Glossary

2×2 **contingency table**. A 2×2 contingency table is used to describe the association between a binary independent variable and a binary response variable of interest.

acceptance region. In hypothesis testing, an acceptance region is a set of sample values for which the null hypothesis cannot be rejected or can be accepted. It is the complement of the rejection region.

actual alpha, **actual significance level**. This is an attained or observed significance level.

allocation ratio. This ratio n_2/n_1 represents the number of subjects in the comparison, experimental group relative to the number of subjects in the reference, control group. Also see [PSS] **unbalanced designs**.

alpha. Alpha, α, denotes the significance level.

alternative hypothesis. In hypothesis testing, the alternative hypothesis represents the counterpoint to which the null hypothesis is compared. When the parameter being tested is a scalar, the alternative hypothesis can be either one sided or two sided.

alternative value, **alternative parameter**. This value of the parameter of interest under the alternative hypothesis is fixed by the investigator in a power and sample-size analysis. For example, alternative mean value and alternative mean refer to a value of the mean parameter under the alternative hypothesis.

analysis of variance, **ANOVA**. This is a class of statistical models that studies differences between means from multiple populations by partitioning the variance of the continuous outcome into independent sources of variation due to effects of interest and random variation. The test statistic is then formed as a ratio of the expected variation due to the effects of interest to the expected random variation. Also see *one-way ANOVA*, *two-way ANOVA*, *one-way repeated-measures ANOVA*, and *two-way repeated-measures ANOVA*.

balanced design. A balanced design represents an experiment in which the numbers of treated and untreated subjects are equal. For many types of two-sample hypothesis tests, the power of the test is maximized with balanced designs.

beta. Beta, β, denotes the probability of committing a type II error, namely, failing to reject the null hypothesis even though it is false.

between-subjects design. This is an experiment that has only between-subjects factors. See [PSS] **power oneway** and [PSS] **power twoway**.

between-subjects factor. This is a factor for which each subject receives only one of the levels.

binomial test. A binomial test is a test for which the exact sampling distribution of the test statistic is binomial; see [R] **bitest**. Also see [PSS] **power oneproportion**.

bisection method. This method finds a root x of a function $f(x)$ such that $f(x) = 0$ by repeatedly subdividing an interval on which $f(x)$ is defined until the change in successive root estimates is within the requested tolerance and function $f(\cdot)$ evaluated at the current estimate is sufficiently close to zero.

case–control study. An observational study that retrospectively compares characteristics of subjects with a certain problem (cases) with characteristics of subjects without the problem (controls). For example, to study association between smoking and lung cancer, investigators will sample subjects with and without lung cancer and record their smoking status. Case–control studies are often used to study rare diseases.

CCT. See *controlled clinical trial*.

cell means. These are means of the outcome of interest within cells formed by the cross-classification of the two factors. See [PSS] **power twoway** and [PSS] **power repeated**.

cell-means model. A cell-means model is an ANOVA model formulated in terms of cell means.

chi-squared test, χ^2 **test**. This test for which either an asymptotic sampling distribution or a sampling distribution of a test statistic is χ^2. See [PSS] **power onevariance** and [PSS] **power twoproportions**.

clinical trial. A clinical trials is an experiment testing a medical treatment or procedure on human subjects.

clinically meaningful difference, **clinically meaningful effect**, **clinically significant difference**. Clinically meaningful difference represents the magnitude of an effect of interest that is of clinical importance. What is meant by "clinically meaningful" may vary from study to study. In clinical trials, for example, if no prior knowledge is available about the performance of the considered clinical procedure, a standardized effect size (adjusted for standard deviation) between 0.25 and 0.5 may be considered of clinical importance.

cohort study. Typically an observational study, a cohort study may also be an experimental study in which a cohort, a group of subjects who have similar characteristics, is followed over time and evaluated at the end of the study. For example, cohorts of vaccinated and unvaccinated subjects are followed over time to study the effectiveness of influenza vaccines.

columns in graph. Think of `power, graph()` as graphing the columns of `power, table`. One of the columns will be placed on the x axis, another will be placed on the y axis, and, if you have more columns with varying values, separate plots will be created for each. Similarly, we use the terms "column symbol", "column name", and "column label" to refer to symbols, names, and labels that appear in tables when tabular output is requested.

comparison value. See *alternative value*.

compound symmetry. A covariance matrix has a compound-symmetry structure if all the variances are equal and all the covariances are equal. This is a special case of the sphericity assumption.

concordant pairs. In a 2×2 contingency table, a concordant pair is a pair of observations that are both either successes or failures. Also see *discordant pairs* and *Introduction* under *Remarks and examples* in [PSS] **power pairedproportions**.

contrasts. Contrasts refers to a linear combination of cell means such that the sum of contrast coefficients is zero.

control group. A control group comprises subjects that are randomly assigned to a group where they receive no treatment or receives a standard treatment. In hypothesis testing, this is usually a reference group. Also see *experimental group*.

controlled clinical trial. This is an experimental study in which treatments are assigned to two or more groups of subjects without the randomization.

critical region. See *rejection region*.

critical value. In hypothesis testing, a critical value is a boundary of the rejection region.

cross-sectional study. This type of observational study measures various population characteristics at one point in time or over a short period of time. For example, a study of the prevalence of breast cancer in the population is a cross-sectional study.

delta. Delta, δ, in the context of power and sample-size calculations, denotes the effect size.

directional test. See *one-sided test*.

discordant pairs. In a 2 × 2 contingency table, discordant pairs are the success–failure or failure–success pairs of observations. Also see *concordant pairs* and *Introduction* under *Remarks and examples* in [PSS] **power pairedproportions**.

discordant proportion. This is a proportion of discordant pairs. Also see *Introduction* under *Remarks and examples* in [PSS] **power pairedproportions**.

dropout. Dropout is the withdrawal of subjects before the end of a study and leads to incomplete or missing data.

effect size. The effect size is the size of the clinically significant difference between the treatments being compared, typically expressed as a quantity that is independent of the unit of measure. For example, in a one-sample mean test, the effect size is a standardized difference between the mean and its reference value. In other cases, the effect size may be measured as an odds ratio or a risk ratio. See [PSS] **intro** to learn more about the relationship between effect size and the power of a test.

effect-size curve. The effect-size curve is a graph of the estimated effect size or target parameter as a function of some other study parameter such as the sample size. The effect size or target parameter is plotted on the y axis, and the sample size or other parameter is plotted on the x axis.

effect-size determination. This pertains to the computation of an effect size or a target parameter given power, sample size, and other study parameters.

equal-allocation design. See *balanced design*.

exact test. An exact test is one for which the probability of observing the data under the null hypothesis is calculated directly, often by enumeration. Exact tests do not rely on any asymptotic approximations and are therefore widely used with small datasets. See [PSS] **power oneproportion** and [PSS] **power twoproportions**.

experimental group. An experimental group is a group of subjects that receives a treatment or procedure of interest defined in a controlled experiment. In hypothesis testing, this is usually a comparison group. Also see *control group*.

experimental study. In an experimental study, as opposed to an observational study, the assignment of subjects to treatments is controlled by investigators. For example, a study that compares a new treatment with a standard treatment by assigning each treatment to a group of subjects is an experimental study.

F test. An F test is a test for which a sampling distribution of a test statistic is an F distribution. See [PSS] **power twovariances**.

factor, **factor variables**. This is a categorical explanatory variable with any number of levels.

finite population correction. When sampling is performed without replacement from a finite population, a finite population correction is applied to the standard error of the estimator to reduce sampling variance.

Fisher–Irwin's exact test. See *Fisher's exact test*.

Fisher's exact test. Fisher's exact test is an exact small sample test of independence between rows and columns in a 2 × 2 contingency table. Conditional on the marginal totals, the test statistic has a hypergeometric distribution under the null hypothesis. See [PSS] **power twoproportions** and [R] **tabulate twoway**.

Fisher's z test. This is a z test comparing one or two correlations. See [PSS] **power onecorrelation** and [PSS] **power twocorrelations**. Also see *Fisher's z transformation*.

Fisher's z transformation. Fisher's z transformation applies an inverse hyperbolic tangent transformation to the sample correlation coefficient. This transformation is useful for testing hypothesis concerning Pearson's correlation coefficient. The exact sampling distribution of the correlation coefficient is complicated, while the transformed statistic is approximately standard normal.

fixed effects. Fixed effects represent all levels of the factor that are of interest.

follow-up study. See *cohort study*.

Greenhouse–Geisser correction. See *nonsphericity correction*.

$\mathbf{H_0}$. See *null hypothesis*.

$\mathbf{H_a}$. See *alternative hypothesis*.

hypothesis. A hypothesis is a statement about a population parameter of interest.

hypothesis testing, **hypothesis test**. This method of inference evaluates the validity of a hypothesis based on a sample from the population. See *Hypothesis testing* under *Remarks and examples* in [PSS] **intro**.

hypothesized value. See *null value*.

interaction effects. Interaction effects measure the dependence of the effects of one factor on the levels of the other factor. Mathematically, they can be defined as the differences among treatment means that are left after main effects are removed from these differences.

Lagrange multiplier test. See *score test*.

likelihood-ratio test. The likelihood-ratio (LR) test is one of the three classical testing procedures used to compare the fit of two models, one of which, the constrained model, is nested within the full (unconstrained) model. Under the null hypothesis, the constrained model fits the data as well as the full model. The LR test requires one to determine the maximal value of the log-likelihood function for both the constrained and the full models. See [PSS] **power twoproportions** and [R] **lrtest**.

lower one-sided test, **lower one-tailed test**. A lower one-sided test is a one-sided test of a scalar parameter in which the alternative hypothesis is lower one sided, meaning that the alternative hypothesis states that the parameter is less than the value conjectured under the null hypothesis. Also see *One-sided test versus two-sided test* under *Remarks and examples* in [PSS] **intro**.

main effects. These are average, additive effects that are associated with each level of each factor. For example, the main effect of level j of a factor is the difference between the mean of all observations on the outcome of interest at level j and the grand mean.

marginal homogeneity. Marginal homogeneity refers to the equality of one or more row marginal proportions with the corresponding column proportions. Also see *Introduction* under *Remarks and examples* in [PSS] **power pairedproportions**.

marginal proportion. This represents a ratio of the number of observations in a row or column of a contingency table relative to the total number of observations. Also see *Introduction* under *Remarks and examples* in [PSS] **power pairedproportions**.

matched study. In a matched study, an observation from one group is matched to an observation from another group with respect to one or more characteristics of interest. Also see *paired data*.

McNemar's test. McNemar's test is a test used to compare two dependent binary populations. The null hypothesis is formulated in the context of a 2×2 contingency table as a hypothesis of marginal homogeneity. See [PSS] **power pairedproportions** and [ST] **epitab**.

MDES. See *minimum detectable effect size*.

mean contrasts. See *contrasts*.

minimum detectable effect size. The minimum detectable effect size is the smallest effect size that can be detected by hypothesis testing for a given power and sample size.

minimum detectable value. The minimum detectable value represents the smallest amount or concentration of a substance that can be reliably measured.

mixed design. A mixed design is an experiment that has at least one between-subjects factor and one within-subject factor. See [PSS] **power repeated**.

negative effect size. In power and sample-size analysis, we obtain a negative effect size when the postulated value of the parameter under the alternative hypothesis is less than the hypothesized value of the parameter under the null hypothesis. Also see *positive effect size*.

nominal alpha, **nominal significance level**. This is a desired or requested significance level.

noncentrality parameter. In power and sample-size analysis, a noncentrality parameter is the expected value of the test statistic under the alternative hypothesis.

nondirectional test. See *two-sided test*.

nonsphericity correction. This is a correction used for the degrees of freedom of a regular F test in a repeated-measures ANOVA to compensate for the lack of sphericity of the repeated-measures covariance matrix.

null hypothesis. In hypothesis testing, the null hypothesis typically represents the conjecture that one is attempting to disprove. Often the null hypothesis is that a treatment has no effect or that a statistic is equal across populations.

null value, **null parameter**. This value of the parameter of interest under the null hypothesis is fixed by the investigator in a power and sample-size analysis. For example, null mean value and null mean refer to the value of the mean parameter under the null hypothesis.

observational study. In an observational study, as opposed to an experimental study, the assignment of subjects to treatments happens naturally and is thus beyond the control of investigators. Investigators can only observe subjects and measure their characteristics. For example, a study that evaluates the effect of exposure of children to household pesticides is an observational study.

observed level of significance. See *p-value*.

odds and odds ratio. The odds in favor of an event are Odds $= p/(1-p)$, where p is the probability of the event. Thus if $p = 0.2$, the odds are 0.25, and if $p = 0.8$, the odds are 4.

The log of the odds is $\ln(\text{Odds}) = \text{logit}(p) = \ln\{p/(1-p)\}$, and logistic regression models, for instance, fit $\ln(\text{Odds})$ as a linear function of the covariates.

The odds ratio is a ratio of two odds: $\text{Odds}_2/\text{Odds}_1$. The individual odds that appear in the ratio are usually for an experimental group and a control group or for two different demographic groups.

one-sample test. A one-sample test compares a parameter of interest from one sample with a reference value. For example, a one-sample mean test compares a mean of the sample with a reference value.

one-sided test, **one-tailed test**. A one-sided test is a hypothesis test of a scalar parameter in which the alternative hypothesis is one sided, meaning that the alternative hypothesis states that the parameter is either less than or greater than the value conjectured under the null hypothesis but not both. Also see *One-sided test versus two-sided test* under *Remarks and examples* in [PSS] **intro**.

one-way ANOVA, **one-way analysis of variance**. A one-way ANOVA model has a single factor. Also see [PSS] **power oneway**.

one-way repeated-measures ANOVA. A one-way repeated-measures ANOVA model has a single within-subject factor. Also see [PSS] **power repeated**.

paired data. Paired data consist of pairs of observations that share some characteristics of interest. For example, measurements on twins, pretest and posttest measurements, before and after measurements, repeated measurements on the same individual. Paired data are correlated and thus must be analyzed by using a paired test.

paired observations. See *paired data*.

paired test. A paired test is used to test whether the parameters of interest of two paired populations are equal. The test takes into account the dependence between measurements. For this reason, paired tests are usually more powerful than their two-sample counterparts. For example, a paired-means or paired-difference test is used to test whether the means of two paired (correlated) populations are equal.

Pearson's correlation. Pearson's correlation ρ, also known as the product-moment correlation, measures the degree of association between two variables. Pearson's correlation equals the variables' covariance divided by their respective standard deviations, and ranges between -1 and 1. Zero indicates no correlation between the two variables.

population parameter. See *target parameter*.

positive effect size. In power and sample-size analysis, we obtain a positive effect size when the postulated value of the parameter under the alternative hypothesis is greater than the hypothesized value of the parameter under the null hypothesis. Also see *negative effect size*.

postulated value. See *alternative value*.

power. The power of a test is the probability of correctly rejecting the null hypothesis when it is false. It is often denoted as $1-\beta$ in the statistical literature, where β is the type II error probability. Commonly used values for power are 80% and 90%. See [PSS] **intro** for more details about power.

power and sample-size analysis. Power and sample-size analysis investigates the optimal allocation of study resources to increase the likelihood of the successful achievement of a study objective. See [PSS] **intro**.

power curve. A power curve is a graph of the estimated power as a function of some other study parameter such as the sample size. The power is plotted on the y axis, and the sample size or other parameter is plotted on the x axis. See [PSS] **power, graph**.

power determination. This pertains to the computation of a power given sample size, effect size, and other study parameters.

power function. The power functions is a function of the population parameter θ, defined as the probability that the observed sample belongs to the rejection region of a test for given θ. See *Hypothesis testing* under *Remarks and examples* in [PSS] **intro**.

power graph. See *power curve*.

probability of a type I error. This is the probability of committing a type I error of incorrectly rejecting the null hypothesis. Also see *significance level*.

probability of a type II error. This is the probability of committing a type II error of incorrectly accepting the null hypothesis. Common values for the probability of a type II error are 0.1 and 0.2 or, equivalently, 10% and 20%. Also see *beta* and *power*.

prospective study. In a prospective study, the population or cohort is classified according to specific risk factors, such that the outcome of interest, typically various manifestations of a disease, can be observed over time and tied in to the initial classification. Also see *retrospective study*.

PSS analysis. See *power and sample-size analysis*.

PSS Control Panel. The PSS Control Panel is a point-and-click graphical user interface for power and sample-size analysis. See [PSS] **GUI**.

p-value. P-value is a probability of obtaining a test statistic as extreme or more extreme as the one observed in a sample assuming the null hypothesis is true.

random effects. Random effects represent a random sample of levels from all possible levels, and the interest lies in all possible levels.

randomized controlled trial. In this experimental study, treatments are randomly assigned to two or more groups of subjects.

RCT. See *randomized controlled trial*.

reference value. See *null value*.

rejection region. In hypothesis testing, a rejection region is a set of sample values for which the null hypothesis can be rejected.

relative risk. See *risk ratio*.

retrospective study. In a retrospective study, a group with a disease of interest is compared with a group without the disease, and information is gathered in a retrospective way about the exposure in each group to various risk factors that might be associated with the disease. Also see *prospective study*.

risk difference. A risk difference is defined as the probability of an event occurring when a risk factor is increased by one unit minus the probability of the event occurring without the increase in the risk factor.

When the risk factor is binary, the risk difference is the probability of the outcome when the risk factor is present minus the probability when the risk factor is not present.

When one compares two populations, a risk difference is defined as a difference between the probabilities of an event in the two groups. It is typically a difference between the probability in the comparison group or experimental group and the probability in the reference group or control group.

risk factor. A risk factor is a variable that is associated with an increased or decreased probability of an outcome.

risk ratio. A risk ratio, also called a relative risk, measures the increase in the likelihood of an event occurring when a risk factor is increased by one unit. It is the ratio of the probability of the event when the risk factor is increased by one unit over the probability without that increase.

When the risk factor is binary, the risk ratio is the ratio of the probability of the event when the risk factor occurs over the probability when the risk factor does not occur.

When one compares two populations, a risk ratio is defined as a ratio of the probabilities of an event in the two groups. It is typically a ratio of the probability in the comparison group or experimental group to the probability in the reference group or control group.

sample size. This is the number of subjects in a sample. See [PSS] **intro** to learn more about the relationship between sample size and the power of a test.

sample-size curve. A sample-size curve is a graph of the estimated sample size as a function of some other study parameter such as power. The sample size is plotted on the y axis, and the power or other parameter is plotted on the x axis.

sample-size determination. This pertains to the computation of a sample size given power, effect size, and other study parameters.

Satterthwaite's t test. Satterthwaite's t test is a modification of the two-sample t test to account for unequal variances in the two populations. See *Methods and formulas* in [PSS] **power twomeans** for details.

score test. A score test, also known as a Lagrange multiplier test, is one of the three classical testing procedures used to compare the fit of two models, one of which, the constrained model, is nested within the full (unconstrained) model. The null hypothesis is that the constrained model fits the data as well as the full model. The score test only requires one to fit the constrained model. See [PSS] **power oneproportion** and [R] **prtest**.

sensitivity analysis. Sensitivity analysis investigates the effect of varying study parameters on power, sample size, and other components of a study. The true values of study parameters are usually unknown, and power and sample-size analysis uses best guesses for these values. It is therefore important to evaluate the sensitivity of the computed power or sample size in response to changes in study parameters. See [PSS] **power, table** and [PSS] **power, graph** for details.

sign test. A sign test is used to test the null hypothesis that the median of a distribution is equal to some reference value. A sign test is carried out as a test of binomial proportion with a reference value of 0.5. See [PSS] **power oneproportion** and [R] **bitest**.

significance level. In hypothesis testing, the significance level α is an upper bound for a probability of a type I error. See [PSS] **intro** to learn more about the relationship between significance level and the power of a test.

size of test. See *significance level*.

sphericity assumption. All differences between levels of the within-subject factor within-subject factor have the same variances.

symmetry. In a 2×2 contingency table, symmetry refers to the equality of the off-diagonal elements. For a 2×2 table, a test of marginal homogeneity reduces to a test of symmetry.

t test. A t test is a test for which the sampling distribution of the test statistic is a Student's t distribution.

A one-sample t test is used to test whether the mean of a population is equal to a specified value when the variance must also be estimated. The test statistic follows Student's t distribution with $N - 1$ degrees of freedom, where N is the sample size.

A two-sample t test is used to test whether the means of two populations are equal when the variances of the populations must also be estimated. When the two populations' variances are unequal, a modification to the standard two-sample t test is used; see *Satterthwaite's t test*.

target parameter. In power and sample-size analysis, the target parameter is the parameter of interest or the parameter in the study about which hypothesis tests are conducted.

test statistic. In hypothesis testing, a test statistic is a function of the sample that does not depend on any unknown parameters.

two-independent-samples test. See *two-sample test*.

two-sample paired test. See *paired test*.

two-sample test. A two-sample test is used to test whether the parameters of interest of the two independent populations are equal. For example, two-sample means test, two-sample variances, two-sample proportions test, two-sample correlations test.

two-sided test, two-tailed test. A two-sided test is a hypothesis test of a parameter in which the alternative hypothesis is the complement of the null hypothesis. In the context of a test of a scalar parameter, the alternative hypothesis states that the parameter is less than or greater than the value conjectured under the null hypothesis.

two-way ANOVA, two-way analysis of variance. A two-way ANOVA model contains two factors. Also see [PSS] **power twoway**.

two-way repeated-measures ANOVA, **two-factor** ANOVA. This is a repeated-measures ANOVA model with one within-subject factor and one between-subjects factor. The model can be additive (contain only main effects of the factors) or can contain main effects and an interaction between the two factors. Also see [PSS] **power repeated**.

type I error. The type I error of a test is the error of rejecting the null hypothesis when it is true; see [PSS] **intro** for more details.

type II error. The type II error of a test is the error of not rejecting the null hypothesis when it is false; see [PSS] **intro** for more details.

type I error probability. See *probability of a type I error*.

type II error probability. See *probability of a type II error*.

unbalanced design. An unbalanced design indicates an experiment in which the numbers of treated and untreated subjects differ. Also see [PSS] **unbalanced designs**.

unequal-allocation design. See *unbalanced design*.

upper one-sided test, **upper one-tailed test**. An upper one-sided test is a one-sided test of a scalar parameter in which the alternative hypothesis is upper one sided, meaning that the alternative hypothesis states that the parameter is greater than the value conjectured under the null hypothesis. Also see *One-sided test versus two-sided test* under *Remarks and examples* in [PSS] **intro**.

Wald test. A Wald test is one of the three classical testing procedures used to compare the fit of two models, one of which, the constrained model, is nested within the full (unconstrained) model. Under the null hypothesis, the constrained model fits the data as well as the full model. The Wald test requires one to fit the full model but does not require one to fit the constrained model. Also see [PSS] **power oneproportion** and [R] **test**.

within-subject design. This is an experiment that has at least one within-subject factor. See [PSS] **power repeated**.

within-subject factor. This is a factor for which each subject receives several or all the levels.

z test. A z test is a test for which a potentially asymptotic sampling distribution of the test statistic is a normal distribution. For example, a one-sample z test of means is used to test whether the mean of a population is equal to a specified value when the variance is assumed to be known. The distribution of its test statistic is normal. See [PSS] **power onemean**, [PSS] **power twomeans**, and [PSS] **power pairedmeans**.