

Glossary

abduction | An approach to research that combines both inductive and deductive elements.

abstract | A short summary of a text written from the perspective of a reader rather than from the perspective of an author.

addition theorem | The theorem addressing the determination of the probability of a given outcome occurring at least once across a series of trials; it is determined by adding the probability of each possible series of outcomes together.

analytic coding | Coding designed to move analysis towards the development of themes and findings.

anecdotalism | When researchers choose particular stories or incidents specifically to illustrate a point rather than because they are representative of the data in general.

ANOVA | A statistical test designed to measure differences between groups.

antecedent variable | A variable that is hypothesized to affect both the independent variable and the dependent variable.

applied research | Research designed to address a specific problem.

archive | A repository of documents, especially those of historical interest.

association | The situation in which variables are able to be shown to be related to one another.

attributes | The possible levels or response choices of a given variable.

bar chart | Also called bar graphs, these graphs display data using bars of varying heights.

basic research | Research designed to increase knowledge, regardless of whether that knowledge may have any practical application.

bell curve | A graph showing a normal distribution—one that is symmetrical with a rounded top that then falls away towards the extremes in the shape of a bell

beta | The standardized regression coefficient. In a bivariate regression, the same as Pearson's r ; in a multivariate regression, the correlation between the given independent variable and the dependent variable when all other variables included in the regression are controlled for.

binary | Consisting of only two options. Also known as dichotomous.

bivariate analyses | Quantitative analyses that tell us about the relationship between two variables.

block quote | A quotation, usually one of some length, which is set off from the main text by being indented on both sides rather than being placed in quotation marks.

CAQDAS | An acronym for "computer-aided qualitative data analysis software," or software that helps to facilitate qualitative data analysis.

causation | A relationship between two phenomena where one phenomenon influences, produces, or alters another phenomenon.

central limit theorem | The theorem that states that if you take a series of sufficiently large random samples from the population (replacing people back into the population so they can be reselected each time you draw a new sample), the distribution of the sample means will be approximately normally distributed.

chronology | A list or diagram of events in order of their occurrence in time.

cliques | An exclusive circle of people or organizations in which all members of the circle have connections to all other members of the circle.

closed coding | Coding in which the researcher developed a coding system in advance based on their theory, hypothesis, or research question.

code tree | A hierarchically-organized coding system.

code weights | Elements of a coding strategy that help identify the intensity or degree of presence of a code in a text.

codebooks | Documents that lay out the details of measurement. Codebooks may be used in surveys to indicate the way survey questions and responses are entered into data analysis software. Codebooks may be used in coding to lay out details about how and when to use each code that has been developed.

codes | Words or phrases that capture a central or notable attribute of a particular segment of textual or visual data.

coding | The process of assigning observations to categories.

coding (in quantitative methods) | Assigning numerical variables to replace the names of variable categories.

cognitive map | Visualizations of the relationships between ideas.

collinearity | The condition where two independent variables used in the same analysis are strongly correlated with one another.

column marginal | The total number of cases in a given column of a table.

concept coding | Coding using words or phrases that represent concepts or ideas.

confidence interval | A range of estimates into which it is highly probable that an unknown population parameter falls.

confidence level | The probability that the sample statistics we observe holds true for the larger population.

continuous variable | A variable measured using numbers, not categories, including both interval and ratio variables. Also called a scale variable.

control variable | A variable that is neither the independent variable nor the dependent variable in a relationship, but which may impact that relationship.

controlling a relationship | Examining a relationship between two variables while eliminating the effect of variation in an additional variable, the control variable.

crosstabulation | An analytical method in which a bivariate table is created using discrete variables to show their relationship.

data cleaning | The process of examining data to find any errors, mistakes, duplications, corruptions, omissions, or other issues, and then correcting or removing data as is appropriate.

data display | Tables, diagrams, figures, and related items that enable researchers to visualize and organize data in ways that permit the perception of patterns, comparisons, processes, or themes.

data management | The process of organizing, preserving, and storing data so that it can be used effectively.

data reduction | The process of reducing the volume of data to make it more usable while maintaining the integrity of the data.

decision tree | A diagram that lays out the steps taken to reach decisions.

deductive | An approach to research in which researchers begin with a theory, then collect data and use that data to test their theory.

deductive coding | Coding in which the researcher developed a coding system in advance based on their theory, hypothesis, or research question.

degrees of freedom | The number of cells in a table that can vary if we know something about the row and column totals of that table, calculated according to the formula $(\# \text{ of columns} - 1) * (\# \text{ of rows} - 1)$.

denominator | The expression below the line in a fraction; the entity used to divide another entity in a formula.

dependent variable | A variable that is affected or influenced by (or depends on) another variable; the effect in a causal relationship.

descriptive coding | Coding that relies on nouns or phrases describing the content or topic of a segment of text.

descriptive statistics | Statistics used to describe a sample.

descriptor | A category in an information storage system; more specifically in Dedoose, a characteristic of an author or entire text. Also, the word used to indicate that category or characteristic.

deviant case | A case that appears to be an exception to commonly-understood patterns or explanations.

dichotomous | Consisting of only two options. Also known as binary.

direction | How categories of an independent variable are related to categories of a dependent variable.

discrete variable | A variable measured using categories rather than numbers, including binary/dichotomous, nominal, and ordinal variables.

dramaturgical coding | Coding that treats texts as if they are scripts for a play.

dummy variable | A two-category (binary/dichotomous) variable that can be used in regression or correlation, typically with the values 0 and 1.

edge | The line connecting nodes in a network diagram; such lines represent real-world relationships or linkages.

elaboration | A term used to refer to the process of controlling for a variable.

elimination of alternatives | In relation to causation, the requirement that for a causal relationship to exist, all possible explanations other than the hypothesized independent variable have been eliminated as the cause of the dependent variable.

emotion codes | Codes indicating emotions discussed by or present in the text, sometimes indicated by the use of emoji/emoticons.

empirical | That which could hypothetically be shown to be true or false; statements about reality rather than opinion.

epistemology | The philosophical study of the nature of knowledge.

ethics (in research) | Standards for the appropriate conduct of research that seek to ensure researchers treat human participants in research appropriately and do not harm them and that scientific misconduct is avoided.

ethnography | A research method in which the researcher is a participant in a social setting while simultaneously observing and collecting data on that setting and the people within it.

evaluation coding | A coding system used to indicate what is or is not working in a program or policy.

exhaustive | The property of a variable which has a category for everyone.

extraneous variable | A variable that impacts the dependent variable but is not related to the independent variable.

face validity | The extent to which measures appear to measure that which they were intended to measure.

feminism | A perspective rooted in the idea that explorations and understandings of gendered power relations should be at the root of inquiry and action.

fieldnotes | Qualitative notes recorded by researchers in relation to their observation and/or participation of participants, social circumstances, events, etc. in which they document occurrences, interactions, and other details they have observed in their observational or ethnographic research.

first-cycle coding | Coding that occurs early in the research process as part of a bridge from data reduction to data analysis.

flow chart | A diagram of a sequence of operations or relationships.

focus group | A research method in which multiple participants interact with each other while being interviewed.

focused coding | Selective coding designed to orient an analytical approach around certain ideas.

frequency distribution | An analysis that shows the number of cases that fall into each category of a variable.

gamma | A measure of the direction and strength of a crosstabulated relationship between two ordinal-level variables.

General Social Survey | A nationally-representative survey on social issues and opinions which has been carried out roughly every other year since 1972. Also known as the GSS.

generalizability | The degree to which a finding based on data from a sample can be assumed to be true for the larger population from which the population was drawn.

genre | A classification of written or artistic work based on form, content, and style.

gerunds | Verb forms that end in -ing and function grammatically in sentences as if they are nouns.

grounded theory | An inductive approach to data collection and data analysis in which researchers strive to generate a conception of how participants understand their own lives and circumstances.

Hawthorne effect | When research participants modify their behavior, actions, or responses due to their awareness that they are being observed.

histogram | A graph that looks like a bar chart but with no spaces between the bars, it is designed to display the distribution of continuous data by creating rectangles to represent equally-sized groups of values.

hypothesis | A statement of the expected or predicted relationship between two or more variables.

In vivo coding | Coding that relies on research participants' own language.

independent variable | A variable that may affect or influence another variable; the cause in a causal relationship.

index variable | A composite variable created by combining information from multiple variables.

inductive | A research approach in which researchers begin by collecting data and then use this data to build theory.

inductive coding | Coding in which the researcher develops codes based on what they observe in the data they have collected.

inferential statistics | Statistics that permit researchers to make inferences (or reasoned conclusions) about the larger populations from which a sample has been drawn.

inter-rater reliability | The extent to which multiple raters or coders assign the same or a similar score, code, or rating to a given text, item, or circumstance.

interaction term | A variable constructed by multiplying the values of other variables together so as to make it possible to look at their combined impact.

interpretivism | A philosophy of research that assumes all knowledge is constructed and understood by human beings through their own individual and cultural perspectives.

interval variable | A variable with adjacent, ordered categories that are a standard distance from one another, typically as measured numerically.

intervening variable | A variable hypothesized to intervene in the relationship between an independent and a dependent variable; in other words, a variable that is affected by the independent variable and in turn affects the dependent variable.

interview | A research method in which a researcher asks a participant open-ended questions.

iterative | A process in which steps are repeated.

Kappa | A measure of association especially likely to be used for testing interrater reliability.

kurtosis | How sharp the peak of a frequency distribution is. If the peak is too pointed to be a normal curve, it is said to have positive kurtosis (or "leptokurtosis"). If the peak of a distribution is too flat to be normally distributed, it is said to have negative kurtosis (or platykurtosis).

latent coding | Interpretive coding that focuses on meanings within texts.

leptokurtosis | The characteristic of a distribution that is too pointed to be a normal curve, indicated by a positive kurtosis statistic.

levels of measurement | Classification of variables in terms of the precision or sensitivity in how they are recorded.

line of best fit | The line that best minimizes the distance between itself and all of the points in a scatterplot.

linear relationship | A relationship in which a scatterplot will produce a reasonable approximation of a straight line (rather than something like a U or some other shape).

logistic regression | A type of regression analysis that uses the logistic function to predict the odds of a particular value of a binary dependent variable.

manifest coding | Coding of surface-level and/or easily observable elements of texts.

margin of error | A suggestion of how far away from the actual population parameter a sample statistic is likely to be.

matrices | Tables with rows and columns that are used to summarize and analyze or compare data.

mean | The sum of all the values in a list divided by the number of such values.

measures of central tendency | A measure of the value most representative of an entire distribution of data.

measures of dispersion | Statistical tests that show the degree to which data is scattered or spread.

median | The middle value when all values in a list are arranged in order.

metadata | Data about other data.

mode | The category in a list that occurs most frequently.

multiple regression | Regression analysis looking at the relationship between a dependent variable and more than one independent variable.

multiplication theorem | The theorem in probability about the likelihood of a given outcome occurring repeatedly over multiple trials; this is determined by multiplying the probabilities together.

multivariate analyses | Quantitative analyses that explores relationships involving more than two variables or examines the impact of other variables on a relationship between two variables.

multivariate regression | Regression analysis looking at the relationship between a dependent variable and more than one independent variable.

mutually exclusive | The characteristic of a variable in which no one can fit into more than one category, such as age categories 5-10 and 11-15 (rather than 5-10 and 10-15, as this would mean ten-year-olds fit into two categories).

network diagram | A visualization of the relationships between people, organizations, or other entities.

NHST | Null hypothesis significance testing.

nodes | Points in a network diagram that represents an individual person, organization, idea, or other entity of the type the diagram is designed to show connections between.

nominal variable | A variable whose categories have names that do not imply any order.

normal distribution | A distribution of values that is symmetrical and bell-shaped.

null hypothesis | The hypothesis that there is no relationship between the variables in question.

null hypothesis significance testing | A method of testing for statistical significance in which an observed relationship, pattern, or figure is tested against a hypothesis that there is no relationship or pattern among the variables being tested

objectivity | The ability to evaluate something without individual perspectives, values, or biases impacting the evaluation.

observational research | A research method in which the researcher observes the actions, interactions, and behaviors of people.

open coding | Coding in which the researcher develops codes based on what they observe in the data they have collected.

ordinal variable | A variable with categories that can be ordered in a sensible way.

organizational chart | A diagram, usually a flow chart, that documents the hierarchy and reporting relationships within an organization.

original relationship | The relationship between an independent variable and a dependent variable before controlling for an additional variable.

p value | The measure of statistical significance typically used in quantitative analysis. The lower the p value, the more likely you are to reject the null hypothesis.

paradigm | A set of assumptions, values, and practices that shapes the way that people see, understand, and engage with the world.

partial | Shorter term for a partial relationship.

partial relationship | A relationship between an independent and a dependent variable for only the portion of a sample that falls into a given category of a control variable.

participant-observation | A research method in which the researcher observes social interaction while themselves participating in the social setting.

participants | People who participate in a research project or from or about whom data is collected.

Pearson's chi-square | A measure of statistical significance used in crosstabulation to determine the generalizability of results.

Pearson's r | A measure of association that calculates the strength and direction of association between two continuous (interval and/or ratio) level variables.

Pie charts | Circular graphs that show the proportion of the total that is in each category in the shape of a slice of pie.

platykurtosis | The characteristic of a distribution that is too flat to be a normal curve, indicated by a negative kurtosis statistic.

population | A group of cases about which researchers want to learn something; generally, members of a population share common characteristics that are relevant to the research, such as living in a certain area, sharing a certain demographic characteristic, or having had a common experience.

population parameter | A quantitative measure of data from a population.

positionality | An individual's social, cultural, and political location in relation to the research they are doing.

positivism | A view of the world in which knowledge can be obtained through logic and empirical observation and the world can be subjected to prediction and control.

pragmatism | A philosophy that suggests that researchers can adapt elements of both objectivist and interpretivist philosophies.

probability | How likely something is to happen; also, a branch of mathematics concerned with investigating the likelihood of occurrences.

probability sample | A sample that has been drawn to give every member of the population a known (non-zero) chance of inclusion.

process coding | Coding in which gerunds are applied to actions that are described in segments of text.

process diagrams | Visualizations that display the relationships between steps in a process or procedure.

qualitative data analysis | Data analysis in which the data is not primarily numeric, for instance based on words or images.

quantification | The transformation of non-numerical data into numerical data.

quantitative data analysis | Data analysis in which the data is numerical.

R squared | The square of the regression coefficient, which tells analysts how much of the variation in the dependent variable has been explained by the independent variable(s) in the regression.

R2 change | The change in the percent of the variance of the dependent variable that is explained by all of the independent variables together when comparing two different regression models

random sample | A sample in which all members of the population have an equal probability of being selected.

range | The highest category in a list minus the lowest category.

ratio level variable | A numerical variable with an absolute zero which can also be multiplied and divided.

reflexivity | A continual reflection on the research process and the researcher's role within that process designed to ensure that researchers are aware of any thought processes that may impact their work.

regression | A statistical technique used to explore how one variable is affected by one or more other variables.

regression line | The line that is the best fit for a series of data, typically as displayed in a scatterplot.

relationship (between variables) | When certain categories of one variable are associated, or go together, with certain categories of the other variable(s).

reliability | The extent to which multiple or repeated measurements of something produce the same results.

repeatability | The extent to which a researcher can repeat a measurement and get the same result.

replicability | The extent to which a research study can be entirely redone and yet produce the same overall findings.

replicate | Repeating a research study with different participants.

representativeness | The degree to which the characteristics of a sample resemble those of the larger population.

reproducibility | The extent to which a new study designed to test the same hypothesis or answer the same question ends up with the same findings as the original study.

respondents | People who participate in a research project or from or about whom data is collected.

rough coding | Coding for data reduction or as part of an initial pass through the data.

row marginal | The total number of cases in a given row of a table.

sample | A subset of cases drawn or selected from a larger population.

sample statistics | Quantitative measures of data from a sample.

sampling error | Measurement error created due to the fact that even properly-constructed random samples are do not have precisely the same characteristics as the larger population from which they were drawn.

saturation | The point in the research process where continuing to engage in data collection no longer yields any new insights. Can also be used to refer to the same point in the literature review process.

scale variable | A variable measured using numbers, not categories, including both interval and ratio variables. Also called a continuous variable.

scatterplot | A visual depiction of the relationship between two interval level variables, the relationship between which is represented as points on a graph with an x-axis and a y-axis.

second-cycle coding | Analytical coding that occurs later in the data analysis process.

significance (statistical) | A statistical measure that suggests that sample results can be generalized to the larger population, based on a low probability of having made a Type 1 error.

simple linear regression | A regression analysis looking at a linear relationship between one independent and one dependent variable.

skewness | An asymmetry in a distribution in which a curve is distorted either to the left or the right, with positive values indicating right skewness and negative values indicating left skewness.

social data analysis | The analysis of empirical data in the social sciences.

social responsibility (in research) | The extent to which research is conducted with integrity, is trustworthy, is relevant, and meets the needs of communities.

spurious | The term used to refer to relationship where variables seem to vary in relation to one another, but where in fact no causal relationship exists.

standard deviation | A measure of variation that takes into account every value's distance from the sample mean.

standard error | A measure of accuracy of sample statistics computed using the standard deviation of the sampling distribution.

standpoint | The particular social position in which a person exists and in which their understandings of the world are rooted.

strength (of relationship) | A measure of how well we can predict the value or category of the dependent variable for any given unit in our sample based on knowing the value or category of the independent variable(s).

string | A data type that represents non-numerical data; string values can include any sequence of letters, numbers, and spaces.

structural coding | Coding that indicates which research question or hypothesis is being addressed by a given segment of text.

subjects | People who participate in a research project or from or about whom data is collected.

summarization | The process of creating abridged or shortened versions of content or texts that still keep intact the main points and ideas they contain.

table | A display that uses rows and columns to show information.

temporal order | The order of events in time; in relation to causation, the fact that independent variables must occur prior to dependent variables.

the elaboration model | A typology developed by Paul Lazarsfeld for the possible analytical outcomes of controlling for a variable.

themes | Concepts, topics, or ideas around which a discussion, analysis, or text focuses.

thick description | A detailed narrative account of social action that incorporates rich details about context and meaning such that readers are able to understand the analytical meaning of the description.

timeline | A diagram that lays out events in order of when they occurred in time.

trace analysis | Research that uses the traces of life people have left behind as data, as in archeology.

triangulation | The use of multiple methods, sites, populations, or researchers in a project, especially to validate findings.

type 1 error | The error made if one infers that a relationship exists in a larger population when it does not really exist; in other words, a false positive error.

type 2 error | The error you make when you do not infer a relationship exists in the larger population when it actually does exist; in other words, a false negative conclusion.

typologies | Classification systems.

univariate | Using one variable.

univariate analyses | Quantitative analyses that tell us about one variable, like the mean, median, or mode.

validity | The degree to which research measurements accurately reflect the real phenomena they are intended to measure.

values coding | Coding that relies on codes indicating the perspective, worldview, values, attitudes, and/or beliefs of research participants.

variable | A characteristic that can vary from one subject or case to another or for one case over time within a particular research study.

variance | A basic statistical measure of dispersion, the calculation of which is necessary for computing the standard deviation.

versus coding | Coding that relies on a series of binary oppositions, one of which must be applied to each segment of text.

voice | The style or personality of a piece of writing, including such elements as tone, word choice, syntax, and rhythm.

word cloud | Visual display of words in which the size and boldness of each word indicates the frequency with which it appears in a body of text.

Yule's Q | A measure of the strength of association use with binary variables

Z score | A way of standardizing data based on how many standard deviations away each value is from the mean.