UNIVERSIDAD DE COSTA RICA

INIE Instituto de Investigación en Educación

2020

# GUIDELINES FOR PRESENTING DATA AND INFORMATION STATISTIC OF THE RESEARCH INSTITUTE IN EDUCATION

**Elaborated by:**

**Melissa Valverde Hernández**

## Table of contents


1. Introduction 2
2. Basic definitions 2
3. Elements for presenting information 2
    3.1 Tables 3
        3.1.1 Elements of a table 3
    3.2 Figures 6
        3.2.1 Graphics 6
        3.2.2 Elements of a graphics 6
        3.2.3 Body types 8
    3.3 Thematic maps 11
        3.3.1 Elements of thematic maps 11
    3.4 Infographics 13
        3.4.1 Elements of an infographic 13
4. Presentation of results of statistical and mathematical material 14
    4.1 Statistical tests 14
    4.2 Statistical Abbreviations 15
    4.3 Some general recommendations 17
5. Bibliography 18

## 1. Introduction

This guide is intended to assist in the presentation of information at the Institute for Research in Education (INIE) of the Universidad de Costa Rica . The main purpose is to achieve a standard model as well as to unify the formats to elaborate tables, graphs, maps, infographics and textual description related to statistical information. In addition, is a support guide for those authors responsible to prepare documents related to the Registered projects partial and Final reports of the Institute and the scientific articles disseminated in the Electronic Journal of Investigative News in Education.

For presentation of some aspects, reference should be elaborated according to the Publications Manual of the American Psychological Association (APA).

## 2. Basic definitions

**Table:** This is a tool to help to summarize the data keeping an organized model, which is presented through rows and columns. The table should facilitate the understanding, interpretation and dissemination of the data.

**Statistical graphic:** This is a tool that helps the visual representation of statistical data to facilitate its understanding . There are different types of graphs, which are used depending on the type of variables (INEC, 2016).

**Percentage:** Represents the magnitude of a proportion. Mathematically, this is the relationship established between the frequency of elements or events of a set, divided by the total of elements that make up the set multiplied by 100. This is represented by the symbol %.

## 3. Elements for presenting information

According to APA, there are two ways to name the elements to present information in a document, they are named as Tables and Figures. Tables refer to information represented in the

form of columns and rows. The figures include graphics, diagrams, schematics, thematic maps, concept maps, photographs, images, among others.

The scientific articles published in the electronic Journal of Investigative News in Education, should preferably contain a maximum of fifteen (15) elements, this counting the tables and figures.

In particular, it is recommended to prioritize a single information representation tool (do not use several types for the same analysis).

### 3.1 Tables

In general, the following information should be considered for the development or construction of tables:

**a.** Sort the rows and columns by some criteria or variable (if possible).
**b.** Include total or averages when required.
**c.** Use rounding to a decimal or two. All figures in the table must have the same number of decimals.
**d.** Use a space as a thousands separator.
**e.** Use a comma or period as a decimal separator, the selected separator must be unified throughout the table and in the document.
**f.** Justify the margin of the numerical values in the table to the right of the column.
**g.** Try not to leave cells without any content or empty.

#### 3.1.1 Elements of a table

For the elaboration of tables, the following elements must be considered (see Table 1):

**Table number:** The table must be identified with Arabic numerals. The enumeration of the tables of the document must be consecutive. Preferably, placed on top and centered.

**Title:** This describes the contents of the table, which must indicate the geographical space and the period to which the information or the set of statistical data refers. It is recommended for the structure of the table to answer the following questions:

• **What?:** Variables, indicators, characteristics or sample information.

• **How?:** It refers to the order in which the information is presented. The use of the words "by" and "according to" is recommended. Example: "by region", "according to department".

• **Where?:** Location or geographical coverage corresponding to the information.

• **When?:** It refers to the period of time covered by the information.

The title should be placed at the top of the table and centered.

•**Notes:** Are used to clarify the general information in the table. They are usually placed bellow the title in parentheses. Only one note should appear per table, which should contain all the necessary clarifications and indications. Commonly, the notes are used to detail the unit of measurement of the data.

**Data matrix header:** Refers to the set of concepts trouhg which the variables will be presented in the clasified table . This should contain only the most important groupings, and should correspond to the "how?" That accompanies the title, that is, the "for" and "according" the mentioned above.

**Body or matrix of the table:** This part of the table refers to the "What?" of the title. In this section the numerical data corresponding to the concepts presented in the matrix heading are placed. In the body of the table the figures or data are destined, which are going to be sorted by means of rows and columns, from top to bottom and from left to right.

**Table footer:** It is placed at the bottom, contains the calls and the source. It is used to write

down those clarifications about the information for a better interpretation.

**Calls**: This is the specific information that was used in the table, it is used in order to make particular clarifications. More than one call can be used, each must be placed according to the order in which the table is read. In addition, it must be preceded by the notes and proceeded by the sources.

**Source:** Its purpose is to grant credit to the entity responsible for producing the information. If this is the case, the name of the document where the information was obtained and the year when it was published should be mentioned, so that the user has the possibility to find the original source.

In case of own elaboration , it is suggested for you to indicate it and to mention the source of the data with its respective year of elaboration. It is recommended to follow the APA format to specify the reference.

Example of a table.

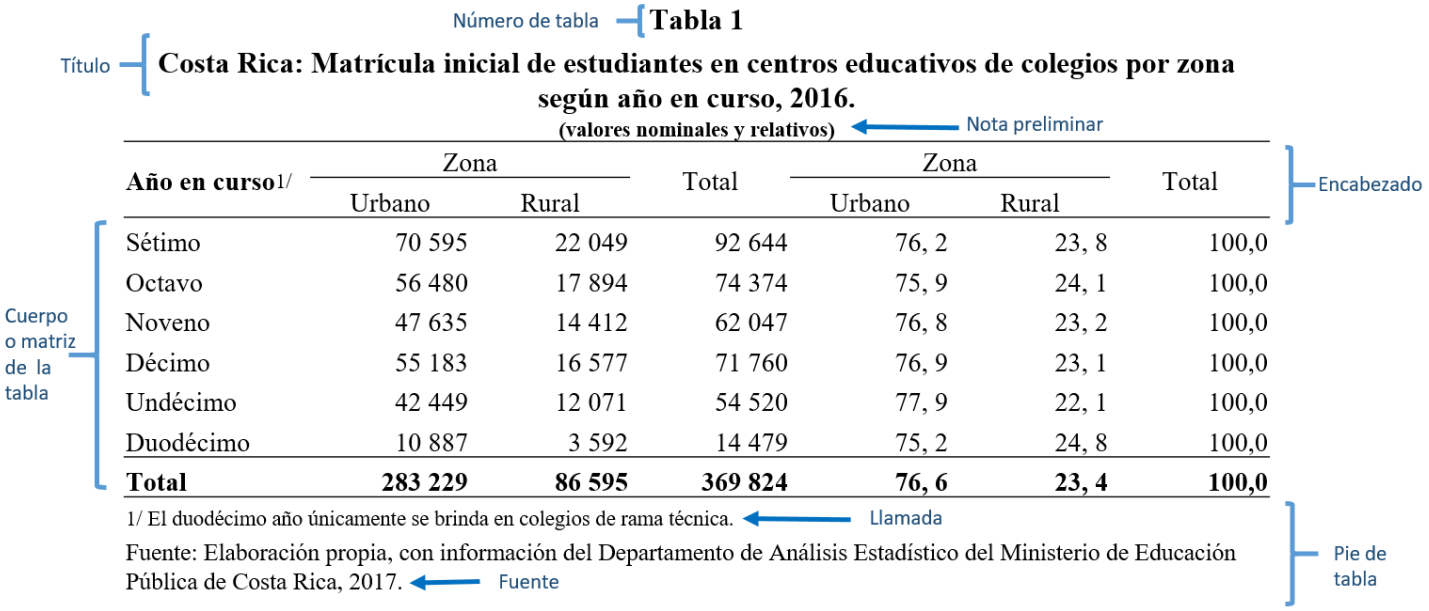



**Tabla 1**

**Costa Rica: Matrícula inicial de estudiantes en centros educativos de colegios por zona según año en curso, 2016.**

**(valores nominales y relativos)**

| **Año en curso**1/ | Zona | | Total | Zona | | Total |
|---|---|---|---|---|---|---|
| | Urbano | Rural | | Urbano | Rural | |
| Sétimo | 70 595 | 22 049 | 92 644 | 76, 2 | 23, 8 | 100,0 |
| Octavo | 56 480 | 17 894 | 74 374 | 75, 9 | 24, 1 | 100,0 |
| Noveno | 47 635 | 14 412 | 62 047 | 76, 8 | 23, 2 | 100,0 |
| Décimo | 55 183 | 16 577 | 71 760 | 76, 9 | 23, 1 | 100,0 |
| Undécimo | 42 449 | 12 071 | 54 520 | 77, 9 | 22, 1 | 100,0 |
| Duodécimo | 10 887 | 3 592 | 14 479 | 75, 2 | 24, 8 | 100,0 |
| **Total** | **283 229** | **86 595** | **369 824** | **76, 6** | **23, 4** | **100,0** |

1/ El duodécimo año únicamente se brinda en colegios de rama técnica.

Fuente: Elaboración propia, con información del Departamento de Análisis Estadístico del Ministerio de Educación Pública de Costa Rica, 2017.

## 3.2. Figures

### 3.2.1 Graphics

The graphics should be named as Figures in the document. Regarding to its preparation, the following should be considered:

**a.** The axes must be properly identified (names and units).

**b.** All categories in the legend must be clearly identified. If some of the categories of a variable have very small percentages, consider grouping categories.

**c.** Do not use 3D format.

**d.** The dimensions of the graphics must have an adequate size which allows their reading and understanding. You can include figures in the chart.

### 3.2.2 Elements of the graphic

For the elaboration of graphics, the following elements must be considered (see Figure 1): Figure number (graphic): Help to identify the graphic. If there is more than one graphic or figure in the document ,the enumeration must be consecutive. Use Arabic numerals for enumeration.

**Title:** It is used to show the variables and characteristics contained.
It must be clear, brief and organized. It must always be below the number of the graph, with bold letters.
It is recommended for the title to answer the following questions:

• **What ?:** Main characteristics that you want to show.

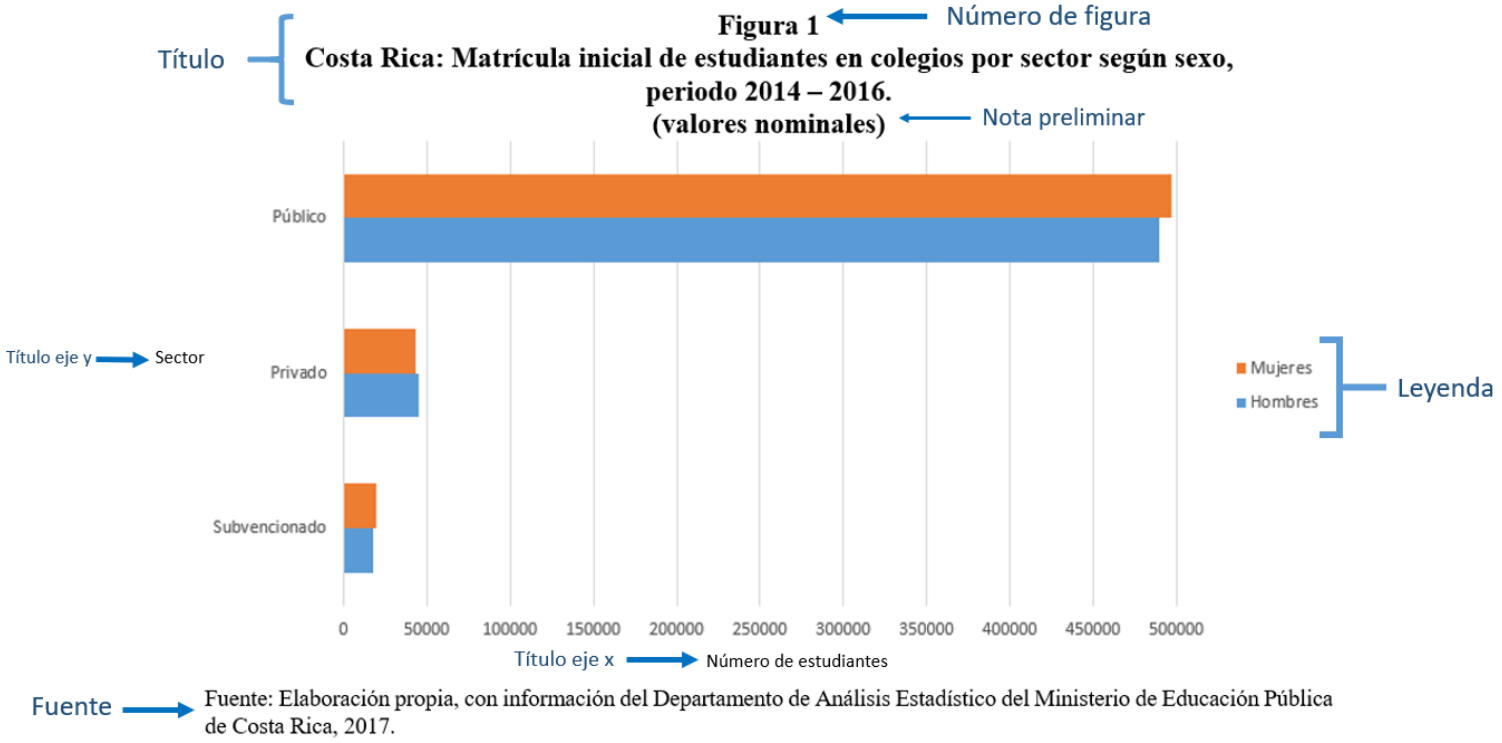
• **How ?:** Refers to the way information is displayed, especially in the types of graphics that use the "x" axis and the "y" axis. The axis of values is presided over by the preposition "by" and the axis of categories is preceded by the preposition "according".

• **Where ?:** Place to which the information corresponds.

• **When ?:** Refers to the time period covered by the information.

Moreover, the title is placed at the top of the graph and centered.

• **Preliminary note:** They are used to make clarifications regarding  to the graph.
They should be placed in parentheses below the title in a small font size. Use in order to clarify the general information of the graph.




**Graphic body**

• Incorporate, in the axes, the titles of the data shown (except pie and radial graphs).
• Legend: This is the description of the symbology used, be it by colors or plot used in the graphic to differentiate the categories to which the statistical data refer.

**Foot of the chart:** It is placed at the bottom of the chart, it contains: calls or notes and the source. It is used to write down those clarifications about the information for a better interpretation by the reader, in addition they should be visible.

•**Calls or notes:** They are used to clarify some detail of the graph. It is possible to add more than one call and these should always be placed before the source.

•**Source:** Its purpose is to grant credit to the entity responsible for producing the information. It must contain the name of the document where the information was obtained and the year when it was published, so that the user has the possibility to find the original source.

### 3.2.3 Body types

#### 3.2.3.1 Bar body

Their bars can be horizontal or vertical, the graphics have two axes, the "x" axis and the "y" axis. Depending on the graph, either horizontal or vertical, one of the axes will represent numerical information, and the other one the study categories. The bars should be the same width, and the spaces between the bars.

**Horizontal bar chart:** Used if working with qualitative or geographic variables. It is recommended that the bars appear from greater to lesser according to the frequency or percentage.

**Vertical bar graph:** It is used if the variables are chronological or quantitative. In this case, the "y" axis will contain the categories and on the "x" axis, the chronological or quantitative values. It is recommended that the graph follow a chronological sequence, in order to facilitate its reading, and the bars are the same color and thickness.

### 3.2.3.2 100% bar chart or stacked bar chart:

It consists of a single bar, either horizontal or vertical, ranging from 0% to 100%, which is divided into parts, where each corresponds to a category and its length depends on the corresponding percentage of the total. Each of its categories must be identified with different colors to achieve a better appreciation and understanding.

### 3.2.3.3 Pie chart

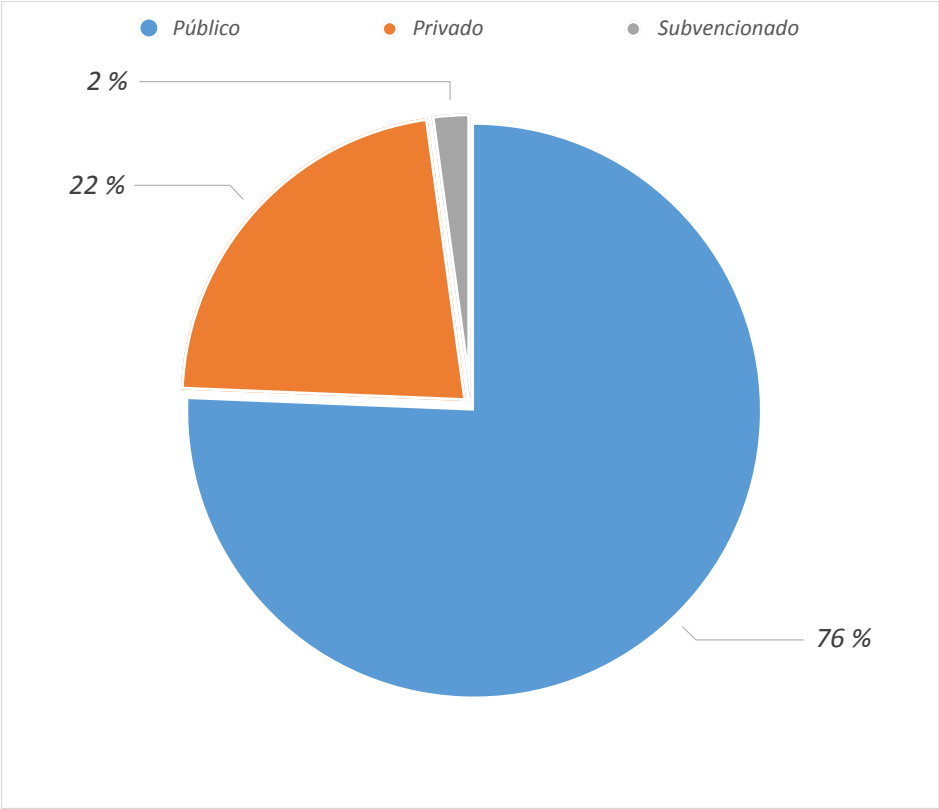
The pie chart, also called "pie," consists on dividing a circle into sectors which surfaces are proportional to the amounts corresponding to each category.

The categories must follow the order of the clock hands and will be organized according to the magnitude of each of them. That is, the category with the highest percentage will be placed first and so on until the last category placed is the one with the lowest percentage (see Figure 2).

If there are options such as: "Do not know", "Other", "Do not respond", these should be placed at the end of the graph.

As in the 100% bar graph, the categories must be in different colors and striking for a better understanding of the data.

**Figura 2**
**Costa Rica: Distribución porcentual de la matrícula inicial de estudiantes en colegios por sector en el año 2016**

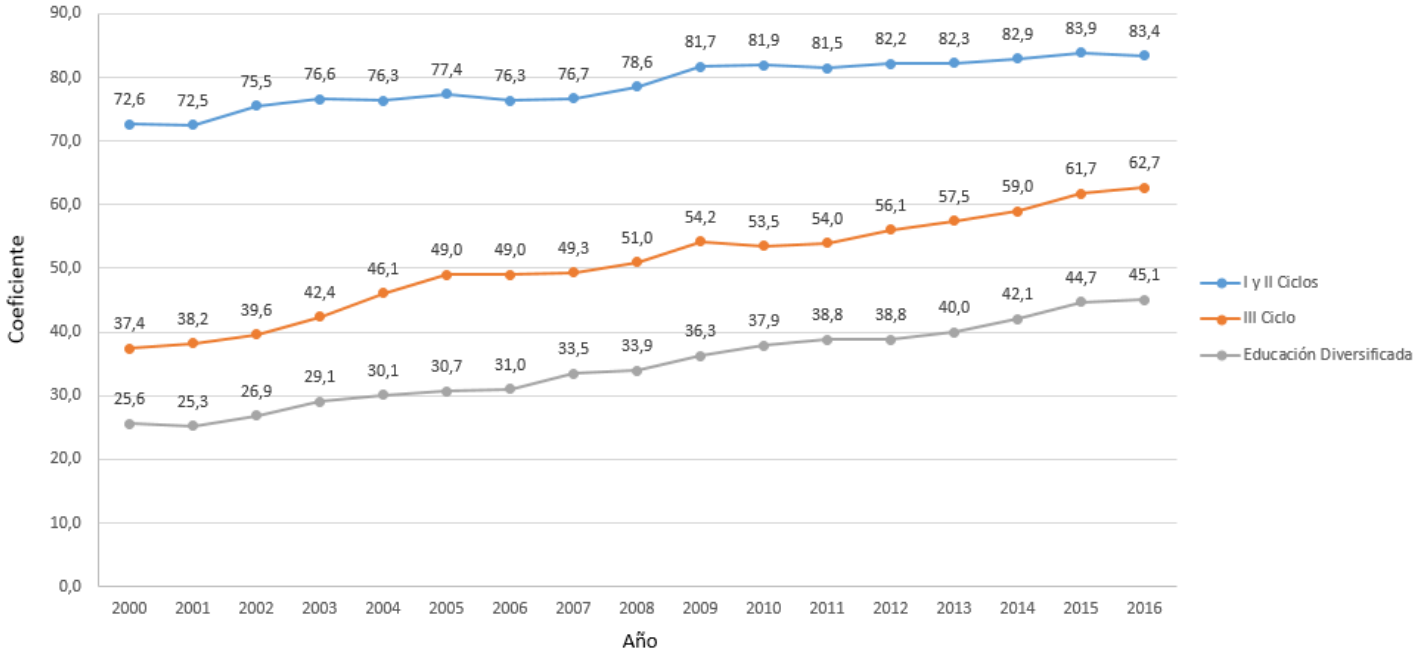



Fuente: Elaboración propia, con información del Departamento de Análisis Estadístico del Ministerio de Educación Pública de Costa Rica, 2017.

### 3.2.3.4 Linear graphs

They are used in chronological series to show the relationship between two continuous variables. Linear graphs cover or can be used for periods, such as minutes, hours, days, months, quarters or years. The time series for each category must be identified with different colors to avoid confusion between them (see Figure 3).

**Figura 3**
**Costa Rica: Coeficiente de permanencia aparente en centros educativos por año según nivel educativo, periodo 2000 – 2016**




Fuente: Elaboración propia, con información del Departamento de Análisis Estadístico del Ministerio de Educación Pública de Costa Rica, 2017.

## 3.3 Thematic maps

Thematic maps are tools that help the representation of the spatial distribution of statistical data related to one or more subjects in a map of territorial division. It is recommended for the representation of geographic data.

The thematic maps do not use the numeric representation of the variables under study, but, rather, use symbols to show the classification found in categories or intervals. For the elaboration of the thematic maps, different striking colors must be used in order to avoid confusion at the time of the map visualization.

### 3.3.1 Elements of thematic maps

To prepare thematic maps, the following elements must be considered (see Figure 4):

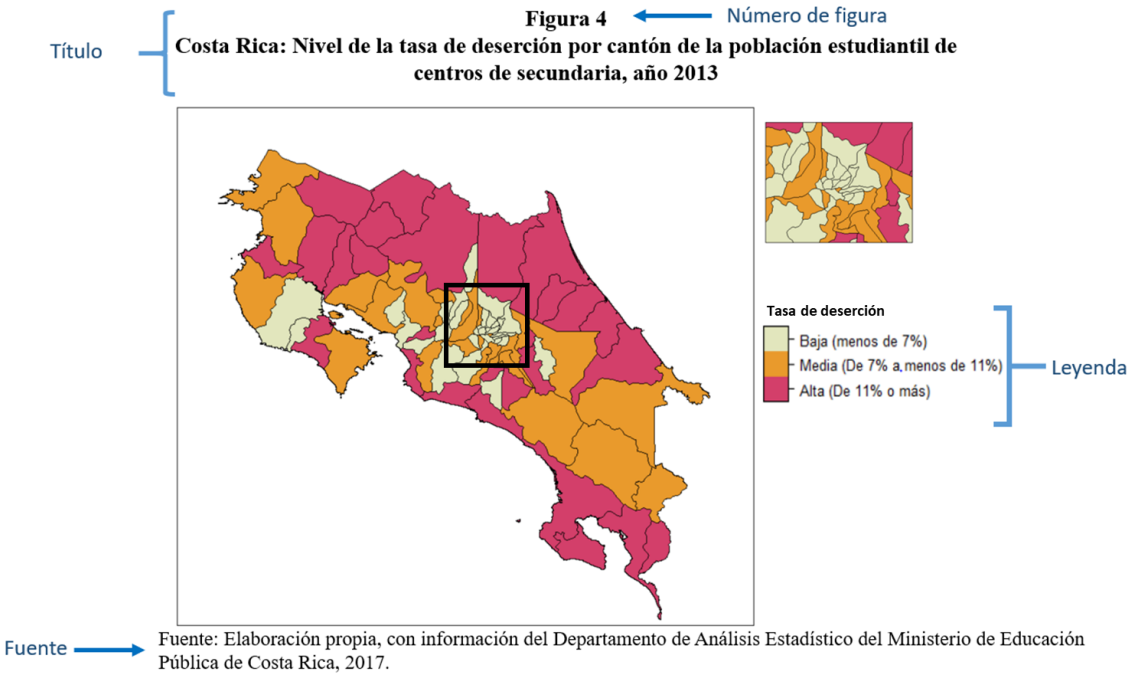
**Figure number (map):** It helps to identify the map. If there is more than one map or listed in the document, the enumeration must be consecutive. Arabic numbers should be used in the enumeration .

**Map title:** Must be short and concise. Shoul contain aspects such as the name of the region and geographic subdivisions, in addition to the displayed information.

**Preliminary note:** Used to add map information that cannot be captured properly in it. It must be placed under the title. They are generally used to specify the unit of measure being used.

**Legend:** Indicates what each key or symbol represents on the map.

**Grid:** It is used to show its precision in terms of latitudes and longitudes (Optional, depends on information).



## 3.4 Infographics

Infographics are relatively a new tool that has revolutionized the presentation of data. This consists on exposing the results and ideas of the information in an easy way in order to became simple to understand for the reader.

To prepare the infographics, you must have knowledge in graphic design and specialized software. An infographic must contain the most relevant information of the study or research, it should not be saturated with information because this could confuse to the readers and thus lose their objective, clarity.

### 3.4.1 Elements of an infographic

The following elements must be considered for the elaboration of infographics (González, 2017):

**Title:** Indicates the theme of the infographic, must be concrete and concise.

**Text:** consists of a brief explanation for the understanding of infographics body.

**Body:** Explanatory graphic visual information is presented. It must contain the unit of measurement of the data.

**Source:** Provides information related to the creation of the map, whether they are bibliographic sources about where the data or information of the creator was collected, in case of own elaboration.

Seguidamente se ofrece el ejemplo de una infografía:

## 4. Presentation of results of statistical and mathematical material

For the presentation of the results, there must be a reference to the Publications Manual of the American Psychological Association (APA).

### 4.1 Statistical tests

Related to the results report of the  of the statistical tests, it is recommended to include the corresponding information  in order to  the  important aspects can be identified in the performance of the hypothesis contrast, and thus corroborate the results. Table 2 details aspects of some frequently used statistical tests.

**Table 2. Some statistical tests for the representation of statistical and mathematical according to Publication Manual of APA**

| *Test* | *Description* | *Example* |
|---|---|---|
| **t Student** | Include in brackets the degree of freedom followed by statistic value and the p-value. | t(177)=3.51, p<.001 |
| **Chi-Squared** | It is written with it's degrees of freedom and the sample size in parentheses, the statistical value with two decimal places and the level of significance. | $\chi 2(1, N=90) = 0.89, p<.05$ |
| **Test F** | Identify with degree of freedom, statistical value and p-value. | F(2,177) = 6.30, p=.002 |

Source: Adapted from Publication Manual of the APA (2010 p.118).

To know how other tests are reported, it is recommended to read the documents found in the following links:

- http://www.bryanburnham.net/wp-content/uploads/2014/01/Reporting-Statistics-in-APA-Format.pdf
- https://depts.washington.edu/psych/files/writing_center/stats.pdf
- http://evc-cit.info/psych018/Reporting_Statistics.pdf

### 4.2 Statistical Abbreviations

In the interpretation of the statistical analyzes, abbreviations are used to detail the magnitudes of the results obtained, therefore, using the correct abbreviations and symbols are relevant for an adequate interpretation. Table 3 shows some frequently used abbreviations and symbols, likewise it is suggested to consult the APA Publications Manual.

**Table 3. Statistical abbreviations and symbols, according to Publication Manual of the APA**

| *Abbreviation / symbol* | *Definition* |
|---|---|
| ANCOVA | Analysis of covariance |
| ANOVA | Analysis of variance |
| SD | Standard deviation |
| MSE | Mean square error |
| SE | Standard error |
| df | Degrees of freedom |
| f | Frequency |
| $H_0$ | Null hypothesis |
| $H_1$ | Alternative hypothesis |
| CI | Confidence interval |
| n | Number of cases (generally in a subsample) |
| N | Total number of cases |
| M | Sample mean, arithmetic average |
| Mdn | Median |
| ns | Not statistically significant |
| P | Porcentage, percentile |
| R | Multiple correlation |
| $r^2$ | Coefficient of determination; measure of strength of relationship; estimate of the Pearson product-moment correlation squared |
| $r_S$ | Spearman rank order correlation |
| R | Multiple correlation |
| $R^2$ | Multiple correlation squared; measure of strength of association |
| s | Sample standard deviation |
| $s^2$ | Sample variance |
| SS | Sum of squares |
| t | Student's t distribution, a statistical test based on the Student t distribution; the sample values of the t-test statistic |

Source: Adapted from Publication Manual of the APA (2010, pp.121-125).

## 4.3 Some general recommendations

Regarding the symbols referring to parameters and statistics, lowercase Greek letters should be used for the population mean μ and population variance σ2.

For the number of people  of a total population, the letter "N" will be used in italics (italics) (N = 156) and to designate samples, lowercase "n" will be used in italics (n = 56).

When a software or statistical language provides a p or significance value of .000, it is recommended to assign an infinitesimal magnitude or it is suggested to write p <0.0001. In addition, it is recommended to read the pronouncement of the American Statistical Association (related to p-value, which is available at this link http://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf)

For some statistical analyzes it is recommended to calculate some aspects such as confidence intervals, effect size and present the coefficients of standardized models.

For relevant descriptive statistics, such as the mean and standard deviation, it is recommended to write it as: "(M = 54, SD = 10.35)".

## 5. Bibliography

-American Psychological Association (2010). APA Publications Manual. Mexico: Editorial The Modern Manual.

-González, Zúñiga. M. (2017). Infographic contributions for the investigation of the Observatory of National and Regional Education. (Research Project No. 724-B4-245). Recovered from: http://repositorio.inie.ucr.ac.cr/handle/123456789/429

-National Institute of Statistics and Informatics. (2006). Manual for the presentation of statistical tables. Lima, Peru: Research and Development Center.

-National Institute of Statistics and Informatics. (2009). Guide for the presentation of statistical graphs. Lima, Peru: Research and Development Center.

-National Institute of Statistics and Censuses. (2016). Guide for the presentation of statistical information. Recovered from: http://www.inec.go.cr/sites/default/files/documetos-biblioteca-virtual/mepresentinfoestadist-28072016.pdf

-Ministry of Labor and Social Security. (2013). Manual for the elaboration of tables and statistical graphs. San José, CR: MTSS, General Directorate of Work Planning, Labor Market Observatory.