson product-moment correlation*. Ranges from -1.0 to 1.0.

counts

Also known as *frequencies*. In a crosstabulation, the *counts* tell how many cases fall into each cell.

covariance

A measure of the strength of association between a pair of continuous variables, based on the sum of the *cross products* of the *deviation* scores of the variables. Unlimited range.

criteria for assessing conceptual definitions

A conceptual definition should include a specification of all essential qualities of the concept; it should be clear, complete, precise, and not circular.

criteria for assessing problem statements

Problem statements should be clear and specific; *empirically verifiable*, phrased affirmatively, and stated simply.

criterion validity

The extent to which your measurement agrees with a second measure you know to be valid.

critical region

The area beyond the *critical value* calculated in a test of *statistical significance*. If your *critical ratio* falls in the critical region, you usually reject the *null hypothesis*.

critical ratio

The ratio obtained by dividing a statistic by its *standard error*; used to compare with the *critical value* to determine whether or not you land in the *critical region*.

critical value

The boundary of the *critical region* for use in a test of *statistical significance*.

cross-product

Multiplication of each value in one set of values by the corresponding value in a second set of values. Used in *correlation* and *covariance*.

cross-sectional

An approach to measurement in which you obtain data from individuals representing all segments of a population. CONTRAST WITH *longitudinal*.

cross tabulation

A procedure by which you make a two-dimensional table with rows corresponding to the values of one variable and columns corresponding to the values of the other variable. Ordinarily used with *discrete* variables.

cumulative percentage

The percentage of cases encountered up to and including a specific point.

curvilinear

A type of relationship which must be plotted as a curve rather than a straight line. An equation that requires values to be raised to higher or lower powers (i.e. square, cube, square root) or that involves logarithms.

deduction

Reasoning from a general statement to a specific case. The logical procedure used to develop hypotheses from theories. CONTRAST WITH *induction*.

degrees of freedom

A number or pair of numbers used in performing chi-square, t-test, and ANOVA. Based on the number of cases in the data or the number of values a variable can take.

denominator degrees of freedom

ANOVA requires two degrees of freedom. The one for the denominator is the number of cases minus the number of groups. Also known as *within-groups* degrees of freedom.

dependent variable

A variable whose value depends on the value of another variable. In a cause-and-effect relationship, the dependent variable is the effect. CONTRAST WITH *independent variable*.

descriptive research

The goal of descriptive research is to obtain a complete and accurate *description* of events, conditions, circumstances, processes, and relationships surrounding the focus of the research. CONTRAST WITH *explanatory* and *exploratory research*.

descriptive statistics

Statistics that summarize the data in your sample, without any reference to the population from which the sample is drawn. CONTRAST WITH *inferential statistics*.

descriptive terms

These parts of problem statements represent classes of phenomena, including *constructs* and *variables*. CONTRAST WITH *operative terms*.

deviation scores

The difference between the original values of a variable and the mean of the set of values. An individual whose score on the variable is above the mean will have a positive deviation score. An individual whose score is below the mean will have a negative deviation score. COMPARE WITH *standard scores*.

direct or positive relationship

A relationship between a pair of variables such that individuals with high scores on one variable tend to have high scores on the other variable,

while individuals with low scores on one variable tend to have low scores on the other variable. CONTRAST WITH *negative* or *inverse relationship*.

directed hypothesis

A hypothesis in which the direction of a difference is predicted. Ex: men drink more coffee than women. For directed hypotheses, you use *one-tailed* significance tests. CONTRAST WITH *undirected hypothesis*.

dispersion

The spread of values in a set of data. The more spread out the values are, the greater the *dispersion*. Commonly measured by *variance*, *interquartile range*, *range*.

double-blind methods

Experimental methods in which neither the person from whom data is collected nor the experimenter know which condition the person is in. Double-blind methods reduce or eliminate *experimenter bias*.

effect

Something that happens later as a result of something that happened earlier.

empirical

Something that can be either observed or *inferred* from observation.

empirical research

Research in which the questions are about things that exist or happen, and in which the answers are obtained by somehow observing things in the world

empirically verifiable

A theory that can be verified by observations of things that exist or happen. A theory whose truth would have real-world implications which can be tested or observed.

error variance

The fluctuation or varying of measures due to chance. Also called 'random' variance. The variance left over in a set of measures after all known sources of systematic variance have been removed from the measures.

essential qualities

The qualities that must be present; the necessary ingredients. Used in conceptual definitions.

estimating population parameters

Because you only have data from a sample (a subset of the population) you cannot measure aspects of the population — you only have access to the sample. So you must *estimate* characteristics of the population.

estimation of standard error of the mean (SEM)

Because the *SEM* is a population *parameter*, it must be estimated. You estimate the *SEM* by dividing an estimate of the population's *standard deviation* by the square root of the sample size.

estimation versus calculation

Because your data describes the whole sample, you can *calculate* summary statistics for the sample. However, because you only have data from a sample (a subset of the population) you cannot measure aspects of the population. So you must *estimate* characteristics of the population.

expected values

The values you would expect to see if a certain assumption is true. When calculating *chi-square*, the assumption, the null hypothesis, is that the row variable is not related to the column variable.

experiment

A study in which you randomly assign the participants to one of two conditions, administer a treatment (stimulus) to the members of one group, take measurements of both groups, and compare the results. Any differences you see are attributed to the treatment you administered to the group.

experimental group

In an experiment, the group to which the treatment (stimulus) is administered. CONTRAST WITH *control group*.

explained variance

If a pair of variables are related to one another, learning the value of one variable will allow you to narrow down the likely range of values of the other one. The stronger the relationship is, the

more variance in one variable is explained by the relationship with the other one. Also called *variance accounted for*.

explanatory research

The goal of explanatory research is to explain causal relationships between events and circumstances. CONTRAST WITH *descriptive* and *exploratory research*.

exponential

A mathematical statement that requires the use of powers higher or lower than 1.

external validity

The extent to which results based on a sample can be generalized to the population the sample came from. Depends on how representative the sample is or how natural is the setting in which the research was conducted.

extreme value

A value that is considerably higher or lower than the great majority of other values.

face validity

The extent to which a measurement seems like it should be a reasonable measure of that which it is supposed to measure.

fail to reject the null hypothesis

When a difference is so small that it could easily be due to *sampling variability* (i.e. the evidence is not sufficient for you to rule out sampling variability as a reasonable explanation for the difference), you *fail to reject the null hypothesis*.

field research

Research conducted in the natural setting where day-to-day life activities take place. CONTRAST WITH *laboratory study*.

Fisher's r to Z transformation

A transformation that you apply to Pearson's r or to Spearman's rho to allow you to perform tests of the significance of the difference between two correlations.

frequencies

Also known as *counts*. The number of times the various values of a variable occur in a set of data.

frequency distribution

A summary of the values for a variable which shows how many times each of the possible values of the variable occurred. Ordinarily used for *categorical* data, sometimes used for *continuous* data which has been collapsed into a set of *bins*.

gamma

A PRE measure of the strength of association between a pair of ordinal variables. This measure tells you the proportional reduction in errors made predicting the value of one variable when you know the value of the other one.

generalizability

The extent to which results based on a sample can be generalized to the population the sample came from. Also the extent to which a person has what it takes to be made a general.

generalization from a sample to a population

If you interview a random sample of 50 residents of a small town and almost all of them say they drink coffee, you are generalizing from the sample to the population if you conclude that almost everyone who lives in the town drinks coffee.

generative forces

Part of an explanation. The causes, rules, or reasons that explain why something happens.

grounded theory

When the scientist begins by making observations and then constructs a theoretical explanation that would account for the observed patterns. This is *grounded theory*. The logical approach is *data-to-theory* and the logical method used here is *inductive*. The logical pattern is to move from the specific to the general.

heterogeneity

The extent to which the members of a population (or sample) are different from one another. The greater the heterogeneity, the larger the sample is needed for a given level of confidence.

heuristic

A theory has *heuristic value* if it sets the stage for further conceptual developments and empirical research.

histogram

A graphic representation of a *frequency distribution* in which each frequency is represented by a vertical bar whose length is proportional to the frequency.

homogeneity

The extent to which the members of a population (or sample) are similar to one another. The greater the homogeneity, the more accuracy you can have at a given level of confidence.

hypothesis

A prediction based on a theory. A hypothesis states what you would expect to see happen if the theory is true. While theories are abstract and general, hypotheses are concrete and specific.

idiographic

An idiographic study is one that explores a single person or event or situation in detail. Although the researcher doing this kind of work would learn a great deal about the idiosyncratic thoughts or behaviors of the person or situation, this information would apply only to the specific situation that was studied. CONTRAST WITH *nomothetic*.

independent variable

A variable whose value determines or influences the value of another variable. The cause. CONTRAST WITH *dependent variable*.

indirect or inverse or negative relationship

A relationship between a pair of variables in which an increase in the value of one is associated with a decrease in the value of the other. The *correlation* between variables negatively related to one another will be negative. CONTRAST WITH *positive* or *direct relationship*.

induction

Reasoning from specific cases to a general statement. The logical procedure used to develop theories from data. CONTRAST WITH *deduction*.

inferential statistics

Statistics based on sample data and used to estimate population parameters. CONTRAST WITH *descriptive statistics*.

inflection points

The points on a curve in which the direction of curvature switches direction. On a *normal curve*, the points at which the *z-score* is -1.0 and 1.0.

information-theoretic uncertainty

A measure of *dispersion* for *nominal* level data, this measure is based on the relative probabilities of the possible outcomes.

intercept

Also called the *Y-intercept*. The point where the *regression line* crosses the Y axis (the vertical axis). The value of the dependent variable where the independent variable equals zero.

internal validity

The consistency of the set of operational and conceptual definitions and the logical relations among them.

interpretive research

An abstract and conceptual approach to research focusing on the deeper meanings underlying events or situations.

interquartile range (IQR)

The difference between the first and third *quartiles*. The range of values that includes the middle 50% of all cases. A measure of *dispersion* for *ordinal data*.

interval scaling

An approach to measurement in which the cases are ordered and placed onto a scale graduated in equal units. Ex: Fahrenheit and Celsius temperature scales. SEE ALSO *nominal*, *ordinal*, *ratio scaling*.

intervening variable

A variable that stands between an independent variable and a dependent variable and alters the size or direction of the effect the independent variable has on the dependent variable.

inverse relationship

A relationship between a pair of variables in which an increase in the value of one is associated with a decrease in the value of the other. The

correlation between variables negatively related to one another will be negative. CONTRAST WITH *positive* or *direct relationship*.

inverted U-shaped relationship

A relationship between a pair of variables in which the value of the dependent variable is low when the value of the independent variable is low, the dependent variable increases as the independent variable increases up to a point beyond which further increases in the independent variable are paired with decreases in the dependent variable. When plotted, this relationship looks like an upside-down U. This is a curvilinear relationship. CONTRAST WITH *U-shaped relationship*.

kurtosis

A measure of the extent to which a distribution is flattened with a wide peak, or narrow, with a thin pointy peak.

laboratory study

Research conducted in an artificial environment, all aspects of which are controlled by the researcher. CONTRAST WITH *field research*.

lambda

A PRE measure of the strength of the association between a pair of nominal variables. The proportional reduction in errors made predicting the value of one variable when you know the value of the other one turns out to be *Yule's Q*.

law

A *theory* that explains patterns by reference to uncontrollable *antecedents* and employs *causal explanation*. A statement of a universally invariant relationship.

linear

A mathematical relationship that can be plotted as a straight line on ordinary graph paper and expressed in the form $y = b_1x_1 + b_2x_2 + \dots + a$, where the *b*'s and *a* are constants and the *x*'s are variables.

linear regression

A type of analysis of the relationship between a pair of variables in which the relationship can be plotted best as a straight line. CONTRAST WITH *curvilinear regression*.

longitudinal

A study in which you collect data from the same individuals over a period of time so you can see how they change. Example: in a longitudinal study of friendship patterns of high school students, you might follow the same students over the four years of high school, collecting data in each of the four years. CONTRAST WITH *cross-sectional*.

magnitude

Size. Measured on a ratio level scale where "0" means "none."

mapping

In measurement, the procedure used to match numbers to the aspect of reality being measured.

marginal counts

The numbers in a *cross-tabulation* that tell how many cases fall in each row or column.

marginal percentages

The percentages in a *cross-tabulation* that tell what portion of the total are in each row or column. SEE ALSO *row* and *column percentages*.

marginals

The *marginal counts* and *percentages* in a *cross-tabulation*.

mean absolute deviation

A measure of *dispersion* obtained by calculating the mean of the absolute values of the deviation scores. The *standard deviation* is almost always used instead of this measure.

mean

Known informally as the "average." A measure of *central tendency* calculated by dividing the sum of all the values by the number of values. Used for *interval* or *ratio scaled* data.

mean square

An alternate name for *variance*, which is the mean of the squared deviation scores. *Mean square* is the term ordinarily used when discussing *Analysis of Variance (ANOVA)*. SEE ALSO *root mean square*, *between-groups mean square*, *within-groups mean square*, *variance*.

measurement

A procedure by which information concerning some aspect of reality is obtained. Includes sorting into categories, arranging in increasing or decreasing order, counting, assessing amounts and distances.

median

A measure of *central tendency* used for ordinal, interval, ratio data. If there are an odd number of cases, the middle value in a rank-ordered list of values; if there are an even number of cases, halfway between the two middle values.

midpoint of range

Halfway between the highest and lowest values a variable takes. A worthless, useless measure of central tendency.

modal value

The value that occurs more often than all other values.

mode

A measure of central tendency for nominal data. The value that occurs more often than other values.

multimodal distribution

A distribution in which there are two or more distinct *modes*.

multiple regression

A *regression analysis* that simultaneously examines the effects of a set of independent variables on a single dependent variable.

multiple methods, multiple measures

Also known as multi-methods multi-measures. An approach used to assess *construct validity* in which several measurement methods are used to perform several different measures of a *construct*.

necessary cause

An *antecedent condition* (something that happens first) that must happen before a consequent *effect* happens. The cause might not be enough by itself to produce the effect, but if the causal event does not happen, the effect will not happen. SEE ALSO *cause, sufficient cause, necessary and sufficient cause, probabilistic cause*.

necessary and sufficient cause

An *antecedent condition* (something that happens first) that must happen before a consequent *effect* happens and that is enough, by itself, to produce the effect. The effect happens whenever and only when the cause happens. SEE ALSO *cause, necessary cause, sufficient cause, probabilistic cause*.

negative finding

A research study in which the *null hypothesis* is not rejected.

nominal scaling

An approach to measurement in which the cases are sorted into a set of categories that have no *mathematical relationship* between them. SEE ALSO *ordinal, interval, ratio scaling*.

nomothetic

Nomothetic research attempts to discover what systems of laws or principles govern different aspects of reality, while idiographic research is interested in describing only a single event, person, or situation.

non-probability sample

A type of sample in which the probability of being included in the sample is not known for all members of the population.

normal curve

A plot of a *normal distribution*. SEE *normal distribution*. Used for a variety of *descriptive* and *inferential statistics* for *interval* and *ratio scaled* data.

normal distribution

Also known as the bell-shape distribution or the bell curve. A symmetrical distribution in which most cases are clustered near the mean and the further away from the mean you go, the fewer cases there are. The shape of this distribution is described in Table 1. The distribution of large numbers of random events and the means of samples in a *sampling distribution* will be normally distributed.

null hypothesis

The hypothesis that the variables you are examining are not related to one another. Any apparent relationship must be due to *sampling variability*.

number

A symbol composed of one or more of the characters from 0 to 9, possibly including a decimal point or a leading minus sign, and indicating how many objects there are, how much of something there is, where something is located along a continuous scale or how many steps it is along a discrete scale. Numbers have *mathematical* meaning, whereas *numerals* do not and *ordinals* only specify ordering relationships.

numeral

A symbol which is used as a label, which has no mathematical meaning, and which is composed of one or more of the characters from 0 to 9. Example: Wayne Gretzky's "99".

numerator degrees of freedom

ANOVA (*analysis of variance*) requires two *degrees of freedom*. The one for the numerator is the number of groups minus one. Also known as *between-groups* degrees of freedom.

observed and expected

Used in the calculation of *chi-squared*. The *observeds* are the numbers of cases falling into the cells of the table. The *expecteds* are the numbers of cases you would expect to fall in the cells if the row variable is independent of the column variable (i.e. if the *null hypothesis* is true).

one-tailed test

The type of *significance test* you perform when your hypotheses are *directed*. Called "one-tailed" because the *critical region* occupies only one tail of the distribution. COMPARE WITH *two-tailed test*.

open-ended

A type of question that doesn't provide a list of possible answers to choose from.

operational definition

A specification of the operations or procedures that will be done in order to measure a concept.

operative terms

Words used in problem statements or hypotheses to clarify the relationship between the specified classes. Examples: "increasing," "decreasing," "is related to increasing," "immediately," "dimin-

ishes," etc. May also be used to specify the type of relationship you expect to observe. Examples: *linear*, *curvilinear*, *exponential*, *positive*, *negative*, *direct*, *inverse*, *inverted U-shaped*, and *U-shaped*.

ordinal

They look a bit like numbers, but they only tell what order things are. Examples are 1st, 2nd, 3rd. No mathematical operations can be performed on ordinals.

ordinal scaling

An approach to measurement in which the cases are ordered into a set of increasing (or decreasing) categories. SEE ALSO *nominal*, *interval*, *ratio scaling*.

parameter

A summary description of a given variable for the entire population. Example: the mean age of citizens in the country.

parsimony

A parsimonious theory has few concepts and relationships; it is simple. It's not complex, complicated, convoluted, or elaborate.

participant observation

An approach to data collection in which the researcher joins or participates in the social system being studied, thereby getting access to information that might otherwise not be available.

Pearson product-moment correlation

Also known as *Pearson's r*. A measure of the strength of a relationship between a pair of continuous interval- or ratio-scaled variables. Defined as the mean of the cross-products of the z-scores of the two variables.

percentage across, compare down

A method of interpreting crosstabulation tables in which you compare the row percents up and down within a single column. The appropriate method to use when the independent variable is the one that determines which row a case is in.

percentage down, compare across

A method of interpreting crosstabulation tables in which you compare the column percents back and forth within a single row. The appropriate

method to use when the independent variable is the one that determines which column a case is in.

perspicuity

A perspicuous theory is precisely stated, clear, lucid, readily understandable, unambiguous. This is an important characteristic of theories, conceptual definitions, problem statements, and hypotheses.

phrased affirmatively

A problem statement is phrased affirmatively if it states or predicts what *is* or *will be* the case, rather than what is not or what will not be the case.

population

The entire set of individuals, events, units with specified characteristics. ex: citizens of Mexico; students; adults; women; newspaper articles.

population parameters vs. sample statistics

Parameters describe populations, can only be estimated, and use Greek letters for symbols. Statistics describe samples, can be calculated directly, and use roman letters for symbols.

positive research

Research about what *is* the case, rather than *normative* research, which is research about what *ought to be* the case.

posttest

A measurement taken after the application of the experimental manipulation/stimulus.

posttest-only design

An experiment in which all measurements are done after the application of the experimental manipulation/stimulus.

pre-experimental design

An experiment in which one or more of the necessary steps (usually random assignment to conditions) is left out.

predictive validity

A measurement whose predictions turn out to be accurate has predictive validity.

pretest

A measurement taken at the beginning of an experiment before the application of the experimental manipulation/stimulus.

pretest-posttest design

An experiment in which all measurements are done both before and after the application of the experimental manipulation/stimulus.

probabilistic cause

An antecedent condition (something that happens first) that is usually followed by a consequent *effect*. When the causal event happens, it is likely that the effect will occur. SEE ALSO *cause*, *sufficient cause*, *necessary cause*.

probability samples

A type of sample in which the probability that any individual will be included in the sample is known. The only kind of sample that can be used to make valid generalizations.

problem statements

A hypothesis that describes a relationship which is the subject of a research study or a question that a research study is designed to answer. SEE ALSO *research hypothesis*, *research question*.

proportionate reduction of error (PRE)

The percentage by which errors of prediction in the value of one variable can be reduced by knowledge of the value of a second variable.

quasi-experimental design

Almost an experiment. An *experiment* missing a critical piece — usually one in which cases are not randomly assigned to conditions.

quota sample

A type of *non-probability* sample in which you select cases in order to match the population's proportions on a number of characteristics. Because it is not possible to match a population on all relevant characteristics, quota samples are not likely to be representative of the population.

r-squared

Pearson's r squared. A measurement of percentage of *variance* in one variable accounted for or explained by the relationship to a second variable.

random variance

Variability in the value of a dependent variable that cannot be explained or accounted for by the independent variable(s).

random assignment to conditions

In experiments, individuals are randomly assigned to control or experimental groups which guarantees that there are no systematic differences between the groups.

random error

The fluctuation or varying of measures due to chance. Also called 'random' variance. The variance left over in a set of measures after all known sources of systematic variance have been removed from the measures.

random sample

A sample in which individuals are selected randomly from the population. Characterized by equal probability and independence of selection.

random selection

A process in which members are drawn from a population where each member of the population has the same chance of being selected as all other members, and where the selection of one person has no influence on anyone else's chances of being selected.

range

The difference between the highest and lowest values for a variable. The weakest measure of *dispersion*. A place where the skies are not cloudy all day.

rank-order correlation

See Spearman's rho.

ratio scaling

An approach to measurement in which the cases are ordered and placed onto a scale graduated in equal units where the value "0" indicates "none." SEE ALSO *nominal, ordinal, interval scaling*.

reason

A type of generative force which refers to the preexisting goals, needs, and desires of a person that explain patterns of behavior.

regression

A method of analysis in which the relationship between independent and dependent variables are expressed as a regression equation. If the

equation is *linear*, it is linear regression. If the equation is *curvilinear*, it is curvilinear regression. There may be several independent variables in a regression equation.

regression line

The line defined by a regression equation.

regression equation

An equation that uses independent variable(s) to predicts the value of a dependent variable. A *linear* regression equation is usually expressed in the form $y = b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + a$, where the *b*'s and *a* are constants and the *x*'s are variables.

reject the null hypothesis

When a difference is so large that it could not be due to *sampling variability*, the *null hypothesis*, which says that the difference is due to sampling variability, is rejected.

reliability

A measure is reliable to the extent to which repetitions of the measurement give the same result.

representativeness

The extent to which a sample has the same distribution of characteristics as the population from which it was selected. A representative sample is necessary if you want to make statements or conclusions about the population with any degree of confidence.

research hypothesis

A problem statement expressed as a hypothesis. A hypothesis that describes a relationship whose validity is the subject of a research study.

research question

A question that a research study is designed to answer. May be a question requiring a description ("What is the non-verbal content of cigarette advertising in national magazines?") or a question requiring a causal explanation ("Are ads that attempt to use fear of cancer and heart disease to convince people to stop smoking more or less effective than ads that take other approaches?").

residuals

The variability in values of the dependent variable that is not accounted for (i.e. left over) after the effect of the independent variable(s) has been taken into consideration. Errors in predictions made with a *regression equation*.

root mean square

The square root of the mean of the squared differences between the individual values and the mean of the values. Also known as *standard deviation*.

row percents

In a *cross-tabulation*, these are the numbers that tell you what percent of the cases in a row fall in a given column. The sum of the row percents in a row should be 100%. SEE ALSO *column percents*.

rule

A social norm, custom, or tradition that specifies expected or required behavior.

rule-based explanation

A theory that explains by reference to norms, social customs, or tradition.

sample

A subset of the population from which data is collected and used as a basis for making statements about the entire population.

sampling distribution

If you draw all possible samples of a given size from a population and calculate the value of a statistic (i.e. the mean) for each sample, the resulting list of values will be the *sampling distribution* of that statistic. Certain characteristics of the sampling distribution allow the calculation of things like confidence intervals or levels. Probability samples are generally required for these calculations and estimates to be made.

sampling distribution of differences between sample means

If you draw all possible pairs of samples of a given size from a population and calculate the difference between the means of each pair of samples, the resulting list of values will be the *sampling distribution of differences between sample means*.

sampling distribution of sample means

If you draw all possible samples of a given size from a population and calculate the mean for each sample, the resulting list of values will be the *sampling distribution of sample means*.

sampling frame

The list or quasi-list of the members of the population from which your sample is to be selected. If the sample is to be truly representative of the population, the sampling frame should include all members of the population.

sampling interval

In *systematic sampling*, one plus the number of individuals you skip before you select the next member of the sample. If you are taking every k^{th} person, you skip $k-1$ and take one, and the sampling interval is k . To determine the sampling interval, divide the population size by the desired sample size.

sampling ratio

The proportion of a population taken into the sample. The sample size divided by the population size.

sampling variability

If more than one sample were taken from a single population, each would have a different value for a given statistic. The differences between samples is *sampling variability*. The more heterogeneous the population is, the greater the sampling variability for samples of a given size.

sampling unit

An element or set of elements considered for selection in a sample. In a simple random sample, the *sampling unit* is the individual. In a cluster sample, the sampling unit will be a cluster of individuals.

self-selection

When individuals volunteer to participate in a research study or decide whether they will be in the experimental or control group, they are said to be self-selected. Self-selection generally is avoided because it leads to bias.

significance test

A test to determine the extent to which the apparent relationship you see in your data could be due to *sampling variability* alone.

skewness

The extent to which the majority of cases are clustered in the high or low end of the scale, rather than in the middle.

slope of the regression line

Indicates how large a change in the dependent variable is predicted for a given change in the independent variable. A positive slope indicates that an increase in the independent variable will be associated with an increase in the dependent variable. A negative slope indicates that an increase in the independent variable will be associated with a decrease in the dependent variable.

Solomon four group design

An experimental design in which there are four groups. Two *control groups* and two *experimental groups*, one of each with and one without a *pretest*. Equivalent to a combination of *posttest only* and *pretest-posttest* designs.

Somers' d

A measure of the strength of association between a pair of ordinal variables. Very similar to gamma, except Somers' d takes tied pairs, ignored by gamma, into account. d is a more conservative measure than gamma, and it does not have a PRE interpretation.

Spearman's rho (r_s)

A measure of the strength of a relationship between a pair of rank-ordered variables. With this measure, there will be a perfect relationship if the object having the highest score on one variable also has the highest score on the other variable; the one having the second-highest score on one variable has the second-highest score on the other variable, and so on. This measure does not have a PRE interpretation. It can be used in the same tests of significance as Pearson's r .

standard deviation

The most often used measure of dispersion for interval- or ratio-scaled data. Defined as the

square root of the mean of the squared deviation scores. Also known as the *root mean square*.

standard deviation of sampling distribution

The standard deviation of a sampling distribution is a *standard error*.

standard errors

The standard deviation of a sampling distribution is a *standard error*. For example, the standard deviation of the sampling distribution of sample means is the *standard error of the mean*.

standard error as measure of sampling variability

Because standard errors are measures of the amount of dispersion of values in a sampling distribution, they are measures of the extent to which the value of a statistic varies from sample to sample—i.e. a measure of *sampling variability*.

standard error of the difference between means

The standard deviation of the sampling distribution of differences between sample means is the *standard error of the differences between means*, sometimes abbreviated SEDBM.

standard error of the mean

The standard deviation of the sampling distribution of sample means is the *standard error of the mean*, sometimes abbreviated SEM.

standard error of the proportion

The standard deviation of the sampling distribution of sample proportions is the *standard error of the proportion*.

standard error as SD of sampling distribution

All standard errors are standard deviations of sampling distributions.

standard scores

When *deviation scores* are divided by the *standard deviation*, you get standard scores. Also known as *z-scores*.

static group design

A two-group *quasi-experimental design* in which individuals are not randomly assigned to conditions.

statistic

A summary description of a given variable for the sample. Example: the mean age of people in the

sample. Statistics are the data upon which estimates of population parameters are based.

statistical significance

The extent to which the apparent relationship you see in your data is not likely to be due to *sampling variability* alone.

statistically significant difference

A difference so large that it could not have been caused by sampling variability.

stimulus

Something that is done to the individuals in the experimental group but not to the individuals in the control group.

stratified sample

A sampling procedure that divides the population into relatively homogeneous groups (strata) before samples are drawn from each group. May use random, systematic, or cluster methods to draw the samples from the strata.

stratum

A subset of a population. Used in stratified sampling methods. To be useful, a stratum must be comprised of individuals that are homogeneous in some way that is relevant to the study being conducted.

sufficient cause

An *antecedent condition* (something that happens first) that is always followed by a consequent *effect*. The cause by itself is enough to produce the effect, but other things may also produce the effect. SEE ALSO *cause, necessary cause, necessary and sufficient cause, probabilistic cause*.

sum of squares (SS)

The sum of the squared differences between the individual values and the mean of the values. Used in the calculation of *variance, standard deviation, and analysis of variance*.

systematic error

In measurement, bias or systematic distortion. When a measurement is biased, it is consistently distorted in a particular direction. CONTRAST WITH *random error*.

systematic variance

Variance in one variable that can be explained or accounted for by a relationship to other variables.

teleological explanation

An explanation that makes reference to goals or subjective reasons for acting.

theory

A set of statements about a number of concepts and the relations between them. A formal statement of definitions and propositions concerning the relations among a set of constructs created for the purposes of explanation, understanding, prediction, and control of phenomena.

tradeoff between precision and confidence

When you estimate a population parameter on the basis of a sample statistic, you can make a very precise prediction with low confidence or you can make a less precise prediction with higher confidence. When the precision is increased, the confidence goes down.

two-tailed test

The type of *significance test* you perform when your hypotheses are *not* directed. Called “two-tailed” because the *critical region* occupies both tails of the distribution. COMPARE WITH *one-tailed test*.

Type I error

The error you make if you reject a true *null hypothesis* — that is, if you conclude that the difference in your data indicates a real difference in the population, when in fact there is none (i.e., when the difference in your sample data was due to *sampling variability*).

Type II error

The error you commit if you fail to reject a false *null hypothesis* — that is, if you conclude that the pattern in your data is due to *sampling variability* when, in fact, it is because your sample came from a population in which the same pattern was present.

U-shaped relationship

A relationship between a pair of variables in which the value of the dependent variable is high

when the value of the independent variable is low, the dependent variable decreases as the independent variable increases up to a point beyond which further increases in the independent variable are paired with increases in the dependent variable. When plotted, this relationship looks like a U. This is a curvilinear relationship. CONTRAST WITH *inverted U-shaped relationship*.

undirected hypothesis

A hypothesis in which the direction of a difference is not specified. Ex: men and women drink different amounts of coffee. CONTRAST WITH *directed hypothesis*.

unexplained variance

SEE *Random variance, error variance, residuals*.

unit of analysis

The smallest unit or individual that is accepted in a sample. Usually our sampling unit will be the individual person, although you may use advertisements, movies, episodes of television programs, conversations, organizations, families, etc.

unobtrusive

An unobtrusive measurement method is one that can be carried out without disturbing or intruding on the people being measured, one that is not noticeable or drawing of attention.

validity

In measurement, the extent to which the measurement is sensitive only to differences in what it is supposed to be measuring.

variability or spread around the mean

If the values in a sample are not all the same, there will be some variability from case to case. The values will be spread out around the mean.

variable

Variables are empirical indicators, or symptoms, of *constructs*. A variable is something concrete that you can observe, and by its appearance you can tell whether the concept is present or absent or to what extent it is present.

variance

A measure of the extent to which a set of values are different from one another, calculated as the

mean of the squared differences between the values and the mean of the set of values. Sometimes called the *mean square*.

variance accounted for

If a pair of variables are related to one another, learning the value of one variable will allow you to narrow down the likely range of values of the other one. The stronger the relationship is, the more variance in one variable is accounted for by the relationship with the other one. ALSO CALLED *explained variance*.

within-groups degrees of freedom

ANOVA (*analysis of variance*) requires two *degrees of freedom*. The one for the denominator is the number of cases minus the number of groups. Also known as *denominator degrees of freedom*.

within-groups mean square

(MS_w) Used in ANOVA (*analysis of variance*); a measure of the variability within the groups, calculated as the *within-groups sum of squares* divided by the *within-groups degrees of freedom*.

within-groups sum of squares (SS_w)

Used in ANOVA (*analysis of variance*); a measure of the variability within the groups. Calculate the ordinary *sum of squares* for each group, and add them together: $SS_1 + SS_2 + \dots + SS_k$

within-groups variance

Same as *within-groups mean square*.

Yule's Q

A measure of the strength of the association between a pair of nominal variables. Like lambda, Q also has a PRE interpretation. It is the proportional reduction in errors made predicting the value of one variable when you know the value of the other one. Q indicates to what degree the pairs tend to fall in one diagonal more than another. When the variables are independent, the pairs are equally likely to fall in each diagonal.

z-score

When *deviation scores* are divided by the *standard deviation*, you get z-scores. Also known as *standard scores*.

Notes