

15.6.3

Use the Two-Sample t-Test

Now that Jeremy has conquered the one-sample t-Test, he's ready for the two-sample t-Test.

The second main form of the t-Test is a two-sample t-Test. Instead of testing whether a sample mean is statistically different from its population mean, the two-sample t-Test determines whether the means of two samples are statistically different. In other words, a two-sample t-Test is used to test the following hypotheses:

- H_0 : There is **no statistical difference** between the two observed sample means.
- H_a : There is **a statistical difference** between the two observed sample means.

There are also five assumptions regarding our input data when using the two-sample t-Test, which are the same as the one-sample t-Test:

1. The input data is numerical and continuous.
2. Each sample data was selected randomly from the population data.
3. The input data is considered to be normally distributed.
4. Each 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.

In R, we use the same `t.test()` function to calculate both a one-sample t-Test and two-sample t-Test. However, the two-sample t-Test arguments are slightly different:

- x is the first numeric vector of sample data.
- y is the second numeric vector of sample data.
- **alternative** tells the `t.test()` function if the hypothesis is one-sided (one-tailed) or two-sided (two-tailed). The options for the alternative argument are "two.sided", "less", or "greater." By default, the `t.test()` function assumes a two-sided t-Test.

Once we have provided the necessary numeric vectors for each sample, the `t.test()` function will calculate our two-sample t-Test and return the same output as before. As practice, let's test whether the mean miles driven of two samples from our used car dataset are statistically different.

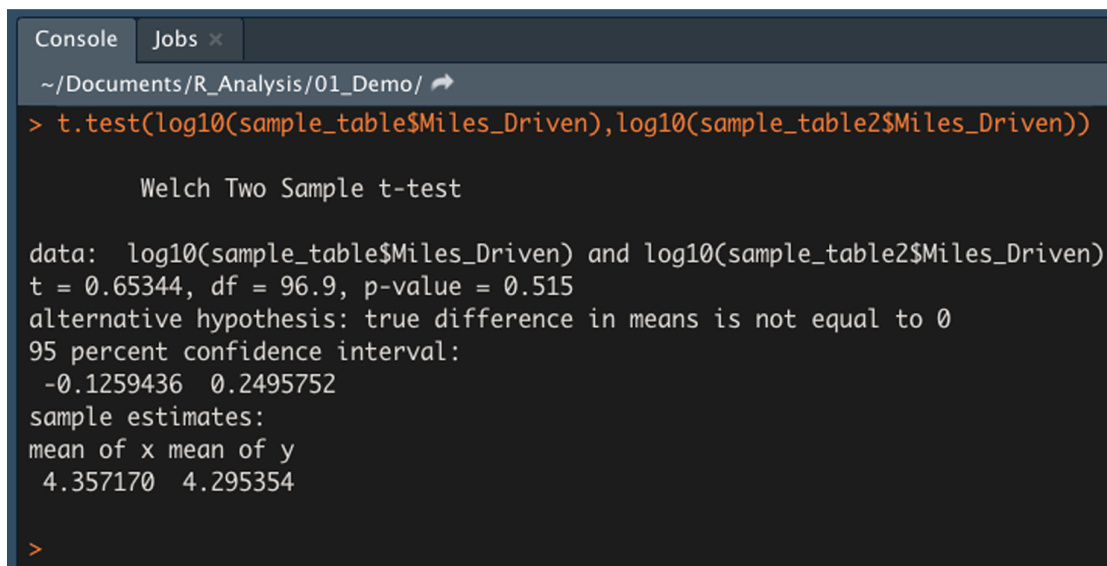
First, we produce our two samples using the following R statements:

```
> sample_table <- population_table %>% sample_n(50) #generate 50 randomly sampled data points
> sample_table2 <- population_table %>% sample_n(50) #generate another 50 randomly sampled data points
```

 [Retake](#)

Because our samples should not contain bias, we would expect our null hypothesis to be true—our samples should not be statistically different. To confirm, we'll use the `t.test()` function as follows:

```
> t.test(log10(sample_table$Miles_Driven),log10(sample_table2$Miles_Driven)) #compare means of two samples
```



```
Console Jobs x
~/Documents/R_Analysis/01_Demo/ ➔
> t.test(log10(sample_table$Miles_Driven),log10(sample_table2$Miles_Driven))

Welch Two Sample t-test

data: log10(sample_table$Miles_Driven) and log10(sample_table2$Miles_Driven)
t = 0.65344, df = 96.9, p-value = 0.515
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1259436  0.2495752
sample estimates:
mean of x mean of y
 4.357170  4.295354
>
```

 [Retake](#)

© 2020 - 2022 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.