An Overview of the Significance of the t-test

DR. CHHAYA JANGAM
Assistant Professor,
Government B.Ed. College, Kachhal, Dist. Surat
Gujarat (India)

Abstract:
In this review article, 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.

Keywords: *Educational research, Significance testing, Statistics, t-test*

1. Introduction

In 1908 William Sealy Gosset, an Englishman publishing under the pseudonym Student developed the *t*-test and *t* distribution. The *t* distribution is a family of curves in which the number of degrees of freedom (the number of independent observations in the sample minus one) specifies a particular curve. As the sample size (and thus the degrees of freedom) increases, the *t* distribution approaches the bell shape of the standard normal distribution. In practice, for tests involving the mean of a sample of size greater than 30, the normal distribution is usually applied.

It is usual first to formulate a null hypothesis, which states that there is no effective difference between the observed sample mean and the hypothesized or stated population mean - i.e., that any measured difference is due only to chance. In an agricultural study, for example, the null hypothesis could be that an application of fertilizer has had no effect on crop yield, and an experiment would be performed to test whether it has increased the harvest. In general, a *t*-test may be either two-sided (also termed two-tailed), stating simply that the means are not equivalent, or one-sided, specifying whether the observed mean is larger or smaller than the hypothesized mean. The test statistic *t* is then calculated. If the observed *t*-statistic is more extreme than the critical value determined by the appropriate reference distribution, the null hypothesis is rejected. The appropriate reference distribution for the *t*-statistic is the *t* distribution. The critical value depends on the significance level of the test (the probability of erroneously rejecting the null hypothesis).

For example, suppose a researcher wishes to test the hypothesis that a sample of size *n* = 25 with mean *x* = 79 and standard deviation *s* = 10 was drawn at random from a population with mean μ = 75 and unknown standard deviation. Using the formula for the *t*-statistic,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

the calculated *t* equals 2. For a two-sided test at a common level of significance α = 0.05, the critical values from the *t* distribution on 24 degrees of freedom are −2.064 and 2.064. The calculated *t* does not exceed these values, hence the null hypothesis cannot be rejected with 95 percent confidence. (The confidence level is 1 − α.)

A second application of the *t* distribution tests the hypothesis that two independent random samples have the same mean. The *t* distribution can also be used to construct confidence intervals for the true mean of a population (the first application) or for the difference between two sample means (the second application).

2. 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.

3. 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:

3.1 Step 1. Develop a Research Question

The researcher develops some research question here.

3.2 Step 2. Find Previous Research

To support, refute, or suggest ways of testing the question.

3.3 Step 3. Construct a Hypothesis

Here the researcher constructs a hypothesis by revising the research question as per table 1

Table 1

Hypothesis	Summary	Type
H_1: A = B	There is no relationship between A and B	Null
H_2: A ≠ 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
H_3: A < B	There is a negative relationship between A and B. Here, the < suggests that the less A is involved, the better B.	Alternate
H_4: A > B	There is a positive relationship between A and B. Here, the > suggests that the more B is involved, the better A.	Alternate

3.4 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 is it 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.

3.5 *Step 5.* Analyze data and draw a conclusion

Testing the null hypothesis leaves two possibilities mentioned as per table 2.

Table 2

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

4. 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:

4.1 *Step 1.* Compose the Research Question

4.2 *Step 2.* Compose a Null and an Alternative Hypothesis

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

4.5 *Step 4.* Conduct a *t*-test

Go to http://www.graphpad.com/quickcalcs/ttest1.cfm and follow the following steps one by one.

* check "Enter mean, SD and N."
* label your groups and enter data. You will need to have mean and SD. N is group size.
* check "Unpaired *t* test."
* 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."

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

t	tells you a t-test was used.
(98)	tells you the degrees of freedom (the sample - # of tests performed).
3.09	is the "t statistic" – the result of the calculation.
$p \leq .05$	is the probability of getting the observed score from the sample groups. **This the most important part of this output to you.**
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 not 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 ≠ B)
difference is significant	
null is incorrect	
"reject the null"	
There is a relationship between A and B.	

6. 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.

7. Part IV Wording Results
7.1 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 ≥ .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 ≥ .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.

7.2 Wording Results in APA Table Format

Table 3 Comparison of CRCT 7th Grade Science Scores by Gender

Gender	N	Mean	SD	t	df	p	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

References
1. Fadem, Barbara (2008). High-Yield Behavioral Science (High-Yield Series). Hagerstwon, MD: Lippincott Williams & Wilkins. ISBN 0-7817-8258-9.
2. Fisher Box, Joan (1987). "Guinness, Gosset, Fisher, and Small Samples". Statistical Science 2 (1): 45–52. Doi: 10.1214/ss/1177013437. JSTOR 2245613.
3. John, A. Rice (2006). Mathematical Statistics and Data Analysis, Third Edition, Duxbury Advanced.
4. Markowski, Carol A; Markowski, Edward P. (1990). "Conditions for the Effectiveness of a Preliminary Test of Variance". The American Statistician 44 (4): 322–326. Doi:10.2307/2684360. JSTOR 2684360.
5. Raju, TN (2005). "William Sealy Gosset and William A. Silverman: two "students" of science". Pediatrics 116 (3): 732–5. Doi:10.1542/peds.2005-1134. PMID 16140715.
6. Zimmerman, Donald W. (1997). "A Note on Interpretation of the Paired-Samples t Test". Journal of Educational and Behavioral Statistics 22 (3): 349–360. JSTOR 1165289.
7. Chaudhary, Asha (2013). Theories of Teaching. International Journal for Research in Education (IJRE). Vol. 2, Issue 3, March 2013. ISSN: 2320-091X (Online) Published By RET Academy for International Journals of Multidisciplinary Research (RAIJMR) at www.raijmr.com