Espiral. Cuadernos del Profesorado
Vol. 17, Num. 36
September 2024
ISSN 1988-7701

How to reference this article:

Ibáñez-López, F. J., Rubio-Aparicio, M., Pedreño-Plana, M., & Sánchez-Martín, M. (2024). Descriptive statistics and basic graphs tutorial to help you succed in statistical analysis. *Espiral. Cuadernos del Profesorado, 17*(36), 88-99. https://doi.org/10.25115/ecp.v17i36.9570

Published on-line: 9 September 2024

Descriptive statistics and basic graphs tutorial to help you succeed in statistical analysis

Descriptive statistics and basic graphs tutorial to help you succeed in statistical analysis

Francisco Javier Ibáñez-López, María Rubio-Aparicio, Marina Pedreño Plana
and Micaela Sánchez-Martín

University of Murcia, Murcia, Spain

"Legolas, what do your elven eyes see?"
(Aragorn, The Lord of the Rings: The Fellowship of the Ring)

Abstract

The previous step to be able to carry out a complete and reliable inferential statistical analysis of the data collected in research is the execution of a correct and exhaustive descriptive statistical analysis. At this stage, we understand the main characteristics of our sample, which allows us to address the complete description of the sample in our research work and, in addition, offers us a first approximation of the variables in which there could be differences depending on others. It is only an approximation. Subsequently, a statistical inferential test will have to say whether or not, statistically speaking, such significant differences exist. This paper presents the main descriptive statistics and their corresponding graphs depending on the type of variable, as well as their calculation and representation using the free statistical software jamovi, a powerful free tool that does not require advanced programming knowledge.

Keywords: Statistical analysis; descriptive; variables; *jamovi*.

Resumen

El paso previo para poder realizar un análisis estadístico inferencial completo y fiable de los datos recolectados en una investigación es la ejecución de un correcto y exhaustivo análisis estadístico descriptivo. En esta etapa, comprendemos las principales características de nuestra muestra, lo que nos permite abordar la descripción completa de la misma en nuestro trabajo de investigación y, además, nos ofrece una primera aproximación sobre las variables en las que podría haber diferencias en función de otras. Solo es una aproximación. Posteriormente, una prueba estadística inferencial tendrá que decir si, estadísticamente hablando, existen o no esas diferencias significativas. En este trabajo se presentan los principales estadísticos descriptivos y sus correspondientes gráficos en función del tipo de variable, así como su cálculo y representación a través del software estadístico libre *jamovi*, como potente herramienta gratuita que no requiere de un conocimiento previo avanzado en programación.

Palabras clave: Análisis estadístico; descriptivos; variables; *jamovi*.

Received: November 10, 2024

Accepted: January 29, 2024

Correspondence: Marina Pedreño-Plana, University of Murcia, Spain

Email: marina.pedreno@um.es

Key ideas

What is known

- Data analysis is one of the main phases in research.
- Descriptive statistics is the preliminary step for the execution of an inferential data analysis in research.
- It allows to understand the main characteristics of the sample

What this work provides

- To know the different types of variables
- To be able to implement a descriptive analysis by calculating descriptive statistics and making graphs of our study using the free software *jamovi*.

Practical case

For the Physical Education subject of the Primary Education Degree, the teacher has asked you to hand in a final paper in which you explain how the work carried out by Aznar-Ballesta and Vernetta (2023) on the satisfaction, boredom and importance attributed by students to the Physical Education subject could be reproduced, and how this perception influences extracurricular sports dropout, associating the factors and determining the reasons for dropping out of the subject. But you have never carried out a descriptive analysis of the data and you do not quite understand where the results of the work to be reproduced come from. How can you then explain how you would carry out the descriptive analysis of the data?

Descriptive statistical analysis of the data

In the research process, the first step is the observation of a need, a problem or a situation to which we want to respond, to offer an explanation. Thus, initially, we can present our hypotheses, elaborate a theory or make predictions aimed at explaining that need or problem. But in order to confirm our hypotheses, theories or predictions, we necessarily have to collect data that, after analysis, confirms that we are right or wrong. This process is known as the research process (Creswell, 2012).

Therefore, data analysis is one of the main phases in research (Field et al., 2012). It is very important to know the different types of variables that we can find in data collection (Ibáñez-López et al., 2023) in order to, after making the data matrix, be able to implement its descriptive analysis, by calculating descriptive statistics and making graphs that help us to understand our data. This paper presents this descriptive analysis process using the *jamovi* statistical software (The jamovi project, 2022).

Descriptive Statistics of Quantitative Variables

. For the description of a quantitative variable, different descriptive statistics are used. The most important ones and their simple derivation by means of *jamovi* are presented below.

Measures of Central Tendency

The mean, median, mode and summation are presented:

- Mean: sum of all observed values of the variable divided by the total number of observations. The mean is very sensitive to distributions with few data in which there are also very extreme values.

- Median: the value that divides the ordered frequency distribution of the variable under study into two equal parts, whereby 50% of the observations will be less than or equal to the median and the other 50% will be equal to or greater than the median.
- Mode: is defined as the most repeated value in the frequency distribution of the variable under study, i.e. the value with the highest absolute frequency. A variable can have a single mode (unimodal) or it can have two, three, ... (bimodal, trimodal, ...).
- Sum: sum of all the observed values of the variable.

We access and calculate these statistics in *jamovi* through the path *Analyses* > *Exploration* > *Descriptives* > *Statistics* > *Central Tendency*.

For qualitative variables it does not make sense to calculate the mean and median. The mode is used as an index of central tendency. However, in ordinal variables, it does not make sense to calculate the mean, and the median and mode are used. In Social Sciences, Likert scale questionnaires are widely used and their items are ordinal variables. Therefore, it is very interesting to provide the median of the data, as well as the minimum and maximum (we will talk about these statistics later on)

Measures of Position

In general, the *n-order* quantile or *n-til* is defined as the n-1 values of the variable that divide its frequency distribution into n parts with the same size. These statistics can be accessed in *jamovi* through the path *Analyses* > *Exploration* > *Descriptives* > *Statistics* > *Percentile Values*. The most commonly used are:

- Percentiles (P): in this case n = 100, so we obtain 99 values that divide the frequency distribution into 100 equal-sized parts. We calculate them by indicating in *jamovi* the value of 10 equal groups. In this way, the 80th percentile is the value of the variable that is equal to or leaves below 80% of the values of its distribution.
- Quartiles (Q): in this case n = 4, so we obtain 3 values that divide the frequency distribution into 4 equal parts (this is how we indicate it in *jamovi*). The first quartile (Q1) leaves below 25% of the observations and above 75%. The second (Q2) leaves 50% below and 50% above (thus coinciding with the median) and the third (Q3) leaves 75% below and 25% above.

Variability Parameters

The variability or dispersion of a frequency distribution refers to the degree of variation of the set of observations (how far the data of our variable are from or close to a given measure). Within these indices, we can distinguish between those that measure dispersion with respect to a measure of central tendency such as the mean (variance and standard deviation, among others) and those that measure dispersion with respect to the degree to which the scores are similar or different from each other (total amplitude).

- Variance: this is the most important measure of dispersion, although not the most widely used. It is defined as the average of the squares of the deviations of the scores from the mean (it can be interpreted as how close or far the data of our variable is from its mean).
- Standard deviation: this is the most commonly used measure of dispersion. It is calculated by taking the positive square root of the variance.
- Total amplitude or range: the range of a set of scores is the distance between their maximum and minimum values.
- Minimum: the minimum value of the observations of the variable. Widely used in ordinal variables.
- Maximum: the lue of the observations of the variable. Also widely used in ordinal variables.
- Standard error of the mean: indicates how far the data deviate from the population mean.
- Interquartile range (IQR): difference between the first and third quartile.

These statistics are accessed in *jamovi* via *Analyses* > *Exploration* > *Descriptives* > *Statistics* > *Dispersion.*

Shape Parameters

The shape of a data distribution is studied by calculating the skewness index and the kurtosis index.

- Skewness: this explains the number of scores in the distribution that lie on either side of the measure of central tendency. If the value of the index is 0, we have zero asymmetry (symmetry); that is, the distribution is symmetrical with respect to the central measure (mean, mode and median coincide). If the index is greater than 0, we obtain positive skewness; that is, the tail of the scores of the distribution moves away to the right (predominance of low scores). Conversely, if the index is negative, we have negative skewness, i.e., the tail of the scores of the distribution moves away to the left (predominance of high scores). There are different indices to measure asymmetry. The most commonly used are Pearson's skewness index and Fisher's skewness index.
- Kurtosis: indicates the degree of skewness of the scores of a distribution taking a normal curve as a reference. Depending on this kurtosis, distributions can be meso-kurtic (the index value is zero), leptokurtic (the index value is greater than zero and indicates positive skewness) and platykurtic (the index value is less than zero).

We access these statistics in *jamovi* through the path *Analyses* > *Exploration* > *Descriptives* > *Statistics* > *Distribution* (Figure 1). *jamovi* provides the value of skewness and kurtosis and the standard error for each of these measures.

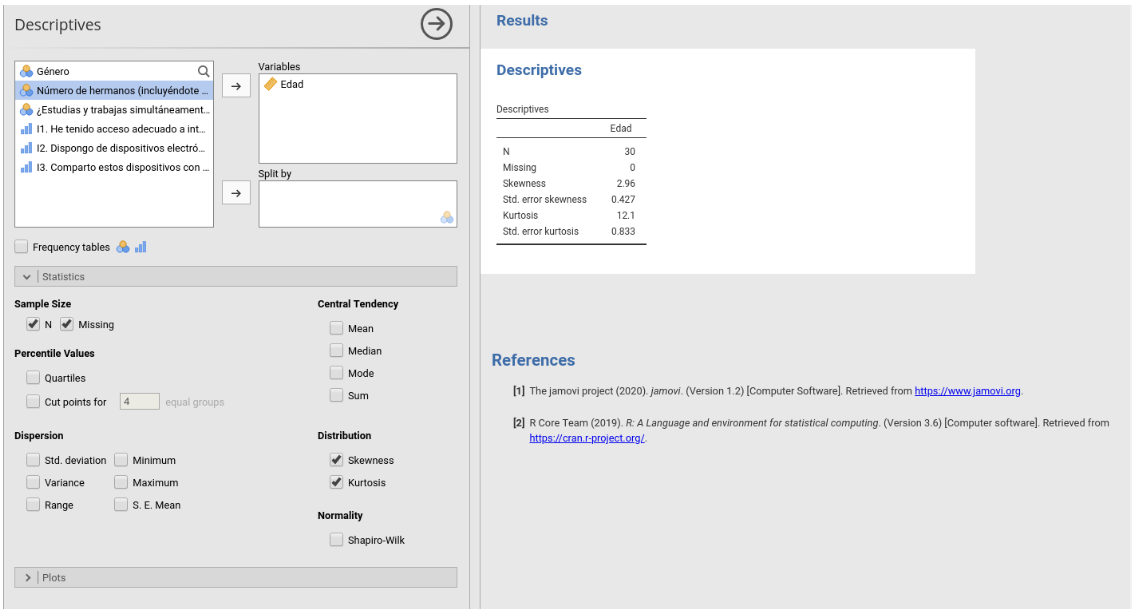
In the example used, as can be seen in the previous figure, a positive skewness (2.96) and a leptokurtic distribution (12.1) are obtained.

Finally, quantitative variables can be represented by histograms, density plots and box plots (the latter are the most commonly used).

Histograms are an extension of bar charts in which the bars are shown joined together with no space between them, indicating the continuity of the variable represented. *jamovi* represents on the abscissa axis the values of the variable distributed in class intervals and on the ordinate axis the density of points per unit (Figure 2).

Figure 1

Shape parameters of the age variable



Descriptives

Género
Número de hermanos (incluyéndote ...
¿Estudias y trabajas simultáneament...
I1. He tenido acceso adecuado a int...
I2. Dispongo de dispositivos electró...
I3. Comparto estos dispositivos con ...

Variables: Edad

Split by

□ Frequency tables

Statistics

Sample Size
☑ N ☑ Missing

Percentile Values
□ Quartiles
□ Cut points for 4 equal groups

Dispersion
□ Std. deviation □ Minimum
□ Variance □ Maximum
□ Range □ S. E. Mean

Central Tendency
□ Mean
□ Median
□ Mode
□ Sum

Distribution
☑ Skewness
☑ Kurtosis

Normality
□ Shapiro-Wilk

Plots

Results

Descriptives

Descriptives

	Edad
N	30
Missing	0
Skewness	2.96
Std. error skewness	0.427
Kurtosis	12.1
Std. error kurtosis	0.833

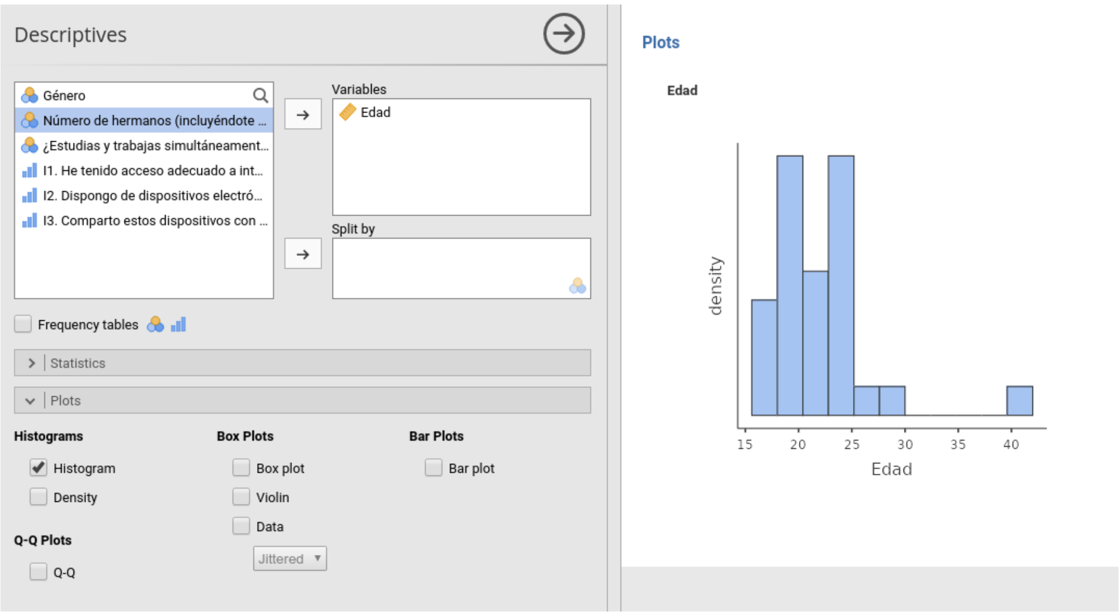
References

[1] The jamovi project (2020). *jamovi*. (Version 1.2) [Computer Software]. Retrieved from https://www.jamovi.org.

[2] R Core Team (2019). *R: A Language and environment for statistical computing*. (Version 3.6) [Computer software]. Retrieved from https://cran.r-project.org/.

Figure 2

Histogram of the age variable



Descriptives

Género
Número de hermanos (incluyéndote ...
¿Estudias y trabajas simultáneament...
I1. He tenido acceso adecuado a int...
I2. Dispongo de dispositivos electró...
I3. Comparto estos dispositivos con ...

Variables: Edad

Split by

□ Frequency tables

> Statistics

⌄ Plots

Histograms

☑ Histogram

□ Density

Q-Q Plots

□ Q-Q

Box Plots

□ Box plot

□ Violin

□ Data

Jittered

Bar Plots

□ Bar plot

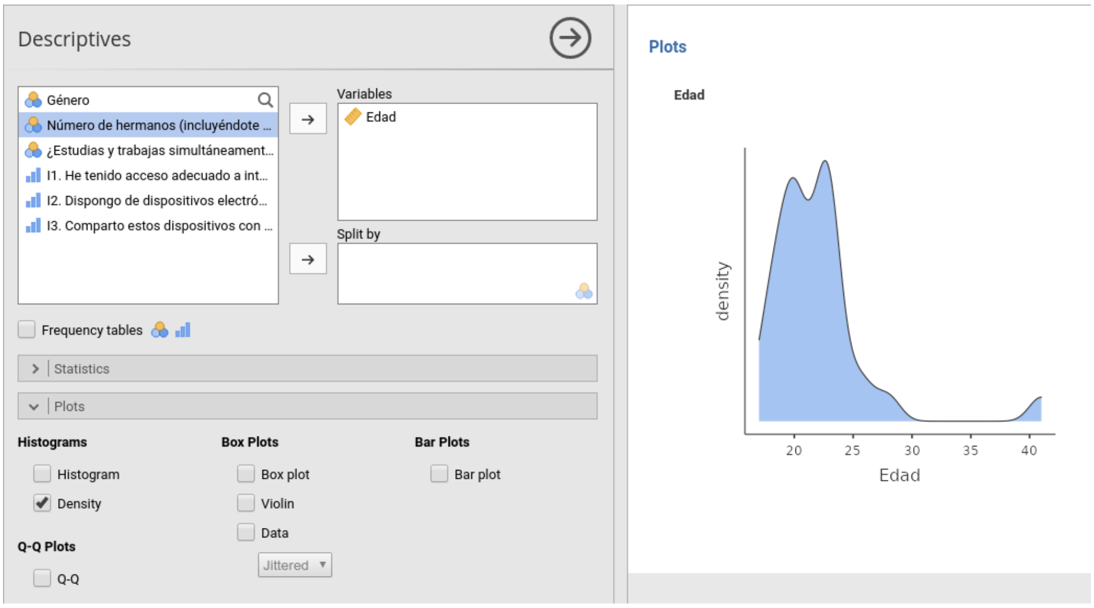
Box plots are often particularly appropriate when dealing with data with asymmetric distributions and outliers, as they dispense with the representation of classical statistics (mean and standard deviation) and use robust statistics resistant to the presence of outliers, such as the median and the interquartile range.

In *jamovi* we get these plots by activating the *Plots* menu at the bottom of the *Analyses > Exploration > Descriptives section.*

A density plot, on the other hand, shows the distribution of data over a continuous time interval or period. It is a variation of the histogram that models the depicted distribution by smoothing out noise, i.e. less frequent data. It can be displayed alone (Figure 3) or combined with its corresponding histogram (Figure 4).

Figure 3

Density plot of the age variable



Descriptives

Género
Número de hermanos (incluyéndote ...
¿Estudias y trabajas simultáneament...
I1. He tenido acceso adecuado a int...
I2. Dispongo de dispositivos electró...
I3. Comparto estos dispositivos con ...

Variables: Edad

Split by

□ Frequency tables

> Statistics

⌄ Plots

Histograms

□ Histogram

☑ Density

Q-Q Plots

□ Q-Q

Box Plots

□ Box plot

□ Violin

□ Data

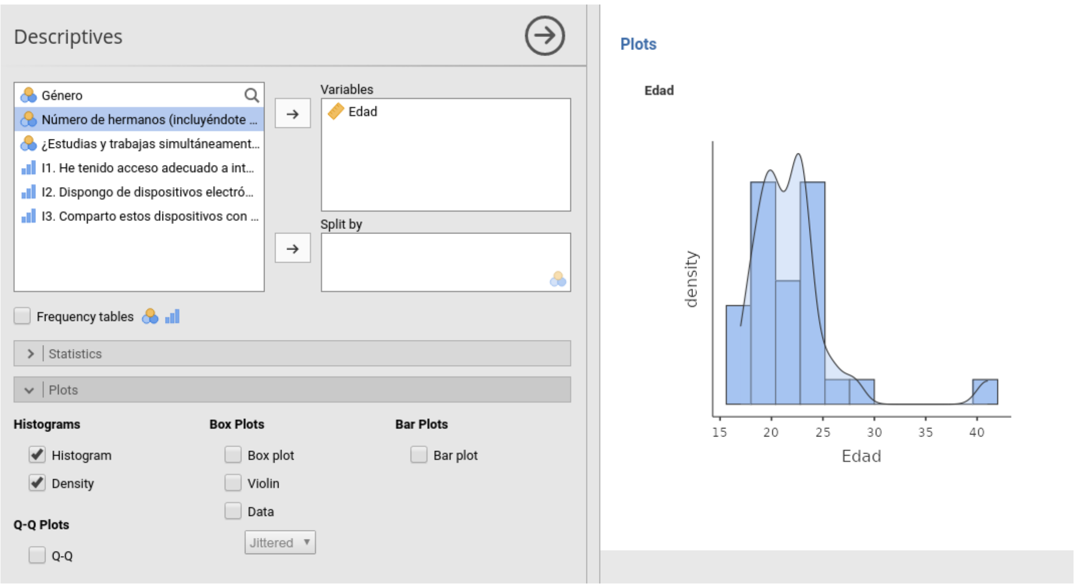
Jittered

Bar Plots

□ Bar plot

Figure 4

Histogram and density plot of the age variable



Descriptives

- Género
- Número de hermanos (incluyéndote ...
- ¿Estudias y trabajas simultáneament...
- I1. He tenido acceso adecuado a int...
- I2. Dispongo de dispositivos electró...
- I3. Comparto estos dispositivos con ...

Variables: Edad

Split by:

☐ Frequency tables

> Statistics

⌄ Plots

Histograms: ☑ Histogram ☑ Density

Q-Q Plots: ☐ Q-Q

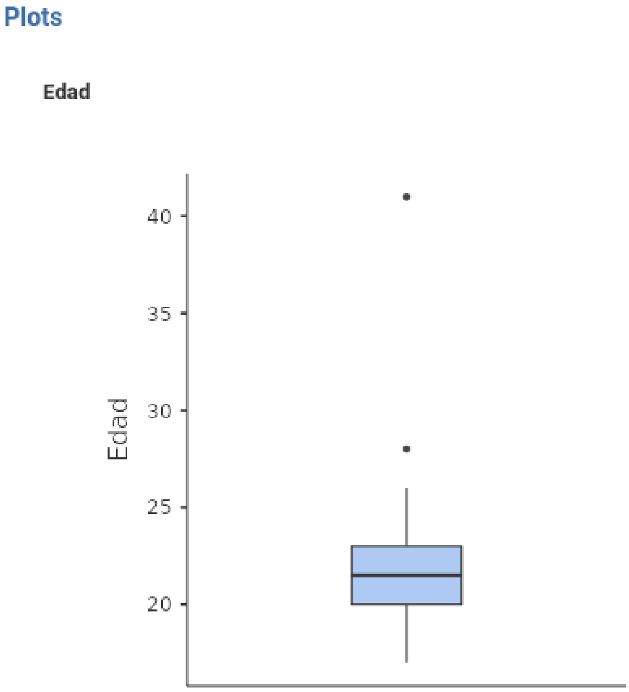
Box Plots: ☐ Box plot ☐ Violin ☐ Data Jittered

Bar Plots: ☐ Bar plot

Finally, the box plot, also known as box and whiskers, is used to check the skewness and outliers of a quantitative variable. It also shows measures of central tendency and dispersion. Figure 5 shows a boxplot of the variable age.

Figure 5

Box plot of the variable age



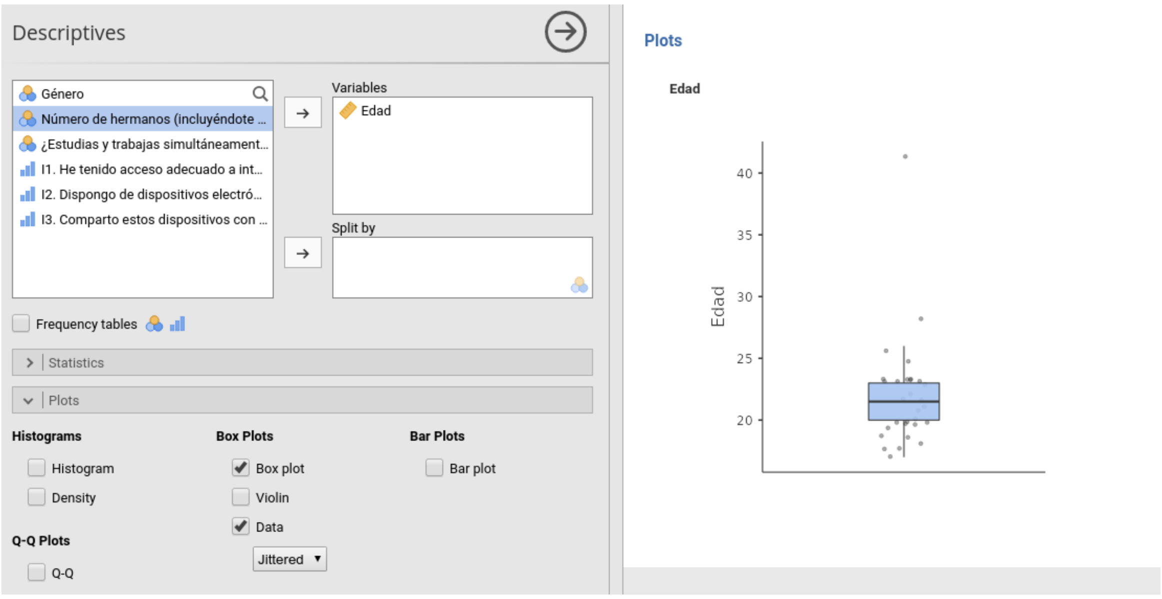
In this graph (Figure 5), we can observe the following elements:

- Median: corresponds to the thickest horizontal centre line. In this example, the median is 21.5.
- First and third quartiles: these are the horizontal lines that make up the bottom and top of the box. In this example, the first quartile is 20 and the third quartile is 23.
- Symmetry: as can be seen, in our example there is more data above the median than below, so there is a positive skewness (tail to the right, as can be seen in Figures 9, 10 and 11). When the distribution is positively skewed it is true that: Q3-Q2 > Q2-Q1.
- Maximum and minimum that are not extreme: the graph shows the maximum and minimum values of the distribution that are not considered extreme values, by means of the thin vertical lines known as whiskers. In this example, they take values of 26 and 17.
- Extreme values: these are represented by dots. In the example, there are two extreme values: 41 and 28.
- Range: for the calculation of the range, the maximum and minimum values of the data distribution are taken into account, whether they are extreme or not. In this example, the maximum value is 41 and is extreme, and the minimum value is 17. Therefore, the range is 24.

In addition, *jamovi* has the option to place the data on the graph (Figure 6) and also offers a *violin plot*, i.e. a boxplot in which the probability density of the data is also marked, which is very popular at the moment (Figure 7).

Figure 6

Boxplot with data for the variable age



Descriptives

Género
Número de hermanos (incluyéndote ...
¿Estudias y trabajas simultáneament...
I1. He tenido acceso adecuado a int...
I2. Dispongo de dispositivos electró...
I3. Comparto estos dispositivos con ...

Variables: Edad

Split by

□ Frequency tables

› Statistics

⌄ Plots

Histograms
□ Histogram
□ Density

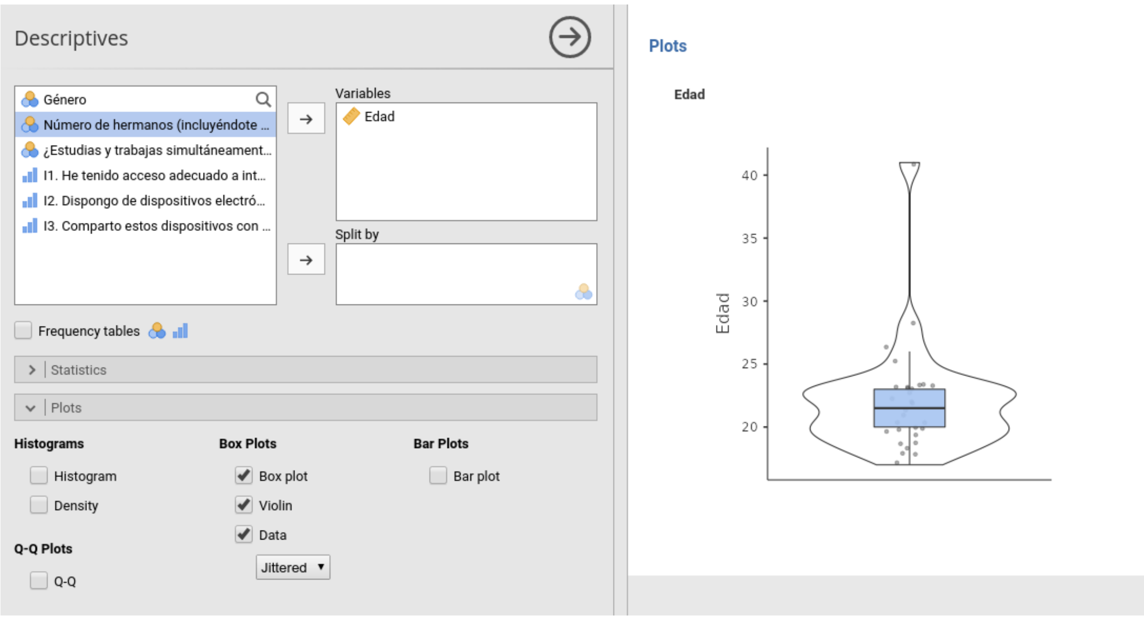
Q-Q Plots
□ Q-Q

Box Plots
☑ Box plot
□ Violin
☑ Data
Jittered ▾

Bar Plots
□ Bar plot

Figure 7

Violin plot with data of the variable age



Descriptives

Variables: Edad

Split by:

□ Frequency tables

Plots

Histograms: □ Histogram □ Density

Box Plots: ☑ Box plot ☑ Violin ☑ Data (Jittered)

Bar Plots: □ Bar plot

Q-Q Plots: □ Q-Q

Género
Número de hermanos (incluyéndote ...
¿Estudias y trabajas simultáneament...
I1. He tenido acceso adecuado a int...
I2. Dispongo de dispositivos electró...
I3. Comparto estos dispositivos con ...
Statistics
Plots
Edad
40
35
30
25
20
Edad

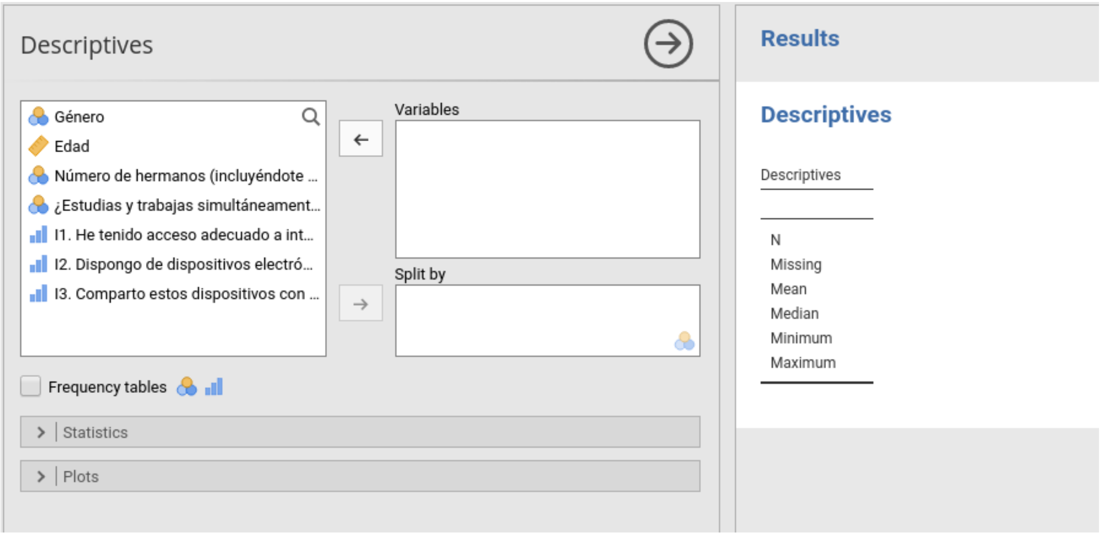
Descriptive Statistics on Qualitative and Ordinal Variables

The description of qualitative and ordinal variables is carried out by drawing up distribution tables of absolute frequencies (counts) and relative frequencies (percentages or proportions). For the graphical representation of these variables we will use bar charts.

To obtain the frequency tables in *jamovi*¸ first go to the menu *Analyses* > *Exploration* > *Descriptives* (Figure 8). Now, select the variable in question, *Gender*, for example, and move it to the right. We will observe how the descriptives of the variable are auto-completed. Finally, we select *Frequency tables* to obtain the frequency table (Figure 9).

Figure 8

Menu of descriptives of a variable



Descriptives

Variables:

Split by:

□ Frequency tables

Results

Descriptives

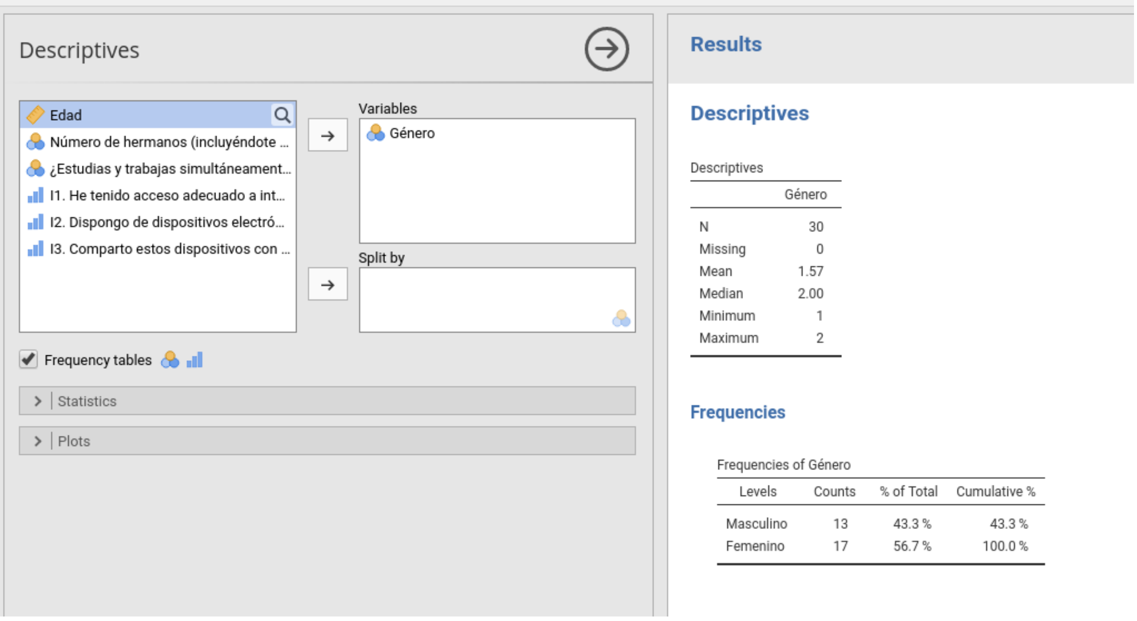
Descriptives

N
Missing
Mean
Median
Minimum
Maximum

Género
Edad
Número de hermanos (incluyéndote ...
¿Estudias y trabajas simultáneament...
I1. He tenido acceso adecuado a int...
I2. Dispongo de dispositivos electró...
I3. Comparto estos dispositivos con ...
Statistics
Plots

Figure 9

Frequency tables


Descriptives
Edad
Número de hermanos (incluyéndote ...
¿Estudias y trabajas simultáneament...
I1. He tenido acceso adecuado a int...
I2. Dispongo de dispositivos electró...
I3. Comparto estos dispositivos con ...
Split by
Statistics
Plots

Variables: Género

☑ Frequency tables

Results

Descriptives

Descriptives

	Género
N	30
Missing	0
Mean	1.57
Median	2.00
Minimum	1
Maximum	2

Frequencies

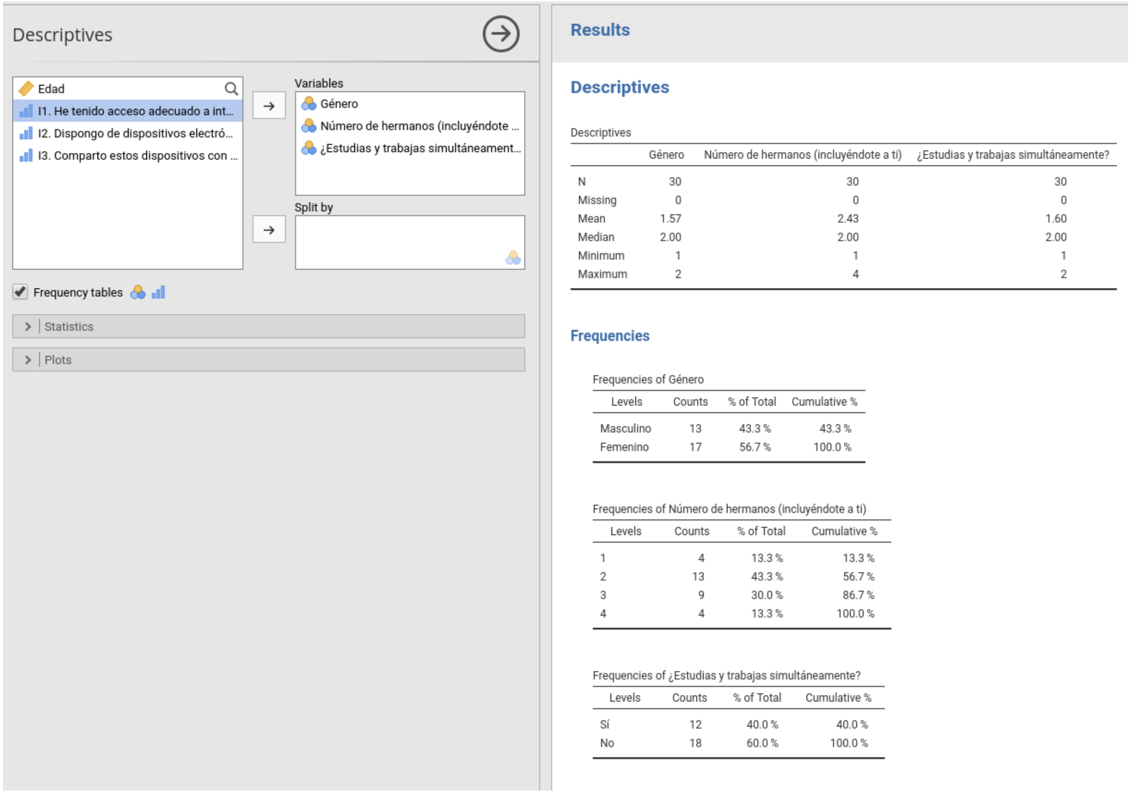
Frequencies of Género

Levels	Counts	% of Total	Cumulative %
Masculino	13	43.3 %	43.3 %
Femenino	17	56.7 %	100.0 %

Tables of frequencies and descriptive statistics can be made simultaneously for several variables at the same time, simply by shifting those variables to the right (Figure 10). In addition, tables created in APA format can be easily copied into any word processor while maintaining the formatting.

Figure 10

Frequency tables of different variables


Descriptives
Edad
I1. He tenido acceso adecuado a int...
I2. Dispongo de dispositivos electró...
I3. Comparto estos dispositivos con ...
Split by
Statistics
Plots

Variables: Género; Número de hermanos (incluyéndote ...; ¿Estudias y trabajas simultáneament...

☑ Frequency tables

Results

Descriptives

Descriptives

	Género	Número de hermanos (incluyéndote a ti)	¿Estudias y trabajas simultáneamente?
N	30	30	30
Missing	0	0	0
Mean	1.57	2.43	1.60
Median	2.00	2.00	2.00
Minimum	1	1	1
Maximum	2	4	2

Frequencies

Frequencies of Género

Levels	Counts	% of Total	Cumulative %
Masculino	13	43.3 %	43.3 %
Femenino	17	56.7 %	100.0 %

Frequencies of Número de hermanos (incluyéndote a ti)

Levels	Counts	% of Total	Cumulative %
1	4	13.3 %	13.3 %
2	13	43.3 %	56.7 %
3	9	30.0 %	86.7 %
4	4	13.3 %	100.0 %

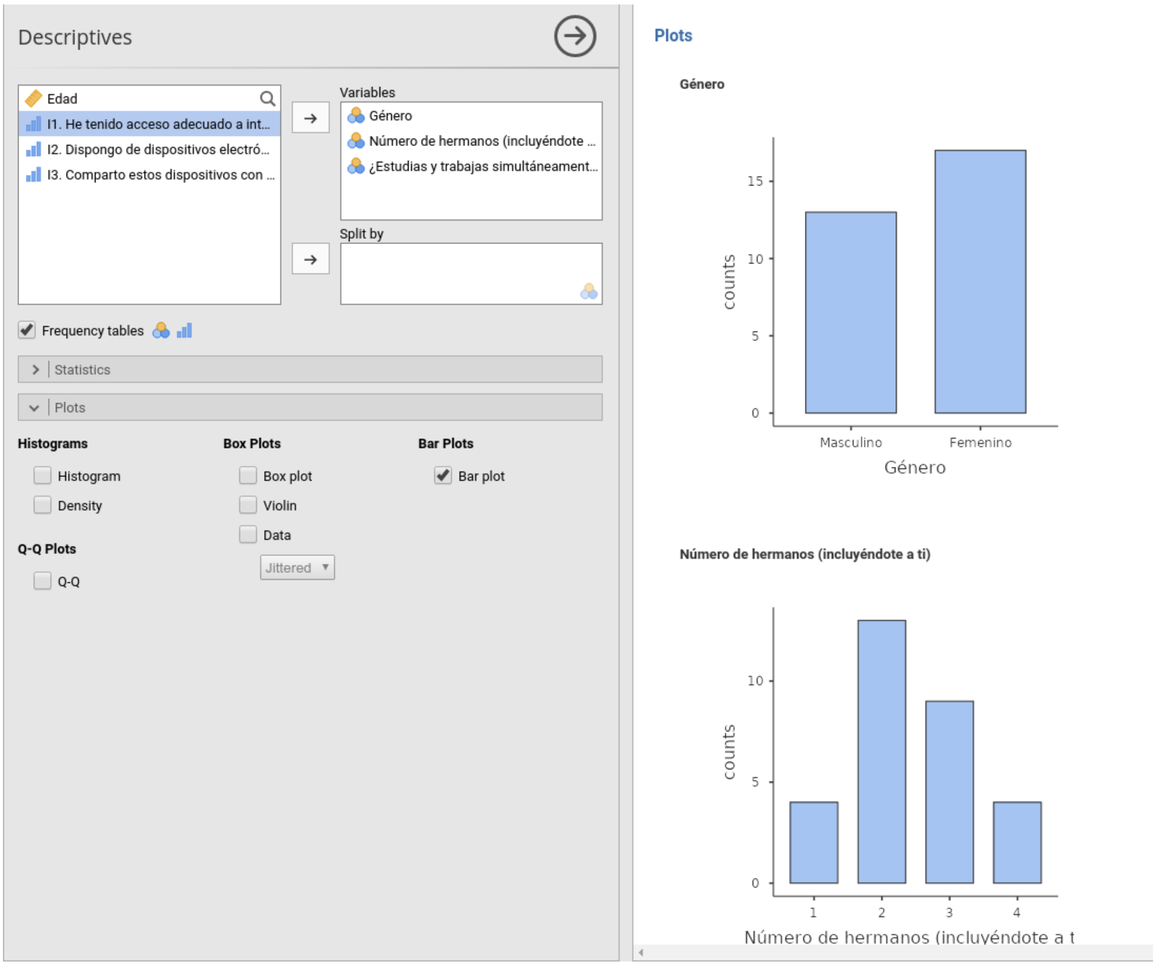
Frequencies of ¿Estudias y trabajas simultáneamente?

Levels	Counts	% of Total	Cumulative %
Sí	12	40.0 %	40.0 %
No	18	60.0 %	100.0 %

The data obtained can be represented by bar charts (Figure 11). In these graphs, the categories of the variable represented are observed on the abscissa axis, projecting a perpendicular bar that indicates the number of elements (the absolute frequency) with its height on the ordinate axis. In *jamovi*, these graphs can be obtained by activating the *Plots* menu at the bottom of the *Analyses* > *Exploration* > *Descriptives* section.

Figure 11

Bar charts of different variables



Descriptives

Edad
I1. He tenido acceso adecuado a int...
I2. Dispongo de dispositivos electró...
I3. Comparto estos dispositivos con ...

Variables
Género
Número de hermanos (incluyéndote ...
¿Estudias y trabajas simultáneament...

Split by

☑ Frequency tables

> Statistics

⌄ Plots

Histograms
□ Histogram
□ Density

Q-Q Plots
□ Q-Q

Box Plots
□ Box plot
□ Violin
□ Data
Jittered ▾

Bar Plots
☑ Bar plot

Plots
Género
15
10
5
0
counts
Masculino
Femenino
Género
Número de hermanos (incluyéndote a ti)
10
5
0
counts
1
2
3
4
Número de hermanos (incluyéndote a t

Solution to the Practical Scenario

Now that you have read this document and done the occasional test in *jamovi*, you have understood perfectly well and you are ready to specify which descriptive statistics you would use for the different variables used in the study that your teacher has proposed for you to use as a reference. On the other hand, you will also be able to indicate which graphs can accompany and contribute to a better understanding of the data and their statistics, which will go beyond the analysis presented in the article. And all this, implemented in a simple way and under the security and confidence at all times that the work you are doing is well done.

Conclusions

In the execution of the descriptive analysis, it is essential to know how to select the type of variable on which one is working, in order to be able to provide the descriptive statistics and the corresponding graphs. Furthermore, this analysis has been presented through a free and multiplatform tool, such as *jamovi*, behind which there is an extensive community of researchers implementing new techniques and new analyses, so that students, teachers and researchers can carry out their statistical analyses without the need to have knowledge of a specific programming language.

Therefore, this tutorial provides a simple explanation on the execution of a basic descriptive analysis in terms of the statistical variables used in research. In this way, the aim is to make these analyses more comprehensible for all those interested in getting started in educational research, through a didactic perspective based on practical assumptions and concrete examples.

Contribution of each author: conceptualisation, FJIL and MSM; manuscript writing, FJIL, MPP and MRA; writing, revising and editing, MPP and MRA; supervision, MSM.

Funding: This research did not receive any external funding.

Conflict of Interests: The authors declare that they have no conflict of interest.

References

Aznar-Ballesta, A. & Vernetta, M. (2023). Influence of the satisfaction and importance of physical education on sports dropout in secondary school. *Espiral. Cuadernos del Profesorado, 16*(32), 18-28. https://doi.org/10.25115/ecp.v16i32.8604

Creswell, J. (2012). *Educational research: planning conducting and evaluating quantitative and qualitative research* (4.ª ed.). Pearson.

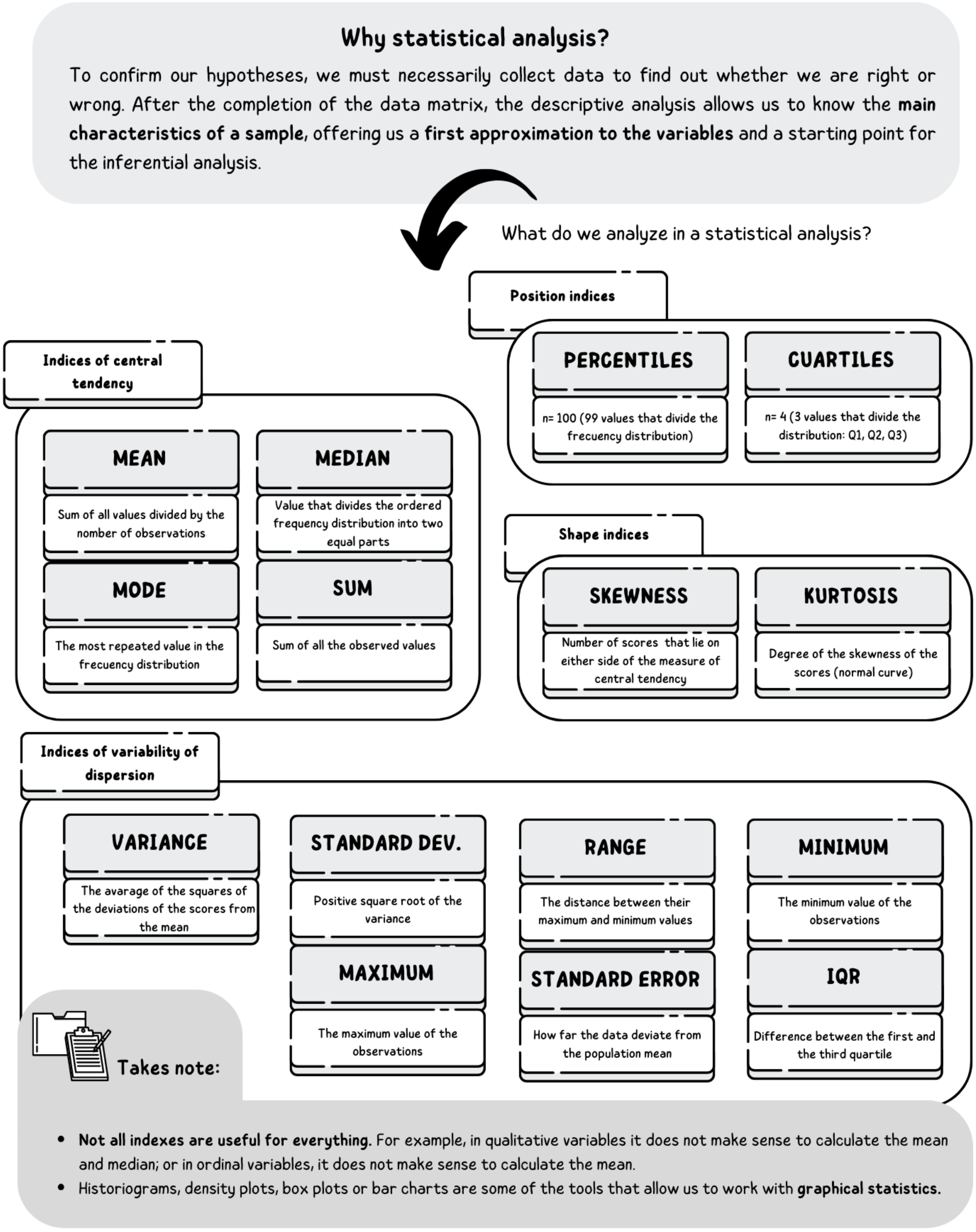
Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R.* SAGE.

Ibáñez-López, F. J., Ponce Gea, A. I., Pedreño-Plana, M., & Sánchez-Martín, M. (2023). Basic survival manual for descriptive statistical analysis. *Espiral. Cuadernos del Profesorado, 16*(32), 118-125. https://doi.org/10.25115/ecp.v16i32.9134

The jamovi project (2022). *Jamovi.* (Version 2.3) [Computer Software]. https://www.jamovi.org

Appendix

Descriptive statistics and basic graphs tutorial to help you succeed in statistical analysis



Why statistical analysis?

To confirm our hypotheses, we must necessarily collect data to find out whether we are right or wrong. After the completion of the data matrix, the descriptive analysis allows us to know the **main characteristics of a sample**, offering us a **first approximation to the variables** and a starting point for the inferential analysis.

What do we analyze in a statistical analysis?

Position indices

PERCENTILES
n= 100 (99 values that divide the frecuency distribution)

CUARTILES
n= 4 (3 values that divide the distribution: Q1, Q2, Q3)

Indices of central tendency

MEAN
Sum of all values divided by the nomber of observations

MEDIAN
Value that divides the ordered frequency distribution into two equal parts

MODE
The most repeated value in the frecuency distribution

SUM
Sum of all the observed values

Shape indices

SKEWNESS
Number of scores that lie on either side of the measure of central tendency

KURTOSIS
Degree of the skewness of the scores (normal curve)

Indices of variability of dispersion

VARIANCE
The avarage of the squares of the deviations of the scores from the mean

STANDARD DEV.
Positive square root of the variance

RANGE
The distance between their maximum and minimum values

MINIMUM
The minimum value of the observations

MAXIMUM
The maximum value of the observations

STANDARD ERROR
How far the data deviate from the population mean

IQR
Difference between the first and the third quartile

Takes note:

- **Not all indexes are useful for everything.** For example, in qualitative variables it does not make sense to calculate the mean and median; or in ordinal variables, it does not make sense to calculate the mean.
- Historiograms, density plots, box plots or bar charts are some of the tools that allow us to work with **graphical statistics.**