Learning Two Sample t-Tests: A Step-by-Step Guide Using the TI-84 Calculator

Authored by

Mohammed looti

November 8, 2025

The **Two Sample t-test** is an indispensable tool in inferential statistics, specifically designed to evaluate whether a statistically significant difference exists between the population means of two distinct, independent groups. This test is fundamentally important across various scientific disciplines, serving as the backbone for analyzing controlled experiments--such as comparing a treatment group against a control--or evaluating inherent variations between naturally occurring datasets. By systematically comparing the sample means and variances derived from the two groups, the t-test provides a robust framework for making reliable inferences about the larger populations from which those samples were drawn.

Navigating complex statistical procedures can often be challenging, but modern graphing calculators simplify the heavy lifting. This comprehensive tutorial is meticulously crafted to provide a step-by-step guide on performing a **Two Sample t-test** using the highly efficient **TI-84 calculator**. We will focus specifically on how to leverage this functionality when working exclusively with summarized data (means, standard deviations, and sample sizes), ensuring that you can execute accurate and efficient hypothesis testing with ease.

The Statistical Rationale of the Two-Sample t-Test

When collecting data from two separate, non-related samples, researchers must employ a method capable of discerning whether the observed numerical differences are genuinely meaningful or merely the product of random sampling fluctuation. The **Two Sample t-test** offers this necessary statistical rigor. It begins by establishing the **null hypothesis** (H0), which serves as the default assumption that the true population means ($\mu 1$ and $\mu 2$) are identical. The subsequent analysis calculates a test statistic (the t-value) that quantifies the evidence against this initial assumption, allowing the researcher to determine if H0 can be convincingly rejected in favor of an alternative hypothesis (Ha).

Successfully applying the Two Sample t-test requires adherence to several critical assumptions regarding the nature of the data. Primarily, the two samples must be independent, meaning the selection or performance of one group does not influence the other. Furthermore, the data distribution within each population should be approximately normal. A key decision point within the test concerns the population variances: are they assumed to be equal (leading to a pooled variance t-test) or unequal (resulting in the more conservative unpooled, or Welch's, test)? The **TI-84 calculator** dramatically streamlines this process, automating the calculation of the **p-value** and degrees of freedom based on the user's explicit input regarding variance pooling.

Understanding the interplay between these assumptions and the calculation method is crucial for accurate interpretation. Choosing the correct pooling option--Yes or No--directly impacts the calculation of the standard error and the degrees of freedom. The Welch's t-test (No pooling) is often preferred in real-world scenarios unless there is robust prior evidence supporting the

assumption of equal variances, as it provides a more robust inference when sample sizes or variances differ significantly.

Illustrative Example: Evaluating Fuel Additive Efficacy

To demonstrate the practical application of the Two Sample t-test, let us analyze a realistic research scenario. Imagine automotive researchers conducting a study to investigate the effectiveness of a newly developed fuel additive. Their primary research question is whether this treatment causes a statistically significant change in a vehicle's average miles per gallon (mpg). To maintain experimental control, they deploy twenty-four identical vehicles, randomly allocating them into two groups of twelve cars each, ensuring independence between the samples.

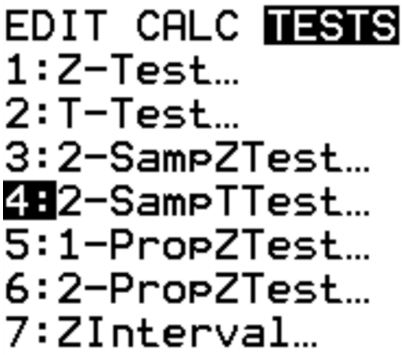
The first group, designated as the **control group** (Group 1), ran on standard fuel. The resulting summary statistics for this group were: a sample size (n1) of 12, an average mpg (x?1) of 21.0, and a **standard deviation** (s1) of 2.73 mpg. The second group, the **treatment group** (Group 2), utilized the new fuel additive. This group also consisted of 12 vehicles (n2 = 12), achieving a higher average mpg (x?2) of 22.75, albeit with a slightly increased **standard deviation** (s2) of 3.25 mpg. Notice that we are working solely with these summarized statistics, eliminating the need to input lengthy lists of raw data points.

Our objective is to use these figures to conduct the **Two Sample t-test**. We seek to formally determine if the observed 1.75 mpg difference (22.75 - 21.0) is large enough to warrant the rejection of the **null hypothesis** (H0: μ1 = μ2). Since the researchers are investigating any "change" (a difference in either direction), we will implement a two-tailed test. This means our alternative hypothesis (Ha) will be μ1 ? μ2, suggesting that the true population means are simply unequal.

Accessing the Two-Sample T-Test Function on the TI-84

The efficiency of the **TI-84 calculator** stems from its specialized statistical functions, which automate the complex calculations based on the t-distribution curve. To initiate the Two Sample t-test procedure, the user must first navigate to the dedicated statistical testing menu. Follow these steps precisely:

Press the Stat button, then navigate right using the arrow keys to the **TESTS** submenu. Scroll down the list of statistical tests until you locate option 4, clearly labeled **2-SampTTest**, and confirm your selection by pressing ENTER. This sequence opens the Two-Sample t-test dialogue screen, prompting the user to input all necessary parameters for the statistical model, as illustrated in the calculator display below.

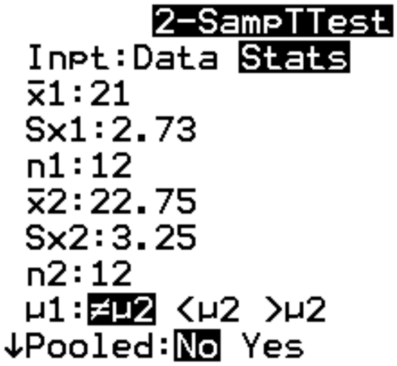


Inputting Summary Statistics and Defining Hypotheses

The input screen requires meticulous entry of the summary statistics derived from our fuel efficiency case study. Accuracy at this stage is paramount, as incorrect inputs will invalidate the resulting t-statistic and **p-value**. Follow the prompts sequentially, using the arrow keys to navigate between fields:

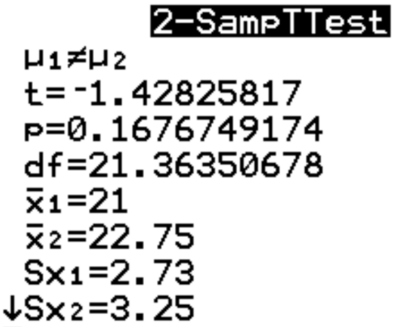
Inpt: This setting controls the input method. Since we possess summarized data rather than a list of raw data points, highlight **Stats** (not Data) and press ENTER.
x1: Enter the sample mean for the first group (Control). Input **21** and press ENTER.
Sx1: Enter the sample **standard deviation** for the control group. Input **2.73** and press ENTER.
n1: Enter the sample size for the control group. Input **12** and press ENTER.
x2: Enter the sample mean for the second group (Treatment). Input **22.75** and press ENTER.
Sx2: Enter the sample **standard deviation** for the treatment group. Input **3.25** and press ENTER.
n2: Enter the sample size for the second group. Input **12** and press ENTER.
μ1: This crucial step defines the alternative hypothesis (Ha). Since we are performing a two-tailed test, looking for any difference, highlight **? μ2** and press ENTER. Remember: <μ2 is for a left-tailed test (μ1 is less than μ2), and >μ2 is for a right-tailed test (μ1 is greater than μ2).
Pooled: This option asks if the population variances are assumed to be equal. Given the lack of strong prior evidence, the generally recommended and more conservative approach is to select **No**, performing the Welch's t-test. Highlight **No** and press ENTER.

Once you have meticulously verified that all fields are correctly populated according to the case study data, scroll down to highlight the **Calculate** option and press ENTER. The calculator will then process the inputs and display the statistical results screen almost instantaneously.



Interpreting the Statistical Output and Key Metrics

The final output screen generated by the **TI-84 calculator** provides all the necessary metrics required to evaluate the research hypothesis formally. Understanding each component of this summary is essential for drawing a valid conclusion.



Below is a detailed breakdown of the critical values displayed:

μ1? μ2: This merely confirms the structure of the test run, verifying that the alternative hypothesis (Ha) was correctly set for a two-tailed difference.
t = -1.42825817: This is the calculated **t test statistic**. It quantifies the difference between the two sample means relative to the standard error of that difference. The negative sign indicates that the mean of the second group (Treatment: 22.75) was numerically larger than the mean of the first group (Control: 21.0).
p = 0.1676749174: This is the highly important **p-value**. It represents the probability of observing a test statistic (t) as extreme as -1.428 (or more extreme) if, in fact, the **null hypothesis** (μ1 = μ2) were true. A small p-value suggests the observed data is unlikely under the null hypothesis.
df = 21.36350678: This is the **degrees of freedom**. Because we chose 'No' for pooling (the Welch's correction), the degrees of freedom are calculated using a complex formula, resulting in a decimal value rather than a simple integer (n1 + n2 - 2).

The remaining metrics (x1, x2, Sx1, Sx2, n1, n2) serve as crucial confirmations, displaying the exact sample means, **standard deviations**, and sample sizes that were fed into the calculation engine.

Drawing Statistical Inferences and Formulating the Conclusion

The final and most crucial step in the hypothesis testing framework is utilizing the calculated metrics, particularly the **p-value**, to reach a formal statistical conclusion. This involves comparing the p-value against the pre-established level of significance, denoted by alpha (α). In social science and applied research, the conventional significance level is set at $\alpha = 0.05$. This threshold means researchers are willing to tolerate a 5% chance of committing a Type I error--incorrectly rejecting a true **null hypothesis**.

The decision rule is straightforward: If the p-value is less than or equal to α ($p \leq \alpha$), we reject the null hypothesis. If the p-value is greater than α ($p > \alpha$), we fail to reject the null hypothesis. In our fuel treatment example, the calculated p-value is approximately 0.1677. Comparing this to our chosen significance level: 0.1677 is greater than 0.05. Consequently, we must **fail to reject the null hypothesis**.

This statistical failure to reject H0 indicates that the 1.75 mpg difference observed between the control group (21.0 mpg) and the treatment group (22.75 mpg) is not statistically large enough to be considered a genuine, population-level effect caused by the fuel additive. Rather, the difference is likely attributable to normal random sampling variability. The final conclusion is that, based on the statistical evidence generated by the **TI-84 calculator**, there is insufficient support to conclude that the new fuel treatment results in a significant change in the average miles per gallon for the vehicle population under study.