How can I perform bootstrapping in R, and what are some examples of its implementation?

Authored by

stats writer

April 22, 2024

Bootstrapping is a statistical resampling technique used to estimate the variability and uncertainty of a population parameter by repeatedly sampling from a given dataset. In R, bootstrapping can be performed using the "boot" package, which allows for the creation of a large number of resampled datasets from the original data. These resampled datasets can then be used to calculate the desired statistic and generate a confidence interval for its true value. Some examples of bootstrapping in R include estimating the mean or median of a population, creating a confidence interval for a regression coefficient, or comparing the means of two groups. Bootstrapping is a useful tool in situations where traditional statistical methods may not be appropriate, such as when the underlying data is not normally distributed or when the sample size is small. It is a versatile and powerful technique that can provide valuable insights into the uncertainty of our data.

Perform Bootstrapping in R (With Examples)

Bootstrapping is a method that can be used to estimate the standard error of any statistic and produce a confidence interval for the statistic.

The basic process for bootstrapping is as follows:

Take *k* repeated samples with replacement from a given dataset.For each sample, calculate the statistic you're interested in.This results in *k* different estimates for a given statistic, which you can then use to calculate the standard error of the statistic and create a confidence interval for the statistic.

We can perform bootstrapping in R by using the following functions from the boot library:

1. Generate bootstrap samples.

boot(data, statistic, R, ...)

where:

data: A vector, matrix, or data framestatistic: A function that produces the statistic(s) to be bootstrappedR: Number of bootstrap replicates

2. Generate a bootstrapped confidence interval.

boot.ci(bootobject, conf, type)

where:

bootobject: An object returned by the boot() functionconf: The confidence interval to calculate. Default is 0.95type: Type of confidence interval to calculate. Options include "norm", "basic", "stud", "perc", "bca" and "all" - Default is "all"

The following examples show how to use these functions in practice.

Example 1: Bootstrap a Single Statistic

The following code shows how to calculate the standard error for the R-squared of a simple linear regression model:

```
set.seed(0)
library(boot)

#define function to calculate R-squared
rsq_function <- function(formula, data, indices) {
d <- data #allows boot to select sample
fit <- lm(formula, data=d) #fit regression modelreturn(summary(fit)$r.square) #return R-squared of model
}
#perform bootstrapping with 2000 replications
reps <- boot(data=mtcars, statistic=rsq_function, R=2000, formula=mpg~disp)

#view results of boostrapping
reps

ORDINARY NONPARAMETRIC BOOTSTRAP

Call:
boot(data = mtcars, statistic = rsq_function, R = 2000,
```

formula = mpg ~
disp)

Bootstrap Statistics :
original bias std. error
t1* 0.7183433 0.002164339 0.06513426

From the results we can see:

The estimated R-squared for this regression model is 0.7183433.The standard error for this estimate is 0.06513426.

We can quickly view the distribution of the bootstrapped samples as well:

plot(reps)