Methods. September, 2020



L

----

**METHODS** 

# Glossary of Statistical Methods

Zad Rafi 1 🕩

Rafi, Z.(2020) Glossary of Statistical Methods. **Methods**. The Paleo Foundation.

<sup>1</sup>Department of Standards, Paleo Foundation, New York, NY

**Correspondence** Zad Rafi Department of Standards, Paleo Foundation, New York, NY

Contact <sup>1</sup>Email: zad@lesslikely.com <sup>1</sup>Twitter: @DailyZad

#### METHODS

# **Glossary of Statistical Methods**

Zad Rafi 1 🕩

<sup>1</sup> Department of Standards, Paleo Foundation, New York, NY

**Correspondence** Karen E. E. Pendergrass Department of Standards, Paleo Foundation, Encinitas, CA

**Contact** <sup>1</sup>Email: zad@lesslikely.com <sup>1</sup>Twitter: @DailyZad

# Glossary

Statistical methods are routinely used to analyze data in nearly every scientific discipline. Despite their wide usage, many descriptions of these methods are inaccurate, stemming from confusion of the concept itself or from incorrect teaching materials. Here we provide a glossary of many of these terms along with selective references and links in which these concepts are meticulously discussed.

#### **KEYWORDS**

Bayesian Statistics, S-Values, P-Values, Statistical Significance, Hypothesis Testing, Classical Statistics

# 1 | INTRODUCTION

S tatistical methods are routinely used to analyze data in nearly every scientific discipline. Despite their wide usage, many descriptions of these methods are inaccurate, stemming from confusion of the concept itself or from incorrect teaching materials.

In a recent survey of introductory psychology textbooks for undergraduates, which often contain brief discussions of research methods, nearly 89% of the sampled 30 textbooks contained incorrect descriptions of P-values and statistical significance. Such errors are not only restricted to psychology but can be found in nearly every discipline that employs statistical methods, being present even within statistics journals themselves. Here we provide a glossary of many of these terms along with selective references and links in which these concepts are meticulously discussed.

# 2 | STATISTICAL CONCEPTS

#### **BAYESIAN STATISTICS:**

A school of statistics that uses Bayes theorem to take into account prior information (priors) along with the likelihood of the data to produce posterior probabilities, and encompasses Bayesian inference, philosophy, and mathematical statistics. There are several variants within Bayesian statistics such as objective Bayes, subjective (personal) Bayes, empirical Bayes, hierarchical Bayes, etc.

# **BAYES FACTOR**

A ratio of the posterior probabilities for the test hypothesis and an alternative, given certain model restrictions/assumptions (the Bayesian variant of the likelihood ratio).

#### **BAYES THEOREM**

A mathematical formula believed to be originally proposed by the reverend Thomas Bayes in which prior information is combined with the data to produce a posterior probability. The theorem is popular in clinical diagnostic testing and in Bayesian statistics.

Likelihood ratio: A ratio comparing the likelihood of the test hypothesis (such as a null hypothesis) to some specified alternative.

#### NULL HYPOTHESIS

Often a statistical hypothesis in which it is assumed that the true parameter value of interest is 0 (for continuous outcomes) or 1 (for binary outcomes).

Alternative hypothesis: A statistical hypothesis, that is often testing a different parameter value from the null and is "accepted" in some contexts (Neyman-Pearson decision theoretic) where the null is tested and rejected. For example, a statistical test of the risk ratio being 2 (alternative hypothesis) as opposed to the null hypothesis (the risk ratio being 1, indicative of no difference).

# STATISTICAL POWER

The pre-study probability of correctly rejecting the test hypothesis (usually the null hypothesis) when the alternative is correct. This quantify often assumes repeated replications using similar alpha levels (the alpha level must be valid) and study designs.

#### **TYPE I ERROR**

Concluding that a result is present/positive, when it is actually negative.

#### TYPE II ERROR

Concluding that a result is absent/negative, when it actually exists.

#### ALPHA LEVEL

The maximum tolerable, type I error rate for a statistical testing procedure that is specified before a study, hence it is fixed. It is often 5% or 0.05 in many research settings, however, this is a result of tradition and not based on empirical findings. Some meta-researchers and statisticians have proposed reducing this threshold to 0.5% or 0.005, to reduce the potential of false positives.

#### **BETA LEVEL**

The maximum tolerable type II error rate for a statistical testing procedure. It is 1 – power (of a testing procedure) \* 100.

# SAMPLING DISTRIBUTION

The distribution of a statistic with a specified probability distribution.

#### STANDARD DEVIATION

A measure (for some probability distributions, referred to as the scale parameter) that describes the spread of the data. In simple scenarios (for example, continuous outcomes where means between two groups are being contrasted), it is the square root of the variance.

#### STANDARD ERROR

The standard deviation of the sampling distribution.

Often used to calculate test statistics and confidence intervals. In simple scenarios (for example, continuous outcomes where means between two groups are being contrasted), it is the standard deviation divided by the square root of the total sample size.

#### SAMPLE SIZE

The number of participants in a particular study. N often refers to the total sample size of the study, whereas n may refer to the number of participants in a particular group/subgroup.

# **META-ANALYSIS**

A statistical technique used to combine information from various studies by pooling the location and scale parameters. There are various forms of meta-analysis, with some procedures pooling effect estimates, and others pooling test statistics (ex: Fisher's method for combining P-values).

# TEST STATISTIC

A data summary/feature that is sometimes computed by taking the effect estimate over the standard error (ex: a Z-test comparing means).

#### **P-VALUE**

The probability of getting a test statistic (a data summary/feature) at least as extreme as what was observed, if every model assumption in addition to the targeted test hypothesis (such as a null hypothesis) were correct.

- Not the probability of the results being a result of chance
- Not the probability of the hypothesis or of it being correct
- Not the probability of a parameter value

#### S-VALUE

Also known as surprisal or Shannon-information, it is the transformation of the p-value by taking the observed P-value p, and taking the base-2 log of it, which results in a measure of information that is unbounded by 0 or 1, stopping it from being confused for a posterior probability.

#### **CONFIDENCE INTERVAL**

A region (with a particular percentile such as 95%) that contains values that are not rejected at the chosen alpha level, and which contain the true parameter value (1-a)\*100 of the time when repeatedly conducting a study and producing an interval estimate at a particular percentile level. The simplest way to produce a CI for a z-test of means between two groups is to take the mean difference and to multiply the standard error by 1.96 and to add and subtract it from the mean difference.

 Not the probability of the true parameter value being within the estimated interval, which is generally a Bayesian posterior interval.

# CONFIDENCE DISTRIBUTION

The frequentist variant of a posterior probability distribution, often includes confidence intervals at every percentile level to form a curve/distribution and encompasses other distributions such as the bootstrap distribution. It is \*not\* a probability distribution.

# STATISTICAL SIGNIFICANCE

When the observed P-value p (to differentiate from the random variable P) from a study falls below the pre-specified alpha level (often 5%). Does not indicate that a study result is meaningful or practical. Statistical significance is often declared when a P-value falls below 5% or 0.05, which is a threshold that is used as a result of tradition. Some meta-researchers and statisticians have proposed reducing this cutoff to 0.005, to reduce the potential of false positives.

# HYPOTHESIS TESTING

Generally, refers to the Neyman-Pearson decision theoretic and procedures for testing null hypotheses and specifying various parameters such as the type I error rate, type II error rate, and accepting alternative hypotheses. The emphasis is not on the exact value of the observed P-value but whether to reject the test hypothesis/accept the alternative based on whether p fell below the specified type I error rate (alpha).

The Bayesian variant is using Bayes factors to quantify the strength of evidence for a particular hypothesis.

# SIGNIFICANCE TESTING

The Fisherian method (in contrast to the Neyman-Pearson decision theoretic) used to compute test statistics and P-values and treat them as continuous measures of evidence again the test hypothesis (usually the null hypothesis). In his early writing, the statistician Ronald Fisher often wrote about the use of 5% as a testing threshold, however, later in his life, he discouraged using this threshold mindlessly.

#### CENSORING

A form of missing data in which the time to event is not observed for various reasons (see right censoring and left censoring).

# MULTIPLE IMPUTATION

A method which uses local residual draws, for example, to impute missing values and create multiple datasets with various imputed values where a model is fit to each dataset and then pooled to produce an estimate.

#### RESAMPLING

A statistical procedure used to redraw observations from a sample of data with or without replacement to form a distribution which is then used for inferential purposes.

# SHRINKAGE ESTIMATION

A form of estimation in which the estimates are penalized based on some hyperparameter which induces a penalty to improve frequentist properties, often used in contexts in which maximum likelihood estimation produces inaccurate results (used in highdimensional settings, variable selection, etc.).

# PENALIZATION

See shrinkage estimation.

#### **REGULARIZATION:**

See shrinkage estimation.

# CLASSICAL STATISTICS

Often called frequentist statistics, it is the dominant school of statistics in which the primary focus is on long-run frequencies and methods and summaries that take these into account (see P-values, alpha level, and confidence interval). Rafi, Z. (2020) Glossary of Statistical Methods. **Methods.** The Paleo Foundation.

This work is licensed under a <u>Creative</u> <u>Commons Attribution 4.0 International License.</u>



