A sampling distribution shows the probability of various sample outcomes when the null hypothesis is true.

The p is always the probability of the statistical test value when H0 is true.

If your test gives you \( p < .05 \), it means that the sample results you actually got occur fewer than 5 times in 100 when the null hypothesis is true.

It is true that adopting a .05 \( \alpha \) level leaves some room for mistaking a chance difference for a real difference.

If you lowering the \( \alpha \) level (to a probability such as .01 or .001) reduces the probability of mistaking a chance difference for a real difference.

Unfortunately, lowering the \( \alpha \) level increases the probability of a different kind of mistake.

<table>
<thead>
<tr>
<th>The true situation in the population</th>
<th>H0 true</th>
<th>H0 false</th>
</tr>
</thead>
<tbody>
<tr>
<td>The decision made on the basis of sample data</td>
<td>Reject H0</td>
<td>1. Type I error</td>
</tr>
<tr>
<td>Retain H0</td>
<td>2. Correct decision</td>
<td>4. Type II error</td>
</tr>
</tbody>
</table>
Cell 1 shows the situation when the null hypothesis is true and you reject it.

You sample data and hypothesis testing have produced a mistake – called Type I error.

If the null hypothesis is false and you retain it (cell 4), you have made a mistake – called Type II error.

The probability of Type I error is $\alpha$, and you control alpha when you adopt a significance level.

If the data produce a $p$ value that is less than $\alpha$, you reject $H_0$ and conclude that the difference observed is statistically significant.

If you reject $H_0$, you could be making a Type I error, but the probability of this error is controlled by your choice of $\alpha$ level.

Rejecting $H_0$ is wrong only when the null hypothesis is true.

A Type II error is possible only if you retain the $H_0$ – the hypothesis test has failed to detect a real treatment effect.

If the $H_0$ is false and you retain it, you have made a Type II error.

The probability of Type II error is symbolized by $\beta$. 
Hypothesis testing with two samples is similar to hypothesis testing with one sample.

The same hypothesis-testing reasoning will be used, but you will have data from two samples or two group experiment.

In a well-designed, well-executed experiment, all imaginable results are included in this statement: Either Treatment A has an effect or it does not have an effect.

Let’s making a tentative assumption that Treatment A does not have an effect.

Gather data.

Using a sampling distribution based on the assumption that Treatment A has no effect, find the probability of the data obtained.

If the probability is low, abandon your tentative assumption and conclude that Treatment A has an effect.

If the probability is not low, you are back where you began: The data have not given you evidence against either of the two logical possibilities.
In the language of experiment:

Begin with two logical possibilities, the null hypothesis, \( H_0 \), and alternative hypothesis, \( H_1 \).

\( H_0 \): Treatment A does not have an effect; that is, the mean of the population scores of those who receive Treatment A is equal to the mean of the population of scores of those who do not receive Treatment A.

The difference between population means is zero. In statistical language:

\( H_0: \mu_A - \mu_{no\ A} = 0 \) or \( H_0: \mu_A = \mu_{no\ A} \)

\( H_1 \): Treatment A does have an effect; that is, the mean of the population of scores of those who receive Treatment A is not equal to the mean of the population of scores of those who do not receive Treatment A.

The alternative hypothesis, which should be chosen before the data are gathered, can be two-tailed or one-tailed.

An alternative hypothesis consists of one of the three \( H_1 \)'s that follow.

Two-tailed alternative: \( H_1: \mu_A \neq \mu_{no\ A} \)
In the example of the sample experiment, this hypothesis says that Treatment A has an effect, but it does not indicate whether the treatment improves or disrupts performance on task Q.

A two-tailed test allows either conclusion.

One-tailed alternative:  
\[ H_1: \mu_A > \mu_{no\ A} \]  
\[ H_1: \mu_A < \mu_{no\ A} \]

The first two-tailed alternative hypothesis allows you to conclude that Treatment A increases scores.

However, no outcome of the experiment can lead to the conclusion that Treatment A produces lower scores than no treatment.

The second alternative hypothesis permits a conclusion that Treatment A reduces scores, but not that it increases scores.

2. Tentatively assume that Treatment A has no effect (that is, assume \( H_0 \)). If \( H_0 \) is true, the two samples will be alike except for the usual chance variation in samples.

3. Decide on a level \( \alpha \) level. (Usually, \( \alpha = .05 \).)

4. Choose an appropriate inferential statistical test.

This test will have
(a) a test statistic that can be calculated from the data,
(b) a sampling distribution of the test statistic that shows its distribution when $H_0$ is true, and
(c) a critical value for the $\alpha$ level.

5. Calculate a test statistic using the sample data.

6. Compare the test statistic to the critical value from the sampling distribution. If the test statistic’s probability is less than $\alpha$, reject $H_0$. If larger, retain $H_0$. (For the $t$ distribution, if the data-based $t$ value is greater than the critical value in the table, reject $H_0$. If the data-based $t$ value is less than the critical value, retain $H_0$.)

7. Write a conclusion that uses the terms of the experiment.

Your conclusion should describe how the levels of the independent variable differ on the dependent variable in the populations the samples came from.

In your use of the $t$ distribution, you have determining degrees of freedom by rule of thumb techniques:

$$N - 1$$ when you determine whether a sample mean $X$ comes from a population with a mean $\mu$, and $N - 2$ when you determine whether a correlation coefficient $r$ is significantly different from .00.
The ‘freedom’ refers to the freedom of a number to have any possible value.

If you are asked to pick two numbers and there is no restrictions, both numbers are free to vary (take any value) and you have two degrees of freedom.

If a restriction is imposed – say, $\Sigma X = 0$ – then one degree of freedom is lost because of that restriction; that is, when you now pick the two numbers, only one of them is free to vary.

For example, if you choose 3 for the first number, the second number must be -3.

In similar way, if you are to pick five numbers with a restriction that $\Sigma X = 0$, you have four degrees of freedom.

When you found $s_x^2$, as required in the formula for the one-sample t test, used

$$s_x = \frac{\bar{X}}{\sqrt{\frac{N}{N-1}}} = \frac{\sqrt{\Sigma (X - \bar{X})^2}}{N-1}$$

The restriction that is built is that $\Sigma (X - \bar{X})$ is always zero and, in order to meet that requirement, one of the $X$’s is determined.

The degree of freedom for $s_x^2$ is $N - 1$. 
An experiment with two groups can be either a correlated-samples design or an independent-samples design.

You must decide what kind of design you have before you analyze the data.

These different design require different $t$ test formulas.

The difference can be tell by knowing whether scores in one group are paired with scores in a second group.

The three types of correlated-samples designs are natural pairs, matched pairs, and repeated measures.

In independent-samples design, a $t$ test is used to decide whether two populations have the same mean.

The null hypothesis is

$$H_0: \mu_1 = \mu_2$$

where the subscripts 1 and 2 are assigned arbitrarily to the two populations.
If this null hypothesis is true, any difference between the two sample means is due to chance.

The task is to establish an α level, calculate a $t$ test value, and compare that value with a critical value of $t$ in Table D.

If the $t$ value calculated from the data is greater than the critical value (that is, less probable than $\alpha$), reject $H_0$ and conclude that the two samples came from populations with different means.

If the data-based $t$ value is not large as critical value, retain $H_0$.

For an independent-samples design, the formula for $t$ test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

The term $s_{\bar{x}_1 - \bar{x}_2}$ is the standard error of a difference.
The *t* Test for Independent Sample Designs

When a *t* test is used to decide whether two populations have the same mean, the null hypothesis is

\[ H_0: \mu_1 = \mu_2; \]

where the subscripts 1 and 2 are assigned arbitrarily to the two populations.

The task is to establish the alpha level, calculate a *t* test value, and compare the value with a critical value of *t* in Table D.

The formula for the *t* test is

\[ t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}} \]

where; \( S_{\bar{X}_1 - \bar{X}_2} \) is the standard error of difference (the s.d. of a sampling distribution of differences between means).

Use this formula when the two samples not have an equal number of scores (\( N_1 \neq N_2 \)).
If N1 = N2, use this formula

$$
S_{X_1 - X_2} = \sqrt{\frac{\Sigma X_1^2 - (\Sigma X_1)^2}{N_1} + \Sigma X_2^2 - (\Sigma X_2)^2}{N_2} \div N_1(N_2 - 1)}
$$

The following example shows the use of an independent-samples design to evaluate the effect of a drug on learning a complex problem solving task.

The drug group (7 monkeys) received pills, while the other group (6 monkeys) was given an inert substance (a placebo).

Pills and training on the task continued for six days.

The total number of errors that each monkeys made was the dependent variable.

$$H_0$$ is the drug has no effect on errors.

A two-tailed test is called for because the investigator is interested in any effect the drug has, either to enhance or to inhibit performance.

The number of errors each monkeys made as follows

<table>
<thead>
<tr>
<th>Drug Group, $X_1$</th>
<th>Placebo Group, $X_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>39</td>
</tr>
<tr>
<td>52</td>
<td>57</td>
</tr>
<tr>
<td>26</td>
<td>68</td>
</tr>
<tr>
<td>47</td>
<td>74</td>
</tr>
<tr>
<td>42</td>
<td>49</td>
</tr>
<tr>
<td>37</td>
<td>57</td>
</tr>
<tr>
<td>40</td>
<td></td>
</tr>
</tbody>
</table>
Applying the t test (because the N’s are unequal for the two samples, the N1#N2 formula is used).

\[
t = \frac{\bar{X}_1 - \bar{X}_2}{\frac{S^2}{n_1} + \frac{S^2}{n_2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}
\]

\[
= \frac{-1.74}{\frac{S^2}{10}} = -2.99
\]

The t value for these data is -2.99, a negative value.

(For a two-tailed test, always use the absolute value of t).

To evaluate a data-based t value of 2.99 with 11 df, you need a critical value from Table D.

The critical value in the column for a two-tailed test with \( \alpha = .05 \) is 2.201.

That is \( t_{.05}(11) = 2.201 \). Thus the null hypothesis can be rejected.
Because 2.99 is also greater than critical value at the .02 level (2.718), the results would usually be reported as “significant at the .02 level.”

The final step is to interpret the results.

“The drug group average fewer errors than the placebo group (37.71 vs 57.33). Thus, the drug facilitated learning (p < .02).”

---

**Exercise.**

A psychologist doing research on how divorce influence a children behavior. The hypothesis that he developed was “Children from divorced parents have an aggressive behavior.” To test this hypothesis, the researcher observed 17 children from divorced parents and 15 children from happy family (live together.)

<table>
<thead>
<tr>
<th>Divorced Family</th>
<th>Happy Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 8</td>
<td>N = 6</td>
</tr>
<tr>
<td>( \sum x = 72.0 )</td>
<td>( \sum x = 210 )</td>
</tr>
<tr>
<td>( 2 \chi^2 = 15.800 )</td>
<td>( 2 \chi^2 = 665.0 )</td>
</tr>
<tr>
<td>( \bar{x} = 4.0 )</td>
<td>( \bar{x} = 7.5 )</td>
</tr>
</tbody>
</table>
Because $t_{.01(13 \; df)} = 3.012$, a $t$ value of 3.78 is significant beyond the .01 level; that is, $p < .01$.

Because the ‘after’ mean is greater than the ‘before’ mean, conclude that racial attitudes were significantly more positive after camp than before.