How can I calculate the weighted standard deviation in R?

Authored by

stats writer

April 28, 2024

Weighted standard deviation is a statistical measure used to determine the spread of a dataset, taking into account the relative importance or weight of each data point. In R, this can be calculated by first assigning weights to each data point and then using the built-in function "weighted.sd". This function takes two arguments - the dataset and the corresponding weights - and returns the calculated weighted standard deviation. It is a useful tool for analyzing data that has varying levels of importance or significance, and can provide more accurate results compared to the traditional standard deviation calculation.

Calculate Weighted Standard Deviation in R

The weighted standard deviation is a useful way to measure of values in a dataset when some values in the dataset have higher weights than others.

The formula to calculate a weighted standard deviation is:

$$\sqrt{\frac{\sum_{i=1}^{N} w_i (x_i - \bar{x}^*)^2}{\frac{(M-1)}{M} \sum_{i=1}^{N} w_i}},$$

where:

N: The total number of M: The number of non-zero weightswi: A vector of weightsxi: A vector of data valuesx: The weighted mean

The easiest way to calculate a weighted standard

deviation in R is to use the wt.var() function from the Hmisc package, which uses the following syntax:

```
#define data values
x <- c(4, 7, 12, 13, ...)

#define weights
wt <- c(.5, 1, 2, 2, ...)

#calculate weighted variance
weighted_var <- wtd.var(x, wt)

#calculate weighted standard deviation
weighted_sd <- sqrt(weighted_var)
```

The following examples show how to use this function in practice.

Example 1: Weighted Standard Deviation for One Vector

The following code shows how to calculate the weighted standard deviation for a single vector in R:

```
library(Hmisc)

#define data values
```

```
x <- c(14, 19, 22, 25, 29, 31, 31, 38, 40, 41)

#define weights
wt <- c(1, 1, 1.5, 2, 2, 1.5, 1, 2, 3, 2)

#calculate weighted variance
weighted_var <- wtd.var(x, wt)

#calculate weighted standard deviation
sqrt(weighted_var)

8.570051
```

The weighted standard deviation turns out to be 8.57.

Example 2: Weighted Standard Deviation for One Column of Data Frame

The following code shows how to calculate the weighted standard deviation for one column of a data frame in R:

```
library(Hmisc)

#define data frame
df <- data.frame(team=c('A', 'A', 'A', 'A', 'A', 'B', 'B', 'C'),
wins=c(2, 9, 11, 12, 15, 17, 18, 19),
```

```
points=c(1, 2, 2, 2, 3, 3, 3, 3))

#define weights
wt <- c(1, 1, 1.5, 2, 2, 1.5, 1, 2)

#calculate weighted standard deviation of points
sqrt(wtd.var(df$points, wt))

0.6727938
```

The weighted standard deviation for the points column turns out to be 0.673.

Example 3: Weighted Standard Deviation for Multiple Columns of Data Frame

```
library(Hmisc)

#define data frame
df <- data.frame(team=c('A', 'A', 'A', 'A', 'A', 'B', 'B', 'C'),
wins=c(2, 9, 11, 12, 15, 17, 18, 19),
points=c(1, 2, 2, 2, 3, 3, 3, 3))

#define weights
wt <- c(1, 1, 1.5, 2, 2, 1.5, 1, 2)

#calculate weighted standard deviation of points and
wins
```

```
sapply(df, function(x) sqrt(wtd.var(x, wt)))

wins points
4.9535723 0.6727938
```

The weighted standard deviation for the wins column is 4.954 and the weighted standard deviation for the points column is 0.673.