

# How to Easily Perform a Cramer-Von Mises Test in R

Authored by  
**stats writer**

November 22, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Easily Perform a Cramer-Von Mises Test in R*.  
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=99636>

The Cramer-Von Mises test is a powerful non-parametric test widely utilized in statistics. Its primary function is to assess the goodness-of-fit, specifically determining whether an observed cumulative distribution function (CDF) differs significantly from a hypothesized theoretical distribution. While it can be adapted to compare two empirical samples, it is most commonly employed to test if a sample comes from a specified distribution, such as the normal distribution.

This test is particularly valued because it is sensitive across the entire range of the distribution, unlike some alternatives that focus solely on the tails or the center. The core idea involves calculating an integral that quantifies the squared difference between the empirical CDF of the sample data and the CDF of the theoretical distribution being tested. The result of this calculation is the test statistic, often denoted as  $\omega^2$  (omega squared).

In the R programming language, executing the Cramer-Von Mises test is straightforward using the dedicated `cvm.test()` function, which is available within the **gofest** package. This function requires the input data vector and the name of the cumulative distribution function (CDF) against which the data should be tested. Mastering this function is essential for researchers and analysts requiring rigorous verification of underlying distributional assumptions.

The **Cramer-Von Mises test** is used to determine whether or not a sample comes from a normal distribution.

## Theoretical Foundations: Normality and Goodness-of-Fit

Verifying that a dataset follows a normal distribution (often referred to as Gaussian distribution) is a critical preliminary step in numerous advanced statistical tests. This assumption of normality forms the basis for parametric methods such as regression, ANOVA, and t-tests. If the data significantly deviates from this distribution, the results derived from these parametric tests may be unreliable or invalid, potentially leading to incorrect conclusions.

The Cramer-Von Mises test provides a formal mechanism for assessing this distributional assumption. It operates under a null hypothesis ( $H_0$ ) which states that the sample data were drawn from the specified distribution (e.g., the normal distribution). The alternative hypothesis ( $H_a$ ) posits that the sample data does not follow that distribution. Like all goodness-of-fit tests, the calculated test statistic measures the discrepancy between the observed data and the null hypothesis model.

To implement this in R, we leverage the `cvm.test()` function found within the powerful **gofest** package. This package is designed specifically for high-quality goodness-of-fit testing. Before running the examples, it is necessary to ensure the package is installed and loaded into the current R session using the `library(gofest)` command.

## Prerequisites and Setup in R

Executing the Cramer-Von Mises test in R requires minimal preparation, primarily installing the correct package. The **goftest** package handles the computation and provides the necessary functionality. Once installed, the function syntax is straightforward: **cvm.test(x, distribution\_function)**, where 'x' is the numeric vector of data, and 'distribution\_function' is a character string specifying the CDF function of the hypothesized distribution.

For testing normality, the required argument is '**pnorm**', which represents the cumulative distribution function for the normal distribution in R. It is crucial to remember that the test assumes the parameters (mean and standard deviation) of the normal distribution are fixed based on the sample estimates, which is standard practice when testing for simple normality.

The following sections demonstrate two practical examples. The first example analyzes a dataset known to be normal, serving as a positive control. The second example tests a dataset generated from a drastically different distribution, illustrating how the test correctly identifies non-normality. We will ensure reproducibility for both examples by utilizing the **set.seed()** function.

### Example 1: Testing Data Generated from a Normal Distribution

To demonstrate the mechanics of the Cramer-Von Mises test, we begin by creating a sample of 100 random values that are explicitly drawn from a normal distribution using R's built-in **rnorm()** function. This simulation allows us to confirm that the test correctly concludes that the data is normally distributed when the assumption is met.

The code block below shows the initialization steps, data generation, and the execution of the **cvm.test()** function, setting the target distribution to the normal CDF ('pnorm').

```
library(goftest)
```

```
#make this example reproducible  
set.seed(0)
```

```
#create dataset of 100 random values generated from a normal distribution  
data <- rnorm(100)
```

```
#perform Cramer-Von Mises test for normality  
cvm.test(data, 'pnorm')
```

Cramer-von Mises test of goodness-of-fit

Null hypothesis: Normal distribution

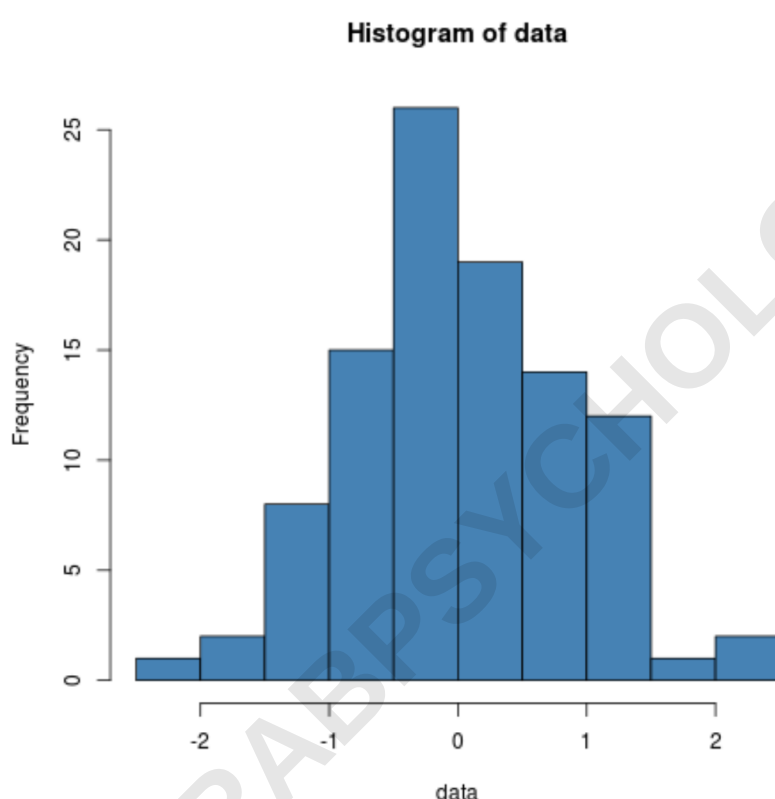
Parameters assumed to be fixed

```
data: data
```

```
omega2 = 0.078666, p-value = 0.7007
```

Following the formal test, it is highly recommended to visualize the data distribution using a histogram. Visualization provides an intuitive check and helps confirm the conclusions drawn from the numerical test results, especially in cases where the test is borderline. We use a simple **hist()** command for this purpose.

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



The resulting histogram clearly displays the characteristic **bell-shape**, with data concentrated symmetrically around a single central peak. This visual evidence strongly supports the assumption that the data is drawn from a normal distribution.

## Interpreting the Results of the Normality Test

The output of the **cvm.test()** provides two crucial values: the test statistic ( $\omega^2 = 0.078666$ ) and the corresponding p-value (0.7007). Interpretation of these results follows standard hypothesis testing procedures. We compare the calculated p-value to a predefined significance level ( $\alpha$ ), conventionally set at 0.05.

In this first example, the calculated p-value is **0.7007**. Since this value is significantly larger than our chosen significance level ( $\alpha=0.05$ ), we lack sufficient evidence to reject the null hypothesis ( $H_0$ ). Therefore, we conclude that there is no statistically significant difference between the empirical distribution of our sample data and the theoretical normal distribution.

This result is consistent with our data generation method, as we explicitly used the **rnorm()** function to create the dataset. A high p-value in a goodness-of-fit test indicates a close match between the observed data and the hypothesized distribution. (For more details on R's normal distribution functions, refer to: [A Guide to dnorm, pnorm, qnorm, and rnorm in R](#)).

## Example 2: Testing Data Generated from a Non-Normal Distribution (Poisson)

To demonstrate the Cramer-Von Mises test's ability to detect non-normality, we now generate a new dataset using the **rpois()** function, drawing 100 random values from a Poisson distribution with a lambda parameter of 3. The Poisson distribution is inherently discrete and often skewed, making it a clear example of data that should violate the normal distribution assumption.

The following R code performs the data generation and executes the **cvm.test()**, again specifying the null hypothesis as the normal distribution ('pnorm').

### library(goftest)

```
#make this example reproducible
set.seed(0)
```

```
#create dataset of 100 random values generated from a Poisson distribution
data <- rpois(n=100, lambda=3)
```

```
#perform Cramer-Von Mises test for normality
cvm.test(data, 'pnorm')
```

Cramer-von Mises test of goodness-of-fit

Null hypothesis: Normal distribution

Parameters assumed to be fixed

data: data

omega2 = 27.96, p-value < 2.2e-16

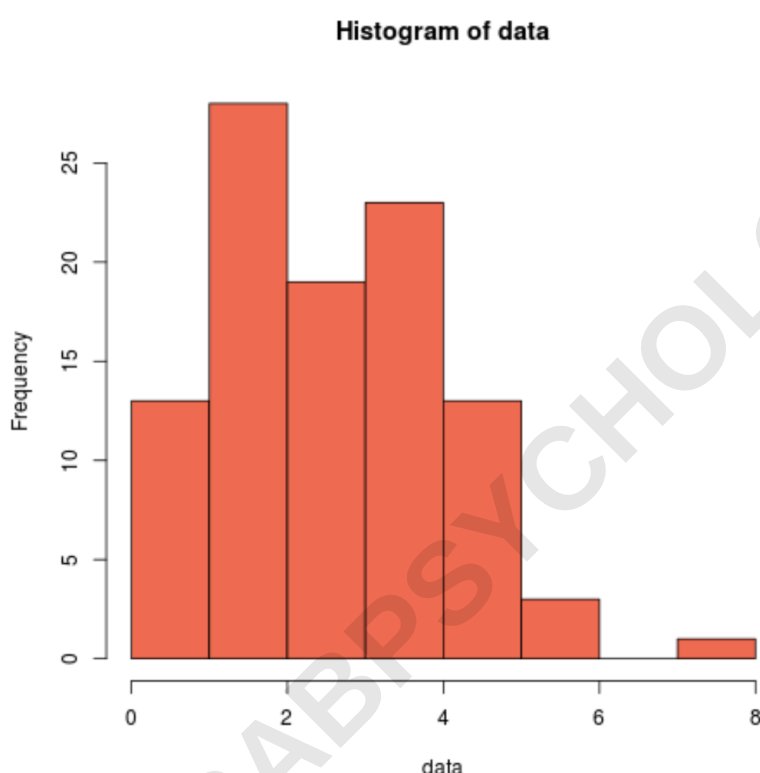
Upon reviewing the output, we observe a drastically different outcome. The calculated test statistic ( $\omega^2 = 27.96$ ) is much larger than in the previous example, and the resulting p-value is extremely small, indicated by the notation **< 2.2e-16**. Since this p-value is far less than the significance level ( $\alpha=0.05$ ), we have overwhelming evidence to reject the null hypothesis of

normality.

This result confirms that the sample data does **not** originate from a population that is normally distributed. This finding aligns perfectly with our knowledge that the data was generated using the **rpois()** function, which is designed to sample from the Poisson distribution. (For more details on R's Poisson distribution functions, refer to: A Guide to dpois, ppois, qpois, and rpois in R).

We visualize this non-normal data using a histogram to confirm the statistical finding:

```
hist(data, col='coral2')
```



As seen in the figure, the distribution is clearly skewed to the right, exhibiting a noticeable lack of symmetry and failing to display the characteristic bell-shape associated with a normal distribution. The histogram thus provides visual confirmation that the numerical results of the Cramer-Von Mises test are correct.

## Addressing Non-Normal Data Through Transformations

When the Cramer-Von Mises test conclusively demonstrates that a dataset is not normally distributed, researchers often face a decision point: either switch to robust non-parametric test methods that do not rely on distributional assumptions, or attempt to transform the existing data to

approximate normality.

Data transformations are mathematical functions applied uniformly to the entire variable, often successfully normalizing skewed data or stabilizing variance, thereby allowing the continued use of powerful parametric statistical tests. The choice of transformation depends heavily on the type and severity of non-normality observed (e.g., positive skew, negative skew, or heavy tails).

Common transformation techniques used to achieve better adherence to the normal distribution include:

**Log Transformation:** Applied by transforming the response variable  $y$  to  $\mathbf{\log(y)}$ . This is highly effective for data that exhibits positive skewness, often encountered in financial or biological variables.

**Square Root Transformation:** Used when the variance of the data is proportional to the mean. It transforms the response variable  $y$  to  $\mathbf{\sqrt{y}}$ . This is a milder transformation than the log transformation.

**Cube Root Transformation:** A relatively mild transformation that can be applied to both positive and negative values, transforming the response variable  $y$  to  $\mathbf{y^{1/3}}$ . This is useful for reducing moderate skewness.

After applying any of these transformations, the researcher should re-run the **cvm.test()** on the transformed data to confirm whether the process has successfully moved the variable closer to normality. If the transformed data satisfies the normality assumption (p-value > 0.05), parametric tests can proceed. (For guidance on implementing these techniques in R, refer to this tutorial to see how to perform these transformations in practice.)

## Alternative Normality Assessment Methods

While the Cramer-Von Mises test is robust, it is only one of several powerful goodness-of-fit tests available to determine adherence to the normal distribution assumption. Depending on the size of the dataset and the specific nature of the deviations expected, other tests might offer complementary or more sensitive results.

The **Shapiro-Wilk test** and the **Anderson-Darling test** are two popular alternatives. The Anderson-Darling test, like Cramer-Von Mises, is based on the empirical distribution function but puts greater weight on the tails of the distribution, making it particularly sensitive to deviations in the extreme values of the data. Often, researchers will perform multiple tests and utilize graphical methods, such as QQ plots, to build a comprehensive case for or against normality.

Understanding and applying a variety of normality tests ensures the highest degree of rigor in

statistical analysis. The following resource provides instruction on performing another key normality test in R:

The following tutorials explain how to perform other normality tests in R:

[How to Conduct an Anderson-Darling Test in R](#)

ARABPSYCHOLOGY.COM