2015

Guide for the Student's *t*-Test

Paul Strode, PhD
Fairview High School
Boulder, Colorado

Ann Brokaw
Rocky River High School
Rocky River, Ohio

Last Updated: September 2015

Statistical Symbols and Equations

Listed below are the universal statistical symbols and equations used in this guide. The calculations can all be done using scientific calculators or the formula function in spreadsheet programs.

N: Total number of individuals in a population (i.e., the total number of butterflies of a particular species)

n: Total number of individuals in a sample of a population (i.e., the number of butterflies in a net)

df: The number of measurements in a sample that are free to vary once the sample mean has been calculated; in a single sample, df = $n - 1$

x_i: A single measurement

i: The i^{th} observation in a sample

Σ: Summation

$\bar{x}$: Sample mean $\bar{x} = \frac{\sum x_i}{n}$

s^2: Sample variance $s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$

s: Sample standard deviation $s = \sqrt{s^2}$

SE: Sample standard error, or standard error of the mean (SEM) $\mathrm{SE} = \frac{s}{\sqrt{n}}$

95% CI: 95% confidence interval 95% CI = $\frac{1.96s}{\sqrt{n}}$

t-Test: $t_{\text{obs}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

Inferential Statistics Used in Biology

Inferential statistics tests statistical hypotheses, which are different from experimental hypotheses. To understand what this means, assume that you do an experiment to test whether “nitrogen promotes plant growth.” This is an experimental explanatory hypothesis because it tells you something about the biology of plant growth. To test this hypothesis, you grow 10 bean plants in dirt with added nitrogen and 10 bean plants in dirt without added nitrogen. You find out that the means of these two samples are

13.2 centimeters and 11.9 centimeters, respectively. Does this result indicate that there is a difference between the two populations and that nitrogen might promote plant growth? Or is the difference in the two means merely due to chance because of the 10 measurements collected? A statistical test is required to discriminate between these possibilities.

Statistical tests evaluate statistical hypotheses. The statistical null hypothesis (symbolized by H_0 and pronounced H-naught) is a statement that you want to test. In this case, if you grow 10 plants with nitrogen and 10 without the null hypothesis is that there will be no difference in the mean heights of the two groups and any observed difference between the two groups would have occurred purely by chance. The alternative hypothesis to H_0 is symbolized by H_1 and usually simply states that there is a real difference between the populations.

The statistical null and alternative hypotheses are statements about the data that should follow from the experimental hypothesis.

Significance Testing: The α (Alpha) Level

Before you perform a statistical test on the plant growth data, you should determine an acceptable significance level of the null statistical hypothesis. That is, ask, "When do I think my results and thus my test statistic are so unusual that I no longer think the differences observed in my data are due simply to chance?" This significance level is also known as "alpha" and is symbolized by α.

The significance level is the **probability** of getting a test statistic rare enough that you are comfortable rejecting the null hypothesis (H_0). (See the "Probability" section of Part 3 for further discussion of probability.) The widely accepted significance level in biology is 0.05. If the p value is less than 0.05 you reject the null hypothesis, if p is greater than or equal to 0.05 you don't reject the null hypothesis.

Comparing Averages: The Student's *t*-Test for Independent Samples

The Student's *t*-Test is used to compare the means of two samples to determine whether they are statistically different. For example, you calculated the sample means of survivor and nonsurvivor finches from Table 1 and you get different numbers. What is the probability of getting this difference in means, if the population means are really the same?

The *t*-Test assesses the probability of getting the observed result (i.e., the values you calculated for the means shown in Figure 1) if the null statistical hypothesis (H_0) is true. Typically the null statistical hypothesis in a t-test is that the mean of the population from which sample 1 was taken (i.e. the mean beak size of survivors) is equal to the mean of the population from which sample 2 was taken (i.e. the mean beak size of the non survivors), or $u_1 = u_2$. Rejecting H_0 supports the alternative hypothesis, H_1, that the population means are significantly different ($u_1 \neq u_2$). In the finch example, the *t*-Test determines whether any observed differences between the means of the two groups of finches (9.67 mm versus 9.11 mm) are statistically significant or have likely occurred simply by chance.

A *t*-Test calculates a single statistic, t, or t_{obs}, which is compared to a critical *t*-statistic (t_{crit}):

$$t_{\text{obs}} = \frac{|\bar{x}_1 - \bar{x}_2|}{SE}$$

To calculate the standard error (SE) specific for the t-Test, we calculate the sample means and the variance (s^2) for the two samples being compared—the sample size (n) for each sample must be known:

$$\mathrm{SE} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Thus, the complete equation for the t-Test is:

$$t_{\mathrm{obs}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Calculation Steps

1. Calculate the mean of each sample population and subtract one from the other. Take the absolute value of this difference.
2. Calculate the standard error, SE. To compute it, calculate the variance of each sample (s^2), and divide it by the number of measured values in that sample (n, the sample size). Add these two values and then take the square root.
3. Divide the difference between the means by the standard error to get a value for t. Compare the calculated value to the appropriate critical t-value in **Table 1**. **Table 1** shows t_{crit} for different degrees of freedom for a significance value of 0.05. **The degrees of freedom is calculated by adding the number of data points in the two groups combined, minus 2**. Note that you do not have to have the same number of data points in each group.
4. If the calculated t-value is greater than the appropriate critical t-value, this indicates that the means of the two samples are significantly different at the probability value listed (in this case, 0.05). If the calculated t is smaller, then you cannot reject the null hypothesis and infer that there is no difference between the population means.

Table 1. Critical *t*-Values for a Significance Level α = 0.05

Degrees of Freedom (df)	t_{crit} (α = 0.05)
1	12.71
2	4.30
3	3.18
4	2.78
5	2.57
6	2.45
7	2.36
8	2.31
9	2.26
10	2.23
11	2.20
12	2.18
13	2.16
14	2.14
15	2.13
16	2.12
17	2.11
18	2.10
19	2.09
20	2.09
21	2.08
22	2.07
23	2.07
24	2.06
25	2.06
26	2.06
27	2.05
28	2.05
29	2.04
30	2.04
40	2.02
60	2.00
120	1.98
Infinity	1.96

Note: There are two basic versions of the t-Test. The version presented here assumes that each sample was taken from a different population and so the samples are therefore independent of one another. For example, the survivor and nonsurvivor finches are different individuals, independent of one another, and therefore considered **unpaired**. If we were comparing the lengths of right and left wings on all the finches, the samples would be classified as **paired**. Paired samples require a different version of the *t*-Test known as a paired *t*-Test, a version to which many statistical programs default. The paired *t*-Test is not discussed in this document.

Application in Biology

After a small population of crayfish was accidentally released into a shallow pond, biologists noticed that the crayfish had consumed nearly all of the underwater plant population; aquatic invertebrates, such as the water flea (*Daphnia* sp.), had also declined. The biologists knew that the main predator of *Daphnia* is the goldfish and they hypothesized that the underwater plants protected the *Daphnia* from the goldfish by providing hiding places. The *Daphnia* lost their protection as the underwater plants disappeared. The biologists designed an experiment to test their hypothesis. They placed goldfish and *Daphnia* together in a tank with underwater plants, and an equal number of goldfish and *Daphnia* in another tank without underwater plants. They then counted the number of *Daphnia* eaten by the goldfish in 30 minutes. They replicated this experiment in nine additional pairs of tanks (i.e., sample size = 10, or n = 10, per group). The results of their experiment and their calculations of experimental error (variance, s^2) are in **Table 2.**

Experimental hypothesis: the underwater plants protect Daphnia from goldfish by providing hiding places
Statistical null hypothesis: there is no real difference between Daphnia that live in the presence or absence of plants, and any difference between the two groups occurs simply by chance
Expected result: an equal number of Daphnia in the two tanks with and without plants
Statistical alternative hypothesis: there is a real difference between the Daphnia that live in the presence or absence of plants

Table 2. Number of *Daphnia* Eaten by Goldfish in 30 Minutes in Tanks with or without Underwater Plants

Tanks	Plants (sample$_1$)	No Plants (sample$_2$)	Plants $(x_i - \bar{x}_1)^2$	No Plants $(x_i - \bar{x}_2)^2$
1 and 2	13	14	$(9.6 - 13)^2 = 11.56$	$(14.4 - 14)^2 = 0.16$
3 and 4	9	12	$(9.6 - 9)^2 = 0.36$	$(14.4 - 12)^2 = 5.876$
5 and 6	10	15	$(9.6 - 10)^2 = 0.16$	$(14.4 - 15)^2 = 0.436$
7 and 8	10	14	$(9.6 - 10)^2 = 0.16$	$(14.4 - 14)^2 = 0.16$
9 and 10	7	17	$(9.6 - 7)^2 = 6.76$	$(14.4 - 16)^2 = 6.76$
11 and 12	5	10	$(9.6 - 5)^2 = 21.16$	$(14.4 - 10)^2 = 19.37$
13 and 14	10	15	$(9.6 - 10)^2 = 0.16$	$(14.4 - 15)^2 = 0.36$
15 and 16	14	15	$(9.6 - 14)^2 = 19.336$	$(14.4 - 15)^2 = 0.36$
17 and 18	9	18	$(9.6 - 9)^2 = 0.36$	$(14.4 - 18)^2 = 12.96$
19 and 20	9	14	$(9.6 - 9)^2 = 0.36$	$(14.4 - 14)^2 = 0.16$
			$\sum (x_i - \bar{x}_1)^2 = 60.4$	$\sum (x_i - \bar{x}_2)^2 = 46.4$
Mean ($\bar{x}$)	$\bar{x}_1 = 9.6$	$\bar{x}_2 = 14.4$	$\frac{\sum (x_i - \bar{x}_1)^2}{n-1} = \frac{60.4}{9}$	$\frac{\sum (x_i - \bar{x}_1)^2}{n-1} = \frac{46.4}{9}$
		Variance (s^2)	$s_1^2 = 6.71$	$s_2^2 = 5.16$

To determine whether the difference between the two groups was significant, the biologists calculated a *t*-Test statistic, as shown below:

$$\text{SE} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{6.71}{10} + \frac{5.16}{10}} = 1.089$$

The mean difference (absolute value) = $|\bar{x}_1 - \bar{x}_2|$ = |9.6 − 14.4| = 4.8

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\text{SE}} = \frac{4.8}{1.089} = 4.41$$

There are (10 + 10 − 2) = 18 degrees of freedom, so the critical value for $p = 0.05$ is 2.10 from **Table 1**. The calculated *t*-value of 4.41 is greater than 2.10, so the students can reject the null hypothesis that the differences in the number of *Daphnia* eaten in the presence or absence of underwater plants were accidental. So what can they conclude? It is possible that the goldfish ate significantly more *Daphnia* in the absence of underwater plants than in the presence of the plants.

Practice!

A teacher noticed that her students' test scores had been improving over the course of the second grading period. She **hypothesized** that her new approach of having students answer warm up questions at the beginning of each class improves their test scores. She decided to **test this hypothesis** during a curriculum unit by having a sample of eleven of her students do warm up questions at the beginning of class while another sample of eleven students were allowed to sit and chat. She **predicted** that the students doing the warm up questions would have significantly higher test scores on the unit exam than students not doing the warm ups. Her exam results are in Table 3.

Table 3. Hypothetical results of a test of the Warm up Hypothesis, plus a worksheet for calculating Descriptive Statistics.

Sample #	Warmups x_1	No Warmups x_2	Squared Difference $(x_i - \bar{x}_1)^2$	Squared Difference $(x_i - \bar{x}_2)^2$
1	89	74		
2	89	78		
3	90	84		
4	90	88		
5	90	88		
6	92	91		
7	92	91		
8	94	91		
9	96	92		
10	97	92		
11	100	94		
Mean ($\bar{x}$)	$\bar{x}_1$ =	$\bar{x}_2$ =		
Sum of Squares (SS)= $\sum (x_i - \bar{x}_1)^2$			SS_1 =	SS_2 =
Variance (s^2) = $\frac{\sum(x_i - \bar{x})^2}{(n-1)}$			s_1^2 =	s_2^2 =
Standard deviation $s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{(n-1)}}$			s_1 =	s_2 =
95% CI = $\frac{2s}{\sqrt{n}}$			95% CI =	95% CI =

Can the teacher conclude that doing warm ups has a statistically significant effect on exam scores? She can do a *t*-Test on the data to determine the probability that the means of the two groups (samples) are actually not really different (called the statistical hypothesis), and that the observed difference occurred by chance, or accident. Try doing the *t*-Test on these data.