Statistical Hypothesis Terminology

Statistical test: A procedure whose inputs are samples and whose result is a hypothesis.

Statistical hypothesis: A statement about the parameters describing a population (not a sample).

Test statistic – We use the test statistic to see if the data observed are in agreement with the model. The test statistic is typically a simple statistic computed from the data and large ,values (positive, negative, or both) indicate unlikely outcomes, or rare events

Simple hypothesis: Any hypothesis which specifies the population distribution completely.

Composite hypothesis: Any hypothesis which does not specify the population distribution completely.

Null hypothesis (H0): A simple hypothesis associated with a contradiction to a theory one would like to propose.

Alternative hypothesis (Ha): A hypothesis (often composite) associated with a theory one would like to propose.

Region of acceptance: The set of values of the test statistic for which we fail to reject the null hypothesis.

Region of rejection / Critical region: The set of values of the test statistic for which the null hypothesis is rejected.

Power of a test (1 - ß): The test's probability of correctly rejecting H0. The complement of the false negative rate, ß.

Size / Significance level of a test (a): For simple hypotheses, this is the test's probability of incorrectly rejecting H0. The false positive rate.

p-value: The probability, assuming the null hypothesis is true, of observing a result at least as extreme as the test statistic.

One-sided Test: A one-sided test is a statistical hypothesis test in which the values for which we can reject the null hypothesis, H0 are located entirely in one tail of the probability distribution.

Two-Sided Test: A two-sided test is a statistical hypothesis test in which the values for which we can reject the null hypothesis, H0 are located in both tails of the probability distribution.

Type I & II Errors: Analogy – Courtroom trial

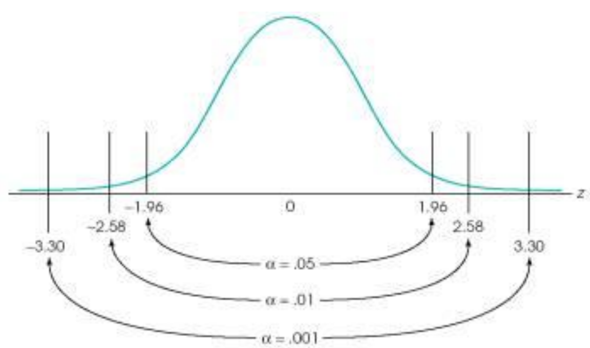
A statistical test procedure is comparable to a criminal trial; a defendant is considered not guilty as long as his or her guilt is not proven. The prosecutor tries to prove the guilt of the defendant. Only when there is enough charging evidence the defendant is convicted.

In the start of the procedure, there are two hypotheses H_0: "the defendant is not guilty", and H_1: "the defendant is guilty". The first one is called *null hypothesis*, and is for the time being accepted. The second one is called *alternative (hypothesis)*. It is the hypothesis one hopes to support.

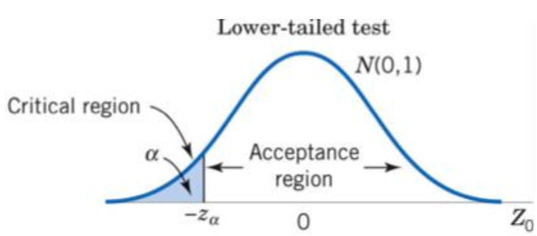
The hypothesis of innocence is only rejected when an error is very unlikely, because one doesn't want to convict an innocent defendant. Such an error is called *error of the first kind* (i.e., the conviction of an innocent person), and the occurrence of this error is controlled to be rare. As a consequence of this asymmetric behaviour, the *error of the second kind* (acquitting a person who committed the crime), is often rather large.

	H_0 is true **Truly not guilty**	**H_1 is true** **Truly guilty**
Accept Null Hypothesis Acquittal	Right decision	Wrong decision Type II Error
Reject Null Hypothesis Conviction	Wrong decision Type I Error	Right decision

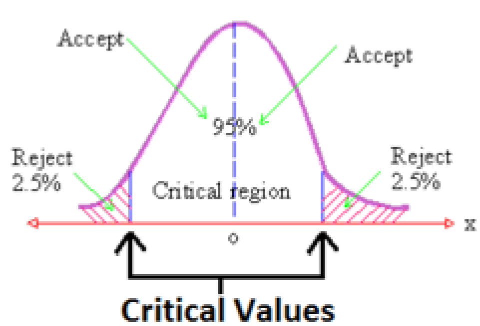
Visualizing progressively increasing significance level:



Visualizing one-sided hypothesis test:



Visualizing two-sided hypothesis test:



Common Test Statistics

Name	Formula	Assumptions or notes
One-sample z-test	$z = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$	(Normal population **or** $n > 30$) **and** σ known. (z is the distance from the mean in relation to the standard deviation of the mean). For non-normal distributions it is possible to calculate a minimum proportion of a population that falls within k standard deviations for any k
Two-sample z-test	$z = \frac{(\overline{x}_1 - \overline{x}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	Normal population **and** independent observations **and** σ_1 and σ_2 are known
One-sample t-test	$t = \frac{\overline{x} - \mu_0}{(s/\sqrt{n})}$, $df = n - 1$	(Normal population **or** $n > 30$) **and** σ unknown
Paired t-test	$t = \frac{\overline{d} - d_0}{(s_d/\sqrt{n})}$, $df = n - 1$	(Normal population of differences **or** $n > 30$) **and** σ unknown or small sample size $n < 30$
Two-sample pooled t-test, equal variances	$t = \frac{(\overline{x}_1 - \overline{x}_2) - d_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$, $df = n_1 + n_2 - 2$	(Normal populations **or** $n_1 + n_2 > 40$) **and** independent observations **and** $\sigma_1 = \sigma_2$ unknown
Two-sample unpooled t-test, unequal variances	$t = \frac{(\overline{x}_1 - \overline{x}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$, $df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$	(Normal populations **or** $n_1 + n_2 > 40$) **and** independent observations **and** $\sigma_1 \neq \sigma_2$ both unknown

One-proportion z-test	$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)}}\sqrt{n}$	$n \cdot p_0 > 10$ **and** $n(1 - p_0) > 10$ **and** it is a SRS (Simple Random Sample).
Two-proportion z-test, pooled for $H_0\colon p_1 = p_2$	$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$ $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$	$n_1 p_1 > 5$ **and** $n_1(1 - p_1) > 5$ **and** $n_2 p_2 > 5$ **and** $n_2(1 - p_2) > 5$ **and** independent observations.
Two-proportion z-test, unpooled for $\lvert d_0 \rvert > 0$	$z = \frac{(\hat{p}_1 - \hat{p}_2) - d_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$	$n_1 p_1 > 5$ **and** $n_1(1 - p_1) > 5$ **and** $n_2 p_2 > 5$ **and** $n_2(1 - p_2) > 5$ **and** independent observations.
Chi-squared test for variance	$\chi^2 = (n - 1)\frac{s^2}{\sigma_0^2}$	Normal population

In general, the subscript 0 indicates a value taken from the null hypothesis, H_0, which should be used as much as possible in constructing its test statistic. ... *Definitions of other symbols:*

- α, the probability of Type I error (rejecting a null hypothesis when it is in fact true)
- n = sample size
- n_1 = sample 1 size
- n_2 = sample 2 size
- $\overline{x}$ = sample mean
- μ_0 = hypothesized population mean
- μ_1 = population 1 mean
- μ_2 = population 2 mean
- σ = population standard deviation
- σ^2 = population variance
- s = sample standard deviation

- s^2 = sample variance
- s_1 = sample 1 standard deviation
- s_2 = sample 2 standard deviation
- t = t statistic
- df = degrees of freedom
- $\overline{d}$ = sample mean of differences
- d_0 = hypothesized population mean difference
- s_d = standard deviation of differences
- χ^2 = Chi-squared statistic

- $\hat{p}$ = *x/n* = sample proportion, unless specified otherwise
- p_0 = hypothesized population proportion
- p_1 = proportion 1
- p_2 = proportion 2
- d_p = hypothesized difference in proportion
- $x_1 = n_1 p_1$
- $x_2 = n_2 p_2$