

Chapter 2: Statistical Tables and Graphical Representations

I. Introduction Statistical tables are a great starting place for summarizing and organizing data. Once have a set of data, one may first want to organize it to see the frequency, or how often each value occurs in the set. Statistical tables can be used to show either quantitative or categorical data.

Graphical representations are tools that help learn about the distribution, or shape of a sample or a population. A graph can be a more effective way of presenting data than a mass of numbers.

II. Statistical Table: In statistics, tables are very useful in presenting data in a structured manner and are more legible. We can distinguish three types of tables: **table of data** (or elementary table), **frequency distribution table** (called also table of counting) and **the table of relative frequencies distribution**.

II.1 Data Table: The raw representation of data is not readable. The information will be more readable whenever they are grouped in a table of data. This is the reason why, in any classic statistical approach, data tables are the first to be drawn up. These are the tables that facilitate and report on the processing of data. Exp: using an Excel file.

Every table is made up of rows and columns. To construct our data table we must therefore draw rows and columns. The columns list the characters studied and the rows correspond to the individuals observed.

	Column 1 = Variable 1	Column 2 = Variable 2	...
Line 1 = Individual 1			
Line 2 = Individual 2		Case = Modality or Variable	
⋮			

Example: Recall the illustrative example of the first chapter on the statistical study conducted on First year students at the departement of Mathematics by the teacher who asked his students to provide responses on:

- The color of their eyes.
- Their behaviour towards morning coffee.
- The number of sisters and brothers they have.
- Their heights in cm.

Responses provided by the students are the data that will be studied. These data are given by in the four statistical series corresponding to the four character respectively:

X: Black, Blue, Blue, Black, Brown, Blue, Black, Blue, Green, Brown, Brown, Green, Brown, Brown, Brown, Black, Blue, Black, Brown, Green.

Y: Sometimes, Often, Sometimes, Always, Often, Always, Often, Always, Sometimes, Always, Often, Sometimes, Sometimes, Never, Often, Never, Sometimes, Always, Never, Sometimes.

Z: 4 3 5 6 1 3 7 4 5 4 2 2 3 3 2 5 3 3 0 4

H: 1.59 1.45 1.53 1.73 1.50 1.72 1.61 1.50 1.71 1.63 1.80 1.58 1.69 1.66 1.69 1.75 1.73 1.65 1.64 1.55.

The four statistical series can be organised in the following table:

	X	Y	Z	H
student 1	Black	Sometimes	4	1.59
student 2	Blue	Often	3	1.45
student 3	Blue	Sometimes	5	1.53
student 4	Black	Always	6	1.73
student 5	Brown	Often	1	1.50
student 6	Blue	Always	3	1.72
student 7	Black	Often	7	1.61
student 8	Blue	Always	4	1.50
student 9	Green	Sometimes	5	1.71
student 10	Brown	Always	4	1.63
student 11	Brown	Often	2	1.80
student 12	Green	Sometimes	2	1.58
student 13	Brown	Sometimes	3	1.69
student 14	Brown	Never	3	1.66
student 15	Brown	Often	2	1.69
student 16	Black	Never	5	1.75
student 17	Blue	Sometimes	3	1.73
student 18	Black	Always	3	1.65
student 19	Brown	Never	0	1.64
student 20	Green	Sometimes	4	1.55

II.2. Frequency distribution table: the distribution table reorganises the data in the data table and presents it in a clearer and more concise manner, without losing any of the information contained in the original statistical series. The construction of the counts table depends on the nature of the characteristic studied. It is done directly within the framework of a qualitative character or quantitative discrete one.

However, in the case of a continuous character, the construction requires passing through classes where the data are grouped into k semi-open intervals, where k is given by one of the two following formulas:

$$\text{Sturge rule } k=1+3.3 \log(n)$$

$$\text{Yule } k=2.5 (n)^{1/4}$$

and The construction of the classes is done as follows:

1. We calculate the range of the statistical series $e = \text{maximum value} - \text{minimum value}$
2. We determine the length l of the classes such as $l > \frac{e}{k}$.

The table of numbers is made up of a column presenting the list of modalities (or values, or classes) of the character studied and the other column corresponding to the number of occurrence for each modality (or value, or classe).

modalities (values or classes)	Counts (or frequency) n_i
...	...

Note: In the same way we define the relative frequency distribution table by replacing the counts n_i by the relative frequencies f_i .

Example: Notice that:

- ✓ X is a qualitative nominal character of four modalities: Black, Blue, Green and Brown.
- ✓ Y is a qualitative ordinal character of four modalities: Never, Often, Sometimes, Always.
- ✓ Z is a quantitative discrete character of values: 0 1 2 3 4 5 6 7.
- ✓ H is a quantitative continuous character of values ranging between 1.50m and 1.80m.

Then, for the three first statistical variables Z, we provide directly their corresponding statistical tables as follows:

➤ For The color of eyes X: (nominal)

x_i	n_i
Black	5
Blue	5
Green	3
Brown	4

➤ For The behaviour towards morning coffee Y. (ordinal)

y_i	n_i
Never	3
Often	7
Sometimes	5
Always	5

➤ For the number of sisters and brothers Z . (discrete)

z_i	n_i
0	1
1	1
2	3
3	6
4	4
5	3
6	1
7	1

➤ For the heights which is a continuous character H (measured in m):

1) Calculate k the number of classes:

$$k=1+3.3 \log(20)\approx 5$$

2) Calculate the range of the statistical series

$$e = \text{maximum value} - \text{minimum value} = 1.80 - 1.45 = 0.35$$

3) Determine the length l

$$l > \frac{e}{k} = \frac{0.35}{5} = 0.07 \Rightarrow l = 0.08$$

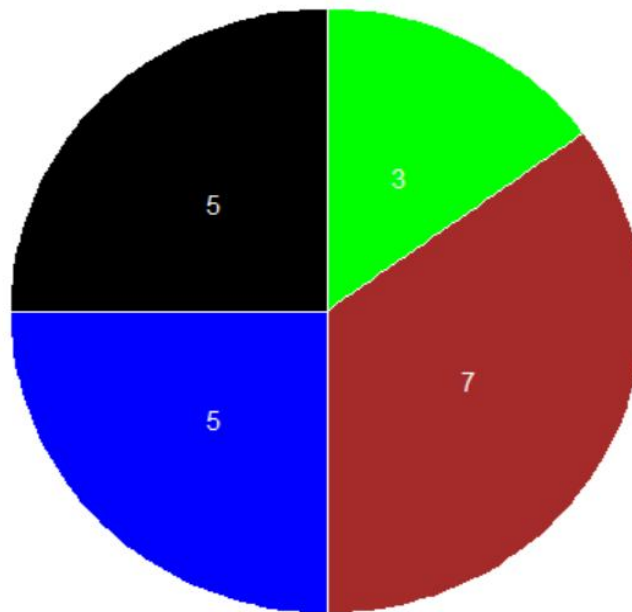
4) Construction of the statistical table:

Classes	n_i
[1.45 , 1.53)	3
[1.53 , 1.61)	4
[1.61 , 1.69)	5
[1.69 , 1.77)	7
[1.77 , 1.85)	1

III. Graphical Representation: It is necessary to draw up a graphical representation in order to bring out part of the information in the data so that it becomes more and more “relevant”. Depending on the nature of the character, the method of graphic representation will be different: **Pie chart** (nominal), **bar chart** (ordinal), **vertical line chart** (discrete) and **histogram** (continuous).

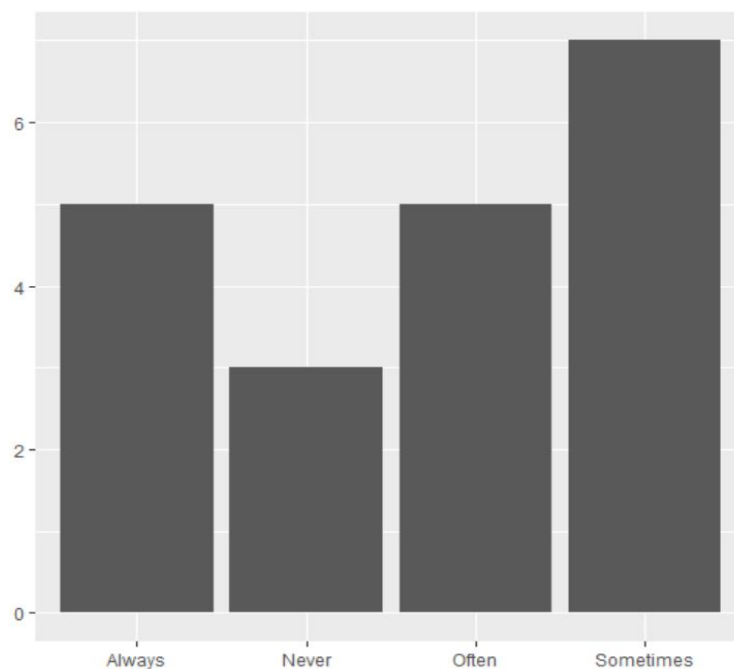
III.1. Pie Chart: A pie chart is a circular statistical graphic divided into slices to illustrate numerical proportion, where each slice represents a percentage of the total whole. It is best used for comparing parts of a single category to a total (100%), often organized from largest to smallest slice for easier interpretation.

Example: For The color of eyes X which is a nominal character, the pie chart is given by:



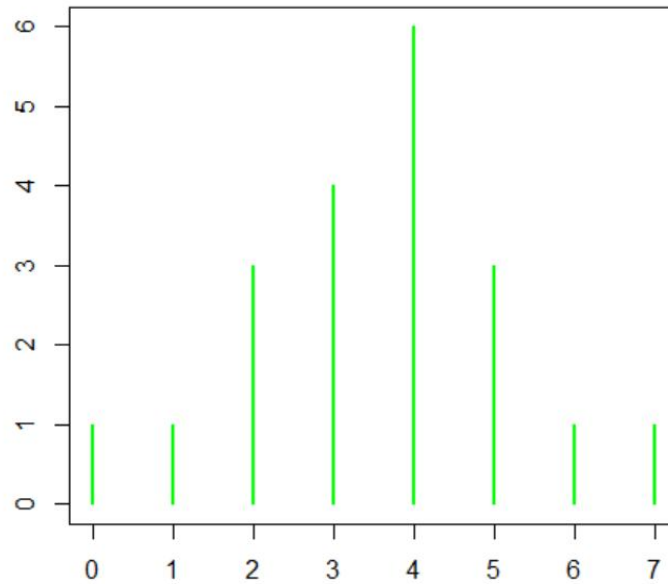
III.2. Bar Chart: A bar chart provides a way of showing data values represented as vertical bars. It is sometimes used to show trend data, and the comparison of multiple data sets side by side.

Example: For The behaviour towards morning coffee Y which is an ordinal character, the bar chart is given by:



III.3. Vertical Lines Chart: A vertical line chart is a specialized visualization primarily used to display discrete. It uses individual vertical lines (also called "stems") to represent the magnitude of a category or a specific data point. Each vertical line corresponds to a specific value on the x-axis.

Example: For the number of sisters and brothers Z , the bar chart is given by:



III.3. Histogram: Histogram is a graphical representation that organizes a group of data points into specified ranges (called bins). It is the most commonly used tool to visualize the distribution of a continuous dataset. In the x-axis are represented the "bins" or intervals (lasses) and in the y-axis corresponding frequencies are represented. Unlike a bar chart, the bars in a histogram usually touch each other.

Example: The continuous character H , the height can be represented by the following histogram:

