15.6.2 Use the One-Sample t-Test

Jeremy is excited to dig into t-tests. He knows this is one of the most popular statistical tests in the world, and the CEO will definitely expect him to be able to do it!

The **Student's t-test** (most commonly referred to as **t-test**) is one of the most basic and popular statistical tests in the world. In statistics, we use a t-test to compare the mean of one dataset to another under a few assumptions.

There are two main forms of the t-test that we use: the **one-sample t-test** and the **two-sample t-test**. The one-sample t-test is used to determine whether there is a statistical difference between the means of a sample dataset and a hypothesized, potential population dataset. In other words, a one-sample t-test is used to test the following hypotheses:

- H_0 : There is **no statistical difference** between the observed sample mean and its presumed population mean.
- H_a : There is **a statistical difference** between the observed sample mean and its presumed population mean.

NOTE

> We can also use a one-sided t-test by changing our alternative hypothesis to state that our sample mean is **significantly less** or **significantly more** than our presumed population mean.

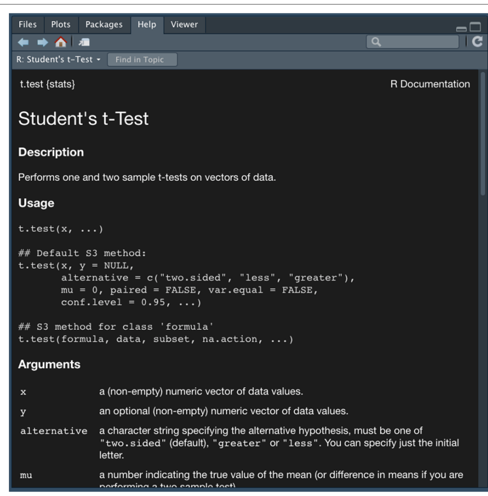
Before we can apply any statistical test to our data, we must check if there are any assumptions regarding our input dataset. When it comes to our one-sample t-test there are five assumptions about our input data:

1. The input data is numerical and continuous. This is because we are testing the distribution of two datasets.
2. The sample data was selected randomly from its population data.
3. The input data is considered to be normally distributed.
4. The sample size is reasonably large. Generally speaking, this means that the sample data distribution should be similar to its population data distribution.
5. The variance of the input data should be very similar.

As long as our input data satisfies (or mostly satisfies) the above assumptions, we can use the one-sample t-test to assert the similarities or differences in our data.

In R, we can implement a one-sample t-test using the built-in stats package `t.test()` function. Type the following code into the R console to look at the `t.test()` documentation in the Help pane:

```
> ?t.test()
```



To use the `t.test()` function to perform our one-sample t-test, we have to use a few arguments:

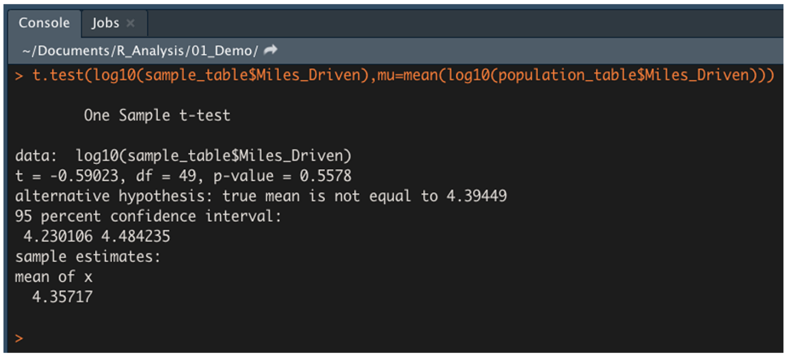
- **x**

- **mu**
- **alternative**

By setting all three of these arguments, the `t.test()` function should produce our test statistic "t" along with our p-value, which we can use to evaluate our null hypothesis.

For example, if we want to test if the miles driven from our previous sample dataset is statistically different from the miles driven in our population data, we would use our `t.test()` function as follows:

```
>t.test(log10(sample_table$Miles_Driven),mu=mean(log10(population_table$Mile
```



```
Console  Jobs
~/Documents/R_Analysis/01_Demo/
> t.test(log10(sample_table$Miles_Driven),mu=mean(log10(population_table$Miles_Driven)))

        One Sample t-test

data:  log10(sample_table$Miles_Driven)
t = -0.59023, df = 49, p-value = 0.5578
alternative hypothesis: true mean is not equal to 4.39449
95 percent confidence interval:
 4.230106 4.484235
sample estimates:
mean of x
  4.35717

>
```

There are a number of metrics produced from the `t.test()` function, but for now we will only concern ourselves with the calculated p-value. Assuming our significance level was the common 0.05 percent, our p-value is above our significance level. Therefore, we do not have sufficient evidence to reject the null hypothesis, and we would state that the two means are statistically similar.

IMPORTANT

> Due to random sampling, your sample dataset may differ from our example and thus your calculations may be different. Therefore, you'll need to compare your calculated p-value to your own significance level. If your p-value is lower than the significance level, you would have sufficient evidence to reject the null hypothesis and state that the two means are statistically different.