

How can one find confidence intervals in R, and what are some examples of using them?

Authored by
stats writer

April 22, 2024

RECOMMENDED CITATION

stats writer (2024). *How can one find confidence intervals in R, and what are some examples of using them?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=138169>

Confidence intervals are a statistical tool used to estimate the range of values within which a population parameter is likely to fall. In R, confidence intervals can be found using various functions such as "confint()" and "t.test()". These functions take a sample of data and calculate the confidence interval based on a specified confidence level. For example, a 95% confidence interval would indicate that there is a 95% chance that the true population parameter falls within the calculated range. This tool is commonly used in hypothesis testing, where the confidence interval can help determine if a null hypothesis can be rejected or not. Additionally, confidence intervals can also be helpful in analyzing data in fields such as market research, social sciences, and medical studies. Overall, confidence intervals in R provide a reliable and versatile method for estimating population parameters and making data-driven decisions.

Find Confidence Intervals in R (With Examples)

A confidence interval is a range of values that is likely to contain a population parameter with a certain level of confidence.

It is calculated using the following general formula:

Confidence Interval = (point estimate) +/- (critical value)*(standard error)

This formula creates an interval with a lower bound and an upper bound, which likely contains a population parameter with a certain level of confidence:

Confidence Interval =

This tutorial explains how to calculate the following confidence intervals in R:

1. Confidence Interval for a Mean

2. Confidence Interval for a Difference in Means

3. Confidence Interval for a Proportion

4. Confidence Interval for a Difference in Proportions

Let's jump in!

Example 1: Confidence Interval for a Mean

We use the following formula to calculate a confidence interval for a mean:

Confidence Interval = $\bar{x} \pm t_{n-1, 1-\alpha/2} * (s/\sqrt{n})$

where:

\bar{x} : sample mean; $t_{n-1, 1-\alpha/2}$: the t-critical values; s : sample standard deviation; n : sample size

Example: Suppose we collect a random sample of turtles with the following information:

Sample size $n = 25$ Sample mean weight $\bar{x} = 300$ Sample standard deviation $s = 18.5$

The following code shows how to calculate a 95% confidence interval for the true population mean weight of turtles:

```
#input sample size, sample mean, and sample standard deviation
```

```
n <- 25
```

```
xbar <- 300
```

```
s <- 18.5
```

```
#calculate margin of error
```

```
margin <- qt(0.975,df=n-1)*s/sqrt(n)
```

```
#calculate lower and upper bounds of confidence interval
```

```
low <- xbar - margin
```

```
low
```

```
292.3636
```

```
high <- xbar + margin
```

```
high
```

```
307.6364
```

The 95% confidence interval for the true population

mean weight of turtles is .

Example 2: Confidence Interval for a Difference in Means

We use the following formula to calculate a confidence interval for a difference in population means:

$$\text{Confidence interval} = (x_1 - x_2) \pm t^* \sqrt{((s_p^2/n_1) + (s_p^2/n_2))}$$

where:

x_1, x_2 : sample 1 mean, sample 2 mean; t : the t-critical value based on the confidence level and $(n_1 + n_2 - 2)$ degrees of freedom; s_p^2 : pooled variance, calculated as $((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2) / (n_1 + n_2 - 2)$; t : the t-critical value; n_1, n_2 : sample 1 size, sample 2 size

Example: Suppose we want to estimate the difference in mean weight between two different species of turtles, so we go out and gather a random sample of 15 turtles from each population. Here is the summary data for each sample:

Sample 1:

$$x_1 = 310 \quad s_1 = 18.5 \quad n_1 = 15$$

Sample 2:

$$x_2 = 300, s_2 = 16.4, n_2 = 15$$

The following code shows how to calculate a 95% confidence interval for the true difference in population means:

```
#input sample size, sample mean, and sample standard deviation
```

```
n1 <- 15
```

```
xbar1 <- 310
```

```
s1 <- 18.5
```

```
n2 <- 15
```

```
xbar2 <- 300
```

```
s2 <- 16.4
```

```
#calculate pooled variance
```

```
sp = ((n1-1)*s1^2 + (n2-1)*s2^2) / (n1+n2-2)
```

```
#calculate margin of error
```

```
margin <- qt(0.975,df=n1+n2-1)*sqrt(sp/n1 + sp/n2)
```

```
#calculate lower and upper bounds of confidence interval
```

```
low <- (xbar1-xbar2) - margin
```

```
low
```

```
-3.055445
```

```
high <- (xbar1-xbar2) + margin
```

```
high
```

```
23.05544
```

The 95% confidence interval for the true difference in population means is .

Example 3: Confidence Interval for a Proportion

We use the following formula to calculate a confidence interval for a proportion:

Confidence Interval = $p \pm z \cdot \sqrt{p(1-p) / n}$

where:

p: sample proportion
z: the chosen z-value
n: sample size

Example: Suppose we want to estimate the proportion of residents in a county that are in favor of a certain

law. We select a random sample of 100 residents and ask them about their stance on the law. Here are the results:

Sample size $n = 100$ Proportion in favor of law $p = 0.56$

The following code shows how to calculate a 95% confidence interval for the true proportion of residents in the entire county who are in favor of the law:

```
#input sample size and sample proportion
```

```
n <- 100
```

```
p <- .56
```

```
#calculate margin of error
```

```
margin <- qnorm(0.975)*sqrt(p*(1-p)/n)
```

```
#calculate lower and upper bounds of confidence interval
```

```
low <- p - margin
```

```
low
```

```
0.4627099
```

```
high <- p + margin
```

```
high
```

0.6572901

The 95% confidence interval for the true proportion of residents in the entire county who are in favor of the law is .

Example 4: Confidence Interval for a Difference in Proportions

We use the following formula to calculate a confidence interval for a difference in proportions:

Confidence interval = $(p_1 - p_2) \pm z \sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$

where:

**p_1, p_2 : sample 1 proportion, sample 2 proportion
 z : the z-critical value based on the confidence level
 n_1, n_2 : sample 1 size, sample 2 size**

Example: Suppose we want to estimate the difference in the proportion of residents who support a certain law in county A compared to the proportion who support the law in county B. Here is the summary data for each sample:

Sample 1:

$n_1 = 100$
 $p_1 = 0.62$ (i.e. 62 out of 100 residents support the law)

Sample 2:

$n_2 = 100$
 $p_2 = 0.46$ (i.e. 46 out of 100 residents support the law)

The following code shows how to calculate a 95% confidence interval for the true difference in proportion of residents who support the law between the counties:

```
#input sample sizes and sample proportions
```

```
n1 <- 100
```

```
p1 <- .62
```

```
n2 <- 100
```

```
p2 <- .46
```

```
#calculate margin of error
```

```
margin <- qnorm(0.975)*sqrt(p1*(1-p1)/n1 + p2*(1-p2)/n2)
```

```
#calculate lower and upper bounds of confidence interval
```

```
low <- (p1-p2) - margin
```

```
low
```

```
0.02364509
```

```
high <- (p1-p2) + margin
```

```
high
```

```
0.2963549
```

The 95% confidence interval for the true difference in proportion of residents who support the law between the counties is .

You can find more R tutorials [here](#).