

Welch's T-Test

Introduction

Thus far, you've seen the traditional Student's t-test for hypothesis testing between two sample means. Recall that z-tests are also appropriate for statistics, such as the mean, which can be assumed to be normally distributed. However, when sample sizes are low ($n_{\text{observations}} < 30$), the t-test is more appropriate, as the t-distribution has heavier tails. Even with this modification, remember that there are still several assumptions to the model. Most notably, traditional t-tests assume that sample sizes and sample variances between the two groups are equal. When these assumptions are not met, Welch's t-test is generally a more reliable test.

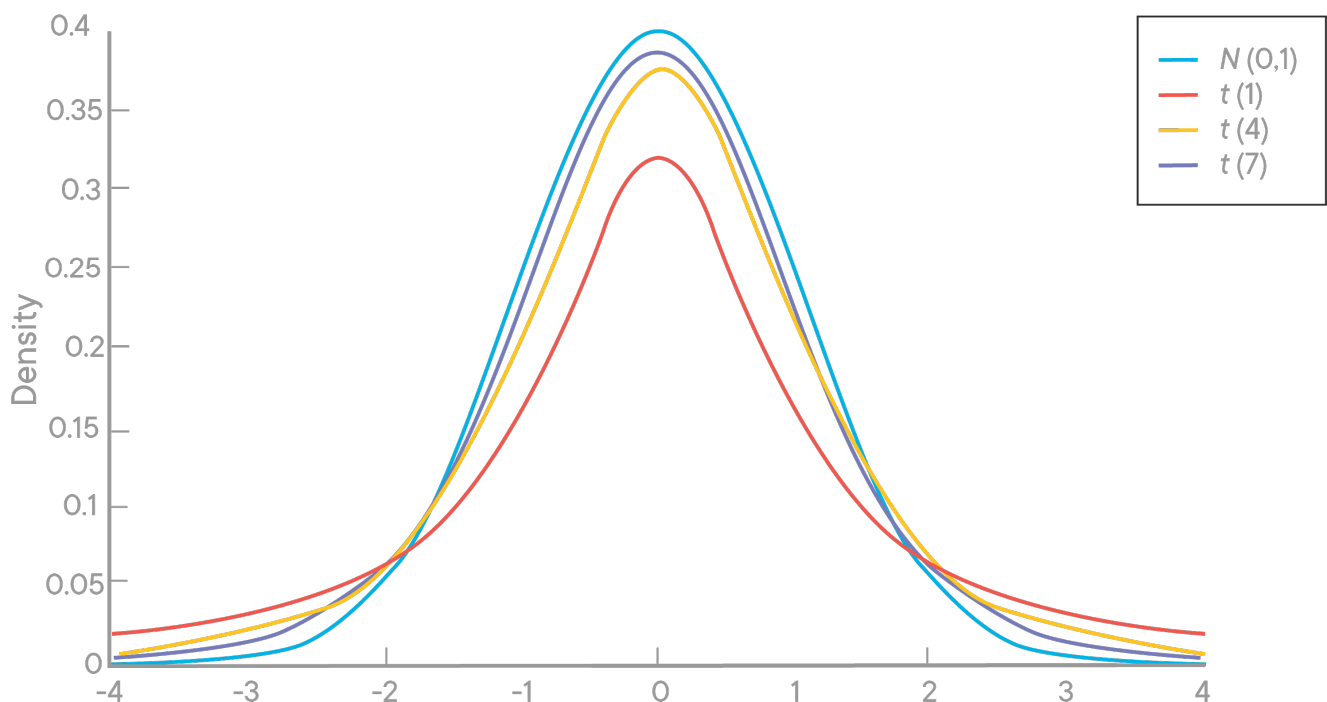
Objectives

You will be able to:

- List the conditions needed to require a Welch's t-test
- Calculate the degrees of freedom for a Welch's t-test
- Calculate p-values using Welch's t-test

T-test review

Recall that t-tests are a useful method for determining whether the mean of two small samples indicate different underlying population parameters. The reasoning behind this begins with the use of z-tests to calculate the likelihood of sampling a particular value from a normal distribution. Furthermore, by the central limit theorem, the mean of a sample is a normally distributed variable centered around the actual underlying population mean. That said, t-tests are more appropriate for small samples ($n_{\text{observations}} < 30$), due to disproportionate tails. Finally, recall that the t-distribution actually converges to a normal distribution as the degrees of freedom continues to increase.



A normal distribution vs. t-distributions with varying degrees of freedom. Note how the t-distribution approaches the normal distribution as the degrees of freedom increases. Recall that when performing a two-sample t-test, assuming that sample variances are equal, the degrees of freedom equals the total number of observations in the samples minus two.

Welch's t-test

Just as Student's t-test is a useful adaptation of the normal distribution which can lead to better likelihood estimates under certain conditions, the Welch's t-test is a further adaptation that accounts for additional perturbations in the underlying assumptions of the

model. Specifically, the Student's t-test assumes that the samples are of equal size and equal variance. When these assumptions are not met, then Welch's t-test provides a more accurate p-value.

Here is how you calculate it:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{se_1^2 + se_2^2}} \text{ where}$$

- \bar{X}_i - mean of sample i
- s_i^2 - variance of sample i
- N_i - sample size of sample i

The modification is related to the **degrees of freedom** in the t-test, which tends to increase the test power for samples with unequal variance. When two groups have equal sample sizes and variances, Welch's t-test tends to give the same result as the Student's t-test. However, when sample sizes and variances are unequal, Student's t-test is quite unreliable, whereas Welch's tends to perform better.

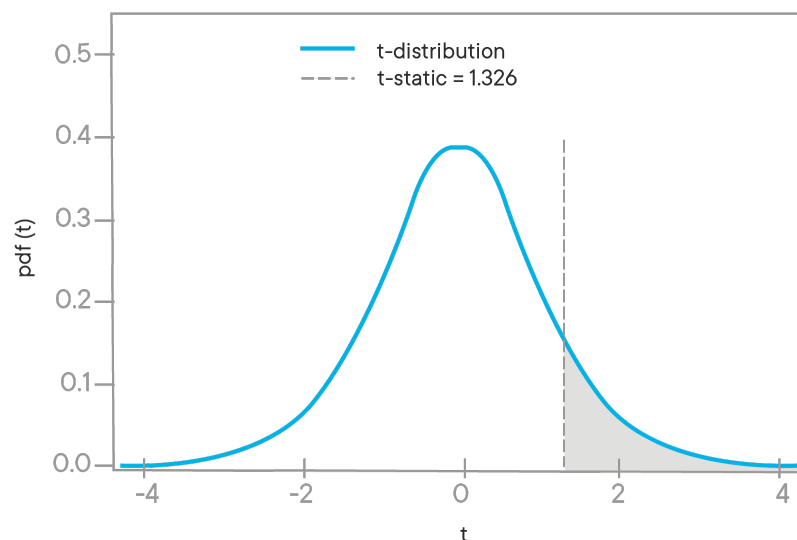
Calculate the degrees of freedom

Once the t-score has been calculated for the experiment using the above formula, you then must calculate the degrees of freedom for the t-distribution. Under the two-sample Student's t-test, this is simply the total number of observations in the samples size minus two, but given that the sample sizes may vary using the Welch's t-test, the calculation is a bit more complex:

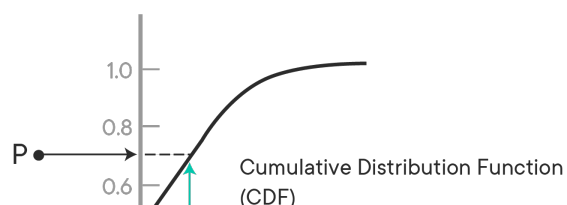
$$\nu \approx \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \right)^2}{\frac{s_1^4}{N_1^3} + \frac{s_2^4}{N_2^3}}$$

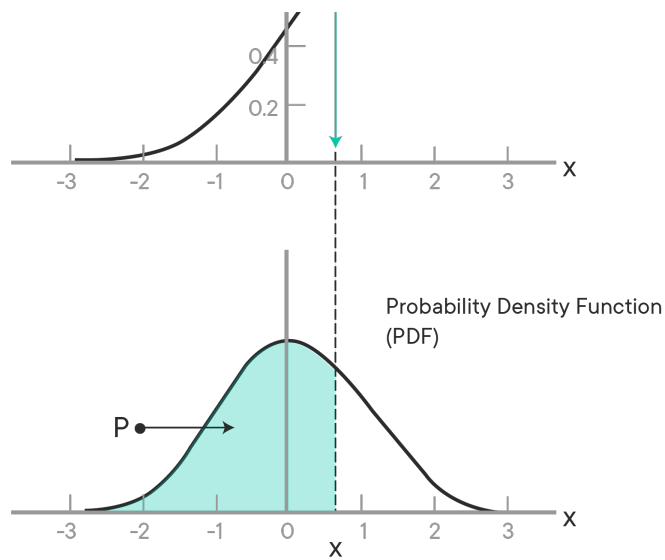
Calculate p-values

Finally, as with the Student's t-test (or a z-test for that matter), you convert the calculated score into a p-value in order to confirm or reject the null-hypothesis of your statistical experiment. For example, you might be using a one-sided t-test to determine whether a new drug had a positive effect on patient outcomes. The p-value for the experiment is equivalent to the area under the t-distribution with the degrees of freedom, as calculated above, and the corresponding t-score.



The easiest method for determining said p-values is to use the `.cdf()` method from `scipy.stats` to find the complement and subtracting this from 1.





Relations Between Two Different Typical Representations of a Population

Here's the relevant code snippet:

```
import scipy.stats as stats
```

```
p = 1 - stats.t.cdf(t, df)
```

Summary

This lesson briefly introduced you to another statistical test for comparing the means of two samples: Welch's t-test. Remember that when your samples are not of equal size or do not have equal variances, it is a more appropriate statistical test than the Student's t-test!