

This website uses cookies to ensure you get the best experience on our website.

[Ok](#) [Learn more](#)

STHDA

Statistical tools for high-throughput data analysis

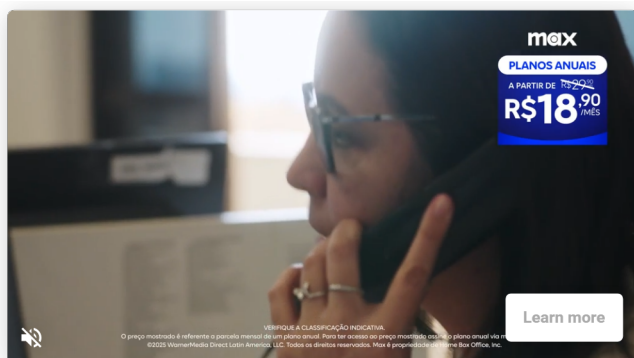
Licence: 

Search... 

[Home](#) [Basics](#) [Data](#) [Visualize](#) [Analyze](#) [Resources](#) [Our Products](#) [Support](#) [About](#)

[Home](#) / [Easy Guides](#) / [R software](#) / [R Basic Statistics](#) / [Comparing Variances in R](#) / [Compare Multiple Sample Variances in R](#)

 Actions menu for module Wiki



Compare Multiple Sample Variances in R

 Tools

- [Statistical tests for comparing variances](#)
- [Statistical hypotheses](#)
- [Import and check your data into R](#)
- [Compute Bartlett's test in R](#)
- [Compute Levene's test in R](#)
- [Compute Fligner-Killeen test in R](#)
- [Infos](#)

This article describes **statistical tests** for comparing the **variances** of two or more samples. Equal variances across samples is called **homogeneity of variances**.

Some statistical tests, such as [two independent samples T-test](#) and [ANOVA test](#), assume that variances are equal across groups. The **Bartlett's test**, **Levene's test** or **Fligner-Killeen's test** can be used to verify that assumption.



Statistical tests for comparing variances

There are many solutions to test for the equality (**homogeneity**) of variance across groups, including:

- ✓ **F-test:** Compare the variances of two samples. The data must be normally distributed.

[Learn more](#)

 The **F-test** has been described in our previous article: [F-test to compare equality of two variances](#). In the present article, we'll describe the tests for comparing more than two variances.

Statistical hypotheses

For all these tests (**Bartlett's test**, **Levene's test** or **Fligner-Killeen's test**),

- the null hypothesis is that all populations variances are equal;
- the alternative hypothesis is that at least two of them differ.

Import and check your data into R

To import your data, use the following R code:

```
# If .txt tab file, use this
my_data <- read.delim(file.choose())
# Or, if .csv file, use this
my_data <- read.csv(file.choose())
```

Here, we'll use ToothGrowth and PlantGrowth data sets:

```
# Load the data
data(ToothGrowth)
data(PlantGrowth)
```

To have an idea of what the data look like, we start by displaying a random sample of 10 rows using the function **sample_n()** [in **dplyr** package]. First, install dplyr package if you don't have it: **install.packages("dplyr")**.

Show 10 random rows:

```
set.seed(123)
# Show PlantGrowth
dplyr::sample_n(PlantGrowth, 10)
```

```
weight group
24  5.50  trt2
12  4.17  trt1
25  5.37  trt2
26  5.29  trt2
2   5.58  ctrl
14  3.59  trt1
22  5.12  trt2
13  4.41  trt1
11  4.81  trt1
21  6.31  trt2
```

```
# PlantGrowth data structure
str(PlantGrowth)
```

```
'data.frame':  30 obs. of  2 variables:
 $ weight: num  4.17 5.58 5.18 6.11 4.5 4.61 5.17 4.53 5.33 5.14 ...
 $ group : Factor w/ 3 levels "ctrl","trt1",...: 1 1 1 1 1 1 1 1 1 1 ...
```

```
# Show ToothGrowth
dplyr::sample_n(ToothGrowth, 10)
```




```
14 17.3 VC 1.0
3 7.3 VC 0.5
18 14.5 VC 1.0
50 27.3 OJ 1.0
46 25.2 OJ 1.0
```

```
# ToothGrowth data structure
str(ToothGrowth)
```

```
'data.frame': 60 obs. of 3 variables:
 $ len : num 4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
 $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 ...
 $ dose: num 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

 Note that, R considers the column "dose" [in ToothGrowth data set] as a numeric vector. We want to convert it as a grouping variable (factor).

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

 We want to test the equality of variances between groups.

Compute Bartlett's test in R

Bartlett's test is used for testing homogeneity of variances in k samples, where k can be more than two. It's adapted for normally distributed data. The **Levene test**, described in the next section, is a more robust alternative to the Bartlett test when the distributions of the data are non-normal.

The R function **bartlett.test()** can be used to compute Bartlett's test. The simplified format is as follow:

```
bartlett.test(formula, data)
```

- **formula**: a formula of the form values ~ groups
- **data**: a matrix or data frame

The function returns a list containing the following component:


- **statistic**: Bartlett's K-squared test statistic
- **parameter**: the degrees of freedom of the approximate chi-squared distribution of the test statistic.
- **p.value**: the p-value of the test

To perform the test, we'll use the *PlantGrowth* data set, which contains the weight of plants obtained under 3 treatment groups.

- **Bartlett's test with one independent variable:**

```
res <- bartlett.test(weight ~ group, data = PlantGrowth)
res
```

```
Bartlett test of homogeneity of variances
data: weight by group
Bartlett's K-squared = 2.8786, df = 2, p-value = 0.2371
```

 From the output, it can be seen that the p-value of 0.2370968 is not less than the significance level of 0.05. This means that there is no evidence to suggest that the variance in plant growth is statistically significantly different for the three treatment groups.

```
Bartlett test of homogeneity of variances
data: len by interaction(supp, dose)
Bartlett's K-squared = 6.9273, df = 5, p-value = 0.2261
```

Compute Levene's test in R

✓ As mentioned above, Levene's test is an alternative to Bartlett's test when the data is not normally distributed.

The function `leveneTest()` [in `car` package] can be used.

```
library(car)
# Levene's test with one independent variable
leveneTest(weight ~ group, data = PlantGrowth)
```

```
Levene's Test for Homogeneity of Variance (center = median)
Df F value Pr(>F)
group 2 1.1192 0.3412
27
```

```
# Levene's test with multiple independent variables
leveneTest(len ~ supp*dose, data = ToothGrowth)
```

```
Levene's Test for Homogeneity of Variance (center = median)
Df F value Pr(>F)
group 5 1.7086 0.1484
54
```

Compute Fligner-Killeen test in R

The **Fligner-Killeen test** is one of the many tests for homogeneity of variances which is most robust against departures from normality.

The R function `fligner.test()` can be used to compute the test:

```
fligner.test(weight ~ group, data = PlantGrowth)
```

```
Fligner-Killeen test of homogeneity of variances
data: weight by group
Fligner-Killeen:med chi-squared = 2.3499, df = 2, p-value = 0.3088
```

Infos

⚠ This analysis has been performed using **R software** (ver. 3.2.4).

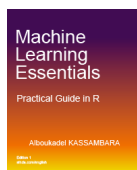
✓ Enjoyed this article? I'd be very grateful if you'd help it spread by emailing it to a friend, or sharing it on Twitter, Facebook or Linked In.

Show me some love with the like buttons below... Thank you and please don't forget to share and comment below!!





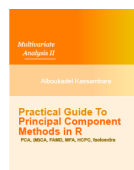
Recommended for You!



Machine Learning Essentials: Practical Guide in R



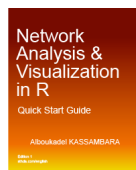
Practical Guide to Cluster Analysis in R



Practical Guide to Principal Component Methods in R



R Graphics Essentials for Great Data Visualization



Network Analysis and Visualization in R



More books on R and data science

✓ This section contains the best data science and self-development resources to help you on your path.

Books - Data Science

Our Books

- [Practical Guide to Cluster Analysis in R](#) by A. Kassambara (Datanovia)
- [Practical Guide To Principal Component Methods in R](#) by A. Kassambara (Datanovia)
- [Machine Learning Essentials: Practical Guide in R](#) by A. Kassambara (Datanovia)
- [R Graphics Essentials for Great Data Visualization](#) by A. Kassambara (Datanovia)
- [GGPlot2 Essentials for Great Data Visualization in R](#) by A. Kassambara (Datanovia)
- [Network Analysis and Visualization in R](#) by A. Kassambara (Datanovia)
- [Practical Statistics in R for Comparing Groups: Numerical Variables](#) by A. Kassambara (Datanovia)
- [Inter-Rater Reliability Essentials: Practical Guide in R](#) by A. Kassambara (Datanovia)

Others

- [R for Data Science: Import, Tidy, Transform, Visualize, and Model Data](#) by Hadley Wickham & Garrett Grolemond
- [Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems](#) by Aurelien Géron
- [Practical Statistics for Data Scientists: 50 Essential Concepts](#) by Peter Bruce & Andrew Bruce
- [Hands-On Programming with R: Write Your Own Functions And Simulations](#) by Garrett Grolemond & Hadley Wickham
- [An Introduction to Statistical Learning: with Applications in R](#) by Gareth James et al.
- [Deep Learning with R](#) by François Chollet & J.J. Allaire
- [Deep Learning with Python](#) by François Chollet

This page has been seen 256870 times

Sign in

Login

Login

Password

Password

Auto connect



Sign in



Register



Forgotten password

Welcome!

Want to Learn More on R Programming and Data Science?

STAY UPDATED

Enter your email

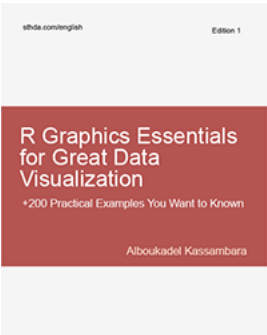
Subscribe

[Click to see our collection of resources to help you on your path...](#)



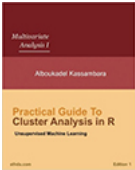
-  [factoextra](#)
-  [survminer](#)
-  [ggpubr](#)
-  [ggcorrplot](#)
-  [fastqcr](#)

Our Books

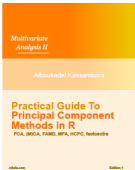


R Graphics Essentials for Great Data Visualization: 200 Practical Examples You Want to Know for Data Science
★ **NEW!!**

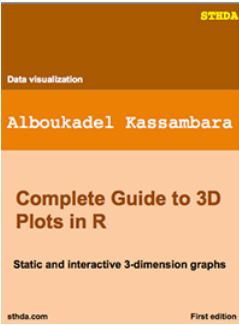
3
D
P
l
o
t
s
i
n
R



Practical Guide to Cluster Analysis in R



Practical Guide to Principal Component Methods in R



Newsletter

Email



Boosted by PHPBoost

