

How can the Empirical Rule be applied in R?

Authored by
stats writer

May 12, 2024

RECOMMENDED CITATION

stats writer (2024). *How can the Empirical Rule be applied in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=143872>

The Empirical Rule is a statistical principle that can be applied in R, a programming language and environment used for statistical analysis and data visualization. This rule states that for a normal distribution, approximately 68% of the data falls within one standard deviation of the mean, 95% within two standard deviations, and 99.7% within three standard deviations. In R, this principle can be applied by using functions such as `mean()`, `sd()`, and `pnorm()` to calculate the mean, standard deviation, and the percentage of data within a certain range. This can be useful for understanding the distribution of data and identifying any outliers. Additionally, the Empirical Rule can be visualized in R using various graphical tools, such as histograms and box plots, to provide a visual representation of the data distribution. Overall, applying the Empirical Rule in R can help in making informed decisions and drawing meaningful insights from data analysis.

Apply the Empirical Rule in R

The Empirical Rule, sometimes called the 68-95-99.7 rule, states that for a given dataset with a normal distribution:

68% of data values fall within one standard deviation of the mean. 95% of data values fall within two standard deviations of the mean. 99.7% of data values fall within three standard deviations of the mean.

In this tutorial, we explain how to apply the Empirical Rule in R to a given dataset.

Applying the Empirical Rule in R

The `pnorm()` function in R returns the value of the cumulative density function of the normal distribution.

This function uses the following basic syntax:

`pnorm(q, mean, sd)`

where:

**q: normally distributed random variable value
mean: mean of distribution
sd: standard deviation of distribution**

We can use the following syntax to find the area under the normal distribution curve that lies in between various standard deviations:

#find area under normal curve within 1 standard deviation of mean

`pnorm(1) - pnorm(-1)`

0.6826895

#find area under normal curve within 2 standard deviations of mean

`pnorm(2) - pnorm(-2)`

0.9544997

#find area under normal curve within 3 standard

deviations of mean

`pnorm(3) - pnorm(-3)`

0.9973002

From the output we can confirm:

68% of data values fall within one standard deviation of the mean. 95% of data values fall within two standard deviations of the mean. 99.7% of data values fall within three standard deviations of the mean.

The following examples show how to use the Empirical Rule with different datasets in practice.

Example 1: Applying the Empirical Rule to a Dataset in R

Suppose we have a normally distributed dataset with a mean of 7 and a standard deviation of 2.2.

We can use the following code to find which values contain 68%, 95%, and 99.7% of the data:

`#define mean and standard deviation values`

`mean=7sd=2.2`

`#find which values contain 68% of data`

mean-2.2; mean+2.2

4.8

9.2

#find which values contain 95% of data

mean-2*2.2; mean+2*2.2

2.6

11.4

#find which values contain 99.7% of data

mean-3*2.2; mean+3*2.2

0.4

13.6

From this output, we can see:

68% of the data falls between 4.8 and 9.2
95% of the data falls between 2.6 and 11.4
99.7% of the data falls between 0.4 and 13.6

Example 2: Finding What Percentage of Data Falls Between Certain Values

Imagine we have a normally distributed dataset with a mean of 100 and standard deviation of 5.

Suppose we want to know what percentage of the data falls between the values 99 and 105 in this distribution.

We can use the `pnorm()` function to find the answer:

```
#find area under normal curve between 99 and 105  
pnorm(105, mean=100, sd=5) - pnorm(99, mean=100,  
sd=5)
```

0.4206045

We see that 42.06% of the data falls between the values 99 and 105 for this distribution.