
Glossary of common statistics terms

ANCOVA – analysis of covariance. ANOVA with 1 or more continuous covariates in addition to factors. The analysis adjusts for significant covariates before testing for factor (group) effects.

ANOVA – analysis of variance. A statistical test to compare the means of several groups simultaneously. It assumes random sampling from independent normally distributed data. Multiple factors (eg, treatments, groups, etc) can be considered simultaneously. ANOVA generalizes the Student *t* test to more than 2 groups.

Bonferroni correction – a simple, but conservative, method of adjusting *p* values in multiple comparisons situations. The idea is to divide *P* by the number of planned comparisons to get a smaller *p* value. This eliminates the problem of increased type I error (ie, testing until you find something significant).

Critical value – a preset value of a test statistic (ie, critical value is 1.96) that reflects the level of probability used for rejecting the null hypothesis. The value determines the probability of a type I error, incorrectly rejecting the null hypothesis, that the investigator is willing to accept. This probability is traditionally called α .

Degrees of freedom – the number of independent observations associated with an estimate of variance or of a component of an ANOVA is free to vary. For example, in an ANOVA with 3 groups (namely, 1, 2, and 3), with a sample of *n* 1 observations, the degrees of freedom is *n* – 1 when you calculate the mean for this group; 1 *df* is used in calculating the mean value from which the variance can be calculated.

Interval level of measurement – requires that data can be ordered (unlike nominal data) but also that the distance between levels is the same. For example, degrees Celsius: the distance between 10° and 15° is the same as the distance between 15° and 20°. Subtraction and addition are meaningful; however, choice of a zero point is arbitrary.

Kruskal-Wallis test – a nonparametric test to compare more than 2 groups simultaneously. It is an extension of the Mann-Whitney 2-sample test.

Mann-Whitney – a nonparametric method for comparing 2 groups. It is based on ranking the raw data and comparing mean ranks between groups.

Multiple comparisons problem – the inflation of type I error (the probability of rejecting the null hypothesis when it is true) that occurs when multiple statistical tests are performed. The more tests that are performed the more likely it is that a test will be declared significant by chance alone when it isn't significant.

Normal distribution – an important probability distribution completely specified by a mean and SD. The data follow a

familiar symmetric bell-shaped distribution. The normal distribution is also called a gaussian distribution because it was first described mathematically by Carl Friedrich Gauss.

Null hypothesis – the hypothesis that assumes no effect of the experiment (ie, no difference between outcome variables for the 2 or more groups being studied).

Parameter – a predefined measurement on a population that is used to characterize a feature of the population. Examples include the population mean and SD.

Population – the complete collection of items or subjects of interest in a study that share a common feature to be measured. A population is usually too large or inaccessible to enable observations on every subject, thus the need for sampling. An example of a population might be all patients with hypertension in a given geographic region.

Power – the probability of correctly rejecting the null hypothesis (eg, concluding that 2 treatments are different when in fact they are). Power is often expressed as a percentage. For study planning, power levels of 80%, 90%, and 95% are commonly used. The power is equal to $1 - \beta$ (see type II error).

P value – the probability of incorrectly rejecting the null hypothesis. When *p* is less than a prespecified number, the critical value, the null hypothesis is rejected. Traditionally, $p < .05$ is grounds for rejecting the null hypothesis, but other values can be used at the judgment of the investigator, depending on the situation.

Random sampling – a method of selecting subjects in a manner that eliminates bias. Random number tables generated by computational methods are commonly used. There are 5 commonly used methods of random sampling: simple random sampling, systematic sampling, stratified sampling, cluster sampling, and multistage sampling.

Repeated-measures ANOVA – a special form of ANOVA in which 1 or more of the factors is measured repeatedly on each subject. Because the repeated measures are correlated, special methods of computing mean square errors are necessary.

Sample – a subset of items or subjects drawn from a population to draw inferences about the population. The most common approach to sampling is random sampling.

SD – standard deviation, a measure of variability in a sample, calculated from the squared distance from the mean for each value and the sample size in a sample.

SE – standard error of the mean, calculated as follows: $SE = SD/\sqrt{n}$, where *S* is the SD and *n* is the sample size.

Statistic – a predefined measure obtained on a sample that characterizes the sample, obtained for the purposes of estimating that measure in the population (see parameter). Examples include the sample mean and sample SD.

Statistically significant – a study conclusion based on a p value less than the prespecified critical value is considered to be statistically significant. This means that the probability of rejecting the null hypothesis by chance is acceptably low. Commonly this means $p < .05$, but other values may be used.

Student t distribution – a probability distribution similar to the normal distribution that is used to estimate the mean of the normal distribution when the sample size is small. The width of the distribution is dependent on the size of the sample and approaches the normal distribution when the sample size gets large.

t test – a test that uses the t statistic with the Student t

distribution to test hypotheses about the mean of a population or for comparing the means of 2 populations.

Type I error – rejecting the null hypothesis when it is actually correct (eg, saying a new drug is better than the standard treatment when in fact the 2 drugs are equivalent). The probability of a type I error is designated as α .

Type II error – failing to reject the null hypothesis when it is not true (eg, saying 2 drugs are the same when in fact they are different). The probability of a type II error is designated as β .

$\bar{X}$ (pronounced “x bar”) – the usual symbol for the mean of a sample.