

What is the complete guide to conducting hypothesis testing in R?

Authored by
stats writer

May 1, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the complete guide to conducting hypothesis testing in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=141713>

The complete guide to conducting hypothesis testing in R is a comprehensive resource that provides a step-by-step approach for analyzing data and making statistical inferences in the R programming language. This guide covers all aspects of hypothesis testing, including formulating research questions, choosing appropriate statistical tests, interpreting results, and assessing the validity and reliability of findings. It also includes practical examples and exercises to help users gain a thorough understanding of the process. By following this guide, individuals can effectively utilize R to conduct rigorous and accurate hypothesis testing in their research and data analysis projects.

The Complete Guide: Hypothesis Testing in R

A is a formal statistical test we use to reject or fail to reject some statistical hypothesis.

This tutorial explains how to perform the following hypothesis tests in R:

One sample t-test Two sample t-test Paired samples t-test

We can use the `t.test()` function in R to perform each type of test:

#one sample t-test

`t.test(x, y = NULL,`

`alternative = c("two.sided", "less", "greater"),`

`mu = 0, paired = FALSE, var.equal = FALSE,`

`conf.level = 0.95, ...)`

where:

x, y: The two samples of data.
alternative: The alternative hypothesis of the test.
mu: The true value of the mean.
paired: Whether to perform a paired t-test or not.
var.equal: Whether to assume the between the samples.
conf.level: The to use.

The following examples show how to use this function in practice.

Example 1: One Sample t-test in R

A is used to test whether or not the mean of a population is equal to some value.

For example, suppose we want to know whether or not the mean weight of a certain species of some turtle is equal to 310 pounds. We go out and collect a simple random sample of turtles with the following weights:

Weights: 300, 315, 320, 311, 314, 309, 300, 308, 305, 303, 305, 301, 303

The following code shows how to perform this one sample t-test in R:

```
#define vector of turtle weights
```

```
turtle_weights <- c(300, 315, 320, 311, 314, 309, 300, 308,  
305, 303, 305, 301, 303)
```

```
#perform one sample t-test
```

```
t.test(x = turtle_weights, mu = 310)
```

One Sample t-test

```
data: turtle_weights
```

```
t = -1.5848, df = 12, p-value = 0.139
```

```
alternative hypothesis: true mean is not equal to 310
```

```
95 percent confidence interval:
```

```
303.4236 311.0379
```

```
sample estimates:
```

```
mean of x
```

```
307.2308
```

From the output we can see:

```
t-test statistic: -1.5848degrees of freedom: 12p-value:  
0.13995% confidence interval for true mean: mean of  
turtle weights: 307.230
```

Since the p-value of the test (0.139) is not less than .05,

we fail to reject the null hypothesis.

Example 2: Two Sample t-test in R

A is used to test whether or not the means of two populations are equal.

For example, suppose we want to know whether or not the mean weight between two different species of turtles is equal. To test this, we collect a simple random sample of turtles from each species with the following weights:

Sample 1: 300, 315, 320, 311, 314, 309, 300, 308, 305, 303, 305, 301, 303

Sample 2: 335, 329, 322, 321, 324, 319, 304, 308, 305, 311, 307, 300, 305

The following code shows how to perform this two sample t-test in R:

```
#define vector of turtle weights for each sample  
sample1 <- c(300, 315, 320, 311, 314, 309, 300, 308, 305,  
303, 305, 301, 303)sample2 <- c(335, 329, 322, 321, 324,  
319, 304, 308, 305, 311, 307, 300, 305)
```

```
#perform two sample t-test  
t.test(x = sample1, y = sample2)
```

Welch Two Sample t-test

```
data: sample1 and sample2  
t = -2.1009, df = 19.112, p-value = 0.04914  
alternative hypothesis: true difference in means is not  
equal to 0  
95 percent confidence interval:  
-14.73862953 -0.03060124  
sample estimates:  
mean of x mean of y  
307.2308 314.6154
```

From the output we can see:

```
t-test statistic: -2.1009degrees of freedom: 19.112p-  
value: 0.0491495% confidence interval for true mean  
difference: mean of sample 1 weights: 307.2308mean of  
sample 2 weights: 314.6154
```

Since the p-value of the test (0.04914) is less than .05, we reject the null hypothesis.

This means we have sufficient evidence to say that the mean weight between the two species is not equal.

Example 3: Paired Samples t-test in R

A is used to compare the means of two samples when each observation in one sample can be paired with an observation in the other sample.

For example, suppose we want to know whether or not a certain training program is able to increase the max vertical jump (in inches) of basketball players.

To test this, we may recruit a simple random sample of 12 college basketball players and measure each of their max vertical jumps. Then, we may have each player use the training program for one month and then measure their max vertical jump again at the end of the month.

The following data shows the max jump height (in inches) before and after using the training program for each player:

Before: 22, 24, 20, 19, 19, 20, 22, 25, 24, 23, 22, 21

After: 23, 25, 20, 24, 18, 22, 23, 28, 24, 25, 24, 20

The following code shows how to perform this paired samples t-test in R:

```
#define before and after max jump heights  
before <- c(22, 24, 20, 19, 19, 20, 22, 25, 24, 23, 22,  
21)after <- c(23, 25, 20, 24, 18, 22, 23, 28, 24, 25, 24, 20)  
  
#perform paired samples t-test  
t.test(x = before, y = after, paired = TRUE)
```

Paired t-test

data: before and after

t = -2.5289, df = 11, p-value = 0.02803

**alternative hypothesis: true difference in means is not
equal to 0**

95 percent confidence interval:

-2.3379151 -0.1620849

sample estimates:

mean of the differences

-1.25

From the output we can see:

t-test statistic: -2.5289degrees of freedom: 11p-value:

0.0280395% confidence interval for true mean difference: mean difference between before and after: -1.25

Since the p-value of the test (0.02803) is less than .05, we reject the null hypothesis.

This means we have sufficient evidence to say that the mean jump height before and after using the training program is not equal.

Use the following online calculators to automatically perform various t-tests: