

Help the Stat Consulting Group by

giving a gift

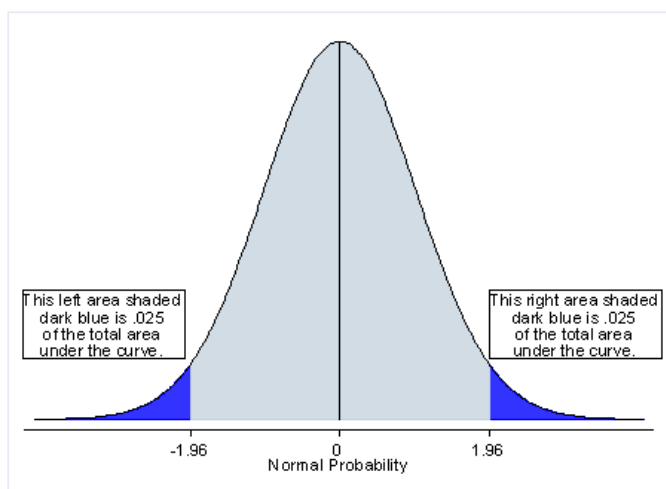
[stat](#) > [mult_pkg](#) > [faq](#) > [general](#) > [tail_tests.htm](#)

FAQ: What are the differences between one-tailed and two-tailed tests?

When you conduct a test of statistical significance, whether it is from a correlation, an ANOVA, a regression or some other kind of test, you are given a p-value somewhere in the output. If your test statistic is symmetrically distributed, you can select one of three alternative hypotheses. Two of these correspond to one-tailed tests and one corresponds to a two-tailed test. However, the p-value presented is (almost always) for a two-tailed test. But how do you choose which test? Is the p-value appropriate for your test? And, if it is not, how can you calculate the correct p-value for your test given the p-value in your output?

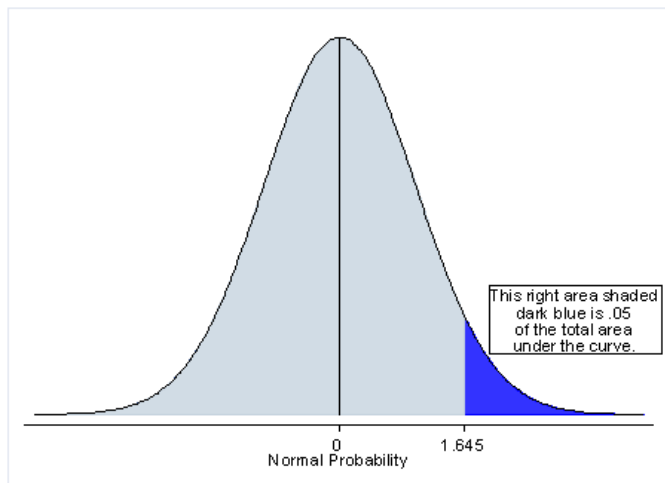
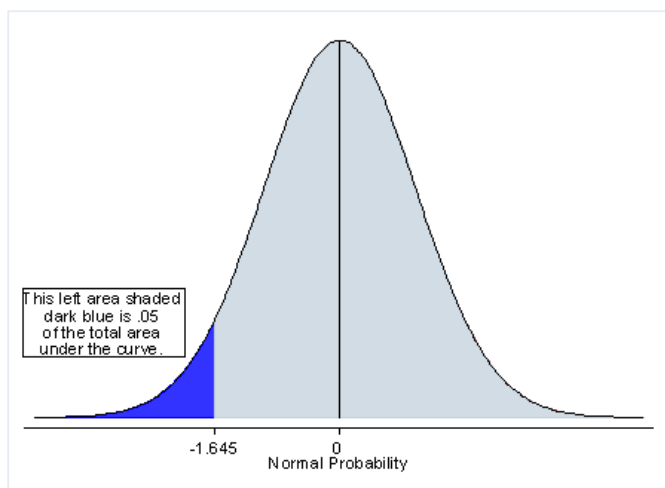
What is a two-tailed test?

First let's start with the meaning of a two-tailed test. If you are using a significance level of 0.05, a two-tailed test allots half of your alpha to testing the statistical significance in one direction and half of your alpha to testing statistical significance in the other direction. This means that .025 is in each tail of the distribution of your test statistic. When using a two-tailed test, regardless of the direction of the relationship you hypothesize, you are testing for the possibility of the relationship in both directions. For example, we may wish to compare the mean of a sample to a given value x using a t-test. Our null hypothesis is that the mean is equal to x . A two-tailed test will test both if the mean is significantly greater than x and if the mean is significantly less than x . The mean is considered significantly different from x if the test statistic is in the top 2.5% or bottom 2.5% of its probability distribution, resulting in a p-value less than 0.05.



What is a one-tailed test?

Next, let's discuss the meaning of a one-tailed test. If you are using a significance level of .05, a one-tailed test allots all of your alpha to testing the statistical significance in the one direction of interest. This means that .05 is in one tail of the distribution of your test statistic. When using a one-tailed test, you are testing for the possibility of the relationship in one direction and completely disregarding the possibility of a relationship in the other direction. Let's return to our example comparing the mean of a sample to a given value x using a t-test. Our null hypothesis is that the mean is equal to x . A one-tailed test will test either if the mean is significantly greater than x or if the mean is significantly less than x , but not both. Then, depending on the chosen tail, the mean is significantly greater than or less than x if the test statistic is in the top 5% of its probability distribution or bottom 5% of its probability distribution, resulting in a p-value less than 0.05. The one-tailed test provides more power to detect an effect in one direction by not testing the effect in the other direction. A discussion of when this is an appropriate option follows.



When is a one-tailed test appropriate?

Because the one-tailed test provides more power to detect an effect, you may be tempted to use a one-tailed test whenever you have a hypothesis about the direction of an effect. Before doing so, consider the consequences of missing an effect in the other direction. Imagine you have developed a new drug that you believe is an improvement over an existing drug. You wish to maximize your ability to detect the improvement, so you opt for a one-tailed test. In doing so, you fail to test for the possibility that the new drug is less effective than the existing drug. The consequences in this example are extreme, but they illustrate a danger of inappropriate use of a one-tailed test.

So when is a one-tailed test appropriate? If you consider the consequences of missing an effect in the untested direction and conclude that they are negligible and in no way irresponsible or unethical, then you can proceed with a one-tailed test. For example, imagine again that you have developed a new drug. It is cheaper than the existing drug and, you believe, no less effective. In testing this drug, you are only interested in testing if it less effective than the existing drug. You do not care if it is significantly more effective. You only wish to show that it is not less effective. In this scenario, a one-tailed test would be appropriate.

When is a one-tailed test NOT appropriate?

Choosing a one-tailed test for the sole purpose of attaining significance is not appropriate. Choosing a one-tailed test after running a two-tailed test that failed to reject the null hypothesis is not appropriate, no matter how "close" to significant the two-tailed test was. Using statistical tests inappropriately can lead to invalid results that are not replicable and highly questionable—a steep price to pay for a significance star in your results table!

Deriving a one-tailed test from two-tailed output

The default among statistical packages performing tests is to report two-tailed p-values. Because the most commonly used test statistic distributions (standard normal, Student's t) are symmetric about zero, most one-tailed p-values can be derived from the two-tailed p-values.

Below, we have the output from a two-sample t-test in Stata. The test is comparing the mean male score to the mean female score. The null hypothesis is that the difference in means is zero. The two-sided alternative is that the difference in means is not zero. There are two one-sided alternatives that one could opt to test instead: that the male score is higher than the female score ($\text{diff} > 0$) or that the female score is higher than the male score ($\text{diff} < 0$). In this instance, Stata presents results for all three alternatives. Under the headings **Ha: diff < 0** and **Ha: diff > 0** are the results for the one-tailed tests. In the middle, under the heading **Ha: diff != 0** (which means that the difference is not equal to 0), are the results for the two-tailed test.

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
male	91	50.12088	1.080274	10.30516	47.97473	52.26703
female	109	54.99083	.7790686	8.133715	53.44658	56.53507
combined	200	52.775	.6702372	9.478586	51.45332	54.09668
diff		-4.869947	1.304191		-7.441835	-2.298059

Degrees of freedom: 198

Ho: mean(male) - mean(female) = diff = 0

Ha: diff < 0

Ha: diff != 0

Ha: diff > 0

t = -3.7341

t = -3.7341

t = -3.7341

P < t = 0.0001

P > |t| = 0.0002

P > t = 0.9999

Note that the test statistic, -3.7341, is the same for all of these tests. The two-tailed p-value is $P > |t|$. This can be rewritten as $P(>3.7341) + P(<-3.7341)$. Because the t-distribution is symmetric about zero, these two probabilities are equal: $P > |t| = 2 * P(<-3.7341)$. Thus, we can see that the two-tailed p-value is twice the one-tailed p-value for the alternative hypothesis that (diff < 0). The other one-tailed alternative hypothesis has a p-value of $P(>-3.7341) = 1 - (P<-3.7341) = 1 - 0.0001 = 0.9999$. So, depending on the direction of the one-tailed hypothesis, its p-value is either 0.5*(two-tailed p-value) or 1-0.5*(two-tailed p-value) if the test statistic symmetrically distributed about zero.

In this example, the two-tailed p-value suggests rejecting the null hypothesis of no difference. Had we opted for the one-tailed test of (diff > 0), we would fail to reject the null because of our choice of tails.

The output below is from a regression analysis in Stata. Unlike the example above, only the two-sided p-values are presented in this output.

Source	SS	df	MS	Number of obs = 200		
Model	7363.62077	2	3681.81039	F(2, 197)	= 46.58	
Residual	15572.5742	197	79.0486001	Prob > F	= 0.0000	
Total	22936.195	199	115.257261	R-squared	= 0.3210	
				Adj R-squared	= 0.3142	
				Root MSE	= 8.8909	
socst	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
science	.2191144	.0820323	2.67	0.008	.0573403	.3808885
math	.4778911	.0866945	5.51	0.000	.3069228	.6488594
_cons	15.88534	3.850786	4.13	0.000	8.291287	23.47939

For each regression coefficient, the tested null hypothesis is that the coefficient is equal to zero. Thus, the one-tailed alternatives are that the coefficient is greater than zero and that the coefficient is less than zero. To get the p-value for the one-tailed test of the variable **science** having a coefficient greater than zero, you would divide the .008 by 2, yielding .004 because the effect is going in the predicted direction. This is $P(>2.67)$. If you had made your prediction in the other direction (the opposite direction of the model effect), the p-value would have been $1 - .004 = .996$. This is $P(<2.67)$. For all three p-values, the test statistic is 2.67.

See also

- [Introduction to Power Analysis](#)

[How to cite this page](#)
[Report an error on this page or leave a comment](#)

The content of this web site should not be construed as an endorsement of any particular web site, book, or software product by the University of California.

IDRE RESEARCH TECHNOLOGY GROUP

High Performance
Computing

Statistical Computing

GIS and Visualization

High Performance Computing

Hoffman2 Cluster

Hoffman2 Account Application

Hoffman2 Usage Statistics

UC Grid Portal

UCLA Grid Portal

Shared Cluster & Storage

About IDRE

GIS

Mapshare

Visualization

3D Modeling

Technology Sandbox

Tech Sandbox Access

Data Centers

Statistical Computing

Classes

Conferences

Reading Materials

IDRE Listserv

IDRE Resources

Social Sciences Data Archive

[ABOUT](#) [CONTACT](#) [NEWS](#) [EVENTS](#) [OUR EXPERTS](#)

© 2016 UC Regents Terms of Use & Privacy Policy