M Test

Weierstrass M-test

In mathematics, the Weierstrass M-test is a test for determining whether an infinite series of functions converges uniformly and absolutely. It applies

In mathematics, the Weierstrass M-test is a test for determining whether an infinite series of functions converges uniformly and absolutely. It applies to series whose terms are bounded functions with real or complex values, and is analogous to the comparison test for determining the convergence of series of real or complex numbers. It is named after the German mathematician Karl Weierstrass (1815–1897).

Box's M test

Box's M test is a multivariate statistical test used to check the equality of multiple variance-covariance matrices. The test is commonly used to test the

Box's M test is a multivariate statistical test used to check the equality of multiple variance-covariance matrices. The test is commonly used to test the assumption of homogeneity of variances and covariances in MANOVA and linear discriminant analysis. It is named after George E. P. Box, who first discussed the test in 1949. The test uses a chi-squared approximation.

Box's M test is susceptible to errors if the data does not meet model assumptions or if the sample size is too large or small. Box's M test is especially prone to error if the data does not meet the assumption of multivariate normality.

Vehicle inspection in the United States

Some states, including Kentucky and Minnesota, have discontinued their testing programs in recent years with approval from the federal government.[citation

In the United States, vehicle safety inspection and emissions inspection are governed by each state individually. Fourteen states have a periodic (annual or biennial) safety inspection program, while Maryland requires a safety inspection and Alabama requires a VIN inspection on sale or transfer of vehicles which were previously registered in another state. Maryland also requires a safety inspection prior to the sale or transfer of ownership of a pre-owned vehicle. Seven states require periodic emissions inspections statewide, and eighteen additional states require periodic emissions inspections only for vehicles registered in heavily-polluted areas.

In 1977, the federal Clean Air Act was amended by Congress to require states to implement vehicle emissions inspection programs, known as I/M programs (for Inspection and Maintenance), in all major metropolitan areas whose air quality failed to meet certain federal standards. New York's program started in 1982, California's program (Smog Check) started in 1984, and Illinois's program started in 1986. The Clean Air Act of 1990 required some states to enact vehicle emissions inspection programs. States impacted were those with metropolitan areas where air quality did not meet federal standards. Some states, including Kentucky and Minnesota, have discontinued their testing programs in recent years with approval from the federal government.

Training, validation, and test data sets

different stages of the creation of the model: training, validation, and test sets. The model is initially fit on a training data set, which is a set of

In machine learning, a common task is the study and construction of algorithms that can learn from and make predictions on data. Such algorithms function by making data-driven predictions or decisions, through building a mathematical model from input data. These input data used to build the model are usually divided into multiple data sets. In particular, three data sets are commonly used in different stages of the creation of the model: training, validation, and test sets.

The model is initially fit on a training data set, which is a set of examples used to fit the parameters (e.g. weights of connections between neurons in artificial neural networks) of the model. The model (e.g. a naive Bayes classifier) is trained on the training data set using a supervised learning method, for example using optimization methods such as gradient descent or stochastic gradient descent. In practice, the training data set often consists of pairs of an input vector (or scalar) and the corresponding output vector (or scalar), where the answer key is commonly denoted as the target (or label). The current model is run with the training data set and produces a result, which is then compared with the target, for each input vector in the training data set. Based on the result of the comparison and the specific learning algorithm being used, the parameters of the model are adjusted. The model fitting can include both variable selection and parameter estimation.

Successively, the fitted model is used to predict the responses for the observations in a second data set called the validation data set. The validation data set provides an unbiased evaluation of a model fit on the training data set while tuning the model's hyperparameters (e.g. the number of hidden units—layers and layer widths—in a neural network). Validation data sets can be used for regularization by early stopping (stopping training when the error on the validation data set increases, as this is a sign of over-fitting to the training data set).

This simple procedure is complicated in practice by the fact that the validation data set's error may fluctuate during training, producing multiple local minima. This complication has led to the creation of many ad-hoc rules for deciding when over-fitting has truly begun.

Finally, the test data set is a data set used to provide an unbiased evaluation of a final model fit on the training data set. If the data in the test data set has never been used in training (for example in cross-validation), the test data set is also called a holdout data set. The term "validation set" is sometimes used instead of "test set" in some literature (e.g., if the original data set was partitioned into only two subsets, the test set might be referred to as the validation set).

Deciding the sizes and strategies for data set division in training, test and validation sets is very dependent on the problem and data available.

Convergence tests

The test is inconclusive if the limit of the summand is zero. This is also known as the nth-term test, test for divergence, or the divergence test. This

In mathematics, convergence tests are methods of testing for the convergence, conditional convergence, absolute convergence, interval of convergence or divergence of an infinite series

?

n

=

1

?

a

n

{\displaystyle \sum _{n=1}^{\infty }a_{n}}

.

Software testing

Software testing is the act of checking whether software satisfies expectations. Software testing can provide objective, independent information about

Software testing is the act of checking whether software satisfies expectations.

Software testing can provide objective, independent information about the quality of software and the risk of its failure to a user or sponsor.

Software testing can determine the correctness of software for specific scenarios but cannot determine correctness for all scenarios. It cannot find all bugs.

Based on the criteria for measuring correctness from an oracle, software testing employs principles and mechanisms that might recognize a problem. Examples of oracles include specifications, contracts, comparable products, past versions of the same product, inferences about intended or expected purpose, user or customer expectations, relevant standards, and applicable laws.

Software testing is often dynamic in nature; running the software to verify actual output matches expected. It can also be static in nature; reviewing code and its associated documentation.

Software testing is often used to answer the question: Does the software do what it is supposed to do and what it needs to do?

Information learned from software testing may be used to improve the process by which software is developed.

Software testing should follow a "pyramid" approach wherein most of your tests should be unit tests, followed by integration tests and finally end-to-end (e2e) tests should have the lowest proportion.

Bechdel test

Bechdel test (/?b?kd?l/ BEK-d?l), also known as the Bechdel-Wallace test, is a measure of the representation of women in film and other fiction. The test asks

The Bechdel test (BEK-d?l), also known as the Bechdel-Wallace test, is a measure of the representation of women in film and other fiction. The test asks whether a work features at least two women who have a conversation about something other than a man. Some versions of the test also require that those two women have names.

A work of fiction passing or failing the test does not necessarily indicate the overall representation of women in the work. Instead, the test is used as an indicator of the active presence (or lack thereof) of women in fiction, and to call attention to gender inequality in fiction.

The test is named after the American cartoonist Alison Bechdel, in whose 1985 comic strip Dykes to Watch Out For the test first appeared. Bechdel credited the idea to her friend Liz Wallace and the writings of Virginia Woolf. Originally meant as "a little lesbian joke in an alternative feminist newspaper", according to

Bechdel, the test became more widely discussed in the 2000s, as a number of variants and tests inspired by it emerged.

Chi-squared test

A chi-squared test (also chi-square or ?2 test) is a statistical hypothesis test used in the analysis of contingency tables when the sample sizes are large

A chi-squared test (also chi-square or ?2 test) is a statistical hypothesis test used in the analysis of contingency tables when the sample sizes are large. In simpler terms, this test is primarily used to examine whether two categorical variables (two dimensions of the contingency table) are independent in influencing the test statistic (values within the table). The test is valid when the test statistic is chi-squared distributed under the null hypothesis, specifically Pearson's chi-squared test and variants thereof. Pearson's chi-squared test is used to determine whether there is a statistically significant difference between the expected frequencies and the observed frequencies in one or more categories of a contingency table. For contingency tables with smaller sample sizes, a Fisher's exact test is used instead.

In the standard applications of this test, the observations are classified into mutually exclusive classes. If the null hypothesis that there are no differences between the classes in the population is true, the test statistic computed from the observations follows a ?2 frequency distribution. The purpose of the test is to evaluate how likely the observed frequencies would be assuming the null hypothesis is true.

Test statistics that follow a ?2 distribution occur when the observations are independent. There are also ?2 tests for testing the null hypothesis of independence of a pair of random variables based on observations of the pairs.

Chi-squared tests often refers to tests for which the distribution of the test statistic approaches the ?2 distribution asymptotically, meaning that the sampling distribution (if the null hypothesis is true) of the test statistic approximates a chi-squared distribution more and more closely as sample sizes increase.

SAT Subject Tests

biology test was the only test to use answers 96–100; questions 1–60 were common to both the E and M tests, in addition, the E used 61–80, and the M used

SAT Subject Tests were a set of multiple-choice standardized tests given by The College Board on individual topics, typically taken to improve a student's credentials for college admissions in the United States. For most of their existence, from their introduction in 1937 until 1994, the SAT Subject Tests were known as Achievement Tests, and until January 2005, they were known as SAT II: Subject Tests. They are still often remembered by these names. Unlike the Scholastic Aptitude Test (SAT) that the College Board offers, which are intended to measure general aptitude for academic studies, the Achievement Tests were intended to measure the level of knowledge and understanding in a variety of specific subjects. Like the SAT, the scores for an Achievement Test ranged from 200 (lowest) to 800 (highest).

Many colleges used the SAT Subject Tests for admission, course placement, and to advise students about course selection. Achievement tests were generally only required by the most selective of colleges. Some of those colleges named one or more specific Achievement Tests that they required for admission, while others allowed applicants to choose which tests to take. Students typically chose which tests to take depending upon college entrance requirements for the schools to which they planned to apply.

Fewer students took achievement tests compared to the SAT. In 1976, for instance, there were 300,000 taking one or more achievement tests, while 1.4 million took the SAT. Rates of taking the tests varied by geography; in 1974, for instance, a half of students taking the SAT in New England also took one or more achievement tests, while nationwide only a quarter did. The number of achievement tests offered varied over

time. Subjects were dropped or added based on educational changes and demand. In the early 1990s, for instance, Asian languages were added so as not to disadvantage Asian-American students, especially on the West Coast.

On January 19, 2021, the College Board discontinued Subject Tests. This was effective immediately in the United States, and the tests were to be phased out by the following summer for international students.

Levene's test

implementations of Levene's test. Bartlett's test F-test of equality of variances Box's M test Levene, Howard (1960). "Robust tests for equality of variances"

In statistics, Levene's test is an inferential statistic used to assess the equality of variances for a variable calculated for two or more groups. This test is used because some common statistical procedures assume that variances of the populations from which different samples are drawn are equal. Levene's test assesses this assumption. It tests the null hypothesis that the population variances are equal (called homogeneity of variance or homoscedasticity). If the resulting p-value of Levene's test is less than some significance level (typically 0.05), the obtained differences in sample variances are unlikely to have occurred based on random sampling from a population with equal variances. Thus, the null hypothesis of equal variances is rejected and it is concluded that there is a difference between the variances in the population.

Levene's test has been used in the past before a comparison of means to inform the decision on whether to use a pooled t-test or the Welch's t-test for two sample tests or analysis of variance or Welch's modified oneway ANOVA for multi-level tests. However, it was shown that such a two-step procedure may markedly inflate the type 1 error obtained with the t-tests and thus is not recommended. Instead, the preferred approach is to just use Welch's test in all cases.

Levene's test may also be used as a main test for answering a stand-alone question of whether two sub-samples in a given population have equal or different variances.

Levene's test was developed by and named after American statistician and geneticist Howard Levene.