Chapter 11
Mann–Whitney U Test and Kruskal–Wallis H Test Statistics in R

11.1 Introduction

The likely effect of the *independent* variable(s) over a *dependent* variable can be analyzed or determined using the "Mann–Whitney U" and "Kruskal–Wallis H" tests. The two tests are used to determine if there exist statistically differences (significant levels usually measured through the p-values, where $p \leq 0.05$) between the independent observations in a given dataset based on the dependent variable or observation. In theory, the tests (Mann–Whitney U and Kruskal–Wallis H) are referred to as *non-parametric* procedures or methods used by the researchers or data analysts to statistically determine whether a group of data comes from the same population by considering the effect of the independent variable on the dependent variable (Frey, 2018; le Cessie et al., 2020; MacFarland et al., 2016; McKight & Najab, 2010; McKnight & Najab, 2010; Nachar, 2008; Okoye et al., 2022; Ortega, 2023; Ostertagová et al., 2014; Vargha & Delaney, 1998).

Furthermore, just like many of the other types of the "non-parametric" procedures, the two tests (Mann–Whitney U and Kruskal–Wallis H) are usually applied when the data sample in question are not normally distributed (i.e., violates the assumption of *t*-distribution) or the data sample size is too small to conduct the parametric methods or procedures (see Chaps. 3 and 4). Thus, while the measurement to establish whether the independent groups of variables being analyzed comes from the same population via the "mean" for the parametric procedures, the non-parametric equivalents or tests (such as the *Mann–Whitney U* and *Kruskal–Wallis H*), on the other hand, are measured by considering the "median" (see Chap. 4).

By definition, the *Mann–Whitney U* test, also known as the *U* test, is used to determine the differences in *median* between *two groups* of an independent variable with no specific distribution on a single ranked scale, and must be ordinal variable data type (McKnight & Najab, 2010; Ramtin, 2023). The test (Mann–Whitney) is often considered as the *non-parametric* version or equivalent of the Independent Samples

K. Okoye and S. Hosseini, *R Programming*,
https://doi.org/10.1007/978-981-97-3385-9_11

t-test (a type of parametric test). Moreover, while the *t*-test (parametric) and Mann–Whitney U (non-parametric) tests may show to serve the same statistical purposes, due to the fact that they are both used to determine if there exists a statistically significant differences between the *two groups of an independent variable*. On the contrary, the Mann–Whitney U test is used with "ordinal" or "ranked" datasets that may have violated the assumptions of normality or small sample size, whereas, the *t*-test is used with "continuous" or "interval" datasets that happen to meet the assumptions of normality or large sample size (MacFarland et al., 2016). Therefore, the Mann–Whitney U test is most suitable when the data that is being analyzed by the researcher or data analyst is in ranked form, deviates from the acceptable *t*-distribution, or the probability that a randomly drawn member(s) of the first group (e.g., group A) of the population will exceed the second group (group B) of the population in a single independent variable or data (see: Chap. 6, Sect. 6.2.6).

Mathematically, the result of applying the Mann–Whitney U test is a U-Statistic or formula represented as follows (Mann & Whitney, 1947; Nachar, 2008):

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

where:

R_1= sum of the ranks for group 1
R_2 = sum of the ranks for group 2
n_1 = number of observations or participants for group 1
n_2 = number of observations or participants for group 2

As seen in the formula above, it is noteworthy to mention that the Mann–Whitney U statistics involves pooling the observations from the two groups of samples (e.g., group A and group B) into one combined sample, done by keeping track of which sample each observation comes from, and then *ranking* them according to lowest to highest, i.e., from 1 to $R_1 + R_2$, respectively.

On the other hand, the *Kruskal–Wallis* test, also referred to as *H* test, is described as an extension of the two-grouped Mann–Whitney U test (McKight & Najab, 2010). Thus, the method (Kruskal–Wallis H) (see Chap. 6, Sect. 6.2.8) is used when the researcher is comparing the median of more than two groups (i.e., three or more categories) of independent samples (Ortega, 2023; Vargha & Delaney, 1998). Just like the Mann–Whitney U test, the Kruskal–Wallis H method uses *ranked* (ordinal) datasets, a powerful alternative (non-parametric version) to the One-way analysis of variance (ANOVA), and proves to be a suitable statistical method when the data sample in question deviates from the acceptable *t*-distribution or is not normally distributed (Ostertagová et al., 2014).

Mathematically, the result of applying the Kruskal–Wallis test is an H-Statistic or formula represented as follows (Kruskal & Wallis, 1952; McKight & Najab, 2010):

$$H = \left(\frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} \right) - 3(n+1)$$

where:

$$\begin{aligned} k &= number\ of\ groups\ being\ compared\ or\ analyzed \\ n &= total\ sample\ size \\ n_j &= sample\ size\ in\ the\ jth\ group \\ R_j &= sum\ of\ the\ ranks\ in\ the\ jth\ jth\ group \end{aligned}$$

As defined in the above formula, it is noteworthy to mention that with the Kruskal–Wallis H statistics, all of the n values or measurements (e.g., $n = n_1 + n_2 + \ldots + n_k$) are jointly ranked (i.e., are treated as one single sample), and one can use the sums of the ranks of the k samples to compare the distributions.

Accordingly, in Table 11.1 the authors provide a summary of the differences and similarities between the Mann–Whitney U test and Kruskal–Wallis H test, including the different conditions that are necessary or required to performing the tests, which are practically demonstrated in R in the next sections of this chapter (Sects. 11.2 and 11.3).

Table 11.1 Differences and similarities between the Mann–Whitney U test and Kruskal–Wallis H tests and Assumptions

Mann–Whitney U	Kruskal–Wallis H
Independent variable must be of two levels or groups, e.g., group A and group B	Independent variable must be more than two levels or groups (i.e., three or more), e.g., group A, group B, group C, …group n^{th}
Used for Ranked or Ordinal datasets	Used for Ranked or Ordinal datasets
Dependent variable should be measured on an ordinal or continuous scale	Dependent variable should be measured on an ordinal or continuous scale
Data sample or observations are not normally distributed, i.e., skewed	Data sample or observations are not normally distributed, i.e., skewed
Data must be independent and randomly drawn from the population, i.e., no relationship should exist between the two groups or within each group	Data must be independent and randomly drawn from the population, i.e., no relationship should exist between the groups (minimum of three or more groups) or within each group
Measures or compares the "median", unlike the parametric counterparts that compare the "mean"	Measures or compares the "median", unlike the parametric counterparts that compare the "mean"
Non-parametric equivalent or version of the Independent sample *t*-test	Non-parametric equivalent or version of the One-way ANOVA test

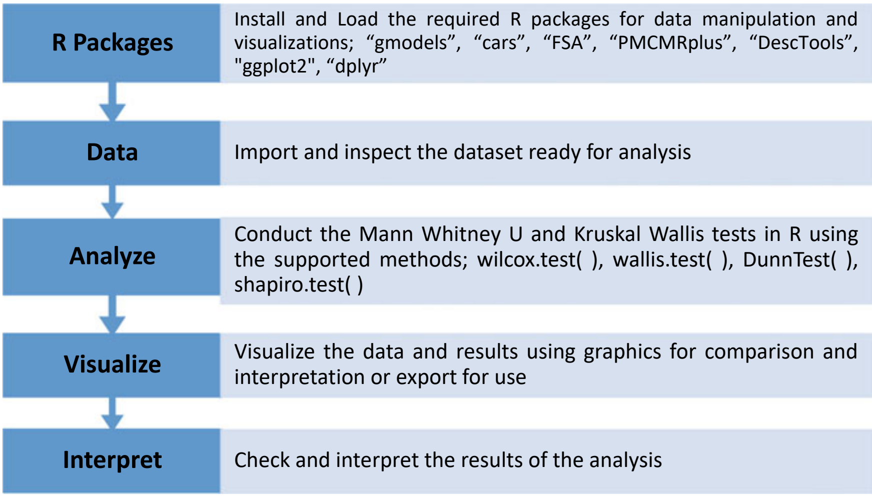

Fig. 11.1 Steps to conducting Mann–Whitney U and Kruskal–Wallis H tests in R

In the next sections of this chapter (Sects. 11.2 and 11.3), the authors will be demonstrating to the readers how to conduct the "Mann–Whitney U" and "Kruskal–Wallis H" tests in R, respectively. We will illustrate the different steps to performing the two tests in R by using the following steps as outlined in Fig. 11.1.

11.2 Mann–Whitney U Test in R

Mann–Whitney U test is used when the data the researcher or analysts wants to analyze are made up of *two groups* and are statistically *independent*. Statistically, the test is used to compare the differences in *median* between an ordinal independent variable, and an ordinal or continuous dependent variable; whereby the independent variable must have two (ranked) levels. As defined earlier in Sect. 11.1, the test (Mann–Whitney U) is distribution free and has the powerful advantage of being used to analyze small sample sizes.

By default, the hypothesis for testing whether there is a *difference in the median* of the two specified groups of independent data (ordinal) against the dependent (usually ordinal or continuous) variable is; *IF* the p-value of the test (Mann–Whitney U) is less than or equal to 0.05 ($p \leq 0.05$), *THEN* we can assume that at least one sample of the two groups being analyzed comes from a population with a different distribution than the other, thus, the *median* of the groups of population (two groups) in the data sample are statistically different (varies), or yet in other words, that the first group are significantly larger than those of the second, or vice and versa, 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 assume that there is no difference in the *median* of the two groups, thus, the two independent groups are homogeneous and have the same distribution (stochastically equal), and any observed difference could only occur by chance (H_0).

Here, we will be demonstrating how to perform the Mann–Whitney U test by using the **wilcox.test( )** function in R. We will do this using the steps outlined in Fig. 11.1.

To begin, **Open RStudio** and **Create a new or Open an existing project**. Once the user have the RStudio and an R Project opened, **Create a new R Script** and name it "**MannWhitneyDemo**" or any name the user chooses (see Chap. 1 if the readers need to refresh on this step or topic).

We will download an example dataset that we will be using to demonstrate both the Mann–Whitney U test in this section (Sect. 11.2) and the Kruskal–Wallis H test in the next section (Sect. 11.3). ***Note: the users can use any dataset or format if they wish to do so***.

As shown in Fig. 11.2, download the example data named "**Sample CSV Files**" from the following link (source: https://www.learningcontainer.com/sample-excel-data-for-analysis/#Sample_CSV_file_download) if the user have not done so in the previous chapter (Chap. 10), and save the downloaded file on the computer. ***Also, the example dataset can accessed and downloaded via the following link (https://doi.org/https://doi.org/10.6084/m9.figshare.24728073) where the authors have uploaded all the example files used in this book.

Once the user have successfully downloaded and/or accessed the example file on the computer, we can proceed to conduct the Mann–Whitney U test in R.

Step 1—Install and Load the required R Packages and Libraries

Install and **Load** the following *R packages* and *libraries* (Fig. 11.3, Step1, Lines 3 to 15) that will be used to call the different R functions, data manipulations, and graphical visualizations for the Mann–Whitney U test.

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

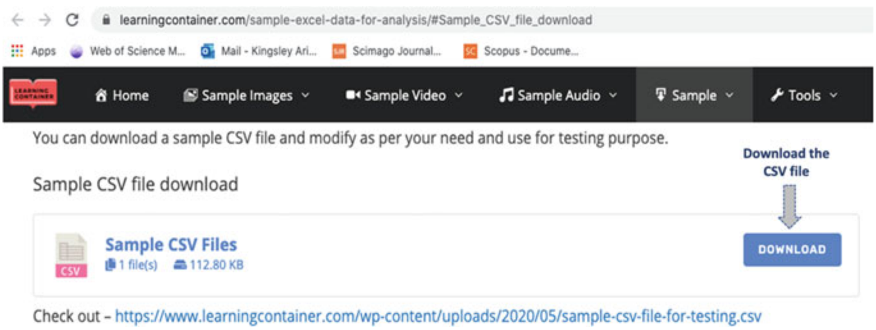

Fig. 11.2 Example of CSV data sample and file download. *Source* https://www.learningcontainer.com/sample-excel-data-for-analysis/#Sample_CSV_file_download

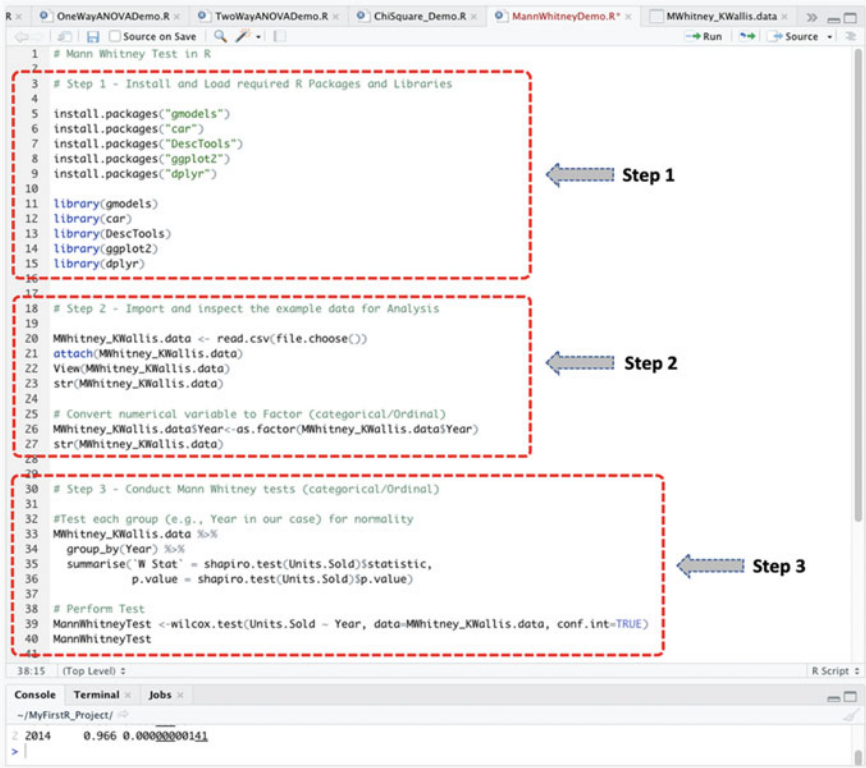

Fig. 11.3 Conducting Mann–Whitney U test in R

```
install.packages("gmodels")
install.packages("car")
install.packages("DescTools")
install.packages("ggplot2")
install.packages("dplyr")

library(gmodels)
library(car)
library(DescTools)
library(ggplot2)
library(dplyr)
```

Step 2—Import and Inspect the example dataset for Analysis.

As shown in the Step 2 in Fig. 11.3 (Lines 18 to 27), import the example dataset named "sample-csv-file-for-testing" (**Sample CSV Files**) that we have downloaded earlier on the computer, and store this as an R object named "**MWhitney_KWallis.data**" in RStudio (***the users can use any name of choice if they wish to do so).

The code for importing and storing of the example file in R is as shown below:

```
MWhitney_KWallis.data <- read.csv(file.choose())
attach(MWhitney_KWallis.data)
View(MWhitney_KWallis.data)
str(MWhitney_KWallis.data)
```

Once the user has successfully imported and stored the dataset in RStudio environment, you will be able to view the details of the example file "sample-csv-file-for-testing" (Sample CSV Files) named "**MWhitney_KWallis.data**" in the file environment as shown in Fig. 11.4 with 700 observations and 16 variables contained in the stored data sample (see data fragment in Fig. 11.4).

Note: the authors have also perfomed an important step by converting the variable named "Year" with two levels or groups (i.e., 2013, 2014) (see: Figs. 11.3 and 11.4) to a Categorial (ordinal) variable as we will be using this to illustrate the Mann–Whitney U test (see code below—Step 2, Fig. 11.3).

```
#Convert numerical variable to Factor (categorical/Ordinal)
MWhitney_KWallis.data$Year<-as.factor(MWhitney_KWallis.data$Year)
str(MWhitney_KWallis.data)
```

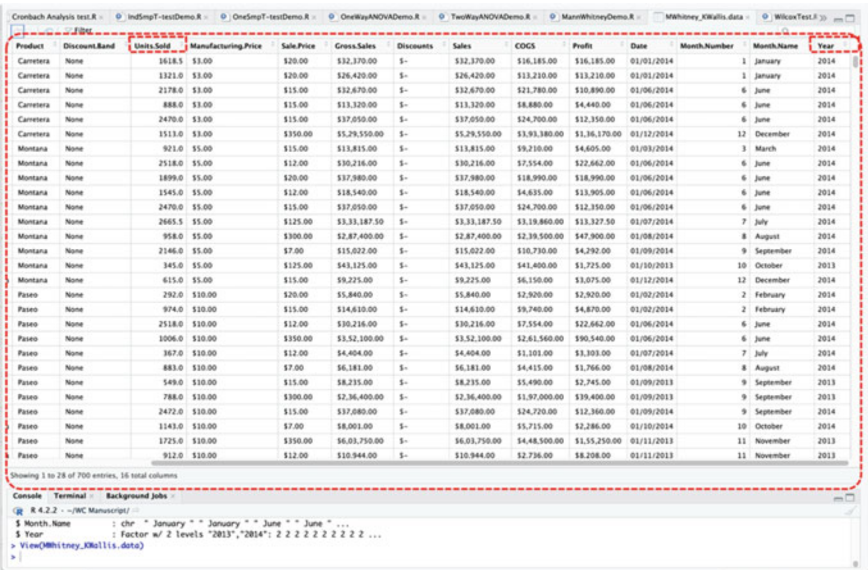

Fig. 11.4 Example of CSV data imported and stored as an R object in RStudio

 # Step 3—Conduct Mann–Whitney U test (used for Categorical/Ordinal variable).

With the example dataset stored and the targeted variables in categorical/ordinal scale, we can proceed to analyze the data which we have stored as **MWhitney_KWallis.data** in R (see: Fig. 11.4).

As defined earlier in the Introduction section (Sect. 11.1):

- **Mann–Whitney U** test or statistics compares the *median or distribution* between two groups of an independent variable against a target dependent variable.
- The targeted independent variable must be an “ordinal” data type.
- The targeted dependent variable must be an “ordinal or continuous” data type.

To demonstrate the Mann–Whitney U test using the **wilcox.test( )** method in R:

- We will test whether the median or distribution of the “**Year**” variable (independent variable with two levels: 2013 and 2014) differ based on the “**Units.Sold**” (dependent variable) in the data.

The syntax and code to conducting the above test (Mann–Whitney) in R is as shown in the code below and represented in Fig. 11.3 (Step 3, Lines 30 to 40).

```
# Step 3 - Conduct Mann Whitney tests (categorical/Ordinal)

#Test each group (e.g., Year in our case) for normality
MWhitney_KWallis.data %>%
  group_by(Year) %>%
  summarise(`W Stat` = shapiro.test(Units.Sold)$statistic,
            p.value = shapiro.test(Units.Sold)$p.value)

# Perform Test
MannWhitneyTest <-wilcox.test(Units.Sold ~ Year, data =
MWhitney_KWallis.data, conf.int=TRUE)
MannWhitneyTest
```

As shown in the code above, we first conducted a data normality test (see Chap. 3) (estimated acceptable p-value > 0.05) by considering each group of the “**Year**” variable before conducting the Mann–Whitney U test. This was done in order to confirm that the dataset does not meet the assumption of normality usually attributed to the Mann–Whitney U test (a non-parametric test), where: neither the 2013 (W=0.951, p=0.00000915) nor the 2014 (W=0.966, p=0.00000000141) groups of the Year variable were normally distributed, otherwise the user would have preferably conducted the Independent sample t-test (parametric equivalent of the Mann–Whitney U) in the event that the data appear to be normally distributed.

Once the user have successfully run the codes provided above (Step 3, Lines 30 to 40, Fig. 11.3), the user will be presented with the results of the Mann–Whitney method in the Console as shown in Fig. 11.5.

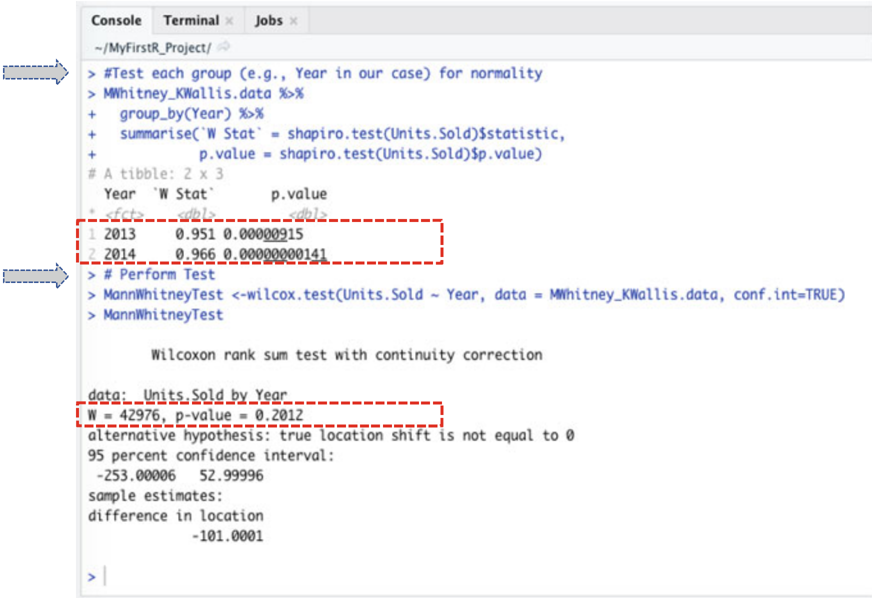

Fig. 11.5 Results of Mann–Whitney U test displayed in the Console in R

To describe the test of assumptions, in Fig. 11.5, we conducted a normality test by considering the two groups (2013, 2014) of the "**Year**" variable against the "**Units.Sold**" variable using the **shapiro.test( )** method or function in R. As highlighted in the figure (Fig. 11.5), the result of the assumption test shows that the dataset taking into account the two groups of the variable was not normally distributed (where a significant level is considered values whereby $p > 0.05$). Therefore, we can assume that the data meets the condition to perform the Mann–Whitney U test with Group A (2013) showing a normality test statistic of `W=0.951, p-value=0.00000915`, and Group B (2014) showing `W=0.966, p-value=0.00000000141`, respectively.

Consequentially, we proceeded to perform the Mann–Whitney U test for the independent variable "**Year**" (with two levels or group) against the dependent variable "**Units.Sold**" as contained in the dataset "sample-csv-file-for-testing" (**Sample CSV Files**) that we stored as an R object named "**MWhitney_KWallis.data**" in R. Accordingly, the result of the Mann–Whitney U statistics was stored as an R object we named or defined as `MannWhitneyTest` (see: Fig. 11.5) which the authors will subsequently discuss in detail in Step 5 in this section.

Step 4—Plot and visualize the data distribution and results.

In Fig. 11.6 (Step 4, Lines 43 to 50), the authors made use of the **ggplot( )** function in R to display a boxplot of the distribution between the two groups (2013, 2014)

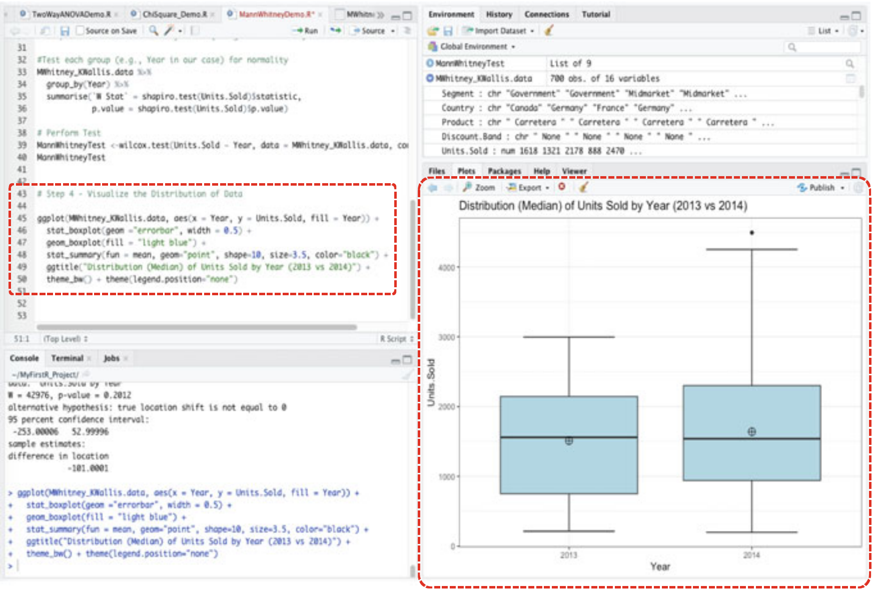

Fig. 11.6 Plot representing the distribution of the two groups of Year variable broken down by Units.Sold using the ggplot() function in R

of the "**Year**" variable against the **"Units.Sold"** as contained in the stored data "**MWhitney_KWallis.data**".

The syntax and code used to plot and visualize the distribution is as shown in the code below, and the resultant plot is represented in Fig. 11.6.

```
# Step 4 - Visualize the Distribution of Data
ggplot(MWhitney_KWallis.data, aes(x = Year, y = Units.Sold, fill = Year)) +
  stat_boxplot(geom ="errorbar", width = 0.5) +
  geom_boxplot(fill = "light blue") +
  stat_summary(fun = mean, geom="point", shape=10, size=3.5, color="black")
+
  ggtitle("Distribution (Median) of Units Sold by Year (2013 vs 2014)") +
  theme_bw() + theme(legend.position="none")
```

 # Step 5—Results Interpretation (Mann–Whitney U).

The final step for the Mann–Whitney U test and analysis is to interpret and understand the results of the test/method.

By default, the hypothesis for conducting the test (Mann–Whitney) considering the analyzed variables "**Year**" and "**Units.Sold**" (see: Fig. 11.4) is;

- **(H$_1$)** *IF* the p-value of the test is less than or equal to 0.05 ($p \leq 0.05$), *THEN* we can assume that there is a difference in the distribution of the two groups of the “Year” variable (2014, 2014) taking into account the “Units.Sold”. Thus, the *median* of the two groups of population (2013, 2014) are statistically different (varies).

- **(H_0)** *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 groups. Thus, the two independent groups (2013, 2014) are homogeneous and have the same distribution (stochastically equal) taking into account the “Units.Sold”.

```
> MannWhitneyTest

Wilcoxon rank sum test with continuity correction
data:  Units.Sold by Year
W = 42976, p-value = 0.2012
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 -253.00006   52.99996
sample estimates:
difference in location
             -101.0001
```

As shown in the results presented above (see: Fig. 11.5), the meaning of the **Mann–Whitney U** test statistics or output can be explained as a list containing the following:

- **Statistics**: `W (U-statistics) = 42976` which represents the value of the distribution test.
- **p-value**: `p-value = 0.2012` is the significance levels of the test.

Statistically, we can see from the result that the p-value (`p-value = 0.2012`) is greater than the stated significance level (i.e., $p \leq 0.05$). Therefore, we reject the H_1 and accept H_0 by concluding that there is no significant difference (i.e., groups distribution are stochastically equal) between the two groups (“**2013**” and “**2014**”) of the “Year” variable taking into account the “Units.Sold”.

11.3 Kruskal–Wallis H Test in R

The *Kruskal–Wallis* H test is an extension of the Mann–Whitney U test. Statistically, the same assumptions or test criteria apply for both tests (Mann–Whitney and Kruskal–Wallis) except that with the Kruskal–Wallis H test, the targeted independent variable must have *more than two groups* or *categories* (i.e., minimum of three or more levels). Therefore, the test is applied by the researchers to test and compare the hypothesis that the *k* (nth) groups (minimum of three) in a data sample have been obtained or drawn from the same population. It is noteworthy to mention that the Kruskal–Wallis test is regarded as an alternative (non-parametric version) to the One-way ANOVA (le Cessie, 2020; Ortega, 2023). Thus, the test (Kruskal–Wallis) is used or applied when assumptions such as the data normality have not been met or the sample size is too small to conduct the parametric test (One-way ANOVA).

Just like the Mann–Whitney U test, by default, the hypothesis for testing whether there is a *difference in the median* of the *k* (nth) groups (minimum of three or more) of an independent data (ordinal) against the dependent (usually ordinal or continuous) variable is; *IF* the p-value of the test result (Kruskal–Wallis H) is less than or equal to 0.05 ($p \leq 0.05$), *THEN* we assume that at least one of the groups (of the *k* (nth) categories or levels) (see Chap. 6, Sect. 6.2.8) being analyzed comes from a population with a different distribution, and therefore, we can then further perform a multiple comparison (Post-Hoc) test to determine where the significant difference may lie across the data. In other words, we can assume that the *median* of the groups of population, *k* (nth), in the data sample are statistically different (varies), 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 say that there is no difference in the *median* of the *k* (nth) groups (three or more). Thus, the *k* (nth) independent groups are homogeneous and have the same distribution (i.e., are stochastically equal), and any difference observed could only occur by chance, and therefore in this scenario, there will be no need to further perform a multiple comparison (Post-Hoc) test (H_0).

Here, the authors will be demonstrating to the readers how to perform the Kruskal–Wallis H test using the **kruskal.test()** function in R. We will do this using the steps outlined earlier in Fig. 11.1.

To begin, **Create a new RScript** and name it "**KruskalWallisDemo**" or any name of your choice.

Also, we will continue to use the same example dataset "sample-csv-file-for-testing" (**Sample CSV Files**) that we have stored earlier as an R object named "**MWhitney_KWallis.data**" to illustrate the Kruskal–Wallis H test. ***The users can also access and download the example dataset via the following link: https://doi.org/https://doi.org/10.6084/m9.figshare.24728073.

Step 1—Install and Load the required R Packages and Libraries

Install and **Load** the following *R packages* and *libraries* (Fig. 11.7, Step1, Lines 3 to 12) that will be used to call the different R functions, data manipulations, and graphical visualizations for the Kruskal–Wallis H test.

The syntax to install and load the necessary R packages and Libraries are as follows:

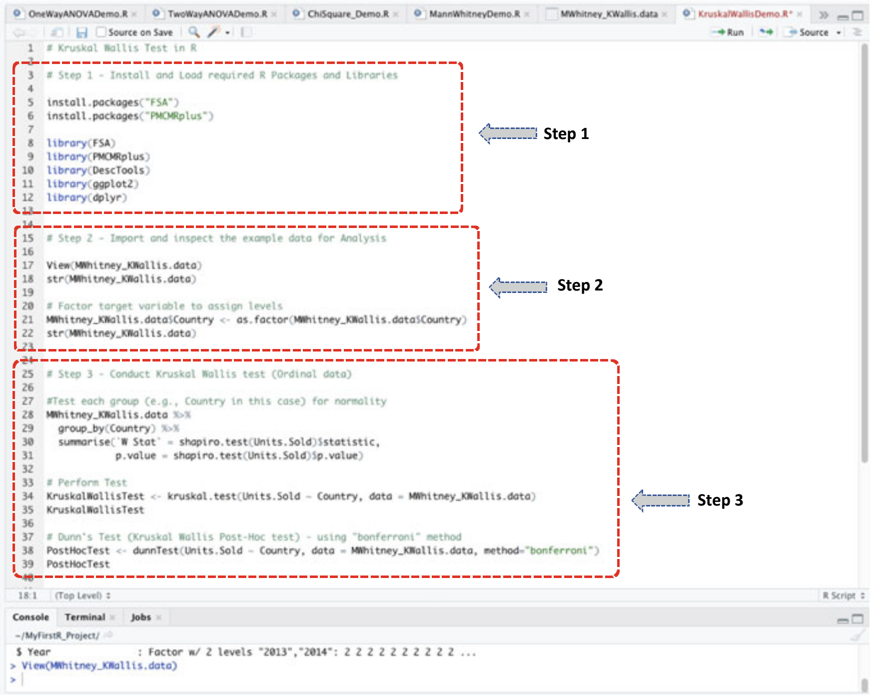

Fig. 11.7 Steps to conducting Kruskal–Wallis H test in R

```
install.packages("FSA")
install.packages("PMCMRplus")

library(FSA)
library(PMCMRplus)
library(DescTools)
library(ggplot2)
library(dplyr)
```

***Note**: as you can see in the code above (see highlighted part) and in Fig. 11.7, we only installed the additional R packages "FSA" and "PMCMRplus" that will be required to perform the Kruskal–Wallis H test, as we have previously installed the other required R packages in R in the previous section or example (see Sect. 11.2). Therefore, we only needed to just load the libraries for the already installed packages (i.e., `DescTools, ggplot2, dplyr`) (see: Lines 10 to 12, Fig. 11.7). ***Note: the users may need to install or re-install the above packages, if necessary, for instance, in the event they have not practiced the previous example in the previous section (Sect. 11.2), or have directly visited this particular section. Please refer to Chap. 2 (Sect. 2.6) on how to install R packages or a refresher on the topic.

 # Step 2—Import or Inspect the example dataset for Analysis.

As shown in Fig. 11.7 (Step 2, Lines 15 to 22), since we have already imported and stored the example dataset "sample-csv-file-for-testing" (**Sample CSV Files**) as an R object we named "**MWhitney_KWallis.data**", we only need to view or inspect the data to make sure we have the variables want to analyze listed there (see: Code below and Fig. 11.8).

```
View(MWhitney_KWallis.data)
str(MWhitney_KWallis.data)

# Factor target variable to assign levels
MWhitney_KWallis.data$Country <-as.factor(MWhitney_KWallis.data$Country)
str(MWhitney_KWallis.data)
```

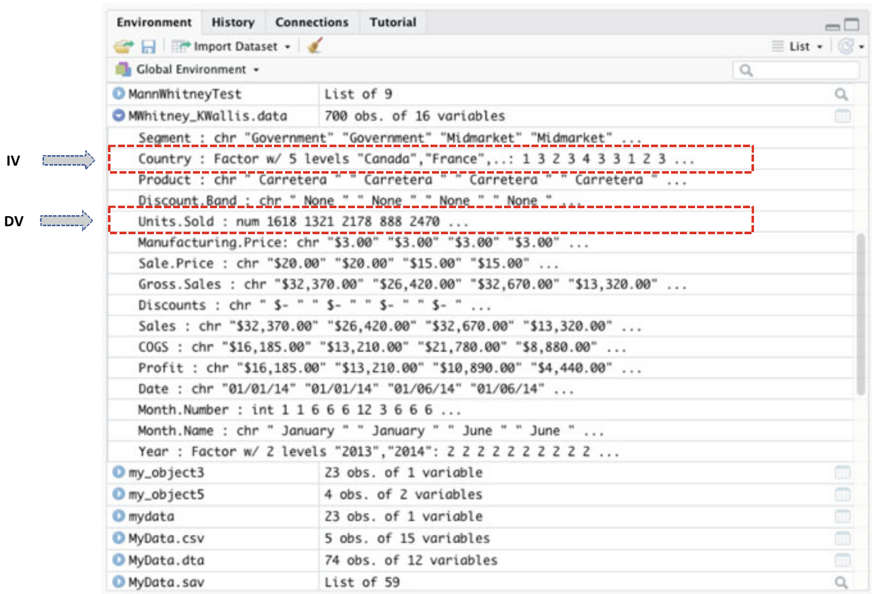

Fig. 11.8 Example of suitable variables for conducting the Kruskal–Wallis H test displayed in R (IV = Independent Variable, DV = Dependent Variable)

Note: if the user have directly visited this specific section or have exited and re-opened RStudio, then, they may need to use the following code below to upload and re-attach the data from their local machine or computer, as the case may be:

```
MWhitney_KWallis.data <- read.csv(file.choose())
attach(MWhitney_KWallis.data)
View(MWhitney_KWallis.data)
str(MWhitney_KWallis.data)
```

Once the user have successfully loaded, inspected, and completed the data conversion (see Step 2, Lines 15 to 22, Fig. 11.7); you will see in the Environment Tab or Console that the variable named "Country" has been factored with 5 levels or groups (see: highlighted part in Fig. 11.8) as we will be using this variable (Country) to illustrate the Kruskal–Wallis H test (i.e., that requires minimum of three levels of an independent variable as a condition to conduct the test).

 # Step 3—Conduct Kruskal–Wallis H test (Ordinal data).

With all the necessary conditions and data format met, we can proceed to analyze the selected variables as highlighted in Fig. 11.8.

As defined earlier in the Introduction section (Sect. 11.1):

- **Kruskal–Wallis H** test or statistics compares the *median or distribution* between three or more groups of an independent variable against a targeted dependent variable.
- The targeted independent variable must be an "ordinal" data type.
- The targeted dependent variable must be an "ordinal or continuous" data type.

To demonstrate to the readers how to conduct the Kruskal–Wallis H test by using the **kruskal.test( )** and **dunnTest( )** method or functions in R:

- We will test whether the *median* or distribution of the "**Country**" (independent variable with 5 levels) differ based on the "**Units.Sold**" (dependent variable)—see Fig. 11.8.

The syntax for conducting the above test (Kruskal–Wallis) in R is as shown in the code below, and as represented in Fig. 11.7 (Step 3, Lines 25 to 39).

```
# Step 3 - Conduct Kruskal Wallis test (Ordinal data)

#Test each group (e.g., Country in this case) for normality
MWhitney_KWallis.data %>%
  group_by(Country) %>%
  summarise(`W Stat` = shapiro.test(Units.Sold)$statistic,
            p.value = shapiro.test(Units.Sold)$p.value)


# Perform Test
KruskalWallisTest <- kruskal.test(Units.Sold ~ Country, data =
MWhitney_KWallis.data)
KruskalWallisTest


# Dunn's Test (Kruskal Wallis Post-Hoc test) - using "bonferroni" method
PostHocTest <- dunnTest(Units.Sold ~ Country, data =
MWhitney_KWallis.data, method="bonferroni")
PostHocTest
```

*****Note**: As defined in the code above; the authors first conducted a normality test by considering each group (five groups) of the "**Country**" variable before conducting the Kruskal–Wallis H test. This was done in order to confirm that the data does not meet the assumption of normality which is commonly a prerequisite to performing the Kruskal–Wallis H test (non-parametric test) (see: Chaps. 3 and 4).

Once the user have successfully run the codes (Step 3, Lines 25 to 39, Fig. 11.7), they will be presented with the results of the assumption test and method in the Console as represented in Fig. 11.9.

As presented in Fig. 11.9, we conducted a normality test considering the five groups of the "**Country**" variable (Canada, France, Germany, Mexico, United States of America) by taking into account the "**Units.Sold**" using the **shapiro.test( )** function in R. As highlighted in the figure (Fig. 11.9), the result of the assumption test shows that the dataset considering the five groups of the "Country" variable against the "Units.Sold" was not normally distributed (whereby significant level is considered values where $p > 0.05$). Therefore, we assume that the data or targeted variables met the condition to perform the Kruskal–Wallis test with Group A (Canada): showing a normality test statistic of `W=0.980, p-value=0.0403`; Group B (France): `W=0.966, p-value=0.00162`; Group C (Germany): `W=0.958, p-value=0.000300`; Group D (Mexico): `W=0.945, p-value=0.0000240`, and Group E (United States of America): `W=0.947, p-value=0.0000369`, respectively.

Therefore, we proceeded to conduct the Kruskal–Wallis H test considering the independent variable "Country" (with five groups) against the dependent variable "Units.Sold" as contained in the example dataset (MWhitney_KWallis.data). Consequentially, we also performed a post-hoc test using the **DunnTest( )** function in R

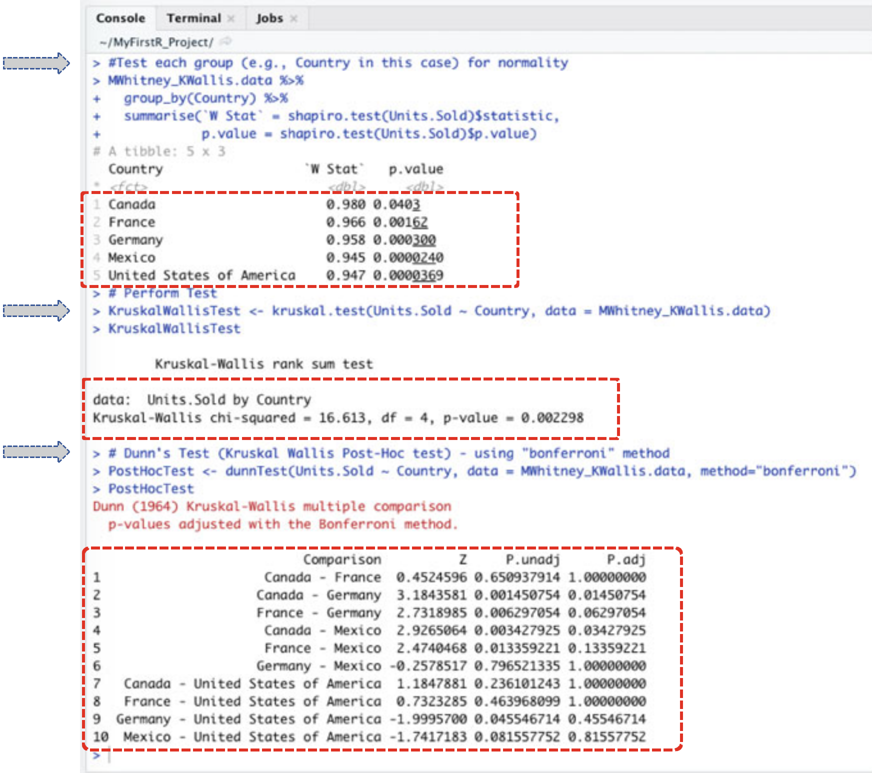

Fig. 11.9 Results of Kruskal–Wallis H test and Post-Hoc test displayed in the Console in R

adjusted with the "bonferroni" method. This is due to the fact that the test (Kruskal–Wallis) results came out significant ($p \leq 0.05$) as we will discuss in detail in Step 5 (Results Interpretation).

Accordingly, the results of the Kruskal–Wallis H test and statistics were stored as an R object we called `KruskalWallisTest,` and the post-hoc test stored as `PostHocTest`, respectively (see: Fig. 11.9).

Step 4—Plot and visualize the data distribution (outliers) and results.

In Fig. 11.10 (Step 4, Lines 42 to 49); we utilized the **ggplot( )** function in R to display a boxplot of the distribution (outliers) for the five groups of the "**Country**" variable (i.e., Canada, France, Germany, Mexico, United States of America) plotted against the **"Units.Sold**". As shown in the figure (Fig. 11.10), the difference in the distribution also confirms the significant difference we found in the *H* test statistics (Step 3) as explained in detail in the next Step 5.

The syntax and used to plot or visualize the distribution of the data or outliers is as shown in the code below, and the resultant chart is represented in Fig. 11.10.

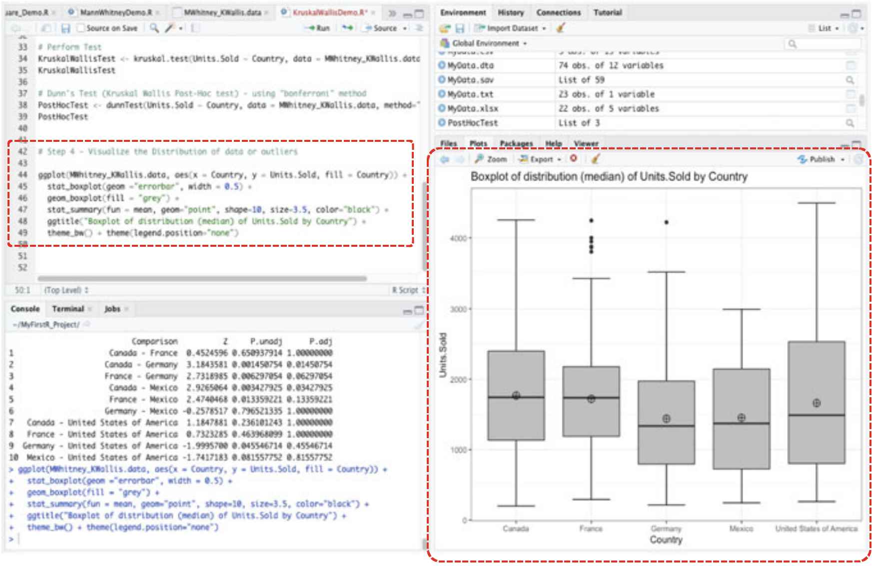

Fig. 11.10 Plot representing the distribution of the five groups of the independent variable broken down by Country using the ggplot() function in R

```
# Step 4 - Visualize the Distribution of data or outliers
ggplot(MWhitney_KWallis.data, aes(x = Country, y = Units.Sold, fill =
Country)) +
  stat_boxplot(geom ="errorbar", width = 0.5) +
  geom_boxplot(fill = "grey") +
  stat_summary(fun = mean, geom="point", shape=10, size=3.5, color="black")+
  ggtitle("Boxplot of distribution (median) of Units.Sold by Country") +
  theme_bw() + theme(legend.position="none")
```

 # Step 5—Results Interpretation (Kruskal–Wallis H).

The final step for the Kruskal–Wallis test and analysis is to interpret and understand the results of the test.

By default, the hypothesis for conducting the test (Kruskal–Wallis) considering the two variables "**Country**" and "**Units.Sold**" (see: Figs. 11.8 and 11.9) is:

- **(H_1)** *IF* the p-value of the test is less than or equal to 0.05 ($p \leq 0.05$), *THEN* we can assume that there is a difference in the distribution of the groups (Canada, France, Germany, Mexico, United States of America) of the "Country" variable taking into account the dependent variable "Units.Sold". Thus, the *median* of the individual group of population (Canada, France, Germany, Mexico, United States of America) are statistically different (varies). Hence, we would consider to further perform a multiple comparison (Post-Hoc) test to determine where the significant differences may lie across the data or groups or variables.

- **(H_0)** *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 five groups of the independent variable taking into account the dependent variable "Units.Sold". Thus, the five groups of the independent variable (Canada, France, Germany, Mexico, United States of America) are homogeneous and have the same distribution (i.e., are stochastically equal). And, if this was the case, then we do not need to further conduct a post-hoc test.

```
> KruskalWallisTest

Kruskal-Wallis rank sum test

data:  Units.Sold by Country

Kruskal-Wallis chi-squared = 16.613, df = 4, p-value = 0.002298
```

As shown in the results of the test presented above (see: Fig. 11.9); the meaning of the **Kruskal–Wallis H** test statistic or output can be explained as a list containing the following:

- **Statistics**: `X`^{2} `(H-statistics) = 16.613` which represents the value of the distribution test.
- **Degrees of freedom**: `df = 4` is the degree of freedom for the *k* (n^{th}) groups of the independent variable.
- **p-value**: `p-value = 0.002298` is the significance level of the test.

Statistically, we can see from the reported result that the p-value is less than the stated significance level (significance, $p \leq 0.05$). Therefore, we reject H_0 and accept H_1 by concluding that there is a significant difference between the five groups of the "Country" variable (Canada, France, Germany, Mexico, United States of America) taking into account the "Units.Sold".

Consequently, having found a significant difference for the analyzed variable or group of countries (`p-value=0.002298`), we do not know which one or where among the countries the differences may lie. Therefore, a post-hoc (multiple comparison) test needs to be conducted, in this case, in order to establish this fact or variations.

```
> PostHocTest
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

                          Comparison          Z      P.unadj      P.adj
1                    Canada - France  0.4524596 0.650937914 1.00000000
2                   Canada - Germany  3.1843581 0.001450754 0.01450754
3                   France - Germany  2.7318985 0.006297054 0.06297054
4                    Canada - Mexico  2.9265064 0.003427925 0.03427925
5                    France - Mexico  2.4740468 0.013359221 0.13359221
6                   Germany - Mexico -0.2578517 0.796521335 1.00000000
7   Canada - United States of America  1.1847881 0.236101243 1.00000000
8   France - United States of America  0.7323285 0.463968099 1.00000000
9  Germany - United States of America -1.9995700 0.045546714 0.45546714
10  Mexico - United States of America -1.7417183 0.081557752 0.81557752
```

As gathered in the above results of the post-hoc (multiple comparison) test using the **DunnTest( )** method adjusted with the “Bonferroni” method in R (see Fig. 11.9); we can now see where among the individual countries (Canada, France, Germany, Mexico, United States of America) the statistical differences we observed lies. For example, we can see that the difference in distribution was exceptionally observed for **`Canada-Germany`** `(Z=3.1843581, P.unadj=0.001450754, P.adj=0.01450754`; **`Canada-Mexico`** `(Z=2.9265064, P.unadj=0.003427925, P.adj=0.03427925).` **`France-Germany`** was also slightly significant with `Z=2.7318985, P.unadj=0.006297054, P.adj =0.06297054`, respectively.

11.4 Summary

In this chapter, the authors explained in detail and practically demonstrated to the readers how to conduct the two most commonly used types of *non-parametric* (or distribution free) tests (Mann–Whitney U and Kruskal–Wallis H) used by the researchers to compare the *median* of “non-normally” distributed data samples in R. In Sect. 11.2, we illustrated how to conduct the Mann–Whitney U test. While in Sect. 11.3 we looked at how to perform the Kruskal–Wallis H test using R.

Also, the chapter covered how to graphically plot the median or distribution (outliers) of the two types of tests (Mann–Whitney U and Kruskal–Wallis H), and then discussed in detail how to interpret and understand the results of the tests in R.

In summary, the main contents covered in this chapter includes:

- Mann–Whitney (U-Statistics) test is a statistical test of hypothesis used to compare the *distribution (in median)* of data samples that are represented in “two independent comparison groups” (usually in ordinal form) and an ordinal or continuous dependent variable.
- Kruskal–Wallis (H-Statistics) test is, on the other hand, applied to compare the *distribution (in median)* of data samples that are represented in “three or more

independent comparison groups" (ordinal form) and an ordinal or continuous dependent variable.

- Mann–Whitney U test is the non-parametric version or alternative (equivalent) to the Independent Sample *t*-test.
- Kruskal–Wallis H test is the non-parametric version or alternative (equivalent) to the One-way ANOVA test.

When choosing whether to conduct a Mann–Whitney U test or Kruskal–Wallis H test? The researcher or data analyst should:

- Perform the "*Mann–Whitney U*" test if the *two groups* come from a single *independently* sampled population, and the distribution of the data sample has been statistically measured or determined to be non-normally distributed.
- Perform the "*Kruskal–Wallis H*" test if the targeted independent variable has *more than two groups* (i.e., minimum of three or more categories), comes from or is drawn from a single *independently* sampled population, and the distribution of the data sample has been statistically measured or determined to be non-normally distributed.
- Perform a post-hoc test (a multiple comparison test) if the result of the Kruskal–Wallis H statistics has shown or appeared to be significant (i.e., $p \leq 0.05$). This is done in order to determine where the significant differences among the groups "k (n^{th})" (minimum of three groups of the independent variable) may lie across the data.

References

Frey, B. B. (2018). Kruskal–Wallis Test. In *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation.* SAGE Publications, Inc. https://doi.org/10.4135/9781506326139.n377.

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. In *Source: Journal of the American Statistical Association* (Vol. 47, Issue 260).

le Cessie, S., Goeman, J. J., & Dekkers, O. M. (2020). Who is afraid of non-normal data? Choosing between parametric and non-parametric tests. In *European journal of endocrinology* (Vol. 182, Issue 2, pp. E1–E3). NLM (Medline). https://doi.org/10.1530/EJE-19-0922.

MacFarland, T. W., Yates, J. M., MacFarland, T. W., & Yates, J. M. (2016). Mann–Whitney U Test. In *Introduction to Nonparametric Statistics for the Biological Sciences Using R* (pp. 103–132). Springer International Publishing. https://doi.org/10.1007/978-3-319-30634-6_4.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics, 18*(1), 50–60. https://doi.org/10.1214/aoms/1177730491

McKight, P. E., & Najab, J. (2010). Kruskal-Wallis Test. In *The Corsini Encyclopedia of Psychology* (pp. 1–1). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470479216.corpsy0491.

McKnight, P. E., & Najab, J. (2010). Mann-Whitney U Test. In *The Corsini Encyclopedia of Psychology* (pp. 1–1). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470479216.corpsy0524.

Nachar, N. (2008). The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. In *Tutorials in Quantitative Methods for Psychology* (Vol. 4, Issue 1).

Okoye, K., Arrona-Palacios, A., Camacho-Zuñiga, C., Achem, J. A. G., Escamilla, J., & Hosseini, S. (2022). Towards teaching analytics: a contextual model for analysis of students' evaluation of teaching through text mining and machine learning classification. *Education and Information Technologies 2021 27:3*, *27*(3), 3891–3933. https://doi.org/10.1007/S10639-021-10751-5.

Ortega, C. (2023). *Kruskal-Wallis test: What it is, advantages and how it is performed*. QuestionPro. https://www.questionpro.com/blog/es/prueba-de-kruskal-wallis/.

Ostertagová, E., Ostertag, O., & Kováč, J. (2014). Methodology and application of the Kruskal-Wallis test. *Applied Mechanics and Materials, 611*, 115–120. https://doi.org/10.4028/www.scientific.net/AMM.611.115

Ramtin, S. (2023). How to choose appropriate bivariate test. In A. E. M. Eltorai, J. A. Bakal, S. F. DeFroda, & B. D. Owens (Eds.), *Translational Sports Medicine* (pp. 145–149). Academic Press. https://doi.org/10.1016/B978-0-323-91259-4.00115-6.

Vargha, A., & Delaney, H. D. (1998). The Kruskal-Wallis Test and Stochastic Homogeneity. *Journal of Educational and Behavioral Statistics, 23*(2), 170–192. https://doi.org/10.3102/10769986023002170