

Research Rundowns > Quantitative Methods > Significance Testing (t-tests)

In this review, we'll look at significance testing, using mostly the *t*-test as a guide. As you read educational research, you'll encounter *t*-test and ANOVA statistics frequently. Part I reviews the basics of significance testing as related to the null hypothesis and *p* values. Part II shows you how to conduct a *t*-test, using an online calculator. Part III deals with interpreting *t*-test results. Part IV is about reporting *t*-test results in both text and table formats and concludes with a guide to interpreting confidence intervals.

What is Statistical Significance?

The terms "significance level" or "level of significance" refer to the likelihood that the random sample you choose (for example, test scores) is not representative of the population. The lower the significance level, the more confident you can be in replicating your results. Significance levels most commonly used in educational research are the .05 and .01 levels. If it helps, think of .05 as another way of saying 95/100 times that you sample from the population, you will get this result. Similarly, .01 suggests that 99/100 times that you sample from the population, you will get the same result. These numbers and signs (more on that later) come from Significance Testing, which begins with the Null Hypothesis.

Part I: The Null Hypothesis

We start by revisiting familiar territory, the [scientific method](#). We'll start with a basic research question: How does variable A affect variable B? The traditional way to test this question involves:

Step 1. Develop a research question.

Step 2. Find previous research to support, refute, or suggest ways of testing the question.

Step 3. Construct a hypothesis by revising your research question:

Hypothesis	Summary	Type
H1: $A = B$	There is no relationship between A and B	Null
H2: $A \neq B$	There is a relationship between A and B. Here, there is a relationship, but we don't know if it is positive or negative.	Alternate
H3: $A < B$	There is a negative relationship between A and B. Here, the $<$ suggests that the less A is involved, the better B.	Alternate
H4: $A > B$	There is a positive relationship between A and B. Here, the $>$ suggests that the more B is involved, the better A.	Alternate

Step 4. Test the null hypothesis. To test the null hypothesis, $A = B$, we use a significance test. The italicized lowercase *p* you often see, followed by $>$ or $<$ sign and a decimal ($p \leq .05$) indicate significance. In most cases, the researcher tests the **null hypothesis, $A = B$** , because it is easier to show there is some sort of effect of A on B, than to have to determine a positive or negative effect prior to conducting the research. This way, you leave yourself room without having the burden of proof on your study from the beginning.

Step 5. Analyze data and draw a conclusion. Testing the null hypothesis leaves two possibilities:

Outcome	Wording	Type
$A = B$	Fail to reject the null. We find no relationship between A and B.	Null
$A =, <, \text{ or } > B$	Reject the null. We find a relationship between A and B.	Alternate

Step 6. Communicate results. See Wording results, below.

Part II: Conducting a *t*-test (for Independent Means)

So how do we test a null hypothesis? One way is with a *t*-test. A *t*-test asks the question, “Is the difference between the means of two samples different (significant) enough to say that some other characteristic (teaching method, teacher, gender, etc.) could have caused it?”

To conduct a *t*-test using an online calculator, complete the following steps:

Step 1. Compose the Research Question.

Step 2. Compose a Null and an Alternative Hypothesis.

Step 3. Obtain two random samples of at least 30, preferably 50, from each group.

Step 4. Conduct a *t*-test:

- Go to <http://www.graphpad.com/quickcalcs/ttest1.cfm>
- For #1, check “Enter mean, SD and N.”
- For #2, label your groups and enter data. You will need to have mean and SD. N is group size.
- For #3, check “Unpaired *t* test.”
- For #4, click “Calculate now.”

Step 5. Interpret the results (see below).

Step 6. Report results in text or table format (see below).

- Get *p* from “P value and statistical significance:” Note that this is the actual value.
- Get the confidence interval from “Confidence interval:”
- Get the *t* and *df* values from “Intermediate values used in calculations:”
- Get *Mean*, and *SD* from “Review your data.”

Part III. Interpreting a *t*-test (Understanding the Numbers)

<i>t</i>	tells you a <i>t</i> -test was used.
(98)	tells you the degrees of freedom (the sample - # of tests performed).
3.09	is the “ <i>t</i> statistic” – the result of the calculation.
$p \leq .05$	is the probability of getting the observed score from the sample groups. <u>This the most important part of this output to you.</u>
If this sign	It means all these things
$p \geq .05$	likely to be a result of chance (same as saying $A = B$)
	difference is <u>not</u> significant
	null is correct
	“fail to reject the null”
	There is no relationship between A and B.
If this sign	It means all these things
$p \leq .05$	not likely to be a result of chance (same as saying $A \neq B$)
	difference is <u>significant</u>
	null is incorrect
	“reject the null”
	There is a relationship between A and B.

Note: We acknowledge that the average scores are different. With a *t*-test we are deciding if that difference is significant (is it due to sampling error or something else?).

Understanding the Confidence Interval (CI)

The [Confidence Interval](#) (CI) of a mean is a region within which a score (like mean test score) may be said to fall with a certain amount of “confidence.” The CI uses sample size and standard deviation to generate a lower and upper number that you can be 95% sure will include any sample you take from a set of data.

Consider Georgia’s AYP measure, the [CRCT](#). For a science CRCT score, we take several samples and compare the different means. After a few calculations, we could determine something like. . .the average difference (mean) between samples is -7.5, with a 95% CI of -22.08 to 6.72. In other words, among all students’ science CRCT scores, 95 out of 100 times we take group samples for comparison (for example by year, or gender, etc.), one of the groups, on average will be 7.5 points lower than the other group. We can be fairly certain that the difference in scores will be between -22.08 and 6.72 points.

Part IV. Wording Results

Wording Results in Text

In text, the basic format is to report: population (*N*), mean (*M*) and standard deviation (*SD*) for both samples, *t* value, degrees freedom (*df*), significance (*p*), and confidence interval (*CI*.₉₅)^{*} .

Example 1: $p \leq .05$, or Significant Results

Among 7th graders in Lowndes County Schools taking the CRCT reading exam ($N = 336$), there was a statistically significant difference between the two teaching teams, team 1 ($M = 818.92$, $SD = 16.11$) and team 2 ($M = 828.28$, $SD = 14.09$), $t(98) = 3.09$, $p \leq .05$, $CI_{.95} -15.37, -3.35$. Therefore, we reject the null hypothesis that there is no difference in reading scores between teaching teams 1 and 2.

Example 2: $p \geq .05$, or Not Significant Results

Among 7th graders in Lowndes County Schools taking the CRCT science exam ($N = 336$), there was no statistically significant difference between female students ($M = 834.00$, $SD = 32.81$) and male students (841.08 , $SD = 28.76$), $t(98) = 1.15$ $p \geq .05$, $CI_{.95} -19.32, 5.16$. Therefore, we fail to reject the null hypothesis that there is no difference in science scores between females and males.

Wording Results in APA Table Format

Table 1. Comparison of CRCT 7th Grade Science Scores by Gender

Gender	<i>n</i>	Mean	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>	95% Confidence Interval
Female	50	834.00	32.81	–	–	–	–
Male	50	841.08	28.76	–	–	–	–
Total	100	837.54	30.90	1.14	98	.2540	-19.32 – 5.16

Note: On the Web site, this appears blocked and should not be. See the .pdf for the correct format.