# Chapter 3
# Test of Normality and Reliability of Data in R

## 3.1 Introduction

In scientific research and statistics, assessment of *normality* and *reliability* of data is commonly a good practice, and a lot of the time, a necessary test performed by the users before conducting the different statistical analysis and procedures.

In theory, *data normality tests* are performed to assess if a data sample is well modeled by a normal distribution. In other words, the probability bell-shaped density curve described by its *mean* and *standard deviation* with the extreme values assumed to have no significant impact on the mean value (Mishra et al., 2019; Vaclavik et al., 2020). The importance and relevance of these measurements and data distribution in determining the direction or type of research carried out by the researchers are discussed in detail in Chap. 4 of this book. The test for normality of data for research is a fundamental assumption and has a significant role in statistics (particularly in parametric tests), especially for the purpose of experimentation and testing, for instance, the researchers or analysts can test whether the hypothesis developed by them fulfills the underlying assumptions or not (Vaclavik et al., 2020).

Some example of the most commonly used types of test for "normality" of data in the available literature include but to mention a few: Kolmogorov–Smirnov (K-S) and Shapiro–Wilk (S-W) tests, Lilliefors corrected K-S test, Anderson–Darling test, Cramer–von Mises test, D'Agostino skewness test, Anscombe–Glynn kurtosis test, D'Agostino–Pearson omnibus test, and Jarque–Bera test, (Ghasemi & Zahediasl, 2012; ÖZTUNA et al., 2006; Peat & Barton, 2005; Thode, 2019), etc.

On the other hand, the *test for reliability* of data or research instruments determines the extent (consistency of measures) to which the scales or variables (items) in the available data are capable of producing a reliable (coherent) result (de Barros Ahrens et al., 2020; Taber, 2018). The reliability and validity of data samples and size is a test mainly carried out by the researchers to demonstrate that the scales that have been construed or adopted in the questionnaires or research instrument are fit for the purpose of experimentation and analysis (Taber, 2018). Some example



K. Okoye and S. Hosseini, *R Programming*,
https://doi.org/10.1007/978-981-97-3385-9_3

of the commonly used types of tests for the “reliability” of data samples in the literature includes: Cronbach’s alpha test (Taber, 2018; Tavakol & Dennick, 2011), Cohen kappa coefficient test (Carpentier et al., 2017), Exploratory Factor Analysis (EFA) (de Barros Ahrens et al., 2020; Goni et al., 2020), Principal Component Factor Analysis (PCA) (Isnainiyah et al., 2019), etc.

It is noteworthy to mention that with large sample sizes ($n > 30$ or 40), the violation of the data normality or assumption should not necessarily be a concrete factor or problem when conducting statistical analysis for research especially the parametric methods (Elliott & Woodward, 2007; Ghasemi & Zahediasl, 2012; Mishra et al., 2019; Pallant, 2007). In other words, the researchers can still use any of the *parametric* methods or procedures that they find fitting for their experiments or data analysis even when the data are not normally distributed (which a lot of the time are attributed to the non-parametric procedures) for the large enough sample sizes ($n > 30$ or 40) (Taber, 2018).

Over the next sections of this chapter, the authors will introduce the readers to how to conduct the most frequently used types of test for data normality and reliability in the literature; Kolmogorov–Smirnov (K-S) and Shapiro–Wilk (S-W) tests for normality (Sect. 3.2), and Cronbach’s alpha test for reliability (Sect. 3.3) in R statistics (Rstudio, 2023).

## 3.2 Test of Data Normality in R: *Kolmogorov–Smirnov (K-S) and Shapiro–Wilk (S-W) Test*

Kolmogorov–Smirnov (K-S) and Shapiro–Wilk (S-W) tests are a great way to determine if variable(s) in a dataset are normally distributed. The default hypothesis for testing whether a given data is normally distributed is *IF* the p-value is greater than 0.05 ($p > 0.05$) and the test statistics above 0.5, *THEN* normality is assumed, *ELSE IF* the p-value is less than or equal to 0.05 ($p \leq 0.05$) THEN normality is not assumed.

We will demonstrate how to conduct the data normality tests using the Kolmogorov–Smirnov (K-S) and Shapiro–Wilk (S-W) methods in R. We will do this in five steps as shown in Fig. 3.1.

To start, **Open RStudio** and **Create a new project** or you can open the existing example project we have created in the previous chapter (Chap. 1) called “**MyFirstR_Project.Rproj**” which the authors will be using to illustrate the examples in this chapter. Once the user have the RStudio and R Project opened, **Create a new RScript** and name it “**NormalityTestDemo**” or any name of your choice (the readers can refer to Chap. 1 on how to do these steps).

Now let’s download an example file that we will be using to demonstrate the Kolmogorov–Smirnov (K-S) and Shapiro–Wilk normality tests in R (***the readers are also welcome to use any pre-existing data or file format of their choice***).

Visit the following link as shown in Fig. 3.2 and download the.csv file named “**Sample CVS Files**” and save this on your computer desktop: https://www.learningcontainer.com/sample-excel-data-for-analysis/#Sample_CSV_file_download.

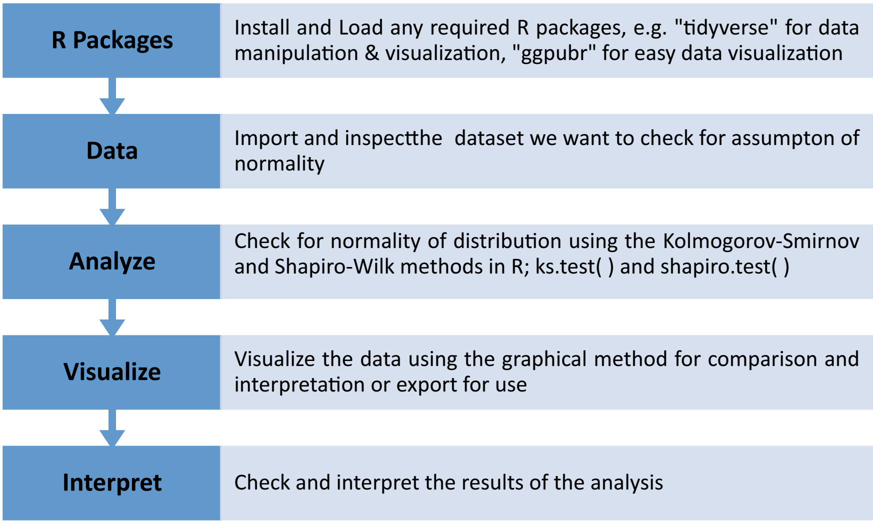



**Fig. 3.1** Steps to conducting Kolmogorov–Smirnov and Shapiro–Wilk tests in R

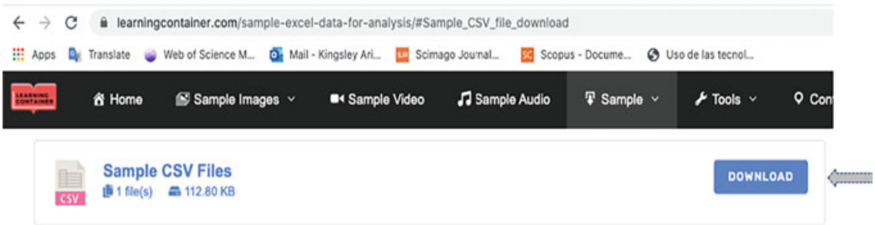



**Fig. 3.2** Example of CSV file format download. (*Source* https://www.learningcontainer.com/sample-excel-data-for-analysis/#Sample_CSV_file_download)

***Note: The example file is also available at the following repository where the authors have uploaded all the example files used in this book: https://doi.org/10.6084/m9.figshare.24728073

Once the user has downloaded and saved the file named "sample-csv-file-for-testing" on the local computer or desktop, we can proceed to conduct the analysis: Kolmogorov–Smirnov (K-S) and Shapiro–Wilk (S-W) tests for normality.

**Install and load the following R packages** ("**dplyr**" and "**ggpubr**") that we will be using to define or call the functions for the test.

The syntax to install and load the "dplyr" and "ggpubr" packages are as follows (see Fig. 3.3).

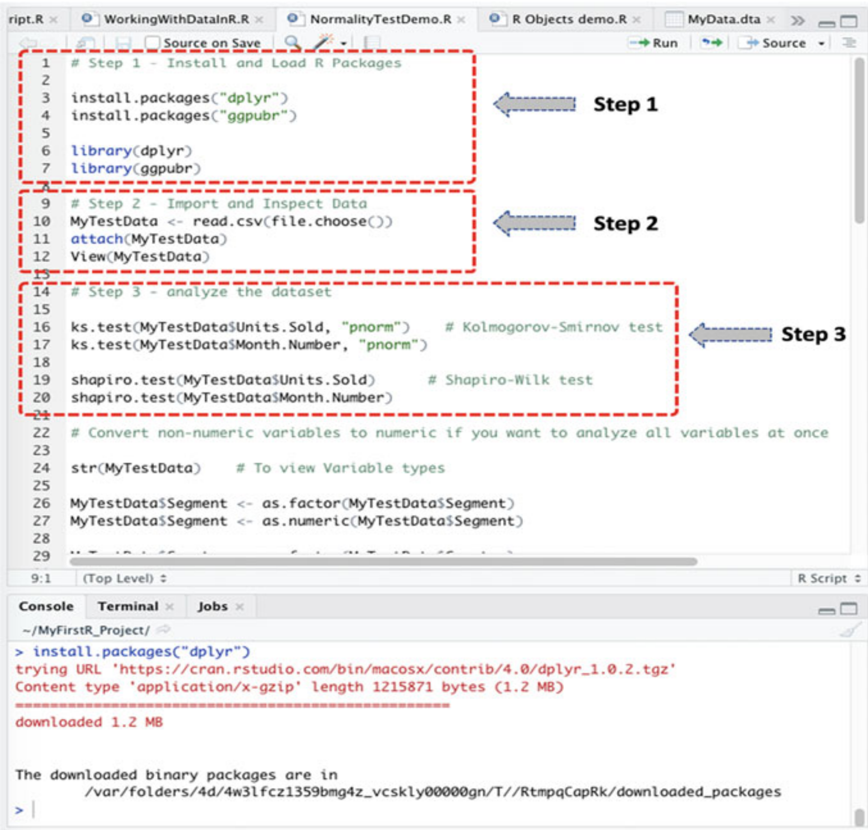



**Fig. 3.3** Conducting Kolmogorov–Smirnov and Shapiro–Wilk tests in R

# # Step 1—Install and Load R Packages

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

library(dplyr)
library(ggpubr)
```

As demonstrated in Fig. 3.3, once you have installed and loaded the required R packages (see Step 1, Fig. 3.3); the next step is to import the dataset for analysis (Step 2—see command below). See also Chap. 2 for more detailed description on how to import datasets into RStudio environment.

# **Step 2—Import and Inspect Data for** conducting the **Normality Test**

```
MyTestData <- read.csv(file.choose())
attach(MyTestData)
View(MyTestData)
```

As defined in Step 2 (see Fig. 3.3), when you run the above codes (see Lines 9–12) the user will be presented with a window through which they can navigate and choose the.csv file named "**Sample CSV files**" (Fig. 3.2) which we have downloaded earlier and stored on the desktop.

Once completed, the data will be stored in an R object named "**MyTestData**" and the user will be able to see the details of the dataset (as shown in Fig. 3.4) with 700 observations and 16 variables contained in the data sample. (*** remember you can use any name of your choice, but for learning purposes and examples described in this book, the authors recommend practicing with the example names and objects created/provided in this book***).

Next, the imported dataset is ready to be analyzed. As defined in Fig. 3.3 (see Step 3, Lines 14–20), we can now perform the Kolmogorov–Smirnov and Shapiro–Wilk normality tests.

The syntax for performing the two tests in R is by using the following functions: **ks.test( )** for Kolmogorov–Smirnov, and **shapiro.test( )** for Shapiro–Wilk tests, respectively.

As shown in the codes provided below, the authors used the methods, i.e., **ks.test( )** and **shapiro.test( )** to check for normality of the distribution for the variables named "**Units.Sold**" and "**Month.Number**" in the "**MyTestData"** dataset.

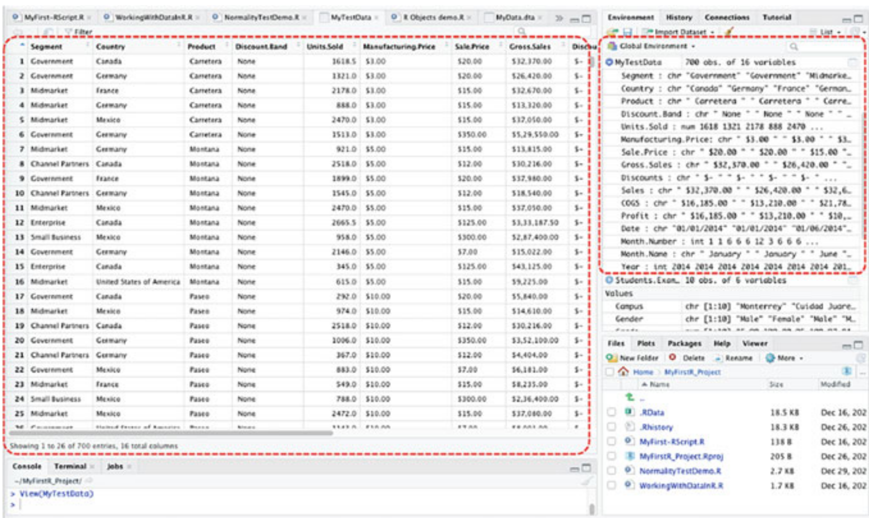


**Fig. 3.4** Example of file or dataset imported in R

# # Step 3—Analyze the Dataset

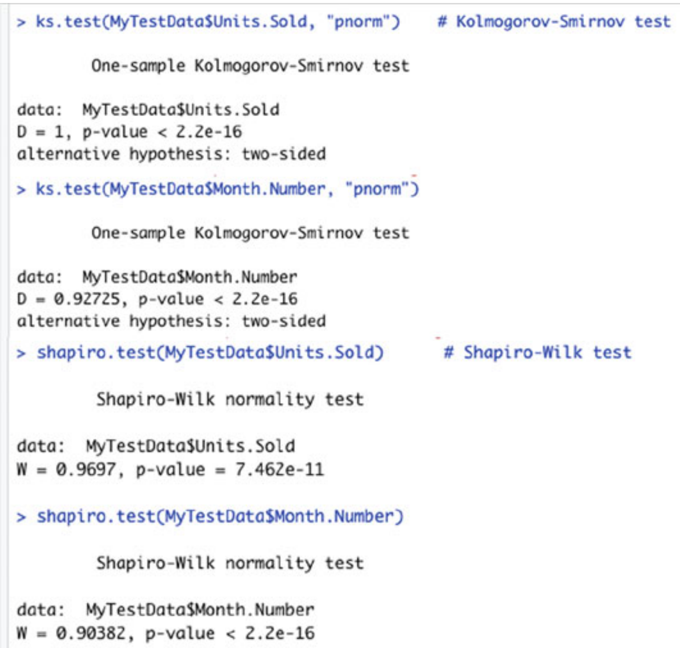
```
ks.test(MyTestData$Units.Sold, "pnorm")    # Kolmogorov-Smirnov test
ks.test(MyTestData$Month.Number, "pnorm")

shapiro.test(MyTestData$Units.Sold)       # Shapiro-Wilk test
shapiro.test(MyTestData$Month.Number)
```

When the user has successfully run the tests (see Lines 14–20, Fig. 3.3), you will be presented with the results in the Console tab, similar to the one shown in Fig. 3.5.

Furthermore, depending on the type of research or variables that are being considered or analyzed by the researchers or analysts (which the book will cover in Part II), the users can decide to check for the normality of distribution of all the variables in the dataset (e.g., MyTestData) all at once.

***To do this, we must carry out an important step which is to inspect the dataset and ensure to convert the *non-numeric* variables to *numeric* form.

As illustrated in Fig. 3.6, the users can use the **str( )** function to view a full list of the different variable(s) names and types, including the total number of observations and the total number of variables (note: this information can also be viewed through the Environment Tab).



```
> ks.test(MyTestData$Units.Sold, "pnorm")    # Kolmogorov-Smirnov test

        One-sample Kolmogorov-Smirnov test

data:  MyTestData$Units.Sold
D = 1, p-value < 2.2e-16
alternative hypothesis: two-sided
> ks.test(MyTestData$Month.Number, "pnorm")

        One-sample Kolmogorov-Smirnov test

data:  MyTestData$Month.Number
D = 0.92725, p-value < 2.2e-16
alternative hypothesis: two-sided
> shapiro.test(MyTestData$Units.Sold)       # Shapiro-Wilk test

        Shapiro-Wilk normality test

data:  MyTestData$Units.Sold
W = 0.9697, p-value = 7.462e-11

> shapiro.test(MyTestData$Month.Number)

        Shapiro-Wilk normality test

data:  MyTestData$Month.Number
W = 0.90382, p-value < 2.2e-16
```

**Fig. 3.5** Results of the Kolmogorov–Smirnov and Shapiro–Wilk tests for data normality displayed in the Console tab in R

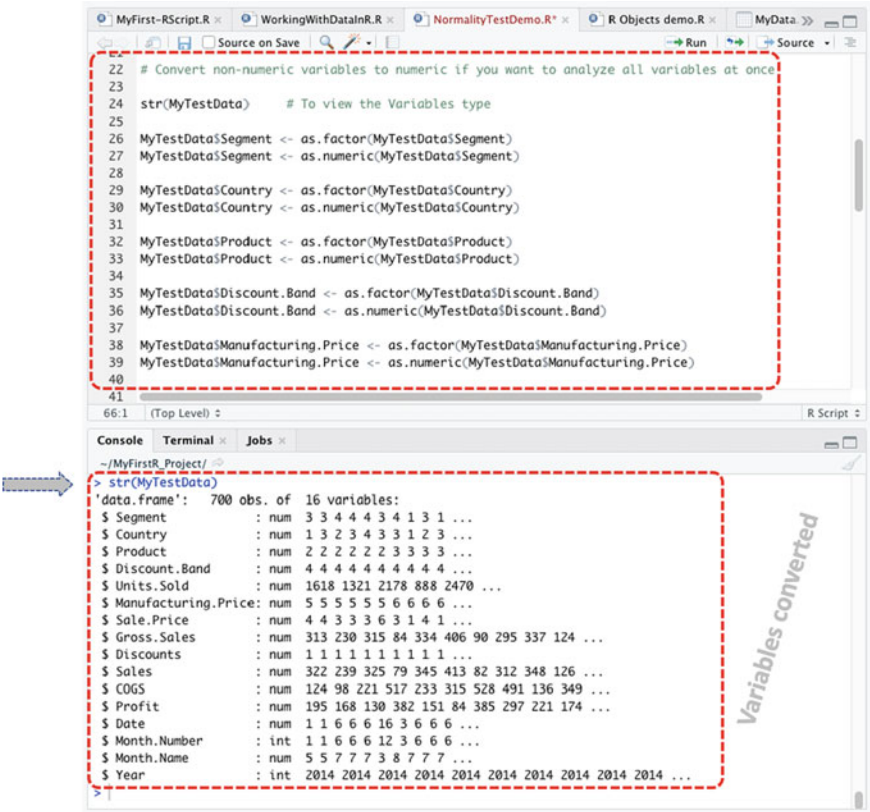



**Fig. 3.6** Converting non-numeric variables or factors to numeric form

Based on the example data that we are working with (see: Figs. 3.2 and 3.4), the authors have implemented the following codes (Fig. 3.6) to convert the "non-numeric (chr)" variables to "numeric (num)" form. (***Note: the readers can refer to Sect. 2.5.5 in Chap. 2 for more detailed description on how to create and perform the *factorization* procedures in R).

# Converting non-numeric variables or factor to numeric form

```
str(MyTestData)     # To view the Variables type

MyTestData$Segment <- as.factor(MyTestData$Segment)
MyTestData$Segment <- as.numeric(MyTestData$Segment)

MyTestData$Country <- as.factor(MyTestData$Country)
MyTestData$Country <- as.numeric(MyTestData$Country)

MyTestData$Product <- as.factor(MyTestData$Product)
MyTestData$Product <- as.numeric(MyTestData$Product)

MyTestData$Discount.Band <- as.factor(MyTestData$Discount.Band)
MyTestData$Discount.Band <- as.numeric(MyTestData$Discount.Band)

MyTestData$Manufacturing.Price <- as.factor(MyTestData$Manufacturing.Price)
MyTestData$Manufacturing.Price <-as.numeric(MyTestData$Manufacturing.Price)

MyTestData$Sale.Price <- as.factor(MyTestData$Sale.Price)
MyTestData$Sale.Price <- as.numeric(MyTestData$Sale.Price)

MyTestData$Gross.Sales <- as.factor(MyTestData$Gross.Sales)
MyTestData$Gross.Sales <- as.numeric(MyTestData$Gross.Sales)

MyTestData$Discounts <- as.factor(MyTestData$Discounts)
MyTestData$Discounts <- as.numeric(MyTestData$Discounts)

MyTestData$Sales <- as.factor(MyTestData$Sales)
MyTestData$Sales <- as.numeric(MyTestData$Sales)

MyTestData$COGS <- as.factor(MyTestData$COGS)
MyTestData$COGS <- as.numeric(MyTestData$COGS)

MyTestData$Profit <- as.factor(MyTestData$Profit)
MyTestData$Profit <- as.numeric(MyTestData$Profit)

MyTestData$Date <- as.factor(MyTestData$Date)
MyTestData$Date <- as.numeric(MyTestData$Date)

MyTestData$Month.Name <- as.factor(MyTestData$Month.Name)
MyTestData$Month.Name <- as.numeric(MyTestData$Month.Name)

str(MyTestData)      # To view the new and converted variables
View(MyTestData)
)
```

Once the user has successfully converted the "**chr**" (character) variable(s) to "**num**" (number) (see Fig. 3.6), we can now run the Kolmogorov–Smirnov and Shapiro–Wilk normality tests for all the elements in the data by using, for instance, the **lapply( )** command as demonstrated code below and results in Fig. 3.7.

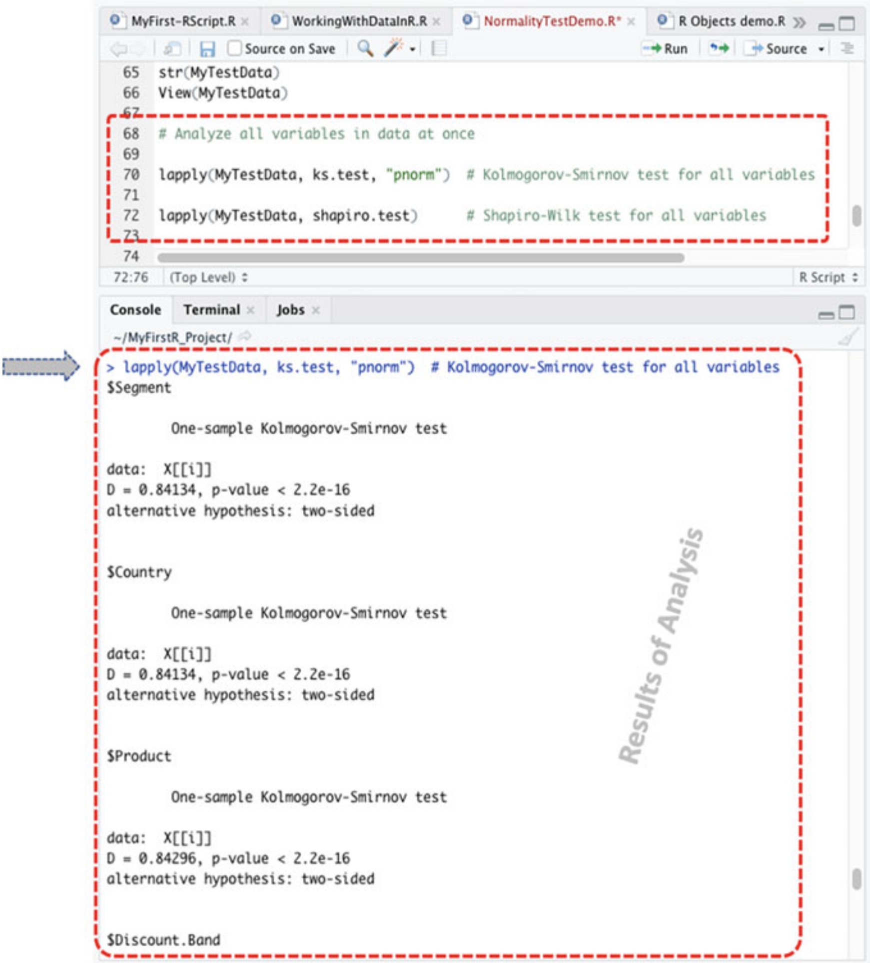



**Fig. 3.7** Conducting Kolmogorov–Smirnov and Shapiro–Wilk normality tests for all elements (variables) in the data

## # Analyze all the variables in the dataset all at once

```
lapply(MyTestData, ks.test, "pnorm") # Kolmogorov-Smirnov test for all variables

lapply(MyTestData, shapiro.test)     # Shapiro-Wilk test for all variables
```

## # Step 4—Visualize Data Normality as a Plot (Graphical Representation)

Another important and useful way to visually check the normality of data in R is by plotting the distribution of the different variables by using, for instance, the “density plot” and “quantile–quantile plot” functions.

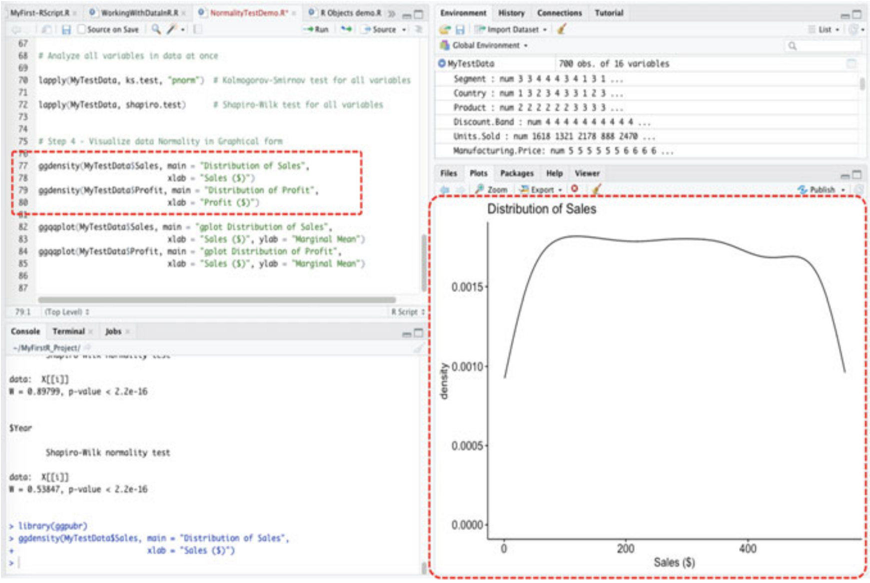



**Fig. 3.8** Density plot of the example dataset

To describe the different functions, the *density plot* (see Fig. 3.8) gives a visual judgment with respect to whether the distribution of the data is "bell-shaped" or "skewed". The *quantile–quantile plot* (Q-Q plot) (Fig. 3.9) tends to draw a correlation between the specified sample and the normality of the distribution. The Q-Q plot also includes a reference line that is usually plotted or mapped at 45 degrees. Thus, each observation is plotted as a single dot, and the dots should form a straight line if the data is normal.

To demonstrate to the readers how to do this, we will use the **ggdensity( )** and **ggqqplot( )** functions, which are supported by the "**ggpubr**" package (see Step 1), to visualize the normality of the example data (MyTestData), respectively. As shown in Figs. 3.8 and 3.9, we will be using the "**Sales**" and "**Profit**" variables in the example data (MyTestData) to demonstrate this. ***Also, feel free to try and plot the other variable(s) in the data using this particular example.

The syntax and code for the two example plots using the ggdensity( ) and ggqqplot( ) functions are as shown below, and the results (distribution graph) are as represented in Figs. 3.8 and 3.9.

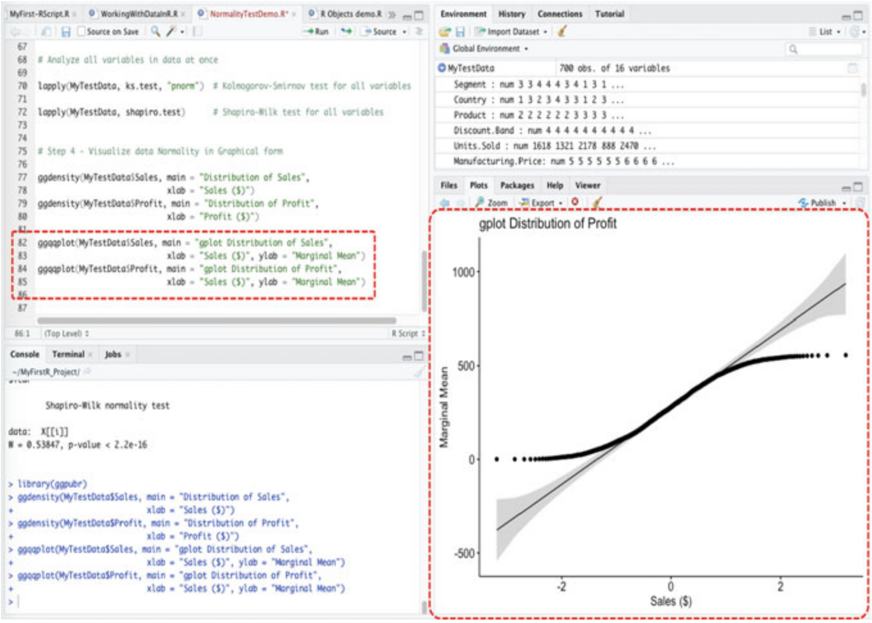


**Fig. 3.9** Quantile–quantile plot (Q-Q plot) of the example data

## # Visualizing the normality of data in Graphical form

```
ggdensity(MyTestData$Sales, main = "Distribution of Sales",
                            xlab = "Sales ($)")
ggdensity(MyTestData$Profit, main = "Distribution of Profit",
                            xlab = "Profit ($)")


ggqqplot(MyTestData$Sales, main = "gplot Distribution of Sales",
                            xlab = "Sales ($)", ylab = "Marginal Mean")
ggqqplot(MyTestData$Profit, main = "gplot Distribution of Profit",
                            xlab = "Sales ($)", ylab = "Marginal Mean")
```

## Useful Tips and Information:

1. The users can use the "**Export**" menu found there on the same **Plot Tab** (see: Figs. 3.8 and 3.9) to save the plot (graphics) as image or pdf file on the local computer for later or further use in presentations or write-up.
2. Use the histogram function, **hist( )** to experiment and view the distribution of the different variables. For example; `hist(MyTestData$Sales)`.

**# Step 5—Results Interpretation for the Data Normality Test**

The last step in the data normality test or analysis is to interpret or understand the results of the test. By default, the null hypothesis for testing whether a dataset is normally distributed is IF the test statistics or value is greater than 0.5, and the p-value greater than the 0.05 threshold (i.e, $p > 0.05$), THEN normality is assumed, ELSE IF $p \leq 0.05$ THEN normality is not assumed.

As shown in the example and results described below by using the variable called "**Segment**" in the example data (MyTestData); the following outputs `D = 0.84134` and `W = 0.89536` represents the value of the statistics for the Kolmogorov–Smirnov and Shapiro–Wilk tests, respectively. While the `p-value < 2.2e-16` (i.e., $p = 0.00$) which was found for both test, represents the significant level (referred to as p-values) of the corresponding tests.

```
> lapply(MyTestData, ks.test, "pnorm")
$Segment
     One-sample Kolmogorov-Smirnov test
data:  X[[i]]
D = 0.84134, p-value < 2.2e-16
alternative hypothesis: two-sided
```

```
> lapply(MyTestData, shapiro.test)
$Segment
     Shapiro-Wilk normality test
data:  X[[i]]
W = 0.89536, p-value < 2.2e-16
```

As reported in the second part of the highlighting in Fig. 3.7 and in the above results, we can see that the **p-values** for each of the variables or test (Kolmogorov: D = 0.84134; Shapiro: W = 0.89536) are less than 0.05 ($p < 0.05$) with the majority of them showing a *p-value of* $p$ < 2.2e-16 (scientifically interpreted as $p = 0.00$). The number or value: 2.2e-16 is the scientific notation of 0.00000000000000022 which means that the values are fundamentally very close to zero.

Therefore, with the above results, we can reject the null hypothesis (i.e., $p > 0.05$), and assume that the example dataset is not normally distributed.

Also, it is important to mention to the readers that datasets which are normally distributed are expected to be "bell-shaped" when graphically represented, whereas non-normally distributed datasets are "skewed". For example, as seen in Figs. 3.8 and 3.9, we can see that the plots are not represented to be bell-shaped (skewed) (Fig. 3.8) or not on a 45 degrees straight line (Fig. 3.9), thus, affirming the statistical results the authors have interpreted earlier (see Fig. 3.7) which shows that the example dataset is not normally distributed.

**Useful Tips and Information:**

1. The Shapiro–Wilk normality test is sensitive and are typically appropriate for small sample sizes.
2. Kolmogorov–Smirnov test is recommended for large sample sizes, for instance, samples greater than 100 ($n > 100$).
3. Shapiro–Wilk test is widely recommended for data normality tests and tends to provide better statistical power than Kolmogorov–Smirnov test. It is based on the correlation between the data and the corresponding normal scores (Ghasemi & Zahediasl, 2012).
4. Parametric procedures or analysis (which the authors will cover in the next chapter—Chap. 4) can still be conducted on large datasets (e.g., $n > 30$ or 40), regardless of whether the specified data violates the assumption of normality (Roscoe, 1975). However, for small sample sizes ($n < 30$) that appear to be non-normally distributed, the *non-parametric* procedures are scientifically recommended.
5. In some situations, the users may get a warning message such as "ties should not be present for the Kolmogorov–Smirnov test" when conducting the Kolmogorov–Smirnov (K-S) test depending on the type of variable(s) being analyzed. K-S test assumes that the datasets are *continuous* (which in some cases are likely not) and therefore tends to generate a warning when it finds the presence of ties. Nonetheless, the diffident sums of parsing or rounding on the variables are still significantly effective on the calculated statistics, and as such are still theoretically estimated to be valid.

## 3.3 Test of Data Reliability in R: Cronbach's Alpha Test

The Cronbach's alpha, α (or coefficient alpha), test is one of the most commonly used method to determine if a dataset or instrument is *reliable*, for instance, for research purposes or the many other types of data analysis and computation. The test (α) is used by researchers to ascertain how closely related a set of item(s) in a dataset are as a group.

By default, the general rule of thumb is that a Cronbach's alpha (α) test result of 0.70 and above is good and acceptable. Meaning that the results or conclusions drawn from analyzing the available data are assumed to be reliable for further analysis and/or drawing scientific conclusions.

Here, we will demonstrate how to conduct Cronbach's alpha test in R. As described in Fig. 3.10, the authors will show two ways or methods of how to perform this test (Cronbach's alpha) using the four defined steps in R:

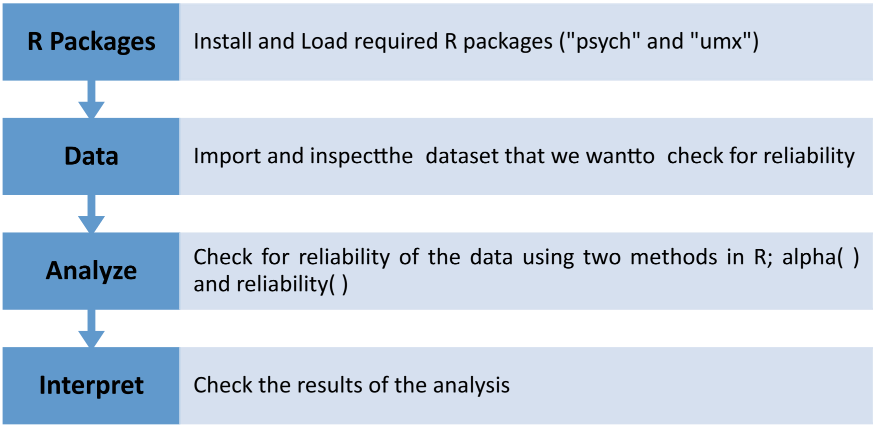



**Fig. 3.10** Steps for conducting Cronbach's alpha test in R

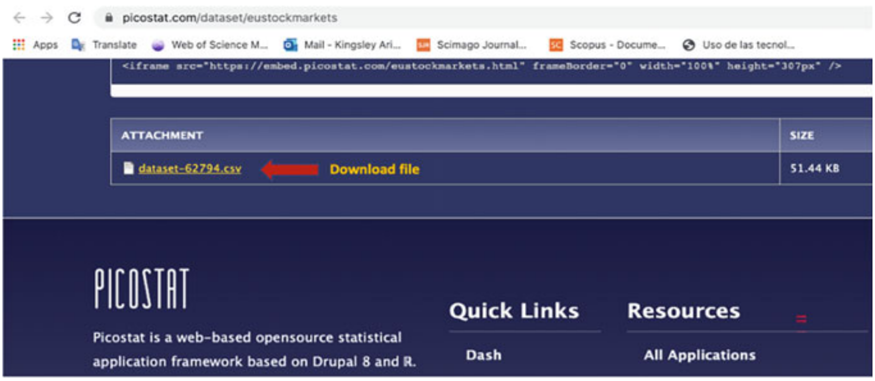



**Fig. 3.11** Example of CVS file download. *Source* https://www.picostat.com/dataset/eustockmarkets)

Now, let's start by creating a new R Script and name it "**CronbachAlphaDemo**" or any name the reader chooses.

Once you have created the R Script, download the example file that we will use to demonstrate the Cronbach's reliability test in R.

As shown in Fig. 3.11, visit the following link (https://www.picostat.com/dataset/eustockmarkets) and download the CVS file named "**dataset-62794.csv**" and save this on your computer desktop. (***the readers are welcome to use any file or format of choice, but for this example, the authors will be using the example CSV file***). The example dataset is also available for download at the following data repository: https://doi.org/https://doi.org/10.6084/m9.figshare.24728073.

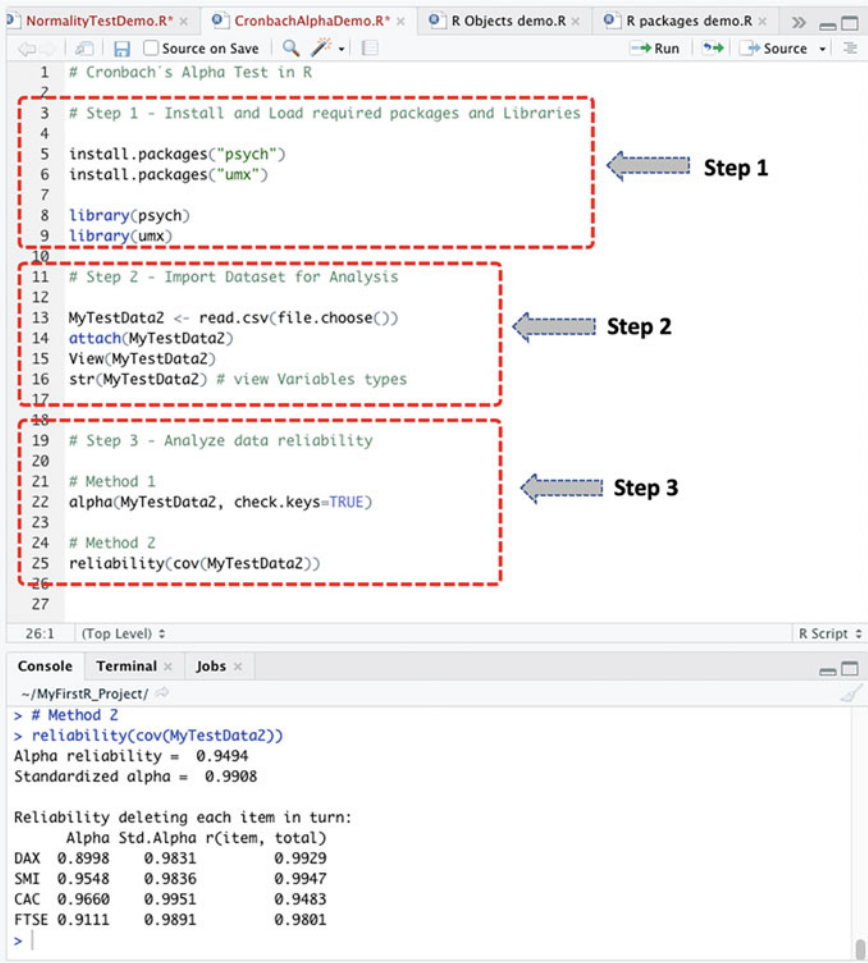



**Fig. 3.12** Cronbach's alpha reliability test in R

Once the file download is completed and the CSV file saved on the local machine or system, e.g., the user's desktop, we can proceed to conduct Cronbach's alpha reliability test by following the steps we have defined in Fig. 3.10.

### # Step 1—Install and Load the Required R Packages

As shown in Fig. 3.12 (see Lines 3–9) **Install and Load** the necessary **R packages** ("**psych**" and "**umx**") that we will be using to conduct Cronbach's alpha test.

The syntax to install and load the "psych" and "umx" packages are as follows:

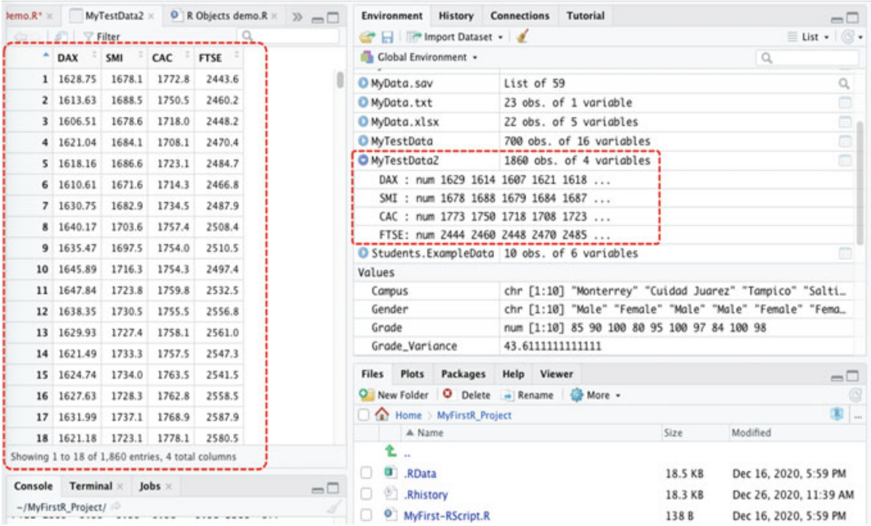


Fig. 3.13 Example of CSV dataset imported in R

```
install.packages("psych")
install.packages("umx")

library(psych)
library(umx)
```

As shown in Fig. 3.12 (Step 1—Lines 5–9), once the user has the necessary packages installed and loaded, the next step is to import the dataset for analysis (Step 2: see code below).

**# Step 2—Import and Inspect Data**

```
MyTestData2 <- read.csv(file.choose())
attach(MyTestData2)
View(MyTestData2)
str(MyTestData2)      # view Variables types
```

As demonstrated in Fig. 3.12 (Step 2), when you run the commands (see Lines 11–16) you will be presented with a window through which you can navigate and choose the.csv file named "**dataset-62794**" that we downloaded earlier and stored on the computer desktop.

Once selected and successful, the data will be imported and stored in an R object defined as "**MyTestData2**" (***remember you can always use any name of your choice***). The users can see the details of the dataset highlighted in Fig. 3.13, with 1860 observations and 4 variables in the data sample.

***Note**: Remember to Convert *non-numeric* variables to *numeric* if using a dataset that contains Character or Categorical values. Users can refer to Sect. 3.2 on how to do this task or Chap. 2 "Working with Data in R (see Sect 2.5.5)" for more details.

**# Step 3—Analyze the Data Reliability using Cronbach's Alpha Methods**

Now we can proceed to analyze the imported dataset (stored as R object we named or defined as—MyTestData2) using the **alpha( )** and **reliability( )** functions in R.

The syntax for performing the reliability tests in R using the two methods is as follows:

**# Method 1**

```
alpha(MyTestData2, check.keys=TRUE)
```

**# Method 2**

```
reliability(cov(MyTestData2))
```

Once the user have successfully run the codes (Lines 19–25), the user will be presented with the results of the tests in the Console, similar to the one illustrated in Fig. 3.14.

**# Step 4–Results Interpretation for Cronbach's Alpha Test**

The final step in the reliability of data analysis is to interpret the results of Cronbach's alpha test. By default, the null hypothesis for testing whether a given dataset or instrument is reliable for research purposes or data analysis is *IF* α (coefficient alpha) is greater than or equal to 0.70 ($\alpha \geq 0.70$), *THEN* reliability of data is statistically good and scientifically acceptable, *ELSE IF* α is less than 0.70 ($\alpha < 0.70$), *THEN* data reliable is questionable. It is also important to mention that these measures can vary based on different context scenarios or research experimental settings.

As highlighted in the results we derived by using the example dataset (defined as MyTestData2) (see: Fig. 3.14), the statistics (reliability result) of the data is $\alpha = 0.95$ (for Method 1) and $\alpha = 0.9494$ (for Method 2) with std. alpha of 0.99 (Method 1) and 0.9908 (Method 2), respectively. Therefore, we can accept the null hypothesis, i.e., $\alpha \geq 0.70$, and assume that the tested or analyzed dataset is reliable for any statistical analysis or research experiments.

## 3.4 Summary

In this chapter, the authors looked at the preliminary, yet, important tests that are conducted by the researchers when carrying out the experimentations or statistical data analysis. This consists of the process of testing the available datasets for

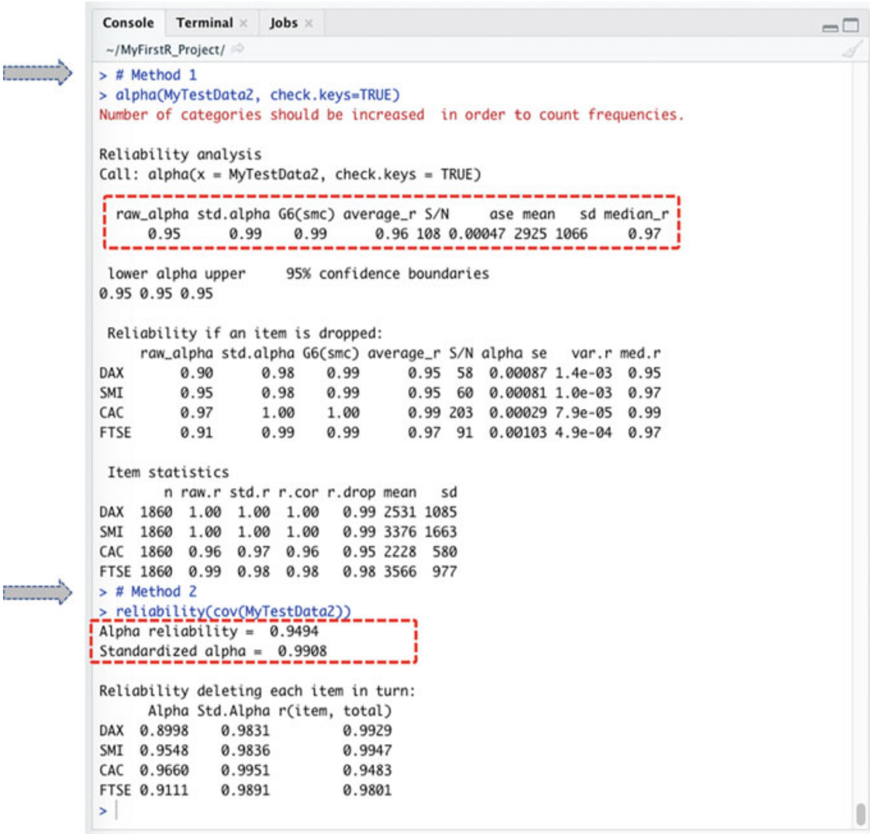



**Fig. 3.14** Results of Cronbach's alpha tests displayed in the Console in R

*normality of distribution* and *reliability of the data sample* for scientific research purposes. In theory, these tests (i.e., normality and reliability) are done to assess the extent to which the data is capable of producing coherent and assertive research results or conclusions. In Sect. 3.2, the authors demonstrated how to conduct the *Kolmogorov–Smirnov* and *Shapiro–Wilk* normality tests in R, which in comparison to the other mentioned types of tests, is the most commonly used method by the researchers to check the distribution of the data. In Sect. 3.3, we illustrated how to conduct *Cronbach's alpha* reliability test using two different methods in R. In each case (Sect. 3.2 and 3.3), we discussed the meaning of the test statistics and how to interpret the results of the different tests (Kolmogorov–Smirnov, Shapiro–Wilk, and Cronbach's alpha) in R.

## References


Carpentier, M., Combescure, C., Merlini, L., & Perneger, T. V. (2017). Kappa statistic to measure agreement beyond chance in free-response assessments. *BMC Medical Research Methodology, 17*(1), 1–8. https://doi.org/10.1186/S12874-017-0340-6

de Barros Ahrens, R., da Silva Lirani, L., & de Francisco, A. C. (2020). Construct validity and reliability of the work environment assessment instrument WE-10. *International Journal of Environmental Research and Public Health, 17*(20), 7364. https://doi.org/10.3390/ijerph17207364

Elliott, A. C., & Woodward, W. A. (2007). *Statistical analysis quick reference guidebook with SPSS examples* (1st ed.). Sage Publications.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism, 10*(2), 486–489. https://doi.org/10.5812/ijem.3505

Goni, M. D., Naing, N. N., Hasan, H., Wan-Arfah, N., Deris, Z. Z., Arifin, W. N., Hussin, T. M. A. R., Abdulrahman, A. S., Baaba, A. A., & Arshad, M. R. (2020). Development and validation of knowledge, attitude and practice questionnaire for prevention of respiratory tract infections among Malaysian Hajj pilgrims. *BMC Public Health, 20*(1), 1–10. https://doi.org/10.1186/s12889-020-8269-9

Isnainiyah, I. N., Yulnelly, Y., & Balqis, A. N. (2019). Usability analysis using principal component analysis (PCA) method for online fish auction application. In *Proceedings–1st international conference on informatics, multimedia, cyber and information system, ICIMCIS 2019* (pp. 231–236). https://doi.org/10.1109/ICIMCIS48181.2019.8985225

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia, 22*(1), 67–72. https://doi.org/10.4103/aca.ACA_157_18

Öztuna, D., Elhan, A. H., & Tüccar, E. (2006). Investigation of four different normality tests in terms of type 1 error rate and power under different distributions. *Turkish Journal of Medical Sciences, 36*(3). The Scientific and Technological Research Council of Turkey.

Pallant, J. (2007). *SPSS survival manual, a step by step guide to data analysis using SPSS for windows* (3rd ed.). McGraw Hill.

Peat, J., & Barton, B. (2005). *Medical statistics: A guide to data analysis and critical appraisal*. Wiley.

Roscoe, J. T. (1975). *Fundamental research statistics for the behavioral sciences* (2nd ed.). Holt, Rinehart, and Winston.

Rstudio. (2023). *RStudio–Statistics*. https://rstudio.com/products/rstudio/

Taber, K. S. (2018). The use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education, 2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Thode, H. C. (2019). *Testing for normality*. CRC Press, Taylor and Francis group.

Vaclavik, M., Sikorova, Z., & Barot, T. (2020). *Skewness in applied analysis of normality* (pp. 927–937). Springer, Cham. https://doi.org/10.1007/978-3-030-63319-6_86