

Part One Concept Reference

Central concepts and notation for Chapters 1–14

Jinsong Chen

Table of contents

Chapter 1: The Quantitative Research Journey	2
Chapter 2: Data, Distributions, and Variation	3
Chapter 3: Estimates, Uncertainty, and Statistical Tests	6
Chapter 4: Group Comparisons with Analysis of Variance	8
Chapter 5: Description and Basic Inference with R	9
Chapter 6: Group Comparisons and Statistical Modelling with R	10
Chapter 7: Simple Linear Regression	10
Chapter 8: Multiple Regression I: Conditioning, Checks, and Coordinated Questions	12
Chapter 9: Multiple Regression II: Categorical Predictors, Model Comparison, Nonlinearity, and Robust Uncertainty	13
Chapter 10: Interaction and Moderation	14
Chapter 11: Binary Logistic Regression	16
Chapter 12: Extending the GLM: Nominal, Ordinal, and Count Outcomes	18
Chapter 13: Prediction, Cross-Validation, and Regularization	20
Chapter 14: Regression Pitfalls and Causal Design	21
Argument Chapter A: How to Read a Quantitative Study as an Argument	23
Argument Chapter B: Extending the Argument-Based Approach to Quantitative Studies	23
Procedures and Their Conditions	24
Shared Notation	27

This reference gathers the Key terms tables of Chapters 1–14 and of Argument Chapters A and B, followed by the procedures of Chapters 3 and 4 and the notation that several chapters share. Each table appears under its chapter’s heading and lists the concepts that chapter explains principally: each term with its common alternatives, and its meaning or the distinction that matters most. The reference is meant to be consulted while reading rather than read through: a term’s first use, its worked example, and the reasoning around it stay in the chapter that teaches it. Chapters 1–9 are available on the website; the tables for Chapters 10–14 point forward to chapter readings that are not yet available.

A few conventions govern the rows. Names that pick out the same concept share one entry, so *mean* and *arithmetic mean* appear together, and *outcome*, *response*, and *dependent variable* are explained as one analytical role. A more specific role, such as treatment, exposure, or

confounder, carries additional design or substantive meaning that the study itself must supply. Related concepts with different meanings, such as mean and median or SD and SE, keep separate entries. A term is listed under the chapter that explains it principally rather than the chapter that first mentions it, so a brief early introduction and a later full account can coexist: Chapter 1 introduces cases and samples through its studies, while Chapter 2 develops their use in reading records and calculations. Argument Chapters A and B, *How to Read a Quantitative Study as an Argument* and *Extending the Argument-Based Approach to Quantitative Studies*, extend the reading of one question-level argument to several statistical strands and coordinated questions; the notation table abbreviates them as *Arg. A* and *Arg. B*.

Chapter 1: The Quantitative Research Journey

Table 1

Key terms of Chapter 1.

Term and common alternatives	Meaning
Question-answer pair	What a study asks and the answer it eventually offers for that question. Each separately owed substantive answer has its own argument.
Traceable information	Specific study content through which an argument component's function can be identified at a documentary location. Identifying that contribution precedes judging its adequacy.
Estimand	The target quantity sought for the named units, conditions, outcome, and scope. It can concern a fixed dataset, a population, a comparison, or prediction performance.
Outcome variable (response variable; dependent variable)	The variable an analysis seeks to describe, compare, or predict. These alternative names share that role in many analyses; a variable's role depends on the question.
Predictor variable (explanatory variable; independent variable)	A variable used to represent an outcome's relationship with other information. These analytical-role labels alone establish neither a causal effect nor statistical independence.
Sampling (recruitment; selection into a study)	How units come to enter a study, by probability selection, volunteering, or availability. It is contrasted with assignment, which concerns the conditions allocated to participating units.
Assignment (allocation)	The allocation of study conditions to participating units, by any mechanism. Actual treatment received may differ from the assigned condition.
Random sampling (probability sampling)	Selecting units from a defined population or sampling frame through a probability mechanism that gives every eligible unit a known, nonzero chance of selection, not necessarily the same chance. "Random selection" alone can describe other chance-based decisions.
Random assignment (random allocation)	Allocating study conditions to participating units through a specified probability mechanism. The assigned units may be individuals or groups; realized groups need not be identical.

Term and common alternatives	Meaning
Causal effect (counterfactual contrast)	The difference between the outcomes that would occur for the same units or target population under two specified conditions, only one of which is observed for each unit. The contrast may be zero; interpreting a comparison causally needs design and identification support.
Confounding	Mixing an exposure's effect with effects of other factors that influence both exposure and outcome. Such a factor is a confounder, a specific causal role.
External validity (generalizability)	Support for extending findings to other people, settings, or conditions. The intended extension determines which similarities and differences matter.
Unit	The entity selected, assigned, measured, or analyzed in a particular operation. The units of assignment, measurement, and analysis can differ, and all are distinct from measurement units such as points or minutes.
Evidence	Observations and contextual records contributing to an answer, including information about how they were produced.
Justification	Reasons and conditions invoked to support a comparison, analytical move, or interpretation.
Inference	Reasoning and analysis connecting evidence and justification to a conclusion.
Bounded claim	The reported conclusion with its uncertainty, conditions, and scope: in the statistical layer, the conclusion about the estimand; in the substantive layer, the question-facing answer with its qualifications.

Chapter 2: Data, Distributions, and Variation

Table 2

Key terms of Chapter 2.

Term and common alternatives	Meaning
Case	The entity or measurement occasion described in a dataset. Depending on the study, a case may be a student, a class, or one measurement occasion for a student.
Variable	A characteristic that can take different values across cases or occasions. A dataset records values of one or more variables for its cases.
Quantitative variable (numerical variable)	A variable expressing an amount, including continuous measurements and discrete counts or scores. Its units and scale determine what numerical differences mean.
Categorical variable	A variable whose values identify categories. Nominal categories have no intrinsic order; ordinal categories have an order; a binary variable has two categories.

Term and common alternatives	Meaning
Population	The collection of units a question concerns. It may be the participating students themselves or a wider collection to which the findings might extend.
Sample	A selected set of units from a larger collection. Calling a set a sample does not by itself explain how it was selected or which population it represents.
Analysis sample	The units whose data contribute to a particular analysis. Eligibility, missing measurements, and exclusions can make this set and its denominator differ across analyses.
Sampling design (sampling scheme)	The specified procedure for selecting a sample, including selection stages, grouping, and, for probability sampling, the selection probabilities.
Simple random sampling	Selection in which every possible sample of the specified size has the same probability of being chosen.
Empirical distribution (observed distribution; data distribution)	How the observed values are distributed across values or intervals, including their frequencies, center, spread, and shape. A probability model is a separate description of possible values.
Skewness and kurtosis (left or negative skew; right or positive skew; tail weight)	Skewness describes asymmetry: in a right (positive) skew the long tail points toward larger values and the mean usually lies above the median; in a left (negative) skew the long tail points toward smaller values and the mean usually lies below the median. Kurtosis describes tail weight relative to a Normal distribution with the same mean and SD, not peakedness.
Mean (arithmetic mean)	The sum of the numerical values divided by their count. Every included value contributes, so unusual values can shift the mean.
Median	The middle value after ordering the data, or the average of the two middle values when the count is even. It describes the center by position.
Quartile	A cut point dividing ordered data into quarters. The first and third quartiles mark approximately the lower and upper edges of the middle half; calculation conventions can differ.
Interquartile range (IQR)	The third quartile minus the first quartile: the width of the middle half of the ordered data, expressed in the variable's original units.
Variance	A measure of spread based on squared deviations from the mean. The usual sample variance divides their sum by $n - 1$ and is expressed in squared units.
Standard deviation (SD)	The square root of the variance, expressing spread around the mean in the variable's original units. Applied to individual observations, it describes variation among their values; applied to a statistic's sampling distribution, it is the SE.
Grand mean	The mean across all included cases. It equals the average of the group means weighted by their group counts, so larger groups contribute more.
Degrees of freedom (df)	A count reflecting freely varying information after constraints or fitted quantities are taken into account. The n deviations from their own mean sum to zero, leaving $n - 1$ free.
Standardized value (z score)	A value minus a specified mean, divided by a specified standard deviation. It expresses distance in SD units; standardizing alone preserves the distribution's shape.

Term and common alternatives	Meaning
Normal distribution (Gaussian distribution)	A symmetric, bell-shaped probability model described by a mean and standard deviation. The standard Normal has mean zero and SD one; observed data need not follow this model.
Conditional proportion	The proportion in one category among cases satisfying a stated condition. Its denominator is the count in that conditioning group, such as students in one education-intention category.
Covariance	A summary based on products of paired deviations from the two means. Its sign shows the direction of joint variation; its magnitude depends on both variables' units and spreads.
Pearson's correlation (Pearson's r ; product-moment correlation)	Covariance divided by the product of the two SDs. For variables with nonzero spread, it describes direction and strength of linear association on a scale from -1 to 1 .
Sampling distribution	The distribution of a statistic, such as a sample mean, across repeated samples generated by a stated sampling procedure or model. Repeated sample summaries in a simulation approximate this distribution.
Sampling error	The part of an estimate's departure from the population quantity that arises from observing a sample rather than the whole population. It varies across repetitions, and its spread is the SE; coverage, nonresponse, and measurement error are separate.
Standard error (SE)	The standard deviation of a statistic's sampling distribution under a stated sampling procedure or model. It describes spread across repeated sample summaries, distinct from individual-value SD and one sample's difference from the population quantity.
Design effect (deff)	The factor by which a sampling design multiplies the variance of a statistic relative to simple random sampling of the same size. For a mean from equal clusters of m units it is about $1 + (m - 1) \text{ICC}$, where the intraclass correlation (ICC) is the share of outcome variance lying between clusters: units in one cluster supply overlapping information.
Data matrix	A rectangular arrangement with cases in rows and variables in columns; x_{ij} denotes case i on variable j .
Binary variable (0/1 coding)	A binary variable has two possible categories; a 0/1 coding marks membership or occurrence. Its sample mean is the proportion coded 1.
Linear transformation	$Y = a + bX$ changes the center to $a + bE(X)$ and the SD to $ b SD(X)$; a negative b also reverses order.
Bernoulli; binomial; categorical; Poisson; lognormal; logistic	Respectively a binary outcome, a count of successes in fixed independent equal-probability trials, one category, an event count under a specified rate model, a positive variable with Normal log, and a continuous symmetric distribution used here as a preview. Logistic regression later models Bernoulli outcomes.
Density; cumulative distribution function; quantile	Density determines probability by area for a continuous variable; the CDF gives $P(X \leq x)$; a quantile is a value reaching the stated cumulative probability.
Rank correlation; point-biserial correlation	Spearman uses ranks; Kendall summarizes concordance; point-biserial is Pearson correlation with one binary variable coded 0/1. Each addresses a specified relationship.

Term and common alternatives	Meaning
Partial correlation	Correlation between two variables after their linear relationships with stated controls are removed by regression; it describes residual association and does not establish causal adjustment.
Parameter	A numerical characteristic of a population or probability model, such as a population mean. A parameter may be the estimand; the question determines which quantity is sought.
Statistic	A quantity calculated from sample data, such as a sample mean or proportion. The same statistic can summarize the sample or serve as an estimate for a wider target.

Chapter 3: Estimates, Uncertainty, and Statistical Tests

Table 3

Key terms of Chapter 3.

Term and common alternatives	Meaning
Estimate	A numerical approximation to an estimand, calculated from the available data. A point estimate expresses the approximation as one value.
Estimator	The rule used to calculate an estimate from data. Its bias and precision concern how the rule behaves across the repetitions specified by the study's procedure or model.
Bias (of an estimator)	The difference between an estimator's expected value over the specified repetitions and the estimand. An unbiased estimator has bias zero; precision describes its spread, a separate quality.
Pooled variance (s_p^2 ; pooled within-group variance)	A weighted average of within-group sample variances, each weighted by its degrees of freedom: $s_p^2 = \{(n_1 - 1)s_1^2 + (n_0 - 1)s_0^2\} / (n_1 + n_0 - 2)$ for two groups. It estimates a variance the groups are assumed to share; its square root is the pooled SD.
Confidence interval (CI)	An interval expressing uncertainty about an estimand. Under its conditions, the procedure covers that estimand at its stated rate across repetitions; the rate describes the procedure.
Critical value	The cut-point of a reference distribution, such as the Normal or t, selected by a confidence or significance level. It multiplies the SE in an interval and marks the rejection region of a test.
Coverage	The proportion of repetitions in which an interval procedure contains the estimand under its conditions. The confidence level is the procedure's target coverage, not a probability for one realized interval.
Null hypothesis	The statistical proposition used as the reference for a test, such as equality of specified population means. The alternative hypothesis states the departure being considered.

Term and common alternatives	Meaning
Test statistic	A quantity calculated from data to assess departure from a null hypothesis. The test compares it with a reference distribution justified under that null and stated conditions.
p-value	Under a specified null hypothesis and conditions, the probability of a test result at least as extreme as the observed one, using the test's stated tail rule.
Significance level (alpha)	The threshold used for rejecting a null hypothesis, chosen to control the probability of rejecting a true null under the test's conditions. This error is called a Type I error.
One-sample z test; one-sample t test	Both compare a sample mean with a stated null mean using its SE; z uses a known process SD under its model, whereas t estimates that SD and uses the appropriate t reference. Neither choice follows an automatic sample-size cutoff.
Paired-samples t test	A one-sample t test of within-pair differences against a specified mean difference, with independent pairs under the stated model.
Equal-variance two-sample t test (pooled t test; Student's two-sample t test)	One of the two independent-samples t tests. It compares two independent group means under the condition that the populations share one variance: its SE is $s_p\sqrt{1/n_1 + 1/n_0}$, and its reference is t with $n_1 + n_0 - 2$ df.
Welch's unequal-variance t test (Welch test)	The other independent-samples t test. It compares two independent group means without assuming a common variance: its SE adds the two estimated mean variances, and its approximate t reference uses Welch-Satterthwaite df. With equal group sizes its statistic equals the equal-variance statistic, and only the df can differ.
Expected count	The count a cell of a two-way table would have under a stated null pattern such as independence, given the observed margins: row total times column total, divided by the overall total.
Pearson's chi-square statistic (X^2)	The sum over cells of squared differences between observed and expected counts, each divided by its expected count, compared with a chi-square reference with $(R - 1)(C - 1)$ degrees of freedom under the stated model.
Multiway table (three-way table; layered two-way tables)	A table that cross-classifies three or more categorical variables: a two-way table holds two variables, a three-way table three, and a four-way table four. A three-way table is read as two-way tables within the layers of the third variable; a chi-square test of one two-way table cannot test the variables jointly, which needs a method that uses all three variables at once, such as a model for multiway tables.
Standardized mean difference	A difference in means divided by a specified SD. The chosen denominator determines the standardized quantity; the raw difference retains the outcome's original units.
Power (statistical power)	The probability of rejecting the null under a specified alternative, design, and test. Failure to reject under that alternative is a Type II error; power is one minus its probability.

Chapter 4: Group Comparisons with Analysis of Variance

Table 4

Key terms of Chapter 4.

Term and common alternatives	Meaning
Factor (grouping factor)	In ANOVA, a categorical variable defining groups or conditions; its categories are levels. This meaning differs from a latent factor used to model shared indicator variation.
Analysis of variance (ANOVA)	A method for studying mean differences through sources of variation. One-way ANOVA uses one grouping factor, and multi-way (factorial) ANOVA uses two or more crossed factors, as two-way and three-way ANOVA do; each factor can have several levels.
Fitted value	The value a fitted model assigns to an observed case. In a one-way group-mean model, it is that case's group mean.
Residual	An observed outcome minus its fitted value. A residual is calculated from the fitted model; an error is the corresponding unknown departure in the population model.
Mean square (MS)	A sum of squares divided by its degrees of freedom. In classical ANOVA, the within-group mean square estimates the common error variance under the model's conditions.
F statistic (F ratio)	In classical ANOVA, a ratio of the mean square for a tested source to the appropriate error mean square. Its null F reference depends on the model and both degrees of freedom.
Omnibus test	A test of one joint null hypothesis about a whole pattern, such as equality of all group means. Rejecting it does not identify which comparisons differ or how large they are.
Eta squared	The proportion of total observed variation attributed to a specified source in an ANOVA decomposition. In one-way ANOVA, it is between-group divided by total sum of squares.
Contrast (linear contrast)	A weighted combination of means whose weights sum to zero, expressing a specific comparison. The means, weights, and target population determine the quantity compared.
Familywise error rate (FWER)	The probability of at least one Type I error among a defined family of tests. Procedures such as Bonferroni allocate the error allowance across that family.
Simultaneous confidence intervals	A set of intervals calibrated so that, under their conditions, all of them cover their estimands together at the stated familywise rate; Bonferroni, Tukey, and Dunnett are different calibrations for different families.
Cell mean and marginal mean	In a two-factor layout, a cell mean is the mean of the cases sharing one level of each factor; a marginal mean averages the cell means across the levels of the other factor, with stated weights.
Main effect	A comparison across one factor's levels averaged over the other factor's levels, using stated weights. It remains an average when the conditional pattern differs across levels.

Term and common alternatives	Meaning
Interaction	A pattern in which a comparison across one factor's levels changes with the other factor's level. A difference of differences expresses one such comparison.
Simple effect	A comparison across one factor's levels at a specified level of the other factor. It describes a conditional comparison within the factorial design.
Three-way interaction (higher-order interaction)	A pattern in which a two-way interaction changes across the levels of a third factor. Each added factor multiplies the cells and adds interactions, so the highest-order interaction is read first and a few comparisons are declared in advance.

Chapter 5: Description and Basic Inference with R

Table 5

Key terms of Chapter 5.

Term and common alternatives	Meaning
Script and Quarto source	A saved <code>.R</code> script records commands so an analysis can be rerun; a <code>.qmd</code> source connects prose, mathematics, executable code, tables, figures, and references, and renders to PDF, HTML, or Word.
Simulation and seed (Monte Carlo simulation)	Generating data from a declared process to see what a statistic or procedure does across repetitions. A seed makes the random draws repeatable; it does not make the process true for an actual study.
Monte Carlo standard error	The numerical uncertainty of a simulated probability or mean that comes from using finitely many repetitions; more repetitions shrink it. It is distinct from a study's SE, which depends on the study's sample size.
Covariance matrix	A square table of variances on the diagonal and covariances off the diagonal for a stated set of variables and observations.
Correlation matrix	A square table of pairwise correlation coefficients for a stated set of variables and observations; its diagonal is 1 for variables with positive variance.

Chapter 6: Group Comparisons and Statistical Modelling with R

Table 6

Key terms of Chapter 6.

Term and common alternatives	Meaning
Model formula (R formula)	Notation such as <code>score ~ group</code> naming an outcome and the variables that explain it for a fitting function. It does not state whether observations are independent or paired; that comes from the design.
Categorical effect scales (difference in proportions; risk ratio; odds ratio; Cramér's V)	Four ways to state the size of an association in a table, each answering its own question: the difference in proportions gives an absolute gap, the risk ratio a relative gap, the odds ratio a ratio of odds that stays the same when rows and columns exchange roles, and Cramér's V a unit-free strength, from 0 to 1, for a table of any size.
Objective function	A numerical criterion evaluated at candidate parameter values for fixed data; fitting seeks a stated minimum or maximum.

Chapter 7: Simple Linear Regression

Table 7

Key terms of Chapter 7.

Term and common alternatives	Meaning
Simple linear regression (simple regression model; regression line)	A model in which an outcome's mean is a straight line in one predictor, $E(Y x) = \beta_0 + \beta_1 x$, with errors around the line that average zero and share one SD σ ; its three parameters are the intercept, the slope and σ . The least-squares line estimates the model's line: its descriptive use concerns the records fitted, and a model-based use needs a declared process.
Intercept	The fitted outcome at predictor value zero. It places the line vertically; when zero is outside the observed support, it is a mathematical anchor rather than a supported prediction. Centering the predictor moves it to a supported reference point.
Slope (regression coefficient)	The difference in fitted outcome per one-unit difference in the predictor, in outcome units per predictor unit. It compares units that differ in the predictor; it is not a change within one unit unless a design supports that reading.
Ordinary least squares (OLS)	The rule that chooses the intercept and slope minimizing the residual sum of squares. With an intercept, the fitted line passes through the mean point and the residuals sum to zero; the slope equals covariance divided by predictor variance.

Term and common alternatives	Meaning
Residual sum of squares (RSS)	The sum of squared residuals from a fitted line, in squared outcome units. Total squared variation about the mean splits into a regression part and RSS when the model has an intercept.
Residual standard deviation (residual standard error)	The square root of RSS divided by its residual degrees of freedom, in outcome units; with an intercept and one predictor the degrees of freedom are $n - 2$. Under the process model it estimates the error SD.
R-squared (coefficient of determination)	One minus RSS divided by total squared variation: the proportion of centered squared outcome variation accounted for by the fit in these records. In simple regression with an intercept it equals the squared correlation.
Confidence interval for a mean response (mean-response interval)	An interval for the process mean outcome at a stated predictor value, using the SE of the fitted value. It concerns an average, narrows near the predictor mean, and does not cover individual outcomes.
Prediction interval (new-case interval)	An interval for one new outcome at a stated predictor value from the same process. Its variance adds the error variance to the fitted-mean variance, so it is wider than the mean-response interval and keeps a nonzero floor as the sample grows.
Leverage	A measure of how unusual a record's predictor value is relative to the others; it grows with distance from the predictor mean and appears in the SE of a fitted value and in influence measures.
Cook's distance (influence)	A summary of how much the fitted values change when one record is omitted, combining its residual and its leverage. It prompts inspection of provenance and a disclosed sensitivity, not automatic deletion.
Heteroskedasticity-consistent covariance (HC3; sandwich estimator)	A coefficient covariance estimate that allows unequal error variance across predictor values under independence. It changes SEs and intervals, not the fitted coefficients, and does not repair curvature, dependence, measurement error, or confounding.
Extrapolation (predictor support)	Using a fitted line at predictor values outside the observed range or where observations are sparse. Software returns a number there; the data do not support it.
Regression toward the mean (regression to the mean)	For a simple linear prediction in standard units, $\hat{z}_y = r z_x$; when $ r < 1$, the predicted value is nearer its mean than the predictor is to its mean. In a stable repeated-measurement model with independent occasion errors, a group selected on an extreme noisy first score is expected to be less extreme on the second. These are statements about predictions or conditional averages, not every case; a narrower spread of predictions does not imply a narrower outcome distribution (the regression paradox).
Regression fallacy	Attributing to an intervention the movement expected from selection on an extreme noisy measurement. Evaluating an intervention requires a credible no-intervention benchmark, supplied by a suitable design or untreated comparison under stated assumptions; a fitted line through treated outcomes alone does not supply it.

Chapter 8: Multiple Regression I: Conditioning, Checks, and Coordinated Questions

Table 8

Key terms of Chapter 8.

Term and common alternatives	Meaning
Multiple regression (multiple regression model; additive model)	A model in which an outcome's mean is an intercept plus one term per predictor, $E(Y x_1, \dots, x_{p-1}) = \beta_0 + \beta_1 x_1 + \dots + \beta_{p-1} x_{p-1}$, with errors of one common SD σ ; its p coefficients leave $n - p$ residual df, and least squares estimates them. Each coefficient compares means at the same values of the other predictors; the additive form assumes each comparison is the same across those values.
Partial slope (conditional coefficient; adjusted coefficient)	The difference in fitted outcome per one-unit difference in one predictor, holding the model's other predictors fixed. Its meaning depends on the conditioning set named with it; it is not an intervention effect or a claim about unmeasured differences.
Indicator variable (dummy variable) and reference category	A 0/1 variable marking membership in one category of a factor; a factor with J levels needs $J - 1$ indicators, and the omitted level is the reference. The intercept refers to the reference category; each indicator coefficient is a difference from it at the same other predictors. Releveling changes coordinates, not fitted values.
Partial regression (residualization; added-variable form)	Regressing both the outcome and one predictor on the other predictors, then relating their residuals. The residual slope equals the partial slope; the residual correlation is the partial correlation, unit-free.
Coefficient covariance matrix	The estimated variances and covariances of the fitted coefficients (<code>vcov()</code> in R). Any linear combination of coefficients, such as a fitted mean or a contrast, has variance $\mathbf{c}^T \widehat{\mathbf{V}} \mathbf{c}$; adding SEs or ignoring covariances is a different, incorrect calculation.
Variance inflation factor (VIF)	One divided by one minus the R^2 from regressing a predictor on the other predictors. Its square root multiplies that coefficient's SE relative to an orthogonal design, holding residual variance fixed; it describes precision, not bias, and is not a deletion rule.
Joint support (covariate hull)	Whether a combination of predictor values lies within the region where observations actually occur, not merely within each variable's separate range. A fitted value outside joint support is extrapolation even if each coordinate is observed somewhere.
Standardization over a named distribution (model-averaged fitted value)	Averaging the model's fitted values after setting one predictor to a chosen value for every profile in a declared set with fixed weights. Its uncertainty uses the coefficient covariance; it is not an observed group mean, a causal effect, or held-out performance.
Learning rule (learning procedure)	A fixed recipe that turns training records into a prediction equation, such as fitting a declared OLS formula. Prediction performance concerns the rule at stated training sizes, not one final fitted equation.

Term and common alternatives	Meaning
Cross-validation (held-out validation)	Dividing records into folds, fitting a rule on the other folds, predicting the held-out fold, and pooling the errors so that no outcome contributes to its own prediction. Resplits describe sensitivity to the partition, not new independent studies.
Mean absolute error (MAE)	The average absolute difference between held-out outcomes and their predictions, in outcome units. It is an evaluation loss chosen separately from the squared-error criterion that OLS minimizes.
Cross-question coherence	The additional evaluation question for coordinated answers: does the synthesis preserve the distinct questions, estimands, methods, uncertainty, and limits of the separate answers rather than merging them into one omnibus claim?

Chapter 9: Multiple Regression II: Categorical Predictors, Model Comparison, Nonlinearity, and Robust Uncertainty

Table 9
Key terms of Chapter 9.

Term and common alternatives	Meaning
ANCOVA (analysis of covariance)	A regression comparing groups while including a quantitative covariate with one common slope; its adjusted group means are read at a stated covariate value within common support. The posttest-only, gain-score and ANCOVA analyses are one regression whose pretest coefficient is 0, 1 or estimated; adjustment cannot supply what random assignment supplies.
Homogeneity of regression slopes (parallel slopes)	ANCOVA's condition that the outcome-covariate slope is the same in every group, checked by testing the group-by-covariate product block; when the slopes differ, the adjusted difference depends on the covariate value at which it is read.
Nested-model F test (extra sum of squares; block test)	Compares a reduced model with a fuller one fitted to the same records and outcome: the RSS reduction per added coefficient, divided by the fuller model's residual mean square, referred to F with the number of restrictions and the fuller model's residual df. It tests the added block jointly, not each coefficient separately.
Information criteria (AIC; BIC)	Penalized in-sample fit: $-2 \log \widehat{L}$ plus $2k$ (AIC) or $k \log n$ (BIC), with k counting the coefficients and the residual variance. Smaller is preferred among models fitted to the same records and outcome scale; neither is held-out error.
Standardized coefficient (beta weight)	A partial slope rescaled by the predictor and outcome SDs, in outcome-SD units per predictor SD, with the other predictors held fixed. It differs from standardization over named profiles and does not measure causal importance.

Term and common alternatives	Meaning
Polynomial term (quadratic term; curvature)	A power of a predictor added as a further column so that the fitted mean can curve. The predictor's comparison then depends on its value; centering before squaring reduces collinearity without changing the fitted curve.
Log-transformed outcome (geometric-mean scale)	Fitting the logarithm of the outcome, so that a coefficient is an additive difference in expected log outcome and $100(e^\beta - 1)$ percent is a multiplicative difference in geometric means. A changed outcome scale is a changed question, and its criteria are not comparable with the original scale's.
Bootstrap (case resampling; percentile interval)	Refitting the model to many resamples of the records drawn with replacement and using the spread of the refitted estimates as an SE, or their middle 95% as a percentile interval. It treats the resampled units as independent and exchangeable; it does not repair dependence, thin support, measurement error, or an observational design.
Cluster-robust standard error (school-clustered covariance)	A coefficient covariance that sums residual contributions within clusters, such as schools, before combining them, allowing dependence within clusters under an approximation that treats clusters as independent. It changes uncertainty, not the coefficient, and does not repair missing outcomes or weighting.

Chapter 10: Interaction and Moderation

Table 10

Key terms of Chapter 10.

Term and common alternatives	Meaning
Product term (interaction term; multiplicative term)	A column formed by multiplying predictors, so that a fitted comparison can depend on their values. A two-way product coefficient is a difference in slopes when no higher-order product makes that difference conditional, and nonadditivity is defined on the modeled scale: a nonlinear change of outcome scale can create or remove it.
Conditional slope (simple slope; conditional effect)	The fitted difference in outcome per one-unit difference in a focal predictor evaluated at stated values or groups of the other predictors. It is a linear combination of coefficients, so its SE uses their covariance, and it describes a conditional mean, not an individual case.
Moderator (moderating variable; moderation)	The variable across whose values or groups a named focal relationship is allowed to vary. The product-term algebra is symmetric in the two predictors, so the label comes from the research question; a moderation reading describes when a relationship differs, not how an effect is transmitted.

Term and common alternatives	Meaning
Hierarchy principle (marginality; well-formulated model)	The rule that a model containing a product retains its component terms and all lower-order products contained in it. Omitting them imposes restrictions tied to the coded origins; a three-way product therefore retains all three two-way products as well as the three component terms.
Centering a predictor (anchoring; reparameterization)	Replacing a predictor by its difference from a chosen value, so that coefficients anchored at zero become comparisons at that value. It changes the intercept and lower-order coefficients while preserving fitted values, residuals and tests of equivalent model restrictions. The highest-order product coefficient stays unchanged under a change of origin; lower-order product coefficients can change in a higher-order model. Centering repairs no design, selection, or measurement problem.
Probe value (pick-a-point value)	A stated value of the moderator at which a conditional slope or contrast is evaluated, chosen from the measurement scale, a threshold, or a decision need rather than by an automatic mean-and-SD rule, and restricted to values the records support.
Johnson–Neyman boundary (region of significance)	A moderator value at which a pointwise interval for a conditional slope touches zero, found as a root of a quadratic whose leading coefficient decides which side is the region. A boundary outside the observed support describes the algebra and not the data, and a pointwise band is not a simultaneous band.
Interaction contrast (difference in differences within a model)	A linear combination of coefficients comparing two conditional comparisons, such as one group’s difference minus another group’s difference. Under reference coding it equals a product coefficient up to sign; its SE uses the coefficient covariance, and its causal reading depends on design, not on the algebra.
Moderated moderation (three-way interaction; higher-order interaction)	A two-way interaction that changes across a second moderator. A hierarchical three-way model retains all component and two-way terms; conditional slopes depend on both moderators, and a direct contrast tests how the named interaction differs rather than comparing separate significance decisions.

Chapter 11: Binary Logistic Regression

Table 11

Key terms of Chapter 11.

Term and common alternatives	Meaning
Event and non-event (the category coded 1)	The category a binary model treats as occurring, named together with the category coded 0. Reversing the event reciprocates the odds ratios and complements fitted event probabilities; probabilities for the original event agree after relabelling.
Prevalence (event rate)	The proportion of analytic rows that are events. Coefficient precision also depends on conditional probabilities, predictor spread and overlap; the rare-outcome approximation of an odds ratio by a risk ratio depends on the risks being compared.
Generalized linear model (GLM)	A response distribution, a linear predictor and a link connecting the conditional mean to the linear predictor; changing the response distribution and link gives the models of Chapters 11 and 12.
Link function; logit link and inverse logit (logistic function)	The logit is the log of the odds and its inverse the S-shaped function bounded in $(0, 1)$, equal to one half and steepest at a linear predictor of zero.
Odds and log-odds	The odds are $\pi/(1 - \pi)$ and the log-odds their logarithm; the three scales describe one quantity with null values 1 and 0, and the log-odds scale is symmetric under reversing the event.
Odds ratio (OR)	The factor e^{β_j} by which the odds differ between units differing by one unit in a predictor and agreeing on the others; it depends on the units, reference, event and adjustment set, is not a risk ratio unless the event is rare, and implies no constant probability difference.
Noncollapsibility (of the odds ratio)	A conditional odds ratio generally differs from the marginal one even with no confounding, so a coefficient that moves under adjustment is not by itself evidence of confounding.
Likelihood and log-likelihood	The probability or density of the observed data under a declared model, viewed as a function of its parameters with the data fixed; the log-likelihood is its logarithm. It is not a probability distribution over the parameters.
Maximum likelihood estimation (MLE)	Choosing parameter values that maximize the likelihood, equivalently the log-likelihood, over the stated parameter space; near the maximum, the curvature of the log-likelihood gives the large-sample SE.
Bernoulli likelihood and score equations (binary-outcome likelihood)	The probability of the observed data as a function of the parameters, maximized to estimate them; for a binary outcome it is a product of Bernoulli probabilities whose score equations have no closed form, and a model with an intercept reproduces the observed event total.
Iteratively reweighted least squares (IRLS)	The algorithm that repeatedly solves a weighted linear regression with weights $\hat{\pi}_i(1 - \hat{\pi}_i)$ until the coefficients stop changing; its convergence flag and iteration count belong in the record and establish no adequacy.

Term and common alternatives	Meaning
Deviance and the likelihood-ratio test (G^2)	For ungrouped binary data the deviance is -2 times the maximized log-likelihood and carries no goodness-of-fit reading alone, while the difference between nested models on the same rows and event is referred to chi-square.
Wald test and profile-likelihood interval	The Wald statistic squares the estimate-to-SE ratio while a profile interval collects the values the log-likelihood does not reject; they diverge for rare events, small samples or near-infinite estimates.
McFadden pseudo R^2	One minus the ratio of the fitted to the intercept-only maximized log-likelihood: bounded in $[0, 1]$, not a proportion of variance explained, and routinely small for useful models.
Empirical logit	The log-odds of a grouped event count with one half added to each count, plotted against a predictor to check linearity of the log-odds in that predictor as entered.
Binned residual	The mean raw residual within a bin of rows sorted by fitted probability, against a two-standard-error band; it replaces the residual-versus-fitted plot, which shows only two curved bands here.
Separation (complete; quasi-complete)	A predictor pattern separates events from non-events completely or on a boundary, so at least one coefficient has no finite ordinary maximum-likelihood estimate. Diverging coefficients, large standard errors and convergence warnings can be symptoms; no one warning or numerical pattern is required.
In-sample calibration and discrimination checks (fitted-probability checks)	Calibration asks whether fitted probabilities match observed event frequencies, discrimination whether they rank events above non-events; the two answer distinct questions, and honest evaluation of new-case performance requires held-out data.
Linear probability model (LPM)	Ordinary least squares on a 0/1 outcome, whose coefficients are probability differences but whose mean function is unbounded and whose conditional variance cannot be constant.
Averaged probability difference (average marginal comparison)	The mean, over the analytic rows with their own other predictor values, of the difference between fitted probabilities at two values of one predictor.

Chapter 12: Extending the GLM: Nominal, Ordinal, and Count Outcomes

Table 12

Key terms of Chapter 12.

Term and common alternatives	Meaning
Outcome triage	The measurement judgment, made before any model is fitted, classifying an outcome as binary, nominal, ordered, a count, a rate, a proportion of known trials, or a duration. The model follows the classification and cannot establish it.
Generalized logit model (baseline-category logit; multinomial logistic regression)	For K unordered categories, $K - 1$ log-ratio equations against one baseline, each with its own intercept and slopes. The fitted probability surface is the same whichever baseline is chosen; the coefficients are not.
Baseline outcome category	The outcome category in the denominator of every generalized-logit equation: a coordinate choice, distinct from the reference level of a categorical <i>predictor</i> , whose change alters every reported ratio.
Relative-risk ratio (RRR)	The exponentiated generalized-logit coefficient: the factor by which one category's probability relative to the baseline's differs per predictor unit. It is not a probability difference, and is incomplete unless both compared categories are named.
Independence of irrelevant alternatives (IIA)	The generalized logit's assumption that the ratio of two categories' probabilities does not depend on which other categories were available: a claim about the choice set, argued rather than settled by one test.
Cumulative probability and cumulative logit	$\gamma_{ij} = P(Y_i \leq j)$ at each of the $K - 1$ cut points, and its log-odds. A category probability is a difference of two adjacent cumulative probabilities.
Threshold (cut point; ordinal intercept)	One of the ordered constants $\zeta_1 < \dots < \zeta_{K-1}$ placing the cut points on the logit scale. Their ordering keeps fitted probabilities nonnegative; they are not comparisons and are not read as effects.
Proportional odds (common-slope restriction; partial proportional odds)	The restriction that one slope vector governs every cut point, so a single odds ratio applies at all of them. It is imposed by the analyst, not conferred by ordering the levels, and is testable against a cut-specific alternative.
Poisson regression (log-linear count model)	A GLM with a Poisson random component and a log link, so the conditional mean is positive and differences are multiplicative on the count scale. Its conditional variance is not free: it is forced to equal the mean.
Incidence-rate ratio (IRR; rate ratio)	For an ordinary log-link count or rate model, the exponentiated coefficient gives the factor by which the expected count, or rate with an exposure, differs per predictor unit. An absolute difference also requires its baseline. In a two-process model, a count-component coefficient alone does not give the overall mean ratio.

Term and common alternatives	Meaning
Exposure and offset	The window over which each unit was observed, and $\log t_i$ entered with its coefficient fixed at one so the model describes a rate. An offset is a decision about the estimand, not an estimate: no SE, no p -value, and freeing its coefficient gives a testable alternative.
Overdispersion and the dispersion statistic	Conditional variance exceeding the conditional mean, assessed by the mean squared Pearson residual $\hat{\phi}$ under a suitable mean model; the deviance ratio is a rough diagnostic and need not be near 1 for sparse counts even under a correct Poisson model. The diagnosis is conditional; a marginal ratio above one is not evidence.
Quasi-Poisson (quasi-likelihood; scaled Poisson)	The Poisson mean with variance $\phi\mu$: identical estimates and every SE multiplied by $\sqrt{\hat{\phi}}$. No full distribution is specified, so there is no log-likelihood and no AIC or BIC.
Negative-binomial regression	A count model with variance $\mu + \mu^2/\theta$, arising when the Poisson mean varies across units sharing predictor values. A different likelihood, not a rescaling: estimates move because large fitted means are down-weighted, and the Poisson comparison sits on the boundary $1/\theta = 0$.
Zero-inflated model	A mixture of an extra-zero component and a count component that can also generate zeros. The count-component mean differs from the overall mean; an observed zero does not identify which component generated it, and fitted extra-zero membership does not establish a substantive structural-zero mechanism.
Hurdle model (two-part count model)	A binary model for zero versus a positive count combined with a zero-truncated count distribution for positive values. The overall mean combines the probability of a positive outcome with its positive-count mean. Its binary coefficient concerns crossing zero, which differs from membership in a zero-inflated model's extra-zero component.
Zero-truncated count distribution	A count distribution conditioned on a positive outcome: its zero probability is removed and the positive probabilities are renormalized. Its mean is the original distribution's mean divided by its probability of a positive count, so it differs from the untruncated mean parameter.

Chapter 13: Prediction, Cross-Validation, and Regularization

Table 13

Key terms of Chapter 13.

Term and common alternatives	Meaning
Generalization error (prediction error; risk)	Expected loss on a new unit from the stated deployment distribution. It can be defined for one fitted model conditional on its training data, or averaged over training samples for a learning rule at a stated training size; held-out observations estimate the chosen target.
Deployment moment (prediction time)	The point at which a prediction must be produced, which decides whether a predictor is available. A variable recorded later may improve measured performance but cannot be used for that earlier prediction task.
Apparent error (training error; resubstitution error)	Loss on the records used to fit the rule. Using those outcomes in fitting generally makes this loss optimistic for new-case performance; greater fitting flexibility can increase that optimism, without guaranteeing an ordering in every sample or for every loss.
Root mean squared error (RMSE)	The square root of the mean squared prediction error, in outcome units. It penalizes a large error far more than MAE does, so the choice between them states which mistakes matter.
Out-of-sample R^2	One minus the held-out squared error divided by the held-out squared error of a named constant reference fitted in the same analysis rows. It can be negative and is not the in-sample R^2 of Chapter 7.
Data leakage (information leakage)	Using information unavailable at the deployment moment, or letting held-out records influence a training or selection step used to predict them. It can make performance estimates optimistic; there is no general proportional rule for the size of that optimism.
Development and test split (holdout set)	A partition in which fitting, screening and tuning use the development records, while test outcomes remain protected until the rule is fixed. Adapting the rule after seeing test results uses the test set for development; repeating an unchanged calculation does not itself do so.
Penalized regression (shrinkage; ridge; lasso; elastic net)	Fitting that adds a coefficient penalty to the loss, trading bias against variance. Ridge's squared penalty generally shrinks coefficients without selecting exact zeros; the lasso's absolute-value penalty can set coefficients to zero, and elastic net combines the penalties.
Tuning parameter (penalty; λ)	A quantity that indexes a family of fits and is chosen from data the fit did not see. Choosing it is itself a data-dependent step and belongs inside the resample.
One-standard-error rule	A convention choosing the largest penalty whose observed cross-validated loss is within the estimated standard error at the minimum. It prefers a simpler fit under that empirical tolerance; it does not bound the increase in true prediction error.
Proper scoring rule (log loss; Brier score)	A score for a predicted probability that is minimized in expectation by the true probability, so that honest uncertainty is rewarded. Accuracy and a single confusion matrix are not proper scores.

Term and common alternatives	Meaning
Calibration	Agreement between predicted probabilities and observed event rates among cases with similar predictions. It is invisible in a ranking measure such as AUC and is checked on held-out predictions.
Discrimination (AUC; concordance)	The probability that a randomly chosen event case receives a higher score than a non-event case, with half credit for tied scores. It depends on ranking and does not establish calibration of the probabilities.
Decision threshold (operating point)	The predicted probability above which an action is taken, determined by the ratio of the costs of the two errors and not by the model. Two decision-makers with different costs read the same probabilities and act differently.
Dataset shift	A difference between the distributions in the development and use settings, potentially involving predictors, outcome frequencies or their relationship. Internal resampling alone does not establish performance in a different setting.

Chapter 14: Regression Pitfalls and Causal Design

Table 14

Key terms of Chapter 14.

Term and common alternatives	Meaning
Directed acyclic graph (causal diagram; DAG)	A diagram whose arrows state assumed direct causes and whose directed paths cannot return to their starting point; it records assumptions that can be questioned, rather than conclusions produced by the data.
Mediator (intermediate variable)	A variable caused by the focal predictor and lying on a path to the outcome; one kind of post-treatment variable, not a synonym for every variable measured after treatment. Conditioning on it changes a total-effect comparison; a direct causal reading needs additional identification and modelling conditions.
Collider (common effect)	A variable caused by two others. A path through a collider is blocked unless the collider or a descendant is conditioned on; such conditioning can induce an association along that path.
Adjustment set	The variables conditioned on for a stated estimand under an assumed structure, justified by that structure and the timing rather than by the available columns or a coefficient's movement.
Aggregation reversal (Simpson's paradox; strict Simpson reversal)	A sign difference between a pooled comparison and the stratified or standardized comparisons, strict when every stratum points the opposite way to the pooled result; it arises from composition, not from a computational error.

Term and common alternatives	Meaning
Within-group, between-group, and pooled comparison	Three estimands from one dataset: inside a level of the grouping variable, across group summaries, and across all rows ignoring the grouping, with the pooled slope a weighted average of the other two by the within and between shares of predictor variation.
Direct standardization (standardized comparison)	Recomputing each group's summary under one declared common distribution of a stratifying variable; a third estimand rather than a compromise between the pooled and conditional comparisons.
Range restriction (restriction of range)	Keeping only part of the observed range; restricting the predictor under a correctly specified linear model costs precision, apparent strength and support outside the window, while restricting on the outcome also moves the slope.
Missingness mechanisms (MCAR; MAR; MNAR)	MCAR means missingness is unrelated to observed and missing values; MAR means it is unrelated to missing values conditional on observed information; MNAR means dependence on missing values remains. Complete-case validity depends on the target and mechanism, and MNAR requires assumptions not identified by the observed data alone.
Attenuation from measurement error	In a single-predictor linear model, classical predictor error multiplies the population slope by reliability. With several predictors the direction can differ; a mismeasured control can leave residual confounding.
Ecological inference (ecological fallacy)	Inferring an individual-level relationship from aggregate records. Aggregate and individual relationships can differ in sign and magnitude; the aggregate relationship alone does not identify the individual comparison.
Specification sensitivity (multiverse analysis; analyst degrees of freedom)	The spread of an estimate across the defensible analyses that could have been run; a disclosure about specification, not evidence about an omitted common cause, selection or measurement shared by every version.
Pseudoreplication (dependence as a design fact)	Treating dependent observations as independent replicates, misrepresenting uncertainty. The relevant dependence, sampling and assignment structure must be represented; its effect on a standard error depends on the comparison.
Identification argument (design warrant)	The explicit reasons a comparison estimates the intended causal quantity, whether assignment or a named quasi-experimental design with its stated assumption; naming a design is not meeting its assumption.
Bounded regression claim (credible-claim template)	A concluding statement whose unit, verb, conditioning set, supported range, design and disclosed sensitivity all match the analysis that produced it.

Argument Chapter A: How to Read a Quantitative Study as an Argument

Table 15

Key terms of Argument Chapter A.

Term and common alternatives	Meaning
Argument component	One of the four functions, evidence, justification, inference, and bounded claim, that a piece of study information performs in the statistical or the substantive layer. A reader identifies the function before judging its adequacy.
Documentary location	The place in a report or its related materials, such as Methods, a protocol, a codebook, or an analytical supplement, where the traceable information for a component is found.
Evaluation question	One of the six questions on alignment, support, and translation used to judge the quality of a question-level argument after it has been read. Each receives a reasoned judgment, not a score.

Argument Chapter B: Extending the Argument-Based Approach to Quantitative Studies

Table 16

Key terms of Argument Chapter B.

Term and common alternatives	Meaning
Statistical strand	An estimand-specific statistical chain inside one argument. Two strands arise only when sibling estimands are jointly required for one integrated substantive answer; each has its own statistical inference and conclusion.
Coordinated questions	Several question-answer pairs in one study whose support depends on one another. Each answer is read on its own, and their coherence across questions is judged separately.

Procedures and Their Conditions

The table below gathers, for each procedure that Chapters 3 and 4 teach, what its result estimates and the conditions it relies on. *Secured by design* lists the conditions a study's design can supply, such as random assignment, random sampling, or pairing; *Model assumptions and how to check them* lists the conditions the statistical model adds and the evidence available about each; *If a condition fails* gives the response the chapters teach; and *Chapter* names the chapter that develops the procedure. The choices made for one particular analysis, such as the sidedness of a test, the confidence level, or the family of comparisons, stay with that analysis in its chapter.

Table 17

Procedures and their conditions.

Procedure and target	Secured by design	Model assumptions and how to check them	If a condition fails	Chapter
One-sample t test: the mean μ of the sampled population, in outcome units, compared with a stated null value	Random sampling from the named population; independent units, each measured once	A Normal population gives an exact t reference; otherwise the sample mean needs a near-Normal sampling distribution. Look for strong skew, heavy tails or one dominating value in the scores; no sample-size cutoff settles the question	Dependent units: an SE that represents the pairs or clusters. Strong skew with few units: disclose the weaker approximation and compare a bootstrap interval. Without random sampling, a population reading needs a stated model	Ch. 3
Paired-samples t test: the mean within-pair difference μ_D , in outcome units, with the subtraction direction stated	Pairs formed by the design, such as one student measured twice or two classes in one school; independent pairs. Random assignment within pairs supports a causal reading	Near-Normal pair differences give the t reference: inspect the differences themselves rather than the two sets of scores; a few pairs say little about shape	Ignoring the pairing: an independent-groups analysis omits the within-pair covariance and misstates the SE. No untreated comparison: a mean change alone cannot be separated from time trends or repeated testing. Skewed differences: resample whole pairs	Ch. 3

Procedure and target	Secured by design	Model assumptions and how to check them	If a condition fails	Chapter
Equal-variance two-sample t test: a difference between two means, in outcome units, either a mean assignment effect over the participants or a difference between two populations' means	Random assignment of individuals, supporting a causal comparison among the participants, or two independent random samples; different units in the two groups	One common population variance and a near-Normal sampling distribution of the difference: compare the group SDs with the group sizes and look for skew or dominating values in each group. After random assignment the t reference is approximate. No interference between units is argued from the setting; the data cannot check it	Unequal spreads with unequal group sizes: the pooled SE is too small when the smaller group varies more, so choose Welch's t in advance from the design, not from a preliminary variance test. With equal group sizes the two SEs coincide	Ch. 3
Welch's unequal-variance t test: a difference between two means, in outcome units, when the two groups' variances may differ	Two independent random samples or randomly assigned groups, with different units in each group	No common variance is assumed. Each group mean needs a near-Normal sampling distribution, and the Welch-Satterthwaite df approximate the reference: inspect each group's distribution, especially the smaller group's	Skew or a dominating value in a small group: disclose the weaker approximation and compare a bootstrap that resamples within each group. Paired or clustered units: an SE that represents them	Ch. 3
Pearson chi-square test of association: whether two categorical variables are associated in the stipulated population or process, against the null of no association (independence); conditional proportions give the direction and size	Independent sampling of units, each counted once in one of mutually exclusive categories	Adequate expected counts for the chi-square approximation: calculate every expected count, row total times column total divided by n , before reading the p-value	Small expected counts: an exact conditional test such as Fisher's (Ch. 6). Repeated or clustered records: the counts overstate the independent information. An observed rather than assigned grouping supports association only	Ch. 3

Procedure and target	Secured by design	Model assumptions and how to check them	If a condition fails	Chapter
One-way ANOVA F test: whether J group means are all equal, an omnibus question about the whole pattern of means	Random assignment of individuals to the groups, or independent random samples; each unit in one group, measured once	Independent Normal errors with a common variance around the group means: compare the group SDs with the group sizes, and inspect residuals against fitted means and in a Normal quantile plot. Normality concerns the residuals, not the pooled scores	Unequal variances, above all with unequal group sizes: Welch's one-way ANOVA, chosen in advance. Classes or repeated measurements: an analysis that represents that structure. Missing outcomes: compare assigned and analysed counts and state what the analysis assumes	Ch. 4
Contrasts with a declared family: comparisons $C = \sum_j c_j \mu_j$ with weights summing to zero, in outcome units, with simultaneous coverage for the declared family	Random assignment of individuals to the groups	The one-way model's common variance and Normal errors, because every SE uses MS_W on $N - J$ df: check the group spreads and residuals as for the F test	Unequal spreads: MS_W misstates the SE of every contrast, tending to understate it for contrasts that include the more variable groups and to overstate it for those that exclude them; take each contrast's SE from the groups' own variances, which for a pair is Welch's t (Ch. 3). Comparisons chosen after inspection: a family that covers the selection, such as Scheffé's for all contrasts	Ch. 4
Two-way (factorial) ANOVA: main effects, interaction and simple effects of two crossed factors, in outcome units, with the averaging weights stated	Random assignment of individuals to the cells of a crossed layout; one score per unit	Independent Normal errors with one common variance in every cell: inspect residuals against the fitted cell means; a few scores per cell say little about shape. A main effect is an average over the other factor's levels; reading it as the effect at every level assumes additivity: examine the interaction row and the cell-mean plot first	Interaction present: report cell means and simple effects rather than main effects alone. Unequal cell counts: state the averaging weights and the hypothesis each sum of squares tests. Unequal cell spreads, above all with unequal cell counts: treat the cells as the groups of one factor and take each stated contrast's SE from the cells' own variances; the joint F tests keep the common-variance model, a limit the report discloses	Ch. 4

Procedure and target	Secured by design	Model assumptions and how to check them	If a condition fails	Chapter
Welch's one-way ANOVA: whether J group means are all equal when the group variances may differ	Independent groups of different units, randomly assigned or randomly sampled	No common variance is assumed: each group mean, weighted by its own precision, needs a near-Normal sampling distribution, and the approximate F reference estimates its denominator df from the group variances. Inspect each group's distribution and size	Clustered or repeated observations: an analysis that represents that structure. Very small or skewed groups: disclose the weaker approximation. Its scope is one factor: crossed factors with unequal cell spreads use contrasts among the cells with the cells' own variances	Ch. 4

Shared Notation

The table below lists the notation that two or more chapters share, with each symbol's meaning, the chapter where it is mainly used, and the alternatives or other symbols it must be kept apart from. Each chapter explains its symbols where they are first used, including the symbols that appear in one chapter only.

Table 18

Notation shared by two or more chapters.

Group	Symbol (preferred form)	Meaning	Main use	Alternatives and notes
Data	n, n_1, n_0, n_j, N	Number of cases in the analysis sample; group sizes; a population or record count	Ch. 2	Named groups take named subscripts, such as n_A and n_B
Data	$i, j, k; x_{ij}$	Locally declared indices; a data-matrix entry	Ch. 2	In the data matrix, i is case and j variable. ANOVA declares factor/group/replicate indices beside its own model.
Data	y_i, x_i	Observed outcome and predictor values of case i ; lowercase marks observed values	Ch. 2	y_{ij} for case i in group j

Group	Symbol (preferred form)	Meaning	Main use	Alternatives and notes
Data	Y_i, Y_{ij}	The outcome as a random variable before observation; uppercase marks random quantities	Ch. 3	Appear where the distinction from observed values matters; in Y_{ij} the second index is the group of case i , whereas the potential outcomes Y_{i1}, Y_{i0} have their own Estimands row
Summaries	$\bar{y}, \bar{x}, \bar{y}_j, \bar{d}$	Observed means; a group mean; the mean of paired differences	Ch. 2	Lowercase marks observed means, including group means
Summaries	$\bar{Y}$	The sample mean as an estimator in variance and SE derivations	Ch. 3	Reserved for the random quantity
Summaries	$s, s^2, s_1, s_0, s_p, s_p^2, s_d$	Sample SD and variance; group SDs; pooled SD and pooled variance; SD of paired differences	Ch. 2; Ch. 3	SD in prose; s_{xy} the sample covariance; the pooled SD and variance enter in Chapter 3, and Chapter 4's MS_W extends s_p^2 to k groups
Summaries	r, ρ	Sample and population correlation	Ch. 2	A partial correlation carries its conditioning set in a subscript (row below)
Relationships	$r_{xy \cdot z}$	Sample partial correlation controlling linearly for z	Ch. 2	Residual association, not causal exclusion
Population	μ, σ, σ^2	Process or population mean, SD, and variance; $\mu_1 - \mu_0$ a mean difference	Ch. 2	$\mu_0 = \beta_0 + \beta_1 x_0$ the mean response at x_0 (Ch. 7), distinct from the group mean μ_0 in $\mu_1 - \mu_0$
Population	p, p_1, p_0	A population proportion and group proportions	Ch. 6	Distinct from the p-value, written p-value in prose and $p = .0059$ for its value
Estimands	Y_{i1}, Y_{i0}	Participant i 's potential outcomes: the score participant i would have under assignment 1 and under assignment 0; over the n stated participants the mean assignment effect is $\Delta\mu = \frac{1}{n} \sum_{i=1}^n \{Y_{i1} - Y_{i0}\}$	Ch. 3; Ch. 5	Each participant reveals only the outcome of the assigned condition; the second index names an assignment, not the observed group of Y_{ij}

Group	Symbol (preferred form)	Meaning	Main use	Alternatives and notes
Estimands	$\Delta\mu, \widehat{\Delta\mu}$	A difference in means as an estimand, $\Delta\mu = \mu_1 - \mu_0$ or the mean assignment effect over the stated participants, and its estimate $\widehat{\Delta\mu} = \bar{y}_1 - \bar{y}_0$; a stipulated value of $\Delta\mu$ serves planning and power calculations	Ch. 3	$\Delta p = p_1 - p_0$ for a difference in proportions (Ch. 6); $\theta, \hat{\theta}$ for a generic estimand and estimator, distinct from the negative-binomial θ of Chapter 12; τ and a bare Δ are not used
Uncertainty	$SE(\cdot), \widehat{SE}(\cdot)$	The standard error of an estimator and its estimate; the hat marks the estimate. Name the statistic: $SE(\bar{y}), SE(\bar{y}_1 - \bar{y}_0), SE(\widehat{C})$	Ch. 3	The hat marks an estimated SE; a chapter that abbreviates the notation says so where it does
Uncertainty	$\text{Var}(\cdot), \text{Cov}(\cdot, \cdot), E(\cdot)$	Variance, covariance, and expectation; upright operator names with round brackets. Chapter 2 defines $E(X)$ for a probability model; Chapter 3 applies the same expectation to an estimator over stated repetitions, $E_{\text{assign}}(\cdot)$ when the repetition needs naming	Ch. 2; Ch. 3	$E[\cdot]$ and $\mathbb{E}$ are not used
Probability	$P(\cdot)$	Probability of an event	Ch. 2	—
Distributions	λ	A Poisson rate	Ch. 2; Ch. 12	Chapter 12 writes a unit's conditional rate per unit exposure as λ_i ; Chapter 13 uses λ for a penalty size
Design	$m, \text{ICC}, \text{deff}$	Units per cluster, such as students per sampled school; the intraclass correlation, the share of outcome variance lying between clusters; the design effect $\text{deff} = 1 + (m - 1)\text{ICC}$ for a mean from equal clusters	Ch. 2; Ch. 3	Chapter 2 introduces all three and Chapter 3 uses deff for a clustered SE. ICC and deff are upright abbreviations, not products of letters; Chapter 14 writes the intraclass correlation as ρ_I
Simulation	B	Number of simulation repetitions or bootstrap resamples, distinct from a study's sample size n	Ch. 5; Ch. 9	Monte Carlo SE shrinks with B ; the study SE does not

Group	Symbol (preferred form)	Meaning	Main use	Alternatives and notes
Reference distributions	z, z^*	A standardized score and a standard Normal critical value	Ch. 2; Ch. 3	—
Reference distributions	t_{df}, t^*	The t distribution with df degrees of freedom; a critical value	Ch. 3	t_{n-2} in regression
Reference distributions	$\chi^2, F(df_1, df_2)$	Chi-square and F reference distributions with their degrees of freedom	Ch. 3; Ch. 4	—
Testing	X^2	Observed Pearson chi-square statistic	Ch. 3	Distinguish it from the χ_{df}^2 reference distribution
Reference distributions	df	Degrees of freedom, upright; subscripts only when two are needed	Ch. 2; Ch. 3	Fractional Welch degrees of freedom follow the precision rule of the <i>Displayed values</i> row
Testing	$H_0, H_A, \alpha, \beta, 1 - \beta$	Null and alternative hypotheses; Type I and Type II error probabilities; power at a specified alternative	Ch. 3	β here is distinct from a regression coefficient, and α from the elastic-net mixing parameter of Chapter 13
Magnitude	d, g	Standardized mean difference (Cohen's d , Hedges' g)	Ch. 3	—
ANOVA	$SS_B, SS_W, SS_T, MS_B, MS_W; SS_A, SS_{AB}, SS_E, MS_E$	Sums of squares and mean squares between, within, and total; factor, interaction and error terms of the two-way table; subscripted plain letters	Ch. 4	In the two-way table SS_B denotes factor B, not all between-cell variation
ANOVA	$\alpha_j, \beta_k, (\alpha\beta)_{jk}$	Factor-model main effects and interaction departures in the two-way model	Ch. 4	Distinct from the unindexed Type I and Type II error probabilities α and β
ANOVA	η^2	Eta squared, an ANOVA effect-size measure	Ch. 4	—
ANOVA	$c_j, C, \widehat{C}$	Contrast coefficients, a contrast $C = \sum_j c_j \mu_j$ with $\sum_j c_j = 0$, and its estimate $\widehat{C} = \sum_j c_j \bar{y}_j$	Ch. 4	L is not used for contrasts; $L(\cdot)$ is a likelihood

Group	Symbol (preferred form)	Meaning	Main use	Alternatives and notes
ANOVA	$DD, \widehat{DD}$	A difference of differences between group means and its estimate	Ch. 6	Chapter 4 develops the object as an interaction contrast declared in advance; Chapter 10 writes the same object as C_{DD} when it is a linear combination of coefficients
Regression	$\beta_0, \beta_1, \beta_j; \hat{\beta}_0, \hat{\beta}_1$	Process intercept and slopes and their OLS estimates	Ch. 7	—
Regression	$\hat{y}_i, e_i, \varepsilon_i$	Fitted value, residual, and error term of case i	Ch. 4; Ch. 7	Chapter 4 uses the grouped forms $\hat{y}_{ij}, e_{ij}, \varepsilon_{ij}$
Regression	$T_i; \varepsilon_{i1}, \varepsilon_{i2}$	A case's stable true score and the independent errors of two measurements of it	Ch. 7; Ch. 14	Distinct from the residual e_i
Regression	S_{xx}, S_{xy}	Centered sums of squares and cross-products	Ch. 7; Ch. 14	—
Regression	$RSS, RegSS, TSS$	Residual, regression, and total sums of squares in regression; upright multi-letter names	Ch. 7	Distinct from the ANOVA SS family
Regression	s_e	Residual standard deviation (residual standard error)	Ch. 7	$\hat{\sigma}_{ML}$ is the maximum-likelihood scale, a different estimate (Ch. 8)
Regression	R^2	Coefficient of determination	Ch. 7	Adjusted R^2 named in words
Regression	D, D_i, D_j	Indicator (0/1) variables; D_i marks case i 's membership, and $J - 1$ indicator columns D_j code a J -level factor with a named reference level	Ch. 2; Ch. 9	Distinct from the deviance D (Ch. 11), Cook's distance D_i (Chs. 7–8), and the diagonal SD matrix $\mathbf{D}$ (Ch. 5)
Regression	$\mathbf{X}, \mathbf{y}, \mathbf{H}, \hat{\boldsymbol{\beta}}$	Design matrix, outcome vector, hat matrix, and coefficient vector; bold, with $\mathbf{X}^T \mathbf{X}$ for the cross-product	Ch. 8; Ch. 9; Ch. 11	Matrices and vectors are bold, with a superscript T for a transpose; an element such as s_{jk} is an italic scalar. $h_{ii} = \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i$, with $\mathbf{x}_i$ case i 's row of $\mathbf{X}$ written as a column, including an initial 1 for the intercept

Group	Symbol (preferred form)	Meaning	Main use	Alternatives and notes
Uncertainty	$\widehat{\mathbf{V}}$	Estimated covariance matrix of the fitted coefficients	Ch. 8	Distinct from the covariance matrix $\mathbf{S}$ of observed variables (Ch. 5)
Checks	h_{ii}, D_i	Leverage and Cook's distance of case i	Ch. 7; Ch. 8	h_i in the one-predictor setting of Chapter 7
Likelihood	$L(\cdot), \ell(\cdot), AIC, BIC$	Likelihood, log-likelihood, and information criteria	Ch. 8; Ch. 9; Ch. 11	Chapter 8 supplies a Normal likelihood for its criteria, Chapter 9 compares models with them, and Chapter 11 develops likelihood fully
De-sign/contrasts	$\delta, \hat{\delta}$	A locally defined group contrast and its estimate	Ch. 8; Ch. 9	Each chapter defines the contrast where it uses it; a causal interpretation requires design and assumptions
Argument diagrams	E, J, I, C	Evidence, justification, inference and bounded claim: functions of information	Ch. 1; Arg. A	They name functions, not report headings or quality scores
Argument diagrams	STAT-E/J/I/C; SUB-E/J/I/C	The four functions in the statistical and substantive layers	Ch. 1; Arg. A/B	SUB-E situates, SUB-J augments and contextualizes, SUB-I interprets, SUB-C bounds the substantive reach of the corresponding statistical contribution
Argument diagrams	Support arrows and cross-layer links	Arrows show logical support within an argument; labelled links connect the meanings of the two layers	Ch. 1; Arg. A/B	Neither is a time sequence or causal path; label cross-question dependencies by the receiving function

Group	Symbol (preferred form)	Meaning	Main use	Alternatives and notes
Numbers	Displayed values	Calculated results rounded once to four decimals with trailing zeros dropped (3.1, not 3.1000), and with no thousands separator; values shown as one set, such as a table column or an interval's limits, share the fewest decimals, up to four, that show every member; the leading zero is kept except for bounded quantities (p , r , R^2 , proportions, probabilities: .0059); a nonzero value that rounds to zero is written as a bound ($p < .0001$), and a proportion or probability that rounds to 1 without equalling 1 keeps four decimals; negative numbers with a true minus sign (−0.2179) in prose, equations, and every authored table; counts, ordinary df , and codes as integers, and a count of 1,000 or more takes a comma in text and tables	Ch. 1	Native software output printed as output and figure axis labels keep their own format; that output and the text R draws in a figure keep R's hyphen before a negative value
GLM	π_i , $\pi(x)$, $\hat{\pi}_i$	Conditional event probability and its fitted value	Ch. 11	Chapter 12 adds a category index (row below)
GLM	K , π_{ik}	Number of outcome categories and the probability that unit i falls in category k	Ch. 2; Ch. 12	Chapter 2 writes the probabilities of one categorical variable as $\pi_1, \dots, \pi_K$
GLM	η_i , logit	Linear predictor; the logit link $\text{logit}(\pi) = \log\{\pi/(1 - \pi)\}$ and its inverse logit^{-1}	Ch. 11; Ch. 12	—
GLM	D , G^2	Deviance and a deviance difference referred to χ_q^2	Ch. 11; Ch. 12	—
Prediction	MAE	Mean absolute error of held-out predictions	Ch. 8; Ch. 13	Distinct from RMSE, which squares the errors before averaging