Chapter 13
Wilcoxon Statistics in R: Signed-Rank Test and Rank-Sum Test

13.1 Introduction

The Wilcoxon test, which is of two types (i) *Wilcoxon Signed-rank* and (ii) *Wilcoxon Rank-sum* (Wilcoxon, 1945), is a non-parametric test and alternative version of the *t*-test (Rey & Neuhäuser, 2011). The test is mostly applied by the researchers to compare two samples by testing whether the median values of the data or variables differ significantly from each other. The resultant models assume that the data comes from two matched, or dependent populations, following the same distribution through time or place (Hayes, 2023). The test (Wilcoxon) can be applied to test the hypothesis that the *median* of a symmetrical distribution equals a given constant. And, as the name implies and as with the many other types of non-parametric tests that we have also previously covered in this book (see Chaps. 4, 6 and 12); this *distribution-free* test is based on ranks (Rey & Neuhäuser, 2011). It is expected that the independent variable in a Wilcoxon test is dichotomous, while the dependent variable is a continuous variable whose measurement is at least ordinal.

The main types and features or summary of the Wilcoxon test include (Hayes, 2023):

- The Wilcoxon test compares two paired or independent groups of variable and comes in two versions depending on the data groupings or scenario: (i) the rank-sum test and (ii) signed-rank test.
- The aim of the tests is to determine if two or more sets of pairs are different from one another in a statistically significant manner.
- Both tests (whether rank-sum or signed-rank) assume that the pairs in the data sample come from the same dependent populations.
- Unlike t-test that calculates the *mean difference* of two variables, the Wilcoxon test is used to calculate the *median difference* between two variables.

The "signed-ranked" version of the Wilcoxon test is calculated based on differences in the samples' median scores but in addition to it taking into account the signs

K. Okoye and S. Hosseini, *R Programming*,
https://doi.org/10.1007/978-981-97-3385-9_13

of the differences, thus, takes into consideration the magnitudes of the observed differences. As the non-parametric equivalent of the *paired t-test*, the signed-rank can be used as an alternative to the t-test when the population data does not follow a normal distribution.

On the other hand, the Wilcoxon "rank-sum" test is often used as the non-parametric version or alternative to the *independent* or *two-sample t-test*.

Thus, the Wilcoxon rank-sum test is used to compare two independent samples, while Wilcoxon signed-rank test is used to compare two related samples.

The value of z in a Wilcoxon test is calculated with the following formula:

$$Z_T = \frac{T - \mu_T}{\sigma_T}$$

where:

- T = the sum of values from calculating the ranges of differences in the sample.

In the next sections of this chapter (Sects. 13.2 and 13.3); the authors will be demonstrating to the readers how to conduct the two main types of the Wilcoxon test (Signed-rank and Rank-Sum) in R. We will explain and illustrate the different steps and functions that are used to perform the test (Wilcoxon) in R by following the outlined steps in Fig. 13.1.

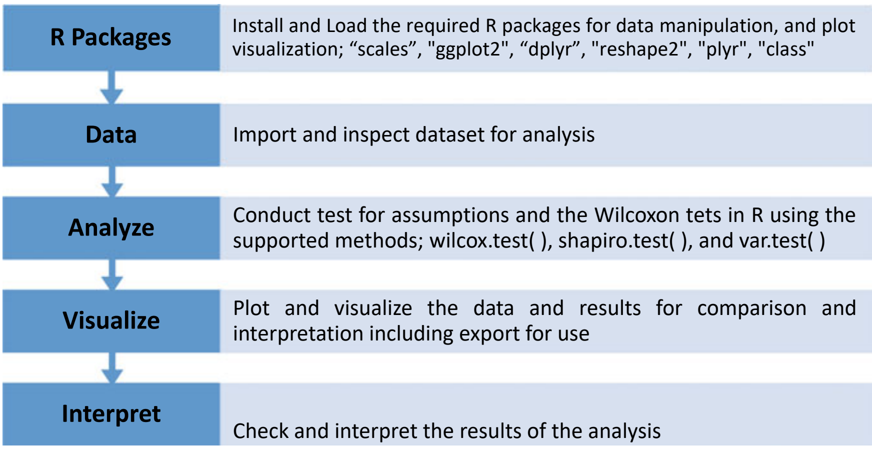

Fig. 13.1 Steps to conducting Wilcoxon tests in R

13.2 Signed-Rank Wilcoxon Test in R

Signed-Rank Wilcoxon test is used by the researchers to determine the median difference between two sets of data. It is used when the researchers or data analysts are interested in knowing the *difference in median* between two measures in a sample (e.g., pre and post tests, or before and after test).

By default, the hypothesis for testing whether there is a *difference in median* of the two (paired) data samples in Signed-Rank Wilcoxon tests is; *IF* the p-value is less than or equal to 0.05 ($p \leq 0.05$), *THEN* we assume that the median of the two sets of data or group of variables are statistically different and that this is not by chance (H_1), *ELSE IF* the p-value is greater than 0.05 ($p > 0.05$) *THEN* we can presume that there is no difference in the median of the two groups and any potential difference could only occur by chance (H_0).

The authors will demonstrate to the readers how to perform the Signed-Rank Wilcoxon tests used for two paired samples in R using the **wilcox.test( )**, **shapiro.test( )** and **var.test( )** functions.

As defined in the previous section (Sect. 13.1), we will do this using the steps outlined in Fig. 13.1.

To begin, **Open RStudio** and **Create a new or open an existing project**. Once the user has the RStudio and an R Project opened, **Create a new R Script** and name it "**Signed-Rank-Wilcoxon**" or any name the user may preferably choose (see Chaps. 1 and 2 if the user requires to refresh on how to do these steps).

Now, let's download an example data that we will use to demonstrate the two types of the Wilcoxon tests (Signed-Rank and Rank-Sum) in R. ***Note: the users are welcome to use any existing data or format they may wish to use for this illustration or analysis***. The example datasets the authors have used here are only for illustration purposes (users can see Chap. 2 for a step-by-step guide on how to work with different data types and format in R).

As shown in Fig. 13.2, download the example file named "**exam_grades.csv**" from the following source (https://www.openintro.org/data/) and save it on the local machine or computer. ***Note: the readers can also visit the following repository (https://doi.org/https://doi.org/10.6084/m9.figshare.24728073) where the authors have uploaded all the example files used in this book to directly access and download the file.

Once the user has downloaded the example file (exam_grades.csv) and saved this on the local machine or computer, we can proceed to conduct the first Wilcoxon test (*Signed-Rank Wilcoxon)* in R.

Step 1—Load the Required R Packages and Libraries

Install and **Load** the following *R packages* and *libraries* (see Fig. 13.3, Step1, Lines 3–20) that will be used to call and run the different R functions, data manipulations, and graphical visualizations for the Signed-Rank Wilcoxon analysis.

The code and syntax to install and load the required R packages are as follows (Fig. 13.3, Step 1, Lines 3–20):

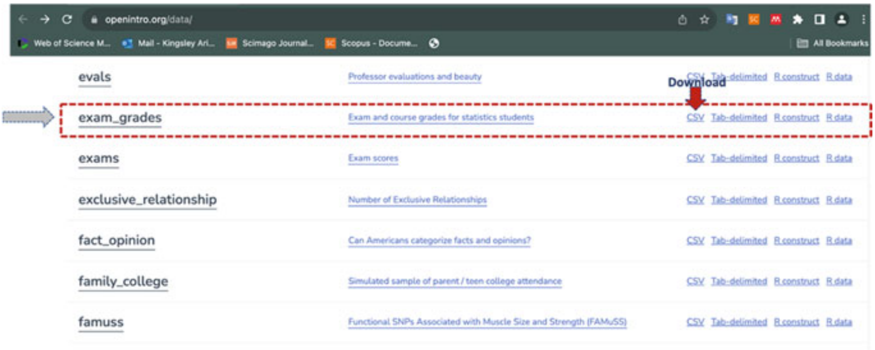

Fig. 13.2 Example of csv data download for Wilcoxon test. *Source* https://www.openintro.org/data/

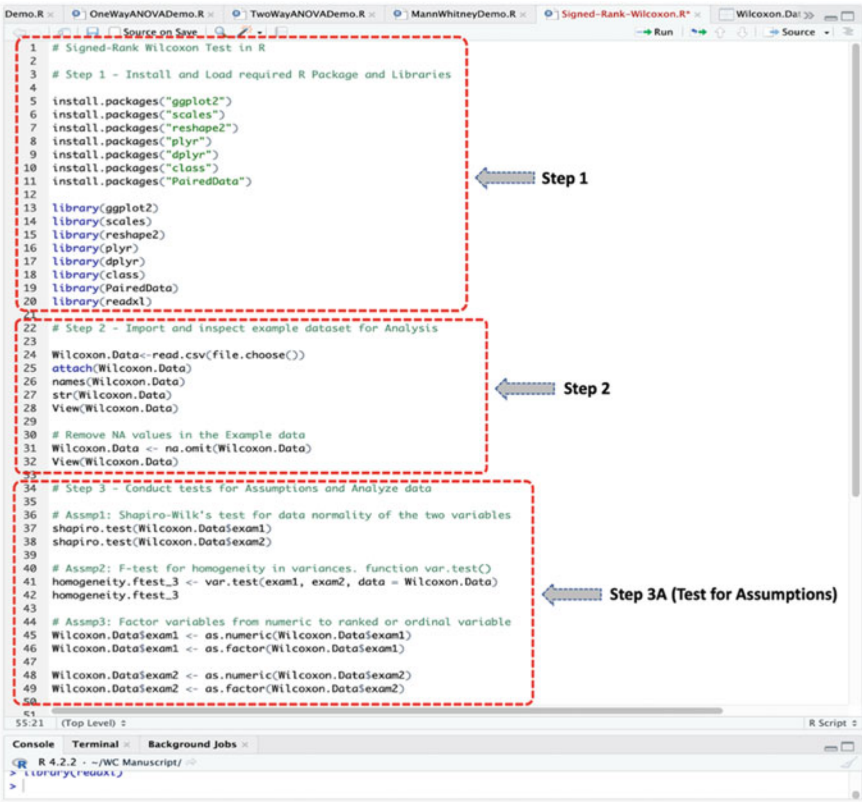

Fig. 13.3 Steps to conducting Wilcoxon test in R

```
install.packages("ggplot2")
install.packages("scales")
install.packages("reshape2")
install.packages("plyr")
install.packages("dplyr")
install.packages("class")
install.packages("PairedData")
```

```
library(ggplot2)
library(scales)
library(reshape2)
library(plyr)
library(dplyr)
library(class)
library(PairedData)
library(readxl)
```

 # Step 2—Import and Inspect the Example Dataset for Wilcoxon Analysis

As illustrated in Fig. 13.3 (Step 2, Lines 22–32), import the dataset named "**exam_grades.csv**" that we have downloaded earlier, and store this in an R object named "**Wilcoxon.Data**" (remember the users can use any name of they may preferably choose if they wish to do so).

Once the user has successfully imported the dataset, they will be able to view the details of the example dataset (**exam_grades.csv**) stored as R object we named "**Wilcoxon.Data**" in R as shown and highlighted in Fig. 13.4 with 233 observations and 6 variables (column) in the data sample.

```
Wilcoxon.Data <- read.csv(file.choose())
attach(Wilcoxon.Data)
names(Wilcoxon.Data)
str(Wilcoxon.Data)
View(Wilcoxon.Data)
```

***Note: a good scientific practice when working with datasets both in R or for research purposes is to clean up the dataset for use, e.g., by removing the NA or empty cells or values to ensure an accurate and reliable calculation or computation. For example, as shown in Step 2 in Fig. 13.3 (see Lines 30 to 32), the authors have used the following syntax and code to "remove the NA values" in the stored dataset (Wilcoxon.Data).

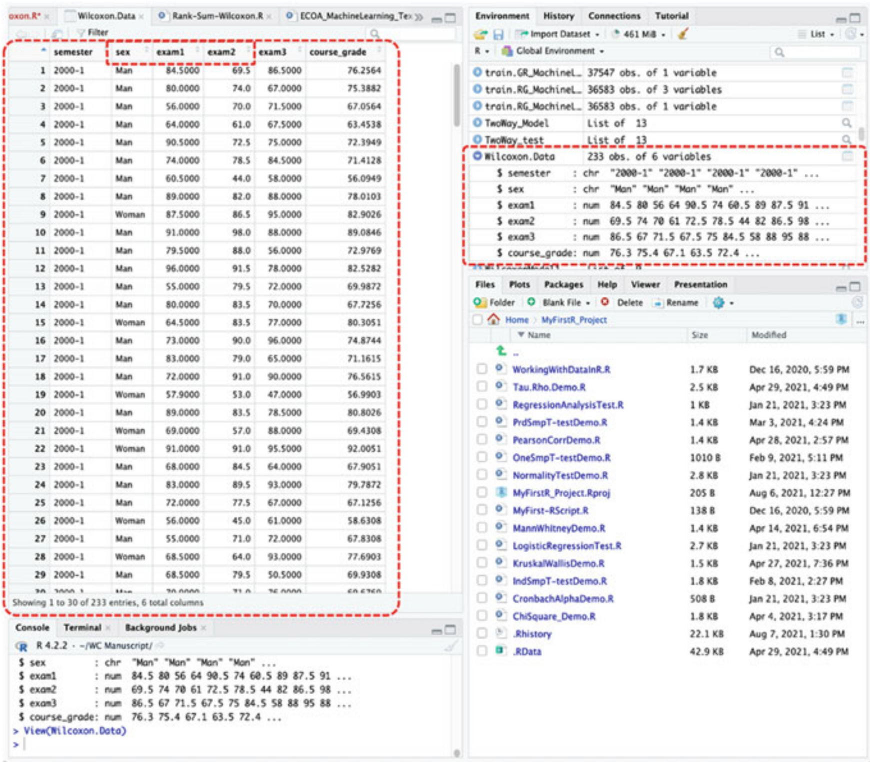

Fig. 13.4 Example dataset.csv imported and stored as an R object in R

```
# Remove NA values in the Example data
Wilcoxon.Data <- na.omit (Wilcoxon.Data)
View(Wilcoxon.Data)
```

***Now, when you use the view function `View(Wilcoxon.Data)` to visualize the example dataset again, you will notice that the program has removed the row that contain the "NA" value under the "exam1" variable. Consequentially, the user will also notice in the Environment Tab that there are now a total of "232 observations" and 6 variables (column) in the cleaned data sample (Wilcoxon.Data).

Step 3—Conduct Tests for Assumptions and Analyze Data

As shown in Fig. 13.5 (Step 3A, Lines 34–49), we first conducted the various necessary tests of assumptions (e.g., data normality and homogeneity of variance) for the selected items or variables (i.e., "**exam1**" and "**exam2**"—see Fig. 13.4) in R before proceeding to perform the main analysis (Signed-Rank Wilcoxon test—Step 3B). The assumption test done in Step 3A (Fig. 13.4) is to ensure that the data does not meet (violates) the data normality or homogeneity of variance condition, which are prerequisite to carrying out the non-parametric tests such as the Wilcoxon test.

The test of assumptions (Step 3A, Fig. 13.5) and Signed-Rank Wilcoxon test defined in Step B (see Fig. 13.5, Lines 51–63) is done by using the **shapiro.test( ), var.test( )**, and **wilcox.test( )** functions in R.

As defined in the Introduction section (Sect. 13.1);

- **Signed-Rank Wilcoxon test** statistics compares the median for two sets of data from a single population but analyzed at different time intervals (e.g., pre and post test, before and after, etc.).
- The targeted variables must be measured in ranked or ordinal scale. Thus, it is assumed that the independent variable in a Wilcoxon test is dichotomous, and the dependent variable is a continuous variable whose measurement is at least ordinal.

To illustrate the Signed-Rank Wilcoxon test using the example dataset we stored as "**Wilcoxon.Data**" (see: highlighted columns in Fig. 13.4) we will:

1. Test whether the median of the grades for the "**exam1**" variable is *equal to* the median of the "**exam2**" variable in the dataset? **(two-tailed test)**.

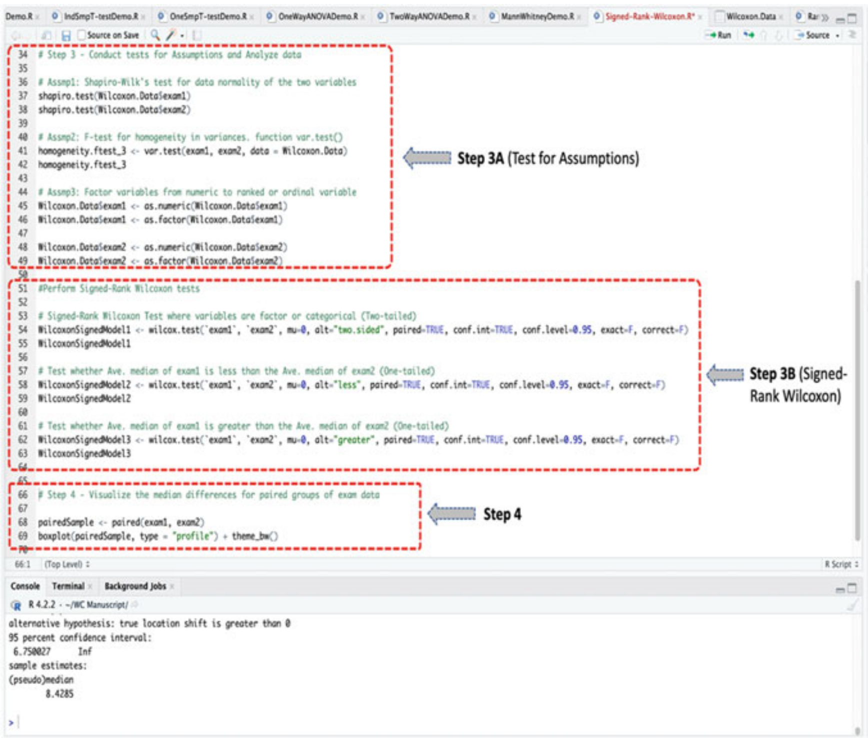

Fig. 13.5 Conducting signed-rank Wilcoxon test in R

2. Test whether the median of the "**exam1**" grades is *less than* the median of the "**exam2**"? **(one-tailed test)**.
3. Test whether the median of the "**exam1**" grades is *greater than* the median of the "**exam2**"? **(one-tailed test)**.

Accordingly, the syntax to performing the above-listed tests in R (see Fig. 13.5, Steps 3A and 3B, Lines 34–63) are as shown in the codes below:

```
# Assmp1: Shapiro-Wilk's test for data normality of the two variables
shapiro.test(Wilcoxon.Data$exam1)
shapiro.test(Wilcoxon.Data$exam2)

# Assmp2: F-test for homogeneity in variances. function var.test()
homogeneity.ftest_3 <- var.test(exam1, exam2, data = Wilcoxon.Data)
homogeneity.ftest_3

# Assmp3: Factor variables from numeric to ranked or ordinal variable
Wilcoxon.Data$exam1 <- as.numeric(Wilcoxon.Data$exam1)
Wilcoxon.Data$exam1 <- as.factor(Wilcoxon.Data$exam1)

Wilcoxon.Data$exam2 <- as.numeric(Wilcoxon.Data$exam2)
Wilcoxon.Data$exam2 <- as.factor(Wilcoxon.Data$exam2)
```

#Perform Signed-Rank Wilcoxon tests

```
# Signed-Rank Wilcoxon Test where variables are factor or categorical (Two-tailed)
WilcoxonSignedModel1 <- wilcox.test(`exam1`, `exam2`, mu=0, alt="two.sided",
paired=TRUE, conf.int=TRUE, conf.level=0.95, exact=F, correct=F)
WilcoxonSignedModel1

# Test whether Ave. median of exam1 is less than the Ave. median of exam2 (One-tailed)
WilcoxonSignedModel2 <- wilcox.test(`exam1`, `exam2`, mu=0, alt="less", paired=TRUE,
conf.int=TRUE, conf.level=0.95, exact=F, correct=F)
WilcoxonSignedModel2

# Test whether Ave. median of exam1 is greater than the Ave. median of exam2 (One-tailed)
WilcoxonSignedModel3 <- wilcox.test(`exam1`, `exam2`, mu=0, alt="greater",
paired=TRUE, conf.int=TRUE, conf.level=0.95, exact=F, correct=F)
WilcoxonSignedModel3
```

Useful Tip:

- As described in the codes and figure above (Fig. 13.5), the users are always required to specify the `paired = TRUE` option when conducting the Signed-Rank Wilcoxon test, which represents as the alternative (non-parametric equivalent) to the Paired Sample *t*-test.
- Use the `alt = "less"` and `alt = "greater"` options to specify a "one-tailed" *t*-test.

Once the user has successfully run the codes defined in **Steps 3A** and **3B** (Lines 34–63) in Fig. 13.5, they will be presented with the results of the "tests for assumptions" (Step 3A) and the "Signed-Rank Wilcoxon test" (Step 3B) in the Console as shown in Figs. 13.6a and b, respectively.

In Fig. 13.6a which represents as the result or outcome of Step 3A (see: Fig. 13.5), we conducted the different necessary assumptions tests for the Wilcoxon test in order to determine if the available dataset and variables are valid to perform the test.

As highlighted in the figure (Fig. 13.6a), the **normality test** (Assmp1) by using the Shapiro–Wilk's method **shapiro.test( )** whereby we hypothetically assume that a

a

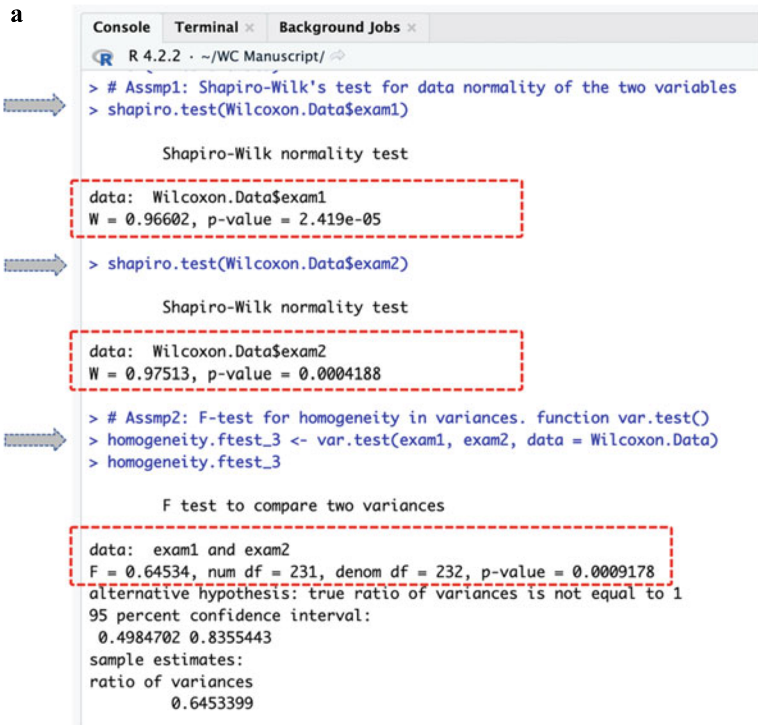

Fig. 13.6 **a** Results of data normality and homogeneity of variance test displayed in the console in R. **b** Results of signed-rank Wilcoxon test displayed in Console in R

b

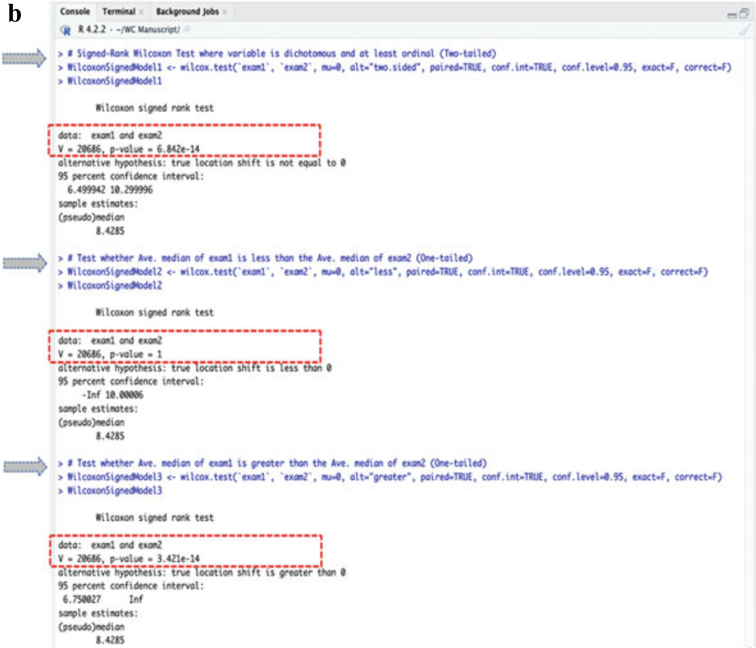

Fig. 13.6 (continued)

score of *p-value* > *0.05* is normal, shows that the distribution of the two sets of data or variables (i.e., `"exam1" where W=0.96602, p-value=2.419e-05`, and `"exam2" where W=097513, p-value=0.0004188`) are not normality distributed.

Also, the **homogeneity of variance** test (Assmp2) for the two variables (`exam1` and `exam2`) using the **var.test( )** method, whereby we assume that a value of $p > 0.05$ indicates equality in variance, shows that there is difference in the variance for the two variables (`exam1`, `exam2`) with `p-value=0.0009178` and `F=0.64534`.

Thus, we presume that the data normality and assumption of equality in variance are not met, and proceed to conduct the Signed-rank Wilcoxon test.

Consequently, Fig. 13.6b is the result of the Signed-Rank Wilcoxon test and statistics as described in Step 3B in Fig. 13.5 (Lines 51–63).

As reported in Fig. 13.6b, we conducted the Signed-Rank Wilcoxon test by testing the median differences for the two target variables (`exam1`, `exam2`). The results of the test were stored in an R object we defined as `"WilcoxonSignedModel1"` for the **two-tailed** analysis, and `"WilcoxonSignedModel2"` and `"WilcoxonSignedModel3"` for the **one-tailed** analysis, respectively. The meaning of the results is discussed in detail in Step 5 in this section.

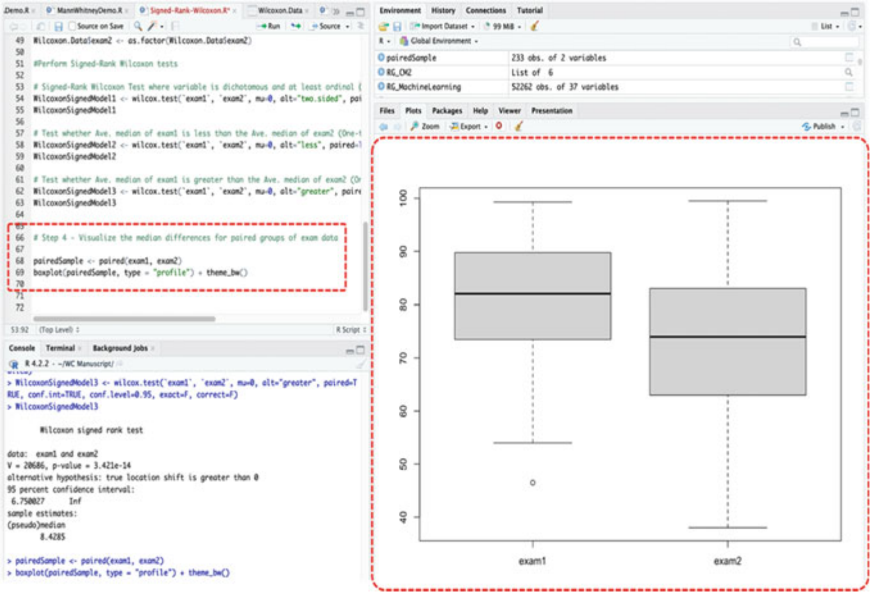

Fig. 13.7 Plot for median difference for two paired group of variables in R

Step 4—Plot and Visualize the Mean Differences for the Two Paired Variables

Another good way to determine the median differences between two variables is to graphically plot or represent it. As defined in Step 4 in Fig. 13.5 (Lines 66–69) and the resultant chart represented in Fig. 13.7; the authors graphically represented or visualized the median between the two paired groups of variables (`exam1`, `exam2`) in the example dataset "**Wilcoxon.Data**" by plotting them using the **paired( )** and **boxplot( )** functions in R.

The code used to plot the median of the two variables (`exam1`, `exam2`) is as shown below, and the resultant graph is as presented in Fig. 13.7.

```
# Visualize median differences for paired groups of data
pairedSample <- paired(exam1, exam2)
boxplot(pairedSample, type = "profile") + theme_bw()
```

Step 5—Results' Interpretation for Signed-Rank Wilcoxon Test

The final step for the "Signed-Rank Wilcoxon" test statistics is to interpret and understand the results of the analysis.

By default, the hypothesis for conducting the test (Signed-Rank Wilcoxon) is; *IF* the p-value is less than or equal to 0.05 ($p \leq 0.05$), *THEN* we assume that the median of the two set of variables or analyzed data are statistically different and not by chance (H_1), *ELSE IF* the p-value is greater than 0.05 ($p > 0.05$) *THEN* we can say that there is no difference in the median of the two sets of data (H_0).

```
> WilcoxonSignedModel1
        Wilcoxon signed rank test
data:  exam1 and exam2
V = 20686, p-value = 6.842e-14
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 6.499942 10.299996
sample estimates:
(pseudo)median
     8.4285
```

As reported in the above statistics and outcome of the Wilcoxon test (Signed-Rank)—Figs. 13.6b; the meaning of the results of the test by using the **wilcox.test( )** function in R can be explained as a list containing the following:

- **Statistics**: `v = 20686` which denotes the value of the *t*-test analysis.
- **p-value**: `p-value = 6.842e-14` is the significance level of the test.
- **Confidence interval:** `Conf.Int(95%, 6.499942 10.299996)` represents the confidence interval for the median assumed to be appropriate to the specified alternative hypothesis.
- **Sample estimates:** `(pseudo)media = 8.4285` is the estimated median of the two groups of variables, e.g., compared by considering the two variables (`exam1`, `exam2`).

Accordingly, the outcome of the **wilcox.test( )** function used for the **Two-tailed** Singed-Rank test or model (`WilcoxonSignedModel1`) shows that there exist a difference in the median between the two sets of analyzed variable (`exam1`, `exam2`). The p-value for the "**Two-tailed**" test was statistically found to be `p=6.842e-14 (V=20686)`, which is significantly less than the scientifically accepted levels ($p \leq 0.05$). Therefore, we conclude that there is a significant difference between the median of the two sets of exam grades (`exam1`, `exam2`) across the analyzed period or data.

Furthermore, as gathered in the results for the **one-tailed** Signed-Rank tests described below (see Fig. 13.6b);

- In our analysis, we also checked whether the median of the "**exam1**" variable is *less than* the median of the "**exam2**" (`WilcoxonSignedModel2`).
- Then checked whether the median of the "**exam1**" is *greater than* the median of the "**exam2**" (`WilcoxonSignedModel3`).

```
> WilcoxonSignedModel2

      Wilcoxon signed rank test

data:  exam1 and exam2
V = 20686, p-value = 1
alternative hypothesis: true location shift is less than 0
95 percent confidence interval:
     -Inf 10.00006
sample estimates:
(pseudo)median
              8.4285
```

```
> WilcoxonSignedModel3

      Wilcoxon signed rank test

data:  exam1 and exam2
V = 20686, p-value = 3.421e-14
alternative hypothesis: true location shift is greater than 0
95 percent confidence interval:
 6.750027      Inf
sample estimates:
(pseudo)median
        8.4285
```

In the results of the **one-tailed** tests presented above, we can see that when we analyzed whether the median of the "**exam1**" variable is *less than* the median of the "**exam2**" that there was no significant difference (`WilcoxonSignedModel2, V=20686, p=1`). But when we analyzed whether the median of the "**exam1**" is *greater than* the median of the "**exam2**" there was a significant difference (`WilcoxonSignedModel3, V=20686, p=3.421e-14`) (i.e., $p \leq 0.05$). Therefore, it can be said from the results of the one-tailed Wilcoxon tests that the median of the "**exam1**" is *greater* and *not less* than the median of the "**exam2**" grades which was statistically differenced (sample estimates) by margin of `8.4285` (pseudo median of the differences) presented in both tests (one-tailed).

Therefore, in summary, we can statistically say that there was a significant difference or variation in the *median* of the grades or targeted variables (exam1 and exam2) across the periods of the exams based on the example dataset that we stored as "**Wilxoxon.Data**" in R. For example, this result suggests a change in the grades of the students or participants across those periods. Moreover, the average median of the "**exam1**" represents to be *greater* (and *not less*) than the median of the "**exam2**" in the analyzed data.

13.3 Rank-Sum Wilcoxon Test in R

Rank-Sum Wilcoxon test (also referred to as the non-parametric equivalent or alternative to the "Unpaired" or "Two-sample" or "Independent-sample" t-test) is used when the dataset the researcher or data analysts wants to analyze are of two types or sample, and are statistically *independent*. In essence, the "Rank-Sum Wilcoxon test" is used to compare the median of two independent groups of variables or data samples. As the non-parametric equivalent of the Independent sample t-test, it (Sum-Rank) allows the researchers to *compare the median* of two distinctive sets of data randomly drawn from two different populations, when the dataset in question violates the necessary conditions to perform the Independent sample t-test or is said to contain outliers.

By default, the hypothesis for testing whether there is a *difference in median* of the two (independent) data samples is; *IF* the p-value is less than or equal to 0.05 ($p \leq 0.05$), *THEN* we assume that the median of the two sets of data or groups of population in the sample are statistically different and that this is not by chance (H_1), *ELSE IF* the p-value is greater than 0.05 ($p > 0.05$) *THEN* we can presume that there is no difference in the median of the two groups of data and any difference observed could only occur by chance (H_0).

In this section, the authors will demonstrate to the readers how to conduct the Rank-Sum (unpaired sample) Wilcoxon test in R using the **shapiro.test( ), var.test( )** and **wilcox.test( )** functions in R.

As defined earlier in the previous section (Sect. 13.1), we will do this by using the computational steps outlined in Fig. 13.1.

To begin, **Create a new or open an existing R project**. Once the user has the RStudio and an R Project opened, **Create a new R Script** and name it "**Sum-Rank-Wilcoxon**" or any name the user may preferably choose (see Chaps. 1 and 2 for step-by-step guide on how to do these steps if required).

Next, we will continue to use the example dataset (**exam_grades.csv**) (see Figs. 13.2 and 13.4) that we downloaded earlier and stored as R object we named or defined as "Wilcoxon.Data" in R to illustrate the Rank-Sum Wilcoxon test. ***Note: the readers can refer to the following repository (https://doi.org/https://doi.org/10.6084/m9.figshare.24728073) where the authors have uploaded all the example files used in this book or if they have not practiced the previous example the authors presented in Sect. 13.2.

 # Step 1—Install and Load the Required R Packages

Since we have previously installed the necessary R packages in our previous example in Sect. 13.2, we do not necessarily need to install the different R packages again, rather we just need to "load" the libraries for the necessary packages (see Lines 12 to 18, Step 1, Fig. 13.8). However, if the user has directly visited this section or has exited or not practiced the previous example in Sect. 13.2, then they would need to "Install and load" the necessary R packages as defined in Lines 5–18 in Step 1 (Fig. 13.8).

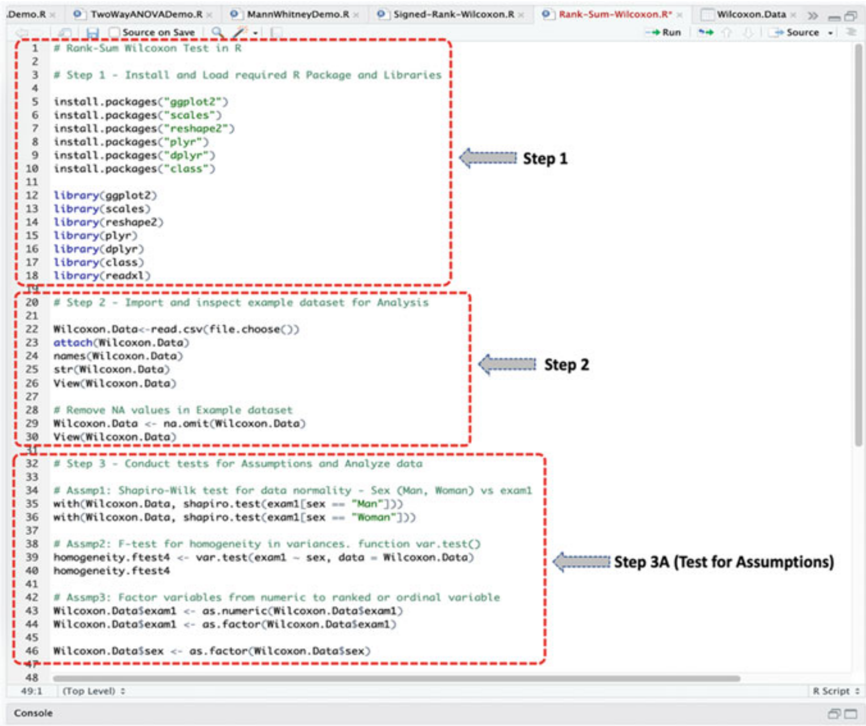

Fig. 13.8 Steps to conducting Rank-Sum Wilcoxon test in R

Depending on the users case scenario, **Install** or **Load** the following *R packages* and *libraries* (Fig. 13.8, Step 1), that we will be using to call the different R functions, data manipulations, and graphical visualizations for the Rank-Sum Wilcoxon test.

The code and syntax to install and load the required R packages are as follows:

```
install.packages("ggplot2")
install.packages("scales")
install.packages("reshape2")
install.packages("plyr")
install.packages("dplyr")
install.packages("class")

library(ggplot2)
library(scales)
library(reshape2)
library(plyr)
library(dplyr)
library(class)
library(readxl)
```

Step 2—Import and Inspect Example Dataset for Analysis

As illustrated in Step 2 in Fig. 13.8 (Lines 20–30); import and/or load the dataset named "**exam_grades.csv**" which we have previously downloaded and stored as an R object we named "**Wilcoxon.Data**" in R (see Figs. 13.2 and 13.4).

Depending on whether the user has continued in our previous example, or has exited or visited this particular section; the code to import and/or attach the example dataset is provided below:

```
Wilcoxon.Data <- read.csv(file.choose())
attach(Wilcoxon.Data)
names(Wilcoxon.Data)
str(Wilcoxon.Data)
View(Wilcoxon.Data)

# Remove NA values in Example dataset
Wilcoxon.Data <- na.omit(Wilcoxon.Data)
View(Wilcoxon.Data)
```

Once the user has successfully imported or loaded the dataset, you will be able to view the details of the dataset (exam_grades.csv) stored as R object "**Wilcoxon.Data**" in the R environment as shown in Figs. 13.2 and 13.4, respectively, with 233 observations and 6 variables (column) in the data sample.

Step 3—Conduct Tests for Assumptions and Analyze Data

Now that we have imported the dataset and/or loaded it ready for analysis (Wilcoxon.Data), we can proceed to analyze the data.

As defined in Step 3A in Fig. 13.8 (Lines 32–46), we first conducted the various necessary tests of assumptions (e.g., data normality and homogeneity of variance) for the chosen items or variables (i.e., "**sex**" and "**exam1**"—see Fig. 13.4) in R before proceeding to perform the analysis (Rank-Sum Wilcoxon test – Step 3B). The assumption tests we performed in Step 3A (Fig. 13.8) are to ensure that the data does not meet or violate the data normality or homogeneity of variance condition which are preconditions to carrying out the non-parametric tests, such as the Wilcoxon test.

The test of assumptions (Step 3A, Fig. 13.8) and Rank-Sum Wilcoxon test described in Step B (see Fig. 13.9, Step 3B, Lines 48–60) is performed by using the **shapiro.test( ), var.test(**), and **wilcox.test( )** functions in R.

As defined earlier in the Introduction section (Sect. 13.1);

- **Rank-Sum Wilxocon test** compares the *median* for two independently sampled groups from two different population whereby the two variables under consideration are independent of each other.
- The targeted grouping "independent" variable (***x***) is often a dichotomous or binary type, while the ***y*** variable is continuous with at least ordinal scale measurement.

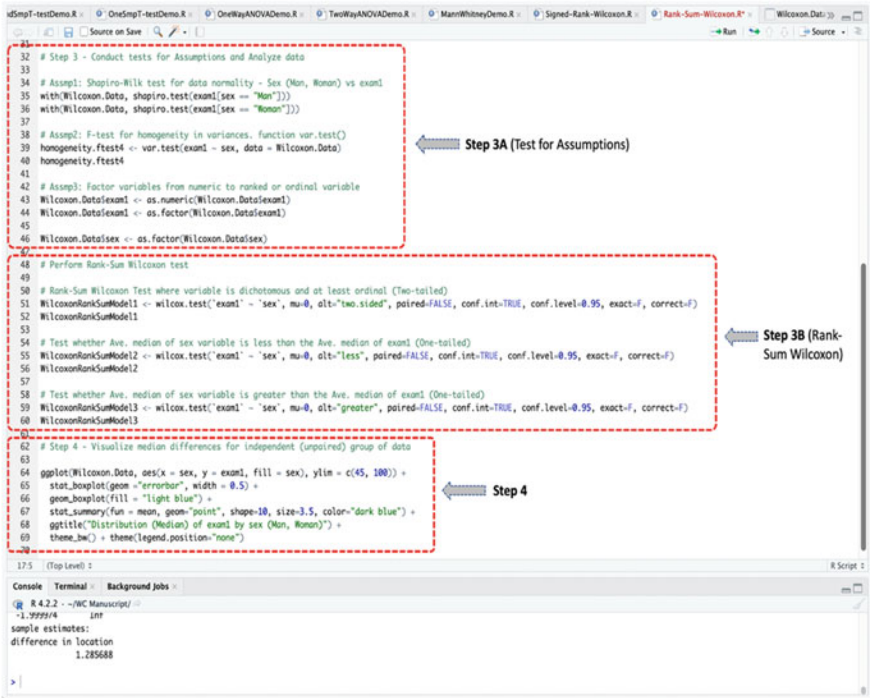

Fig. 13.9 Conducting Rank-Sum (Unpaired) Wilcoxon Test in R

To illustrate this test (Rank-Sum) using the example dataset "**Wilxocon.Data"** in R (see: highlighted columns in Fig. 13.4) we will:

1. Test whether the median of the **group A** (Man) of the "**sex**" variable is *equal to* the mean of the **group B** (Woman) of the "**sex**" variable considering the "**exam1**" period in the data? (**two-tailed test)**
2. Also, we will check whether the *median* of the **group A** (Man) is *less than* the median of **group B** (Woman) in the "**sex**" variable? (**one-tailed test)**
3. Then we will check whether the *median* of **group A** (Man) is *greater than* the median of **group B** (Woman) in the "**sex**" variable? (**one-tailed test)**

The syntax and code to perform the above tests in R (see Fig. 13.9, Steps 3A and 3B, Lines 32–60) is as shown in the codes provided below:

Test of Assumption (Step 3A):

```
# Assmp1: Shapiro-Wilk test for data normality - Sex (Man, Woman) vs exam1
with(Wilcoxon.Data, shapiro.test(exam1[sex == "Man"]))
with(Wilcoxon.Data, shapiro.test(exam1[sex == "Woman"]))

# Assmp2: F-test for homogeneity in variances. function var.test()
homogeneity.ftest4 <- var.test(exam1 ~ sex, data = Wilcoxon.Data)
homogeneity.ftest4

# Assmp3: Factor variables from numeric to ranked or ordinal variable
Wilcoxon.Data$exam1 <- as.numeric(Wilcoxon.Data$exam1)
Wilcoxon.Data$exam1 <- as.factor(Wilcoxon.Data$exam1)

Wilcoxon.Data$sex <- as.factor(Wilcoxon.Data$sex)
```

Perform Rank-Sum Wilcoxon Test (Step 3B):

```
# Rank-Sum Wilcoxon Test where variable is dichotomous and at least ordinal (Two-tailed)
WilcoxonRankSumModel1 <- wilcox.test(`exam1` ~ `sex`, mu=0, alt="two.sided",
paired=FALSE, conf.int=TRUE, conf.level=0.95, exact=F, correct=F)
WilcoxonRankSumModel1

# Test whether Ave. median of sex variable is less than the Ave. median of exam1 (One-tailed)
WilcoxonRankSumModel2 <- wilcox.test(`exam1` ~ `sex`, mu=0, alt="less",
paired=FALSE, conf.int=TRUE, conf.level=0.95, exact=F, correct=F)
WilcoxonRankSumModel2

# Test whether Ave. median of sex variable is greater than the Ave. median of exam1 (One-
tailed)
WilcoxonRankSumModel3 <- wilcox.test(`exam1` ~ `sex`, mu=0, alt="greater",
paired=FALSE, conf.int=TRUE, conf.level=0.95, exact=F, correct=F)
WilcoxonRankSumModel3
```

Useful Tips

- As shown and illustrated in the code and figure above (Fig. 13.9), the users must always use the `paired = FALSE` option to specify the Rank-Sum (unpaired) Wilcoxon test.
- Whereas the `paired = TRUE` option is used to specify the Signed-Rank (paired) Wilcoxon test (which the authors have covered in the previous Sect. 13.2).
- Users need to use the `exact = F` and `correct = F` options to specify unequal variances for both the Signed-Rank and Rank-Sum Wilcoxon test or analysis.
- Then use the `alt = "less"` and `alt = "greater"` option to specify a "one-tailed" Wilcoxon analysis.

Accordingly, once the user has successfully run the codes as defined in **Steps 3A** and **3B** (see Fig. 13.9, Lines 32–60), they will be presented with the results of the "tests for assumptions" (Step 3A) and the "Rank-Sum Wilcoxon test" in the Console as shown in Figs. 13.10a and b, respectively.

In Fig. 13.10a, which represents the outcome of the Step 3A (see Figs. 13.8 and 13.9); we conducted the different tests for assumptions for the Rank-Sum Wilcoxon test in order to determine if the analyzed dataset is fitting and valid for the test (Independent Samples t-test). In other words, whether the dataset in question does not meet the criteria to perform the parametric test or version (i.e., Independent Sample t-test).

As highlighted in the figure (Fig. 13.10a), we can see that the *normality test* (Assmp1) by using the **shapiro.test( )** method (where we assume a value of $p > 0.05$ is normal) shows that the distribution of the two groups of the independent variable (i.e., **Man** and **Woman** of the "**sex**") are not normality distributed when analyzed against the "**exam1**" variable, which is the second targeted variable in our analysis, whereby (exam1[sex == "Man"], p-value=0.0004686, W=0.96992) and (`exam1[sex == "Woman"], p-value=0.01099, W=0.93194`), respectively.

Also, in the second assumption test (Assmp2, Fig. 13.10a); we tested the *homogeneity of variance* for the two targeted variables (**exam1** ~ **sex**) using the **var.test( )** method, whereby we assume that a value of $p > 0.05$ indicates that "equality in variance" is met. Consequentially, as highlighted in Fig. 13.10a, we note that there are slightly differences (i.e., close to equality of variance being met) in the variance for the two sets of analyzed variables with `p=0.06007.` and `F = 0.65852`, respectively.

Thus, with no data normality met and the slight differences in homogeneity of variance, we can proceed to conduct the "Rank-Sum Wilcoxon test" as defined in Step 3B (Fig. 13.9) and the results represented in Fig. 13.10b.

As gathered the results presented in Fig. 13.10b, we conducted the Rank-Sum (Two-sample) Wilcoxon test by considering the two variables (**exam1** ~ **sex**). The results of the tests are stored in R object we named or defined as "`WilcoxonRankSumModel1`" for the **two-tailed** analysis, and "`WilcoxonRankSumModel2`" and "`WilcoxonRankSumModel3`" for the **one-tailed** analysis, respectively. The meaning of the results is interpreted in detail in Step 5 described in this section.

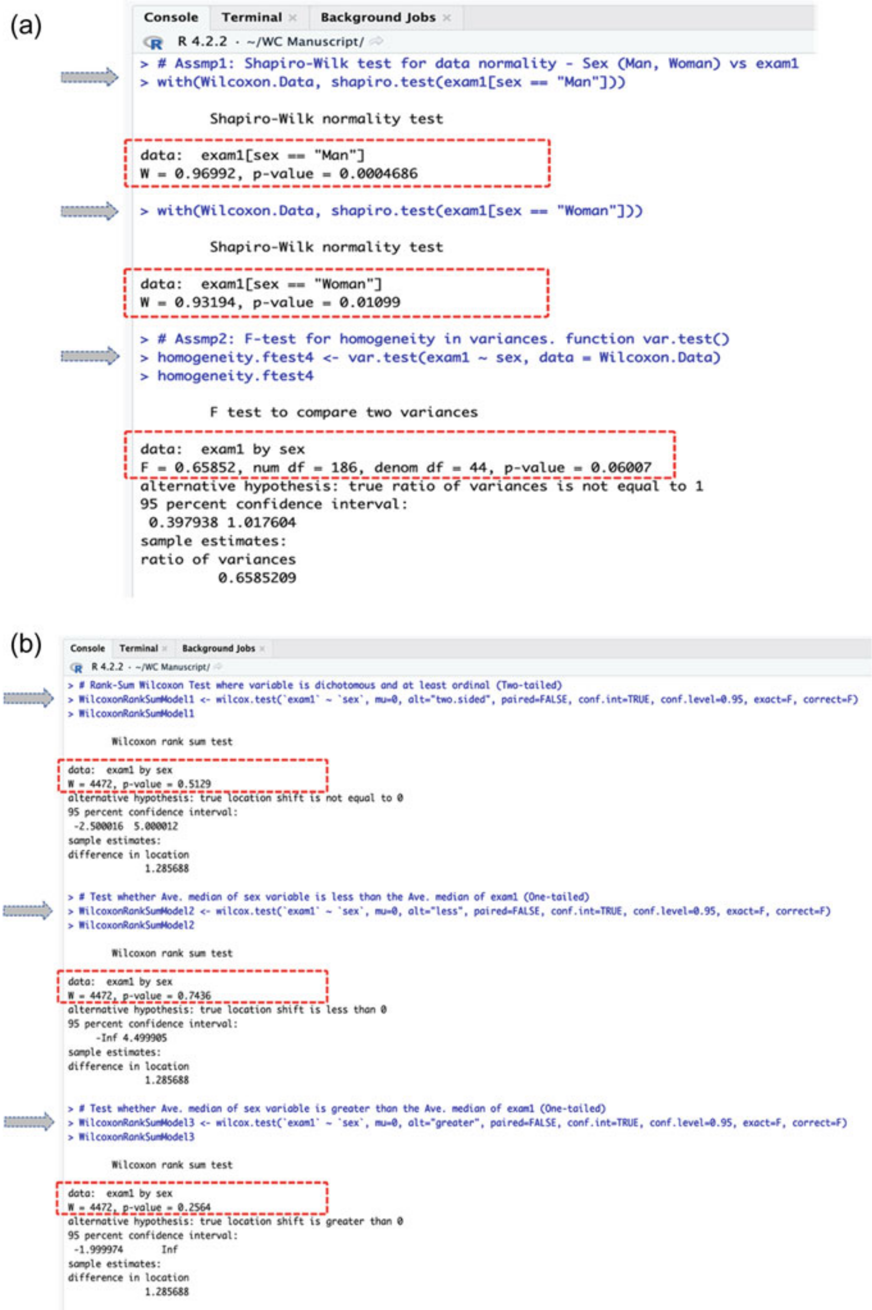

Fig. 13.10 **a** Results of data normality and homogeneity of variance tests displayed in the console in R. **b** Result of rank-sum (two-sample or unpaired) Wilcoxon test in R

Step 4—Plot and Visualize the Median Differences for Two (Independent) Variables

As illustrated previously in Sect. 13.2, another great way to check whether there is a difference in median of the two independent group of variables (e.g., sex and exam1) is by plotting them as graph. By so doing, the users will be able to visualize the difference in the median (if any) between the two data sample.

To do this, in Step 4 in Fig. 13.9 (Lines 62–69) and the resultant chart represented in Fig. 13.11; the authors used the **ggplot( )** function in R to visualize the *median* between the two groups of variables "**sex**" (**Man, Woman**) by taking into account or plotting it against the second variable **"exam1**" in the example data "**Wilcoxon.Data**".

The code and syntax used in plotting the median for the two variables is shown in the code below, and the resultant chart is as represented in Fig. 13.11.

Visualize Median Difference for Two Independent (Unpaired) Variables

```
ggplot(Wilcoxon.Data, aes(x = sex, y = exam1, fill = sex), ylim = c(45, 100)) +
  stat_boxplot(geom ="errorbar", width = 0.5) +
  geom_boxplot(fill = "light blue") +
  stat_summary(fun = mean, geom="point", shape=10, size=3.5, color="dark blue") +
  ggtitle("Distribution (Median) of exam1 by sex (Man, Woman)") +
  theme_bw() + theme(legend.position="none")
```

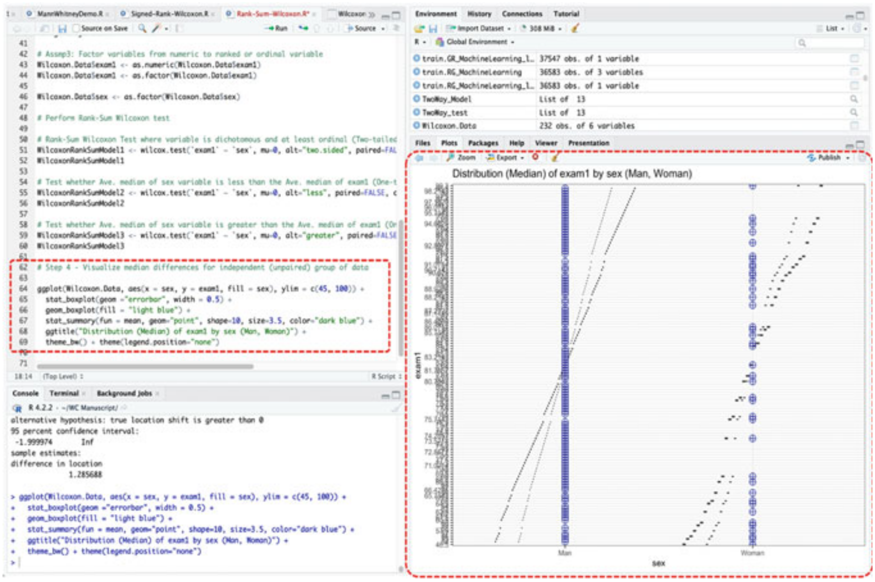

Fig. 13.11 Plotting the median differences for two sets of independent variables in R using the ggplot() function

Step 5—Results' Interpretation (Sum-Rank Wilcoxon Test)

The final step for the Sum-Rank Wilcoxon test or statistics is to interpret and understand the results of the test.

By default, the hypothesis for conducting the test (Sum-Rank Wilcoxon) is; *IF* the p-value is less than or equal to 0.05 ($p \leq 0.05$), *THEN* we assume that the median of the two sets of data or groups of variables is statistically different and that this is not by chance (H_1), *ELSE IF* the p-value is greater than 0.05 ($p > 0.05$) THEN we can conclude that there is no difference in the median of the two sets of variable and that any difference observed may only occur by chance (H_0).

```
> WilcoxonRankSumModel1 <- wilcox.test(`exam1` ~ `sex`, mu=0,
alt="two.sided", paired=FALSE, conf.int=TRUE, conf.level=0.95, exact=F,
correct=F)
> WilcoxonRankSumModel1

    Wilcoxon rank sum test

data:  exam1 by sex
W = 4472, p-value = 0.5129
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 -2.500016  5.000012
sample estimates:
difference in location
              1.285688%
```

As shown in the results of the test presented above or the outcome of the Rank-Sum Wilcoxon test (`WilcoxonRankSumModel1`) (see (Fig. 13.10b); the meaning of the results of the Rank-Sum test for the **two-tailed** (**exam1 ~ sex**) by using the **wilcox.test( )** function in R can be explained as a list containing the following:

- **Statistics**: `W = 4472`, which denotes the value of the Rank-Sum test statistics.
- **p-value**: `p-value = 0.5129` is the p-value or significance level of the test.
- **Confidence interval:** `Conf.Int(95%, -2.500016 5.000012)` represents the confidence interval for the median assumed to be appropriate to the specified alternative hypothesis.
- **Sample estimates:** `1.285688` is the median difference in location between the two groups of data or population that is compared considering the two variables (exam1 ~ sex).

As seen in the results described above, statistically it can be said that the p-value of the Rank-Sum Wilcoxon test (`WilcoxonRankSumModel1`), i.e., the **two-tailed** analysis, is `p=0.5129`. As we can see, the p-value of the test is greater than the scientifically acceptable significance levels ($p \leq 0.05$). Therefore, we can statistically conclude that there is no statistically significant difference between the medians of the two groups of data (Man vs Woman) of the "**sex**" variable taking into account the "**exam1**" grades of the students or participants in the analyzed data. In other words, the *median* of the **group A (Man)** of the "**sex**" variable is *equal to* the median of **group B (Woman)** of the "**sex**" when analyzed against the "**exam1**" grades (**two-tailed test).**

Furthermore, as shown in the next results for the "one-tailed" Rank_Sum tests presented below and as reported in Fig. 13.10b.

- We also checked whether the median of the **group A** (i.e., Man) is *less than* the median of the **group B** (Woman) of the "**sex**" variable (`WilcoxonRankSumModel2`), and
- Whether the median of the **group A** (Man) is *greater than* the median of **group B** (Woman) of the "**sex**" variable (`WilcoxonRankSumModel3`), respectively.

```
> WilcoxonRankSumModel2 <- wilcox.test(`exam1` ~ `sex`, mu=0,
alt="less", paired=FALSE, conf.int=TRUE, conf.level=0.95, exact=F,
correct=F)
> WilcoxonRankSumModel2

      Wilcoxon rank sum test

data:  exam1 by sex
W = 4472, p-value = 0.7436
alternative hypothesis: true location shift is less than 0
95 percent confidence interval:
 -Inf  4.499905
sample estimates:
difference in location
              1.285688
```

```
> WilcoxonRankSumModel3 <- wilcox.test(`exam1` ~ `sex`, mu=0,
alt="greater", paired=FALSE, conf.int=TRUE, conf.level=0.95, exact=F,
correct=F)
> WilcoxonRankSumModel3

      Wilcoxon rank sum test

data:  exam1 by sex
W = 4472, p-value = 0.2564
alternative hypothesis: true location shift is greater than 0
95 percent confidence interval:
  -1.999974 Inf
sample estimates:
difference in location
              1.285688
```

As gathered in the above results of the one-tailed tests (`WilcoxonRankSumModel2, where W=4472, p=0.7436`; and `WilcoxonRankSumModel3, where W=4472, p=0.2564`), we can see that there are no significant differences in the medians of the two groups of data (i.e., the value of p-value is above the threshold of $p \le 0.05$, and the `difference in location (sample estimates) = 1.285688`), as also evidenced in the result of the “two-tailed test (see Fig. 13.10b). Therefore, we can statistically conclude that there are no differences in the *median* of the scores for the group of Men vs Women in the “exam1” variable.

13.4 Summary

In this final chapter of the book and PART II, the authors illustrated to the readers how to conduct the two main types of Wilcoxon test (Signed-Rank and Rank_Sum) in R. In Sect. 13.2, we explained and practically illustrated how to perform the Signed-rank (paired) Wilcoxon test. While Sect. 13.3 covers how to conduct the Rank-Sum (unpaired or two-sample) Wilcoxon test.

The chapter also covered how to graphically plot the median of the different analyzed variables in the data and/or results of the Wilcoxon tests. In addition, it discussed in detail how to interpret and understand the results of the two main types of the Wilcoxon tests (Signed-Rank and Rank_Sum) in R.

To summarize the contents of this chapter:

- *Wilcoxon test* is one of the inferential (non-parametric) statistics that are used for hypothesis testing, and for determining if there are any significant differences (where applicable) between the *medians* of two independent groups of data or paired variables in a given data sample.

When choosing whether to conduct the ”Signed-Rank“ or ”Rank-Sum“ Wilcoxon test? The researcher or data analyst should:

- Perform “*Signed-Rank (paired) Wilcoxon test”* if the targeted group or variable comes from a single population but is analyzed at different time intervals (e.g., *pre* and *post* tests, before and after an intervention or treatment for the same group of people, place, or thing, etc). It is also important to mention that this test (Signed-Rank) is done when the dataset in question violates the necessary conditions or assumptions for conducting the Paired-Sample t-test. Thus, is (Signed-Rank) considered as the non-parametric alternative to the parametric Paired-sample t-test.
- Perform the “*Rank-Sum (independent or two sample) Wilcoxon test”* if the groups of data to be analyzed come from *two different populations* (i.e., are statistically independent). For example, two different categories of people, thing or place (gender: male and female, state: on and off, region: north and south, etc.). In addition, it is also important to mention that the test (Rank-Sum) is done when the dataset in question violates the necessary conditions or assumptions for conducting the Independent Sample t-test. Thus, is (Rank-Sum) considered as the non-parametric alternative to the parametric Independent-sample t-test.

Finally, the users will need to perform the "*two-tailed*" test or statistics, if they only want to determine whether the median of the two data samples is different from one another. Whereas, on the other hand, they will require to perform a "*one-tailed test*" if their goal includes also to determine whether the median of a specific sample or variable is *less than* or *greater than* the median of another variable, as the case may be.

References

Hayes, A. (2023). *Wilcoxon Test: Definition in Statistics, Types, and Calculation.* https://www.investopedia.com/terms/w/wilcoxon-test.asp.

Rey, D., & Neuhäuser, M. (2011). Wilcoxon-Signed-Rank Test. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 1658–1659). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_616.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1*(6), 80–83. https://doi.org/10.2307/3001968