How can I use the fitdistr() function in R to fit distributions?

Authored by

stats writer

June 23, 2024

The fitdistr() function in R is a powerful tool that allows users to fit different statistical distributions to their data. This function takes in a dataset as input and uses maximum likelihood estimation to find the best parameters for the specified distribution. It also provides useful information such as the estimated parameters, standard errors, and confidence intervals. This allows users to compare different distributions and choose the one that best fits their data. Overall, the fitdistr() function is a useful tool for analyzing and understanding the underlying distribution of a dataset.

Use fitdistr() in R to Fit Distributions

You can use the fitdistr() function from the MASS package in R to estimate the parameters of a distribution by maximizing the likelihood function.

This function uses the following basic syntax:

fitdistr(x, densefun, ...)

where:

x: A numeric vector representing the values of the distributiondensefun: the distribution to estimate the parameters for

Note that the densefun argument accepts the following potential distribution names: beta, cauchy, chi-squared, exponential, gamma, geometric, lognormal, logistic, negative binomial, normal, Poisson, t and Weibull.

The following example shows how to use the fitdistr() function in practice.

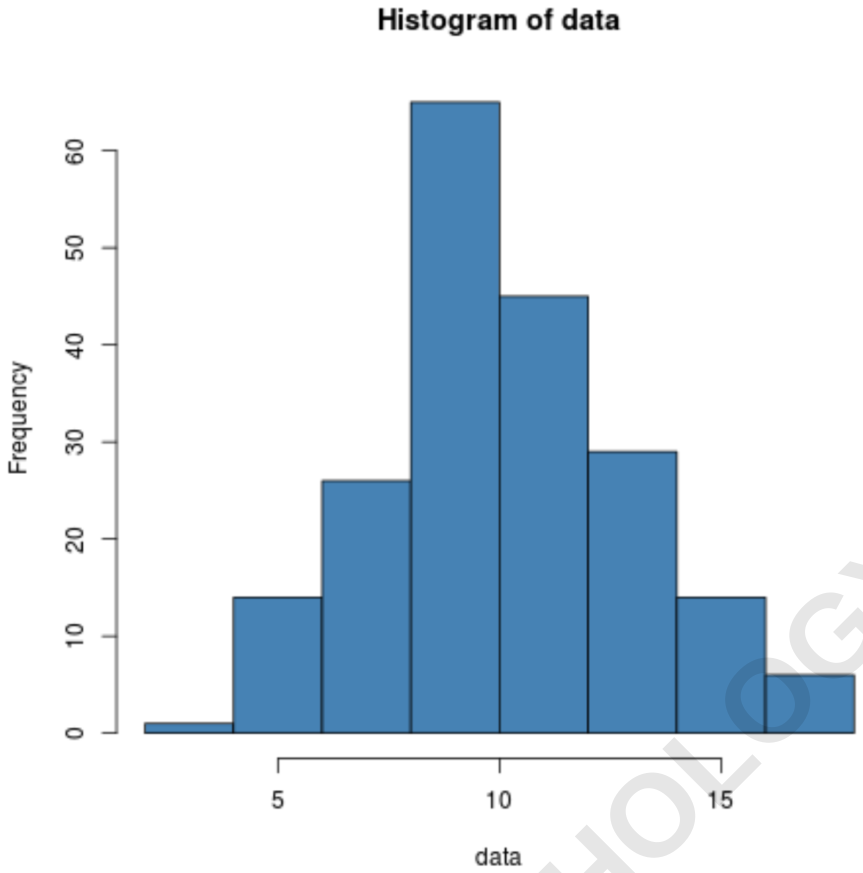
Example: How to Use fitdistr() Function to Fit Distributions in R

Suppose we use the rnorm() function in R to generate a vector of 200 values that follow a normal distribution:

```
#make this example reproducible
set.seed(1)

#generate sample of 200 observations that follows normal dist with mean=10 and sd=3
data <- rnorm(200, mean=10, sd=3)

#view first 6 observations in sample
head(data)

8.120639 10.550930 7.493114 14.785842 10.988523 7.538595
```

We can use the hist() function to create a histogram to visualize the distribution of data values:

```
hist(data, col='steelblue')
```



We can see that the data does indeed look normally distributed.

We can then use the fitdistr() function to estimate the parameters of this distribution:

library(MASS)

#estimate parameters of distribution
fitdistr(data, "normal")

mean sd

10.1066189 2.7803148
(0.1965979) (0.1390157)

The fitdistr() function estimates that the vector of values follows a with a mean of 10.1066189 and standard deviation of 2.7803148.

How to Plot a Normal Distribution in R

How to Perform a Shapiro-Wilk Test for Normality in R