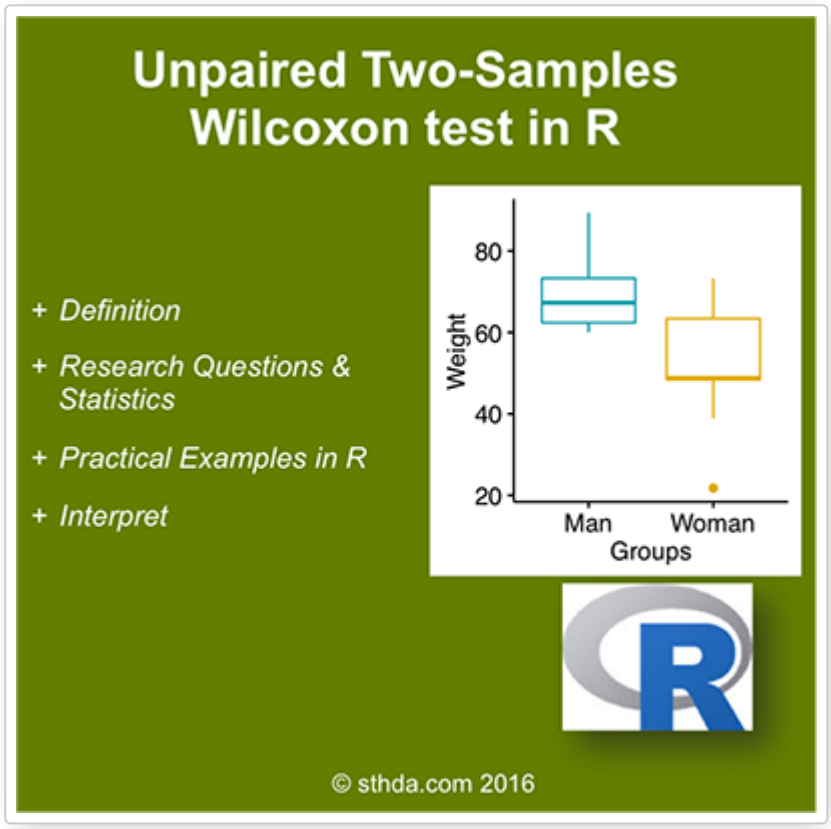
Unpaired Two-Samples Wilcoxon Test in R

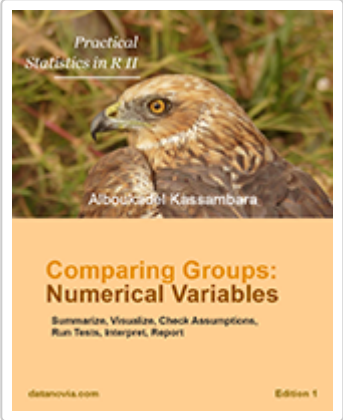
Tools

- Visualize your data and compute Wilcoxon test in R
  - R function to compute Wilcoxon test
  - Import your data into R
  - Check your data
  - Visualize your data using box plots
  - Compute unpaired two-samples Wilcoxon test
- Online unpaired two-samples Wilcoxon test calculator
- See also
- Infos

The **unpaired two-samples Wilcoxon test** (also known as **Wilcoxon rank sum test** or **Mann-Whitney** test) is a non-parametric alternative to the unpaired two-samples t-test, which can be used to compare two independent groups of samples. It's used when your data are not normally distributed.



Related Book:



Practical Statistics in R for Comparing Groups: Numerical Variables

This article describes how to compute two samples Wilcoxon test in R.

Visualize your data and compute Wilcoxon test in R

R function to compute Wilcoxon test

To perform two-samples Wilcoxon test comparing the means of two independent samples (x & y), the R function **wilcox.test**() can be used as follow:

```
wilcox.test(x, y, alternative = "two.sided")
```

- **x,y**: numeric vectors
- **alternative**: the alternative hypothesis. Allowed value is one of "two.sided" (default), "greater" or "less".

Import your data into R

1. **Prepare your data** as specified here: Best practices for preparing your data set for R
2. **Save your data** in an external .txt tab or .csv files
3. **Import your data into R** as follow:

```
# 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 an example data set, which contains the weight of 18 individuals (9 women and 9 men):

```
# Data in two numeric vectors
women_weight <- c(38.9, 61.2, 73.3, 21.8, 63.4, 64.6, 48.4, 48.8, 48.5)
men_weight <- c(67.8, 60, 63.4, 76, 89.4, 73.3, 67.3, 61.3, 62.4)
# Create a data frame
my_data <- data.frame(
                group = rep(c("Woman", "Man"), each = 9),
                weight = c(women_weight,  men_weight)
                )
```

? We want to know, if the median women's weight differs from the median men's weight?

Check your data

```
print(my_data)
```

```
   group weight
1  Woman   38.9
2  Woman   61.2
3  Woman   73.3
4  Woman   21.8
5  Woman   63.4
6  Woman   64.6
7  Woman   48.4
8  Woman   48.8
9  Woman   48.5
10   Man   67.8
11   Man   60.0
12   Man   63.4
13   Man   76.0
14   Man   89.4
15   Man   73.3
16   Man   67.3
17   Man   61.3
18   Man   62.4
```

It's possible to compute summary statistics (median and interquartile range (IQR)) by groups. The dplyr package can be used.

- To install **dplyr** package, type this:

```
install.packages("dplyr")
```

- Compute summary statistics by groups:

```
library(dplyr)
group_by(my_data, group) %>%
  summarise(
    count = n(),
    median = median(weight, na.rm = TRUE),
    IQR = IQR(weight, na.rm = TRUE)
  )
```

```
Source: local data frame [2 x 4]
   group count median   IQR
  (fctr) (int)  (dbl) (dbl)
1    Man     9   67.3  10.9
2  Woman     9   48.8  15.0
```

Visualize your data using box plots

You can draw R base graphs as described at this link: R base graphs. Here, we'll use the **ggpubr** R package for an easy ggplot2-based data visualization

- Install the latest version of ggpubr from GitHub as follow (recommended):

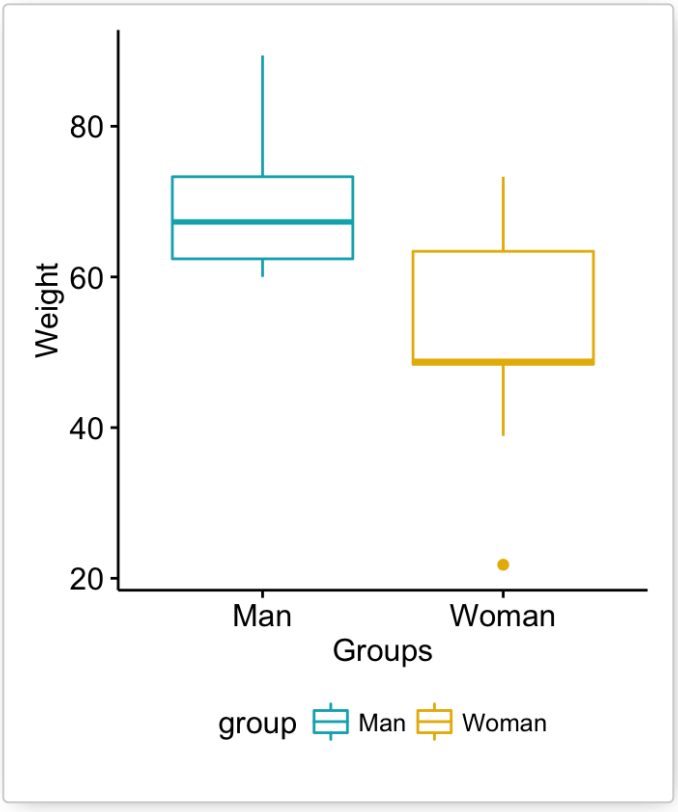
```
# Install
if(!require(devtools)) install.packages("devtools")
devtools::install_github("kassambara/ggpubr")
```

- Or, install from CRAN as follow:

```
install.packages("ggpubr")
```

- Visualize your data:

```
# Plot weight by group and color by group
library("ggpubr")
ggboxplot(my_data, x = "group", y = "weight",
          color = "group", palette = c("#00AFBB", "#E7B800"),
          ylab = "Weight", xlab = "Groups")
```



Compute unpaired two-samples Wilcoxon test

❓ Question : Is there any significant difference between women and men weights?

1) Compute two-samples Wilcoxon test - Method 1: The data are saved in two different numeric vectors.

```
res <- wilcox.test(women_weight, men_weight)
res
```

```
    Wilcoxon rank sum test with continuity correction
data:  women_weight and men_weight
W = 15, p-value = 0.02712
alternative hypothesis: true location shift is not equal to 0
```

⚠ It will give a warning message, saying that "cannot compute exact p-value with tie". It comes from the assumption of a Wilcoxon test that the responses are continuous. You can suppress this message by adding another argument **exact = FALSE**, but the result will be the same.

2) Compute two-samples Wilcoxon test - Method 2: The data are saved in a data frame.

```
res <- wilcox.test(weight ~ group, data = my_data,
                   exact = FALSE)
res
```

```
    Wilcoxon rank sum test with continuity correction
data:  weight by group
W = 66, p-value = 0.02712
alternative hypothesis: true location shift is not equal to 0
```

```
# Print the p-value only
res$p.value
```

```
[1] 0.02711657
```

❗ As you can see, the two methods give the same results.

✔ The **p-value** of the test is 0.02712, which is less than the significance level alpha = 0.05. We can conclude that men's median weight is significantly different from women's median weight with a **p-**

value = 0.02712.

Note that:

- if you want to test whether the median men's weight is less than the median women's weight, type this:

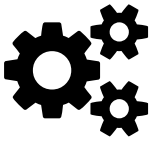
```
wilcox.test(weight ~ group, data = my_data,
        exact = FALSE, alternative = "less")
```

- Or, if you want to test whether the median men's weight is greater than the median women's weight, type this

```
wilcox.test(weight ~ group, data = my_data,
        exact = FALSE, alternative = "greater")
```

Online unpaired two-samples Wilcoxon test calculator

You can perform unpaired **two-samples Wilcoxon test**, **online**, without any installation by clicking the following link:



Online two-samples Wilcoxon test calculator

See also

- Compare one-sample mean to a standard known mean
  - One-Sample T-test (parametric)
  - One-Sample Wilcoxon Test (non-parametric)
- Compare the means of two independent groups
  - Unpaired Two Samples T-test (parametric)

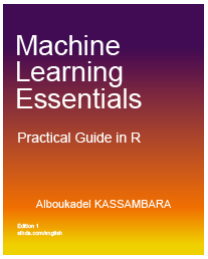
Infos

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

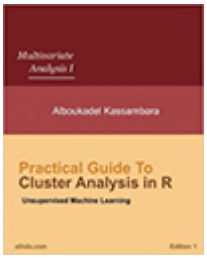
Enjoyed this article? I'd be very grateful if you'd help it spread by emailing it to a friend, or sharing it on Twitter, Facebook or Linked In.

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

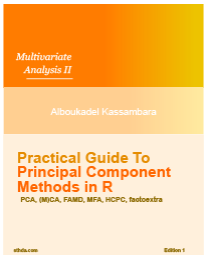
Recommended for You!



Machine Learning Essentials: Practical Guide in R

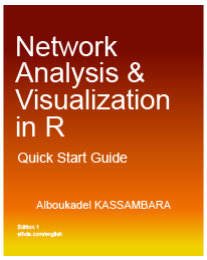


Practical Guide to Cluster Analysis in R



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

Recommended for you

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

Coursera - Online Courses and Specialization

Data science

- Course: Machine Learning: Master the Fundamentals by Standford
- Specialization: Data Science by Johns Hopkins University
- Specialization: Python for Everybody by University of Michigan
- Courses: Build Skills for a Top Job in any Industry by Coursera
- Specialization: Master Machine Learning Fundamentals by University of Washington
- Specialization: Statistics with R by Duke University
- Specialization: Software Development in R by Johns Hopkins University
- Specialization: Genomic Data Science by Johns Hopkins University

Popular Courses Launched in 2020

- Google IT Automation with Python by Google
- AI for Medicine by deeplearning.ai
- Epidemiology in Public Health Practice by Johns Hopkins University
- AWS Fundamentals by Amazon Web Services

Trending Courses

- The Science of Well-Being by Yale University
- Google IT Support Professional by Google
- Python for Everybody by University of Michigan
- IBM Data Science Professional Certificate by IBM
- Business Foundations by University of Pennsylvania
- Introduction to Psychology by Yale University
- Excel Skills for Business by Macquarie University
- Psychological First Aid by Johns Hopkins University
- Graphic Design by Cal Arts

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 Grolemund
- 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 Grolemund & 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

Want to Learn More on R Programming and Data Science?

Follow us by Email

Subscribe
by FeedBurner

On Social Networks:
on Social Networks

Get involved :

Click to **follow us** on Facebook and Google+ :

Comment this article by clicking on "Discussion" button (top-right position of this page)

Sign in

Login

Login

Password

Password

Auto connect

Sign in

Register

Forgotten password

Welcome!

Want to Learn More on R Programming and Data Science?

Follow us by Email

Subscribe

by FeedBurner

on Social Networks

Click to see our collection of resources to help you on your path...

Course & Specialization

Recommended for You (on Coursera):

- Course: Machine Learning: Master the Fundamentals
- Specialization: Data Science
- Specialization: Python for Everybody
- Course: Build Skills for a Top Job in any Industry
- Specialization: Master Machine Learning Fundamentals
- Specialization: Statistics with R
- Specialization: Software Development in R
- Specialization: Genomic Data Science

See More Resources

factoextra

survminer

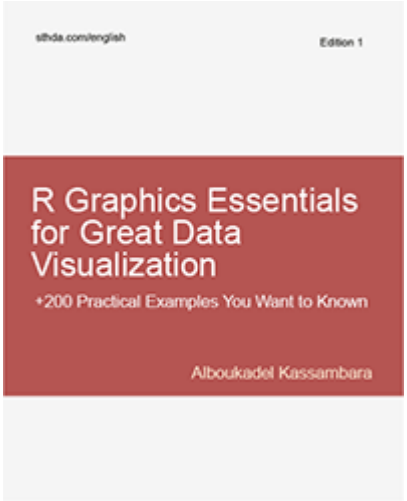
ggpubr

ggcorrplot

fastqcr

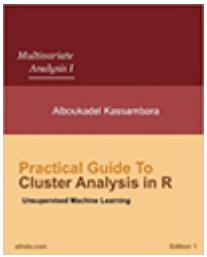
Our Books

3D Plots in R

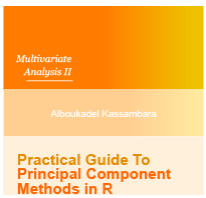


R Graphics Essentials for Great Data Visualization: 200 Practical Examples You Want to Know for Data Science

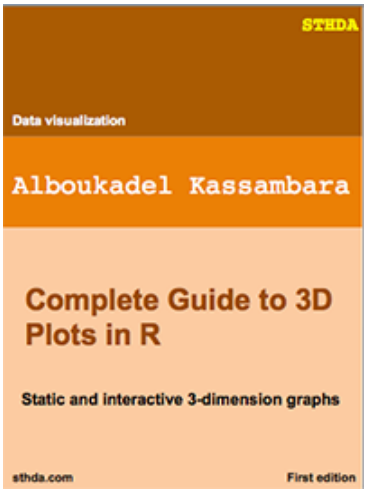
★ NEW!!



Practical Guide to Cluster Analysis in R



Practical Guide to Principal Component Methods in R



Datanovia: Online Data Science Courses

R-Bloggers

Newsletter Email

Recommended for you

ggplot2 legend : Easy steps to change the position and the appe...

www.sthda.com

ggplot2 scatter plots : Quick start guide - R software and data vis...

www.sthda.com

Reading Data From TXT|CSV Files: R Base Functions - Easy Guid...

www.sthda.com