*A statistical test provides a mechanism for making quantitative decisions about a process or processes. The intent is to determine whether there is enough evidence to "reject" a conjecture or hypothesis about the process. The conjecture is called the null hypothesis. Not rejecting may be a good result if we want to continue to act as if we "believe" the null hypothesis is true. Or it may be a disappointing result, possibly indicating we may not yet have enough data to "prove" something by rejecting the null hypothesis.*

*Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter. The methodology employed by the analyst depends on the nature of the data used and the reason for the analysis. Hypothesis testing is used to infer the result of a hypothesis performed on sample data from a larger population.*

*Null hypothesis is a type of hypothesis used in statistics that proposes that no statistical significance exists in a set of given observations. The null hypothesis attempts to show that no variation exists between variables or that a single variable is no different than its mean. It is presumed to be true until statistical evidence nullifies it for an alternative hypothesis.*

*The p-value is the level of marginal significance within a statistical hypothesis test representing the probability of the occurrence of a given event. The p-value is used as an alternative to rejection points to provide the smallest level of significance at which the null hypothesis would be rejected. A smaller p-value means that there is stronger evidence in favor of the alternative hypothesis.*

*How Is P-Value Calculated?*
*P-values are calculated using p-value tables or spreadsheet/statistical software. Because different researchers use different levels of significance when examining a question, a reader may sometimes have difficulty comparing results from two different tests.*

*For example, if two studies of returns from two particular assets were undertaken using two different significance levels, a reader could not compare the probability of returns for the two assets easily.*

*For ease of comparison, researchers often feature the p-value in the hypothesis test and allow the reader to interpret the statistical significance themselves. This is called a p-value approach to hypothesis testing.*

### *P-Value Approach to Hypothesis Testing*

*The p-value approach to hypothesis testing uses the calculated probability to determine whether there is evidence to reject the null hypothesis. The null hypothesis, also known as the conjecture, is the initial claim about a population of statistics.*

*The alternative hypothesis states whether the population parameter differs from the value of the population parameter stated in the conjecture. In practice, the p-value, or critical value, is stated in advance to determine how the required value to reject the null hypothesis.*

### *Type I Error*

*A type I error is the false rejection of the null hypothesis. The probability of a type I error occurring or rejecting the null hypothesis when it is true is equivalent to the critical value used. Conversely, the probability of accepting the null hypothesis when it is true is equivalent to 1 minus the critical value.*

### *Fast Facts*

- *In a statistical hypothesis test, p-value is the level of marginal significance representing a given event's probability of occurrence.*
- *To calculate p-values, you can use p-value tables or spreadsheet/statistical software.*
- *A smaller p-value indicates that there is stronger evidence favoring the alternative hypothesis.*

### *Real World Example of P-Value*

*Assume an investor claims that their investment portfolio's performance is equivalent to that of the Standard & Poor's (S&P) 500 Index. In order to determine this, the investor conducts a two-tailed test. The null hypothesis states that the portfolio's returns are equivalent to the S&P 500's returns over a specified period, while the alternative hypothesis states that the portfolio's returns and the S&P 500's returns are not equivalent. If the investor conducted a one-tailed test, the alternative hypothesis would state that the portfolio's returns are either less than or greater than the S&P 500's returns.*

*One commonly used p-value is 0.05. If the investor concludes that the p-value is less than 0.05, there is strong evidence against the null hypothesis. As a result, the investor would reject the null hypothesis and accept the alternative hypothesis.*

*Conversely, if the p-value is greater than 0.05, that indicates that there is weak evidence against the conjecture, so the investor would fail to reject the null hypothesis. If the investor finds that the p-value is 0.001, there is strong evidence against the null hypothesis, and the portfolio's returns and the S&P 500's returns may not be equivalent.*

# *Z*-test

*A Z-test is any statistical test for which the distribution of the test statistic under the null hypothesis can be approximated by a normal distribution.*

*Because of the central limit theorem, many test statistics are approximately normally distributed for large samples.*

*For each significance level, the Z-test has a single critical value (for example, 1.96 for 5% two tailed) which makes it more convenient than the Student's t-test which has separate critical values for each sample size.*

*Therefore, many statistical tests can be conveniently performed as approximate Z-tests if the sample size is large or the population variance is known.*

*If the population variance is unknown (and therefore has to be estimated from the sample itself) and the sample size is not large ($n < 30$), the Student's t-test may be more appropriate.*

*If T is a statistic that is approximately normally distributed under the null hypothesis, the next step in performing a Z-test is to estimate the expected value $\theta$ of T under the null hypothesis, and then obtain an estimate s of the standard deviation of T.*

*After that the standard score $Z = (T - \theta) / s$ is calculated, from which one-tailed and two-tailed p-values can be calculated as $\Phi(-Z)$ (for upper-tailed tests), $\Phi(Z)$ (for lower-tailed tests) and $2\Phi(-|Z|)$ (for two-tailed tests) where $\Phi$ is the standard normal cumulative distribution function.*

*LOCATION TESTING*

The term "*Z*-test" is often used to refer specifically to the one-sample location test comparing the mean of a set of measurements to a given constant when the sample variance is known. If the observed data $X_1$, ..., $X_n$ are (i) independent, (ii) have a common mean $\mu$, and (iii) have a common variance $\sigma^2$, then the sample average $X$ has mean $\mu$ and variance $\sigma^2 / n$.

The null hypothesis is that the mean value of X is a given number $\mu_0$. We can use $X$ as a test-statistic, rejecting the null hypothesis if $X - \mu_0$ is large.

To calculate the standardized statistic $Z = (X - \mu_0) / s$, we need to either know or have an approximate value for $\sigma^2$, from which we can calculate $s^2 = \sigma^2 / n$. In some applications, $\sigma^2$is known, but this is uncommon.

If the sample size is moderate or large, we can substitute the sample variance for $\sigma^2$, giving a *plug-in* test. The resulting test will not be an exact *Z*-test since the uncertainty in the sample variance is not accounted for—however, it will be a good approximation unless the sample size is small.

A *t*-test can be used to account for the uncertainty in the sample variance when the data are exactly normal.

There is no universal constant at which the sample size is generally considered large enough to justify use of the plug-in test. Typical rules of thumb: the sample size should be 50 observations or more.

For large sample sizes, the *t*-test procedure gives almost identical *p*-values as the *Z*-test procedure.

Other location tests that can be performed as *Z*-tests are the two-sample location test and the paired difference test.

The term "*Z*-test" is often used to refer specifically to the one-sample location test comparing the mean of a set of measurements to a given constant when the sample variance is known. If the observed data $X_1$, ..., $X_n$ are (i) independent, (ii) have a common mean $\mu$, and (iii) have a common variance $\sigma^2$, then the sample average $X$ has mean $\mu$ and variance $\sigma^2 / n$.

The null hypothesis is that the mean value of X is a given number $\mu_0$. We can use $X$ as a test-statistic, rejecting the null hypothesis if $X - \mu_0$ is large.

To calculate the standardized statistic $Z = (X - \mu_0) / s$, we need to either know or have an approximate value for $\sigma^2$, from which we can calculate $s^2 = \sigma^2 / n$. In some applications, $\sigma^2$is known, but this is uncommon.

If the sample size is moderate or large, we can substitute the sample variance for $\sigma^2$, giving a *plug-in* test. The resulting test will not be an exact *Z*-test since the uncertainty in the sample variance is not accounted for—however, it will be a good approximation unless the sample size is small.

A *t*-test can be used to account for the uncertainty in the sample variance when the data are exactly normal.

There is no universal constant at which the sample size is generally considered large enough to justify use of the plug-in test. Typical rules of thumb: the sample size should be 50 observations or more.

For large sample sizes, the *t*-test procedure gives almost identical *p*-values as the *Z*-test procedure.

Other location tests that can be performed as *Z*-tests are the two-sample location test and the paired difference test.