

# The Satterthwaite Approximation: Definition & Example

Authored by  
**Mohammed loot**

November 6, 2025

## RECOMMENDED CITATION

Mohammed loot (2025). *The Satterthwaite Approximation: Definition & Example*.  
PSYCHOLOGICAL STATISTICS. Retrieved from  
<https://statistics.arabpsychology.com/?p=11492>

## Introduction to the Satterthwaite Approximation

The **Satterthwaite approximation** is a critical mathematical tool in [inferential statistics](#), specifically designed to calculate the "effective [degrees of freedom](#)" (df) when comparing two independent samples. This formula addresses a fundamental challenge in hypothesis testing, ensuring that statistical inferences remain robust even when underlying population assumptions are violated.

It is most commonly employed as a core component of the [Welch's t-test](#). Unlike the standard pooled [two-sample t-test](#), the Welch approach does not assume that the populations from which the samples are drawn have equal variances, a condition known as heterogeneity of variance.

When variances are unequal, the standard calculation for degrees of freedom becomes unreliable. The [Satterthwaite approximation](#) provides a pragmatic, non-integer estimate for the degrees of freedom, which is essential for accurately determining the critical value and corresponding p-value of the test statistic.

## The Challenge of Unequal Variances in T-Tests

The traditional [two-sample t-test](#) relies on the assumption of homoscedasticity--that the variances of the two populations being compared are statistically equal. If this assumption holds true, the variances can be "pooled" to calculate a single standard error, and the degrees of freedom are simply calculated as the total sample size minus two ( $n_1 + n_2 - 2$ ).

However, real-world data frequently exhibits unequal variances (heteroscedasticity). Ignoring this inequality can lead to significantly inaccurate results, particularly if the sample sizes ( $n_1$  and  $n_2$ ) are also different. If the larger variance is associated with the smaller sample size, the Type I error rate (false positives) can become inflated.

This is why the [Welch's t-test](#) is preferred in such scenarios. By using the unpooled standard error, it correctly accounts for the unequal variances. But this necessary correction means the resulting test statistic no longer precisely follows the standard t-distribution with  $n_1 + n_2 - 2$  degrees of freedom.

The Satterthwaite approximation provides a method to calculate the effective [degrees of freedom](#). This effective value is generally lower than the pooled value, reflecting the increased statistical uncertainty introduced by estimating separate variances.

## The Satterthwaite Formula and Components

The [Satterthwaite approximation](#) is a complex ratio involving the weighted variances of the two samples. It calculates the effective degrees of freedom (often denoted  $df$  or  $\nu$ ), which is

necessary for consulting the t-distribution table or calculating the p-value accurately.

The formula for the Satterthwaite approximation is as follows:

**Degrees of freedom:  $(s_1^2/n_1 + s_2^2/n_2) / \{ \}$**

where the variables are defined as:

**s<sub>1</sub><sup>2</sup>, s<sub>2</sub><sup>2</sup>:** These represent the respective [sample variance](#) of the first and second sample. These values quantify the spread of data around the mean within each sample.

**n<sub>1</sub>, n<sub>2</sub>:** These denote the sample size (the number of observations) collected for the first and second sample, respectively.

Understanding the structure of the formula reveals its reliance on weighted variance terms. The weighting mechanism ensures that the sample with the larger variance contributes disproportionately to the reduction in effective degrees of freedom, correcting the distribution under unequal variance conditions.

### Example: Calculating the Effective Degrees of Freedom

Suppose we want to know if the mean height of two different plant species is equal. We collect two simple random samples of each species and measure the height of the plants in inches. This scenario requires a [two-sample t-test](#) where unequal variances are anticipated.

The raw data collected for each species is presented below:

**Sample 1 (Species A):** 14, 15, 15, 15, 16, 18, 22, 23, 24, 25, 25

**Sample 2 (Species B):** 10, 12, 14, 15, 18, 22, 24, 27, 31, 33, 34, 34, 34

The calculated descriptive statistics necessary for the Satterthwaite approximation are:

**x<sub>1</sub>** = 19.27 (Mean height for Species A)

**x<sub>2</sub>** = 23.69 (Mean height for Species B)

**s<sub>1</sub><sup>2</sup>** = 20.42 (Sample variance for Species A)

**s<sub>2</sub><sup>2</sup>** = 83.23 (Sample variance for Species B)

**n<sub>1</sub>** = 11 (Sample size for Species A)

**n<sub>2</sub>** = 13 (Sample size for Species B)

We plug these values for the variances and sample sizes into the Satterthwaite approximation formula to find the effective degrees of freedom:

$$df = (s_1^2/n_1 + s_2^2/n_2)^2 / \{ \frac{s_1^4}{n_1^2} + \frac{s_2^4}{n_2^2} \}$$

$$df = (20.42/11 + 83.23/13)^2 / \{ \frac{20.42^2}{11^2} + \frac{83.23^2}{13^2} \}$$

$$df \approx 18.129$$

The effective [degrees of freedom](#) turns out to be **18.129**. If we were performing a manual calculation using a standard statistical table, we would conservatively round this value down to 18. This value is substantially lower than the 22 degrees of freedom calculated using the traditional pooled method, confirming the impact of the unequal [variances](#).

## Interpreting the Results and Conclusion

With the effective degrees of freedom established, we proceed with the final steps of the Welch's t-test: determining the critical value and calculating the test statistic. Assuming a significance level of  $\alpha = 0.05$  for a two-tailed test, we use the rounded degrees of freedom ( $df=18$ ) to find the appropriate critical threshold.

We consult the t-distribution table that corresponds to a two-tailed test with  $\alpha = .05$  for 18 degrees of freedom:

|           | P     |       |        |        |        |         |         |
|-----------|-------|-------|--------|--------|--------|---------|---------|
| one-tail  | 0.1   | 0.05  | 0.025  | 0.01   | 0.005  | 0.001   | 0.0005  |
| two-tails | 0.2   | 0.1   | 0.05   | 0.02   | 0.01   | 0.002   | 0.001   |
| DF        |       |       |        |        |        |         |         |
| 1         | 3.078 | 6.314 | 12.706 | 31.821 | 63.656 | 318.289 | 636.578 |
| 2         | 1.886 | 2.92  | 4.303  | 6.965  | 9.925  | 22.328  | 31.6    |
| 3         | 1.638 | 2.353 | 3.182  | 4.541  | 5.841  | 10.214  | 12.924  |
| 4         | 1.533 | 2.132 | 2.776  | 3.747  | 4.604  | 7.173   | 8.61    |
| 5         | 1.476 | 2.015 | 2.571  | 3.365  | 4.032  | 5.894   | 6.869   |
| 6         | 1.44  | 1.943 | 2.447  | 3.143  | 3.707  | 5.208   | 5.959   |
| 7         | 1.415 | 1.895 | 2.365  | 2.998  | 3.499  | 4.785   | 5.408   |
| 8         | 1.397 | 1.86  | 2.306  | 2.896  | 3.355  | 4.501   | 5.041   |
| 9         | 1.383 | 1.833 | 2.262  | 2.821  | 3.25   | 4.297   | 4.781   |
| 10        | 1.372 | 1.812 | 2.228  | 2.764  | 3.169  | 4.144   | 4.587   |
| 11        | 1.363 | 1.796 | 2.201  | 2.718  | 3.106  | 4.025   | 4.437   |
| 12        | 1.356 | 1.782 | 2.179  | 2.681  | 3.055  | 3.93    | 4.318   |
| 13        | 1.35  | 1.771 | 2.16   | 2.65   | 3.012  | 3.852   | 4.221   |
| 14        | 1.345 | 1.761 | 2.145  | 2.624  | 2.977  | 3.787   | 4.14    |
| 15        | 1.341 | 1.753 | 2.131  | 2.602  | 2.947  | 3.733   | 4.073   |
| 16        | 1.337 | 1.746 | 2.12   | 2.583  | 2.921  | 3.686   | 4.015   |
| 17        | 1.333 | 1.74  | 2.11   | 2.567  | 2.898  | 3.646   | 3.965   |
| 18        | 1.33  | 1.734 | 2.101  | 2.552  | 2.878  | 3.61    | 3.922   |
| 19        | 1.328 | 1.729 | 2.093  | 2.539  | 2.861  | 3.579   | 3.883   |
| 20        | 1.325 | 1.725 | 2.086  | 2.528  | 2.845  | 3.552   | 3.85    |
| 21        | 1.323 | 1.721 | 2.08   | 2.518  | 2.831  | 3.527   | 3.819   |
| 22        | 1.321 | 1.717 | 2.074  | 2.508  | 2.819  | 3.505   | 3.792   |
| 23        | 1.319 | 1.714 | 2.069  | 2.5    | 2.807  | 3.485   | 3.768   |
| 24        | 1.318 | 1.711 | 2.064  | 2.492  | 2.797  | 3.467   | 3.745   |
| 25        | 1.316 | 1.708 | 2.06   | 2.485  | 2.787  | 3.45    | 3.725   |
| 26        | 1.315 | 1.706 | 2.056  | 2.479  | 2.779  | 3.435   | 3.707   |
| 27        | 1.314 | 1.703 | 2.052  | 2.473  | 2.771  | 3.421   | 3.689   |
| 28        | 1.313 | 1.701 | 2.048  | 2.467  | 2.763  | 3.408   | 3.674   |
| 29        | 1.311 | 1.699 | 2.045  | 2.462  | 2.756  | 3.396   | 3.66    |
| 30        | 1.31  | 1.697 | 2.042  | 2.457  | 2.75   | 3.385   | 3.646   |
| 60        | 1.296 | 1.671 | 2      | 2.39   | 2.66   | 3.232   | 3.46    |
| 120       | 1.289 | 1.658 | 1.98   | 2.358  | 2.617  | 3.16    | 3.373   |
| 1000      | 1.282 | 1.646 | 1.962  | 2.33   | 2.581  | 3.098   | 3.3     |
| Inf       | 1.282 | 1.645 | 1.96   | 2.326  | 2.576  | 3.091   | 3.291   |

The t critical value is determined to be **\$pm\$2.101**.

Next, we calculate the observed test statistic using the unpooled standard error calculation central to the [Welch's t-test](#):

**Test statistic:**  $(\bar{x}_1 - \bar{x}_2) / (\sqrt{s_1^2/n_1 + s_2^2/n_2})$

Substituting the descriptive statistics:

Test statistic:  $(19.27 - 23.69) / (\sqrt{20.42/11 + 83.23/13}) = -4.42 / 2.873 \approx \mathbf{-1.538}$

Since the absolute value of our test statistic (1.538) is not larger than the t critical value (2.101), the

result falls within the acceptance region. We therefore fail to reject the null hypothesis of the test.

The final interpretation is that there is not sufficient evidence to conclude that the mean heights of the two populations of plant species are significantly different at the chosen level of significance.

## The Satterthwaite Approximation in Statistical Software

In practice, you will rarely have to calculate the [Satterthwaite approximation](#) by hand. The primary value of manual computation is conceptual understanding rather than operational necessity.

Instead, common statistical software packages are programmed to use the Satterthwaite approximation automatically whenever the assumption of equal variances is relaxed. Software provides a more precise answer because it utilizes the fractional [degrees of freedom](#) (e.g., 18.129) rather than the rounded integer (18) required for traditional tables.

Statistical environments like R, Python (using SciPy), SAS, and Stata all offer robust implementations of the Welch's t-test, ensuring that researchers can analyze data exhibiting heteroscedasticity with accuracy and confidence. This automated application makes the Satterthwaite method the modern standard for robustly comparing two population means.