# Polling liberals and conservatives

50xp

You are responsible for conducting polls to understand voters' preferences for a particular political candidate.

In the first poll, you want to understand how preferences vary between liberals and conservatives. You ask a group of liberals and a group of conservatives to each rate their likelihood of voting for the candidate.

Which test should you use to determine if there's a significant difference in preferences between these two groups?

#### Possible Answers

* 

Single sample t-test

* 

Dependent t-test

* 

Independent t-test

[**Take Hint (-15xp)**](https://campus.datacamp.com/courses/intro-to-statistics-with-r-students-t-test/chapter-one-introduction-to-t-tests?ex=2)

# More political polls

50xp

In the previous exercise, you decided to conduct an independent t-test to compare liberals and convervatives, since you were comparing two independent samples at a single point in time. The poll was so successful that you decide to conduct two more.

In the second poll, you want to understand the effect of a campaign speech on voters' preferences. You ask a single group of voters to rate their likelihood of voting for the candidate before the speech and again after the speech.

In the third poll, you want to understand if voters from a particular neighborhood are likely to vote differently when compared to the overall voting population. You poll voters from this neighborhood and compare the results to a recent and trustworthy national poll.

Which tests should you use for the second and third polls, respectively?

#### Possible Answers

* 

Dependent t-test, independent t-test

* 

Independent t-test, single sample t-test

* 

Single sample t-test, dependent t-test

* 

Dependent t-test, single sample t-test

[**Take Hint (-15xp)**](https://campus.datacamp.com/courses/intro-to-statistics-with-r-students-t-test/chapter-one-introduction-to-t-tests?ex=3)

# Significance tests

50xp

Which of the following are commonly referred to as significance tests?

#### Possible Answers

* 

z-tests

* 

t-tests

* 

p-values

* 

Options 1 & 2 above

* 

None of the above

[**Take Hint (-15xp)**](https://campus.datacamp.com/courses/intro-to-statistics-with-r-students-t-test/chapter-one-introduction-to-t-tests?ex=5)

# What's a summary statistic?

50xp

A sampling distribution is a hypthotical distribution of a summary statistic from...

#### Possible Answers

* 

Multiple samples of different sizes, each from the same underlying population.

* 

Multiple samples of the same size, each from the same underlying population.

* 

Multiple samples of different sizes, each from a different underlying population.

* 

Multiple samples of the same size, each from a different underlying population.

* 

A single sample from the underlying population.

[**Take Hint (-15xp)**](https://campus.datacamp.com/courses/intro-to-statistics-with-r-students-t-test/chapter-one-introduction-to-t-tests?ex=6)

# Single sample t-tests

50xp

You want to know how a group of people from a particular geographic region perform on a well-known test of intelligence. In particular, you are interested in finding out whether or not this group scores significantly higher than the overall population on an IQ test. This is a form of Null Hypothesis Significance Testing (NHST), where the null hypothesis is that there's no difference between this group and the overall population.

Using a random sample of 10 subjects from the group, you find that their mean IQ is 1.5 standard errors higher than the mean IQ of the population. What do you conclude (using a significance level of 5%) regarding the group's intelligence? (Recall that the critical value for the z-distribution at a significance level of 5% is 1.96.)

#### Possible Answers

* 

There is not enough information to make a conclusion.

* 

The group's IQ is not significantly different from the mean IQ of the population.

* 

The group's IQ is significantly different from the mean IQ of the population.

* 

The group's IQ is equal to the mean IQ of the population.

[**Take Hint (-15xp)**](https://campus.datacamp.com/courses/intro-to-statistics-with-r-students-t-test/chapter-one-introduction-to-t-tests?ex=8)

# Understanding the t-distribution

100xp

When performing a t-test, you first calculate your t-statistic using the familiar formula:

t=X−MSEt=X−MSE

XX is the observed value, MM is the expected value under the null hypothesis (or population mean), and SESE is the standard error. Once you've computed the t-statistic, you then compare it to the so-called critical value, which comes from the relevant t-distribution.

The shape of a t-distribution, and thus the critical value, is determined entirely by its degrees of freedom. To demonstrate this, let's draw some density plots for t-distributions using different degrees of freedom.

## Instructions

* Create a vector x that contains a sequence of length 100 between -4 and 4. See ?seq for help.
* Use dt() to generate t-distributions with 4, 6, 8, 10, and 12 degrees of freedom (in that order). The first argument to dt() is the vector of values at which to evalute the t-distribution (x from above) and the second argument (df) is the degrees of freedom.
* Plot each of the t-distributions. Once the inital plot() is created, you'll use lines() to plot each additional distribution. The two arguments to lines() are the same as the first two arguments to plot(), except that you'll have to substitute the appropriate y-values. Use the color black for 4 degrees of freedom, red for 6, orange for 8, green for 10, and blue for 12.
* Add a legend() to your plot. The legend should be situated at the top right corner of your plot and should have the title "t-distributions". This is done by setting the first argument to "topright" and the title argument to "t-distributions".

# The working memory dataset

100xp

In the following exercises, you will conduct a dependent (or paired) t-test on the "working memory" dataset. This dataset consists of the intelligence scores for subjects before and after training, as well as for a control group. Our goal is to assess whether intelligence training results in significantly different intelligence scores for the individuals.

The observations of individuals before and after training are two samples from the same group at different points in time, which calls for a dependent t-test. This will test whether or not the difference in mean intelligence scores before and after training are significant.

The working memory dataset has been loaded into your workspace as the object wm. It contains the data for both the group who received training and the group who did not.

## Instructions

* Print wm to the console to get a feel for the data
* Create a subset of wm that includes only the training group and store the result in wm\_t. A value of 1 in the train column indicates that a subject received training, while a value of 0 indicates that they did not.
* View summary statistics for wm\_t with the [describe()](http://www.rdocumentation.org/packages/psych/functions/describe) function.
* Use the [boxplot()](http://www.rdocumentation.org/packages/graphics/functions/boxplot) function to create a boxplot of the pre and post column of wm\_t. Give the x-axis the label "Pre- and Post-Training" and the y-axis the label "Intelligence Score".