

Glossary

alternative hypothesis | A statement that is accepted as true only if there is convincing evidence in favor of it.

ANOVA F -test | A test based on an F statistic to check whether several population means are equal.

binomial random variable | A random variable that counts successes in a fixed number of independent, identical trials of a success/failure experiment.

box plot | For a data set, a diagram constructed using the five-number summary, which graphically summarizes the distribution of the data.

chi-square goodness-of-fit test | A test based on a chi-square statistic to check whether a sample is taken from a population with a hypothesized probability distribution.

chi-square random variable | A random variable that follows a chi-square distribution.

chi-square test | A test based on a chi-square statistic to check whether two factors are independent.

coefficient of determination | A number that measures the proportion of the variability in y that is explained by x .

complement of an event | The event does not occur.

conditional probability | The probability of the event A taking into account the fact that event B is known to have occurred.

confidence interval | An interval with endpoints $\bar{x} \pm E$, computed from the sample data in such a way that a specified proportion of all intervals constructed by this process will contain the parameter of interest.

continuous | A random variable whose possible values contain an interval of decimal numbers.

critical value | The number or one of a pair of numbers that determines the rejection region.

data list | An explicit listing of all the individual measurements made on a sample.

degrees of freedom | A number that specifies a particular t -distribution and that is computed based on the sample size.

density function | The function $f(x)$ such that probabilities of a continuous random variable X are areas of regions under the graph of $y=f(x)$.

Descriptive statistics | The organization, display, and description of data.

error | Using $y - y^A$, the actual y -value of a data point minus the y -value that is computed from the equation of the line fitting the data.

event | Any set of outcomes.

expected value | Its mean.

extrapolation | The process of using the least squares regression equation to estimate the value of y at an x value not in the proper range.

F -distribution | A particular probability distribution specified by two degrees of freedom, df_1 and df_2 .

factors | A variable with several qualitative levels.

F random variable | A random variable following an F -distribution.

five-number summary | Of a data set, the list $\{x_{\min}, Q_1, Q_2, Q_3, x_{\max}\}$.

frequency | Of a class of measurements, the number of measurements in the data set that are in the class.

hypothesis | A statement about a population parameter.

Hypothesis testing | A statistical procedure in which a choice is made between a null hypothesis and a specific alternative hypothesis based on information in a sample.

independent | Events whose probability of occurring together is the product of their individual probabilities.

Inferential statistics | Drawing conclusions about a population based on a sample.

interquartile range (IQR) | Of a data set, the difference between the first and third quartiles.

intersection of events | Both events occur.

least squares regression equation | The equation $y^A = \beta^A_1 x + \beta^A_0$ of the least squares regression line.

least squares regression line | The line that best fits a set of sample data in the sense of minimizing the sum of the squared errors.

level of confidence | The proportion of confidence intervals which, if under repeated random sampling were always constructed according to the formula of the text, would contain the parameter of interest.

level of significance of the test | The probability α that defines an event as "rare;" the probability that the test procedure will lead to a Type I error.

linear correlation coefficient | A number computed directly from the data that measures the strength of the linear relationship between the two variables x and y .

most conservative estimate | The estimate obtained using $p^A = 0.5$, which gives the largest estimate of n .

mutually exclusive | Events that cannot both occur at once.

normal distribution | Assignment of probabilities to a continuous random variable using a bell curve for the density function.

normal random variable | A continuous random variable whose probabilities are determined by a bell curve.

NOVA | Analysis of variance.

null hypothesis | The statement that is assumed to be true unless there is convincing evidence to the contrary.

observed significance or p -value | The probability, if H_0 is true, of obtaining a result as contrary to H_0 and in favor of H_a as the result observed in the sample data.

percentile rank | Of a measurement x , the percentage of the data that are less than or equal to x .

percentile rank | Of a measurement x , the percentage of the data that are less than or equal to x .

population mean | The familiar average of a population data set.

population regression line | The line with equation $y = \beta_1 x + \beta_0$ that gives the mean of the variable y over the sub-population determined by x .

population standard deviation | The variability of population data as measured by the number $\sigma^2 = \Sigma(x - \mu)^2 / N$.

probability distribution | A list of each possible value and its probability.

probability of an event | A number that measures the likelihood of the event.

probability of an outcome | A number that measures the likelihood of the outcome.

Qualitative data | Measurements for which there is no natural numerical scale.

Quantitative data | Numerical measurements that arise from a natural numerical scale.

quartiles | Of a data set, the three numbers Q_1, Q_2, Q_3 that divide the data approximately into fourths.

random variable | A numerical value generated by a random experiment.

range | The variability of a data set as measured by the number $R = x_{\max} - x_{\min}$.

rejection region | An interval or union of intervals such that the null hypothesis is rejected if and only if the statistic of interest lies in this region.

relative frequency histogram | A graphical device showing how data are distributed across the range of their values by collecting them into classes and indicating the proportion of measurements in each class.

sample | The objects examined.

sample data | The measurements from a sample.

sample mean | The familiar average of a sample data set.

sample median | The middle value when data are listed in numerical order.

sample mode | The most frequent value in a data set.

sample standard deviation | The variability of sample data as measured by the number $\Sigma(x - \bar{x})^2 / (n - 1)$.

sample standard deviation of errors | The statistic s_e .

sampling distribution | The probability distribution of a sample statistic when the statistic is viewed as a random variable.

standard deviation | The number $\Sigma(x - \mu)^2 / 2P(x)$ (also computed using $\Sigma x^2 P(x) - \mu^2$), measuring its variability under repeated trials.

standard normal random variable | The normal random variable with mean 0 and standard deviation 1.

standardized test statistic | The standardized statistic used in performing the test.

statistic | A number computed from the sample data.

Statistics | Collection, display, analysis, and inference from data.

tail | The region under a density curve whose area is either $P(X < x^*)$ or $P(X > x^*)$ for some number x^* .

Type II error | Failure to reject a false null hypothesis.

union of events | One or the other event occurs.

z-score | Of a measurement x , the distance of x from the mean in units of standard deviation.