

How can I calculate the pooled standard deviation in R?

Authored by
stats writer

May 5, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the pooled standard deviation in R?*

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=143114>

The pooled standard deviation in R is a statistical measure used to calculate the overall variability of a set of data points from different groups. It is useful in comparing the variability between groups and determining if there is a significant difference in the means of these groups. To calculate the pooled standard deviation in R, one must first obtain the standard deviation for each group, then combine them using a weighted average based on the sample sizes of each group. This can be done using the "pooled.sd" function in R, which takes in the standard deviations and sample sizes of each group as inputs. The output of this function is the pooled standard deviation, which can be interpreted as the average amount of deviation from the mean across all groups.

Calculate Pooled Standard Deviation in R

A pooled standard deviation is simply a weighted average of standard deviations from two or more independent groups.

In statistics it appears most often in the two sample t-test, which is used to test whether or not the means of two populations are equal.

The formula to calculate a pooled standard deviation for two groups is as follows:

$$\text{Pooled standard deviation} = \sqrt{(n_1-1)s_1^2 + (n_2-1)s_2^2 / (n_1+n_2-2)}$$

where:

n_1 , n_2 : Sample size for group 1 and group 2, respectively. s_1 , s_2 : Standard deviation for group 1 and

group 2, respectively.

The following examples show two methods for calculating a pooled standard deviation between two groups in R.

Method 1: Calculate Pooled Standard Deviation Manually

Suppose we have the following data values for two samples:

Sample 1: 6, 6, 7, 8, 8, 10, 11, 13, 15, 15, 16, 17, 19, 19, 21
Sample 2: 10, 11, 13, 13, 15, 17, 17, 19, 20, 22, 24, 25, 27, 29, 29

The following code shows how to calculate the pooled standard deviation between these two samples:

```
#define two samples  
data1 <- c(6, 6, 7, 8, 8, 10, 11, 13, 15, 15, 16, 17, 19, 19,  
21)  
data2 <- c(10, 11, 13, 13, 15, 17, 17, 19, 20, 22, 24, 25, 27,  
29, 29)  
  
#find sample standard deviation of each sample  
s1 <- sd(data1)
```

```
s2 <- sd(data2)
```

```
#find sample size of each sample
```

```
n1 <- length(data1)
```

```
n2 <- length(data2)
```

```
#calculate pooled standard deviation
```

```
pooled <- sqrt(((n1-1)*s1^2 + (n2-1)*s2^2) / (n1+n2-2))
```

```
#view pooled standard deviation
```

```
pooled
```

```
5.789564
```

The pooled standard deviation turns out to be 5.789564.

Method 2: Calculate Pooled Standard Deviation Using a Package

Another way to calculate the pooled standard deviation between two samples in R is to use the `sd_pooled()` function from the `effectsize` package.

The following code shows how to use this function in practice:

```
library(effectsize)
```

```
#define two samples
```

```
data1 <- c(6, 6, 7, 8, 8, 10, 11, 13, 15, 15, 16, 17, 19, 19,  
21)
```

```
data2 <- c(10, 11, 13, 13, 15, 17, 17, 19, 20, 22, 24, 25, 27,  
29, 29)
```

```
#calculate pooled standard deviation between two  
samples
```

```
sd_pooled(data1, data2)
```

```
5.789564
```

Note that this matches the value that we calculated manually in the previous example.

The following tutorials provide more information on calculating a pooled standard deviation: