

Learn How to Perform the Cramer-Von Mises Test in R with Examples

Authored by
Mohammed looti

May 2, 2026

RECOMMENDED CITATION

Mohammed looti (2026). *Learn How to Perform the Cramer-Von Mises Test in R with Examples*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=3541>

The **Cramer-Von Mises test** is a powerful and widely respected [statistical test](#) used primarily to determine whether an observed sample of data deviates significantly from a specified theoretical [cumulative distribution function](#) (CDF). Most frequently, this test is applied as a [goodness-of-fit test](#) to assess the critical assumption of the [normal distribution](#). By quantifying the discrepancy between the empirical distribution function (EDF) of the sample and the theoretical CDF, the test provides a robust measure of fit, making it indispensable in modern data analysis.

Establishing whether a dataset adheres to the [normal distribution](#) is a foundational step in countless statistical procedures. Many advanced analytical techniques, including [regression analysis](#), [ANOVA](#), and [t-tests](#), rely on the assumption that the data (or the error terms) are normally distributed. Failure to meet this requirement can severely compromise the reliability of your results, leading to inflated Type I errors, inaccurate confidence intervals, and ultimately, flawed conclusions. Therefore, utilizing a reliable diagnostic tool like the [Cramer-Von Mises test](#) ensures the statistical integrity of your research.

In the context of the [R programming language](#), executing the Cramer-Von Mises test is streamlined through specialized, community-maintained packages. The specific function required is `cvm.test()`, which is conveniently housed within the [gofest package](#). This function provides a highly efficient and accurate method for conducting this goodness-of-fit assessment against a specified theoretical distribution, making it an essential component in the toolkit of any statistician or data scientist working in R.

The remainder of this guide will provide detailed, practical examples illustrating the application of the `cvm.test()` function in [R](#). We will systematically explore two distinct scenarios: one involving data intentionally drawn from a [normal distribution](#), and another utilizing data that clearly deviates from normality. By analyzing the resulting [p-values](#) and accompanying visual evidence via [histograms](#), readers will gain a comprehensive understanding of how to interpret and effectively apply the test to their own datasets.

Example 1: Analyzing Data Drawn from a Normal Distribution

To establish a foundational understanding of the Cramer-Von Mises test output, we begin with an ideal case: a dataset explicitly generated to follow a normal distribution. This controlled example serves as a crucial baseline, demonstrating how the test correctly accepts the [null hypothesis](#) when the data perfectly aligns with the required distribution. We will generate a synthetic sample and then proceed to apply the `cvm.test()` function for assessment.

The following R programming language snippet details the necessary steps. We first ensure reproducibility using `set.seed(0)`, and then generate 100 random values using the [rnorm\(\) function](#). Since this function is specifically designed to produce random variates from a normal distribution, we expect the subsequent test to confirm normality. The `'pnorm'` argument in

cvm.test() specifies that we are testing the observed data against the theoretical normal CDF.

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

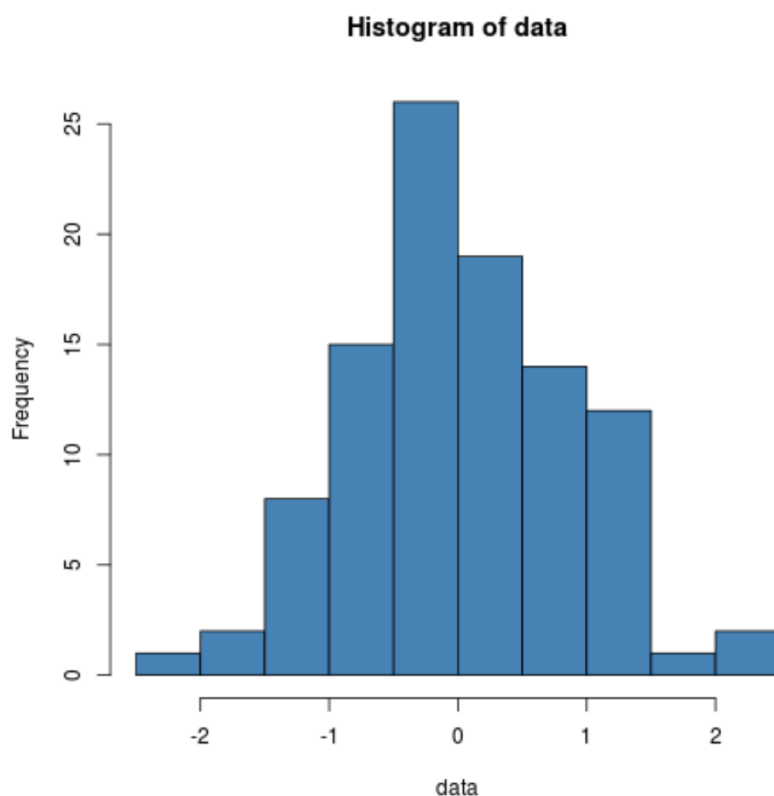
Upon execution, the output provides the calculated test statistic (omega2) and the corresponding [p-value](#), which is approximately **0.7007** in this instance. The core interpretation revolves around comparing this [p-value](#) to a chosen significance level (alpha), typically set at 0.05. The underlying assumption, the [null hypothesis](#), asserts that the data indeed follows a normal distribution.

Since the calculated [p-value](#) (0.7007) is significantly larger than the alpha level (0.05), we do not possess sufficient statistical evidence to reject the [null hypothesis](#). The conclusion drawn is that the sample data is statistically consistent with being sampled from a normally distributed population. This result validates the functionality of the test when presented with data that perfectly meets the assumed distributional shape.

Related: [A Guide to dnorm, pnorm, qnorm, and rnorm in R](#)

While quantitative tests provide statistical rigor, visual validation is essential for comprehensive data analysis. By generating a [histogram](#) of the sample data, we can visually confirm the distribution's shape, offering an intuitive complement to the numerical result obtained from the Cramer-Von Mises test.

hist(data, col='steelblue')



The visual assessment of the [histogram](#) clearly displays a characteristic bell-shaped curve that is symmetric around a central peak. This visual evidence perfectly aligns with the definition of a normal distribution. The graphical representation strongly supports the non-rejection of the null hypothesis, reinforcing the conclusion that the sample originated from a normal population.

Example 2: Analyzing Data That Deviates from Normality

Having successfully confirmed normality in the first example, we now shift our focus to a scenario where the data intentionally violates the normality assumption. This second example is critical for demonstrating the sensitivity and discriminative power of the Cramer-Von Mises test when identifying significant departures from the normal distribution. We will generate a synthetic dataset from a known non-normal distribution and observe how the test reacts.

To create a distinctly non-normal sample, the following R code generates 100 random values using the [rpois\(\) function](#), thereby drawing the data specifically from a [Poisson distribution](#) (which models count data and is often skewed). Since the [Poisson distribution](#) is inherently different from the normal distribution, this setup guarantees a significant result, allowing us to accurately interpret the test's rejection of the null hypothesis.

```
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
 $\omega^2 = 27.96$, p-value < $2.2e-16$

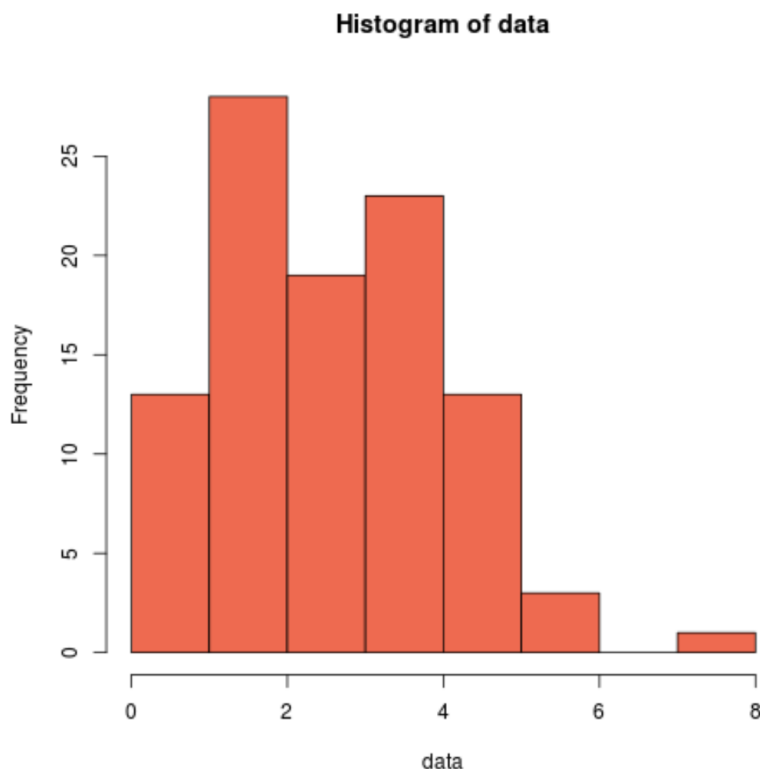
Analyzing the output of **cvm.test()** for this [Poisson-distributed](#) sample reveals a dramatically different outcome compared to the first example. The reported [p-value](#) is extremely small--less than $2.2e-16$. This minuscule value indicates a profound statistical disagreement between the observed data distribution and the theoretical normal distribution specified in the [null hypothesis](#).

Because this p-value is effectively zero, and thus substantially below the conventional significance threshold of 0.05, we have overwhelming statistical evidence to reject the [null hypothesis](#). We must conclude that the sample data does *not* originate from a population that is normally distributed. This finding is entirely expected, given that we deliberately generated the sample using the [rpois\(\)](#) [function](#) to capture the characteristics of a Poisson distribution.

Related: [A Guide to dpois, ppois, qpois, and rpois in R](#)

To visually solidify this conclusion, we generate a [histogram](#). Visual confirmation is a crucial step that helps analysts intuitively understand the magnitude and nature of the data's deviation from normality, providing context for the significant p-value obtained from the Cramer-Von Mises test.

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



Upon inspecting the [histogram](#), it is immediately apparent that the distribution is significantly skewed and does not exhibit the characteristic "bell-shape" associated with a normal distribution. Instead, we observe a shape typical of a Poisson distribution, often characterized by a peak at lower values and a tail extending towards higher values. This visual evidence perfectly corroborates the results of the Cramer-Von Mises test, unequivocally confirming that our sample data does not originate from a normal population.

Strategies for Addressing Non-Normal Data

A finding of non-normality, confirmed by tests like the Cramer-Von Mises criterion, is not an endpoint but rather a call for careful methodological adjustment. When a dataset violates the assumptions of normality, simply proceeding with parametric tests designed for normal data can undermine the validity and power of the entire analysis. Therefore, researchers must employ strategies to either transform the data or utilize appropriate non-parametric alternatives.

One primary strategy involves applying mathematical transformations to the response variable. The goal of these transformations is multi-faceted: they can stabilize variance, improve linearity in regression models, and, most importantly, reshape the distribution of the data to approximate normality more closely. Successful normalization allows the continued use of powerful parametric [statistical tests](#).

Below are common and highly effective transformations used to mitigate positive skewness and approach a normal distribution:

- 1. Log Transformation:** This technique converts the variable y to $\log(y)$. It is particularly useful for highly skewed positive data, as it compresses the range of larger values much more aggressively than smaller values, pulling the distribution closer to symmetry.
- 2. Square Root Transformation:** This involves transforming the variable y to $\sqrt{y}$. It offers a moderate solution for reducing right-skewness and is frequently recommended when analyzing count data, especially where the variance is related to the mean (such as in Poisson models).
- 3. Cube Root Transformation:** For data exhibiting severe skewness where the log transformation might be too strong, applying the cube root, transforming y to $y^{1/3}$, provides an intermediate level of compression. This transformation is highly versatile for positive skewed distributions.

By judiciously selecting and applying an appropriate transformation, analysts can significantly adjust the shape of the variable's distribution toward the desired normality. It is crucial to remember, however, that the final interpretation of any statistical model must be carefully conducted within the context of the transformed scale, ensuring that conclusions remain meaningful and relevant to the original phenomenon being studied.

For detailed instructions and practical code examples on how to implement these data manipulations within the [R programming environment](#), readers are encouraged to consult [this comprehensive tutorial](#), which provides the necessary guidance for applying these techniques effectively to real-world datasets.

Further Reading and Related Goodness-of-Fit Tests

The Cramer-Von Mises test is just one of several powerful goodness-of-fit tools available to statisticians. Expanding your knowledge of alternative tests is essential for developing a robust analytical strategy, as different tests may be more sensitive to specific types of distributional departures (e.g., departures in the tails versus the center).

To enhance your expertise in normality testing and related statistical procedures within R, we suggest exploring the following supplementary resources:

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

Mastering various normality tests and data transformation methodologies is fundamental to conducting rigorous and robust statistical analysis, enabling researchers to confidently derive accurate and defensible conclusions from their data.