

Statistical processing of clinical information

Volume 10 Issue 3 - 2025

Juan José García García

Head of the Pediatric Service at Sant Joan de Déu Hospital

Correspondence: Juan José García García, Head of the Pediatric Service at Sant Joan de Déu Hospital, Tel 34632532100

Received: June 9, 2025 | **Published:** September 10, 2025

Introduction

For the purposes of clinical or clinical-epidemiological research, the data of a person receiving medical care, which are collected by interrogation, such as those offered by the identification card, the various types of history, and the set of signs and symptoms related to the current condition, those obtained by physical examination, as well as the information that comes from auxiliary diagnostic tests, they constitute, in statistical terms, a set of variables that can be classified as qualitative or quantitative, and practically all the processing that is done of them depends on the type to which they correspond, so it must be clear about them, so it will be the axis of development of this work.

Qualitative variables

In this first group, two major types are identified, depending on the possibility of identifying the presence or not of a hierarchy between the modalities they take. For example, if with respect to a variable it can only be established whether it is of one variety or another (sex), or whether it is present or absent (such as a hereditary family antecedent or a sign or symptom), they are called *dichotomous nominals*. In the event that there are several options (such as marital status or religion), and there is no relationship of order between them, they are called polytomic, and both situations correspond to an elementary level of measurement.

In the event that there is or can be defined a gradation or hierarchy between the modalities that the characteristic may take, such as schooling, the degree of dyspnea or jaundice, then we speak of *ordinal qualitative variables*.

With respect to qualitative variables, health personnel may ask themselves about the main reasons for consultation that they attend?, what are the most common clinical manifestations of a given condition?, how many of the patients with an established diagnosis present a particular clinical data?, how many evolve satisfactorily with the prescribed management?, what are the adverse reactions of a certain pharmacological treatment?. These, and many other questions, lead to the use of the corresponding absolute figures when carrying out the classification and counting of the data of interest. Along with the above, various types of measurements can be calculated to summarize the information and express its magnitude. For example, **ratio measures** relate to each other, through a quotient, independent subsets, in such a way that when placing the largest figure in the numerator, the denominator acts as a reference value, (the unit), and the result is read, for example, 2: 1 (two to one). This may be that for every two cases of a disease in males there is one in the female sex. The measures that can be calculated in the various types of epidemiological studies in order to establish whether there is an association between a variable considered as exposure and the occurrence of a particular health damage, are within this group: relative risk, odds ratio, prevalence ratio, incidence rate ratio. It should be noted that in this type of indicators, a value greater than one suggests that a factor analyzed may increase the probability that an unfavorable outcome will occur; a value equal

to 1 reflects that the frequency of a disease is equal between a group exposed to a factor and a group not exposed; On the other hand, a result lower than 1 suggests that the factor studied acts as a protector against the damage analyzed, since the frequency of this is higher in the unexposed group.

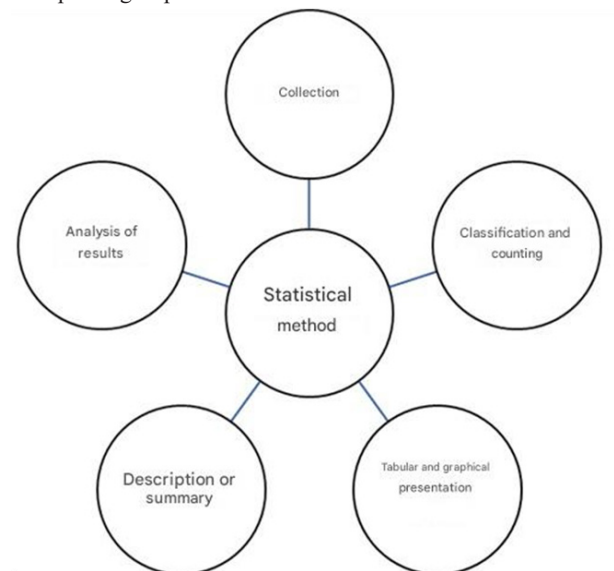


Figure 1 The above figure shows the stages involved in the statistical method.

Another way to summarize qualitative data is through the calculation of proportions (usually presented as percentages). These indicate the magnitude of a fraction with respect to the whole. For example, what proportion of the total number of cases treated in the Medical Unit during a given period corresponded to obesity? If the answer is 0.2, it means that 1 in 5 people had such a diagnosis, that is, 20%.

Some clinical-epidemiological indicators correspond to this type of summary measure: prevalence, cumulative incidence, sensitivity, specificity, predictive values of a diagnostic test.

In order to display the information collected, it is advisable to construct charts and/or graphs. The choice of one way or the other depends on intentionality. For example, if you want the reader to have an overview of the behavior of a variable, you can use a graph, while if you want to provide more detailed information, a table is preferable. In any case, they are required to have a degree that answers four basic questions: what is presented? (main variable), how is its behavior?

(by age, sex), where does the data come from? (geographical area, hospital service, state), and when did they occur? (moment or time to which they correspond).

The body of the table or graph itself must conform to a principle of clarity and simplicity, which means that it must have the respective headings in columns and lines, on the corresponding axes, and not overload with information.

Thirdly, in each case, the source from which the data comes must be indicated, as well as explanatory notes regarding keys or acronyms that would have been used in the previous sections.

Different types of bar graphs are used to present this type of data:

Simple, if only one variable is shown. Alternatively, you can build a pie chart.

If you want to show the simultaneous behavior of two qualitative variables, through the absolute values found, you can use associated bars.

If you prefer to present the information together of two variables by means of relative frequencies, subdivided bars can be used.

Quantitative variables

When the information is numerical, but the data can only be expressed as integer values, there are so-called **discrete** variables, such as the number of pregnancies, respiratory and heart rate, and blood pressure. Age, for practical purposes, is considered in this modality.

When values can be expressed with whole and decimal numbers, we speak of continuous variables, such as anthropometric measurements, and various laboratory results.

In reviewing the data once captured, it is important to identify those cases in which a decimal point was omitted or placed where it did not belong.

With regard to the graphical presentation of this type of variables, it is recommended to use histograms to process discrete data, and frequency polygons for those that are continuous.

If you want to show the simultaneous behavior of two quantitative variables, you can use scatter plots.

In the case of a quantitative and a qualitative variable, histograms or polygons of overlapping frequencies can be used, or a modality of the type of population pyramids.

As for the summary measures for quantitative variables, there are two groups: those of central tendency (mode, median, mean or arithmetic average), and those of dispersion (rank, percentiles or another variant, and standard deviation).

It is worth emphasizing that, when describing the behavior of a variable, both types of summary measures should be used, in order to have a clearer idea of how it is distributed. For each measure of central tendency there is a measure of dispersion.

Although they can all be calculated, and statistical programs can provide them automatically, the decision of which are the most appropriate for each case depends on the consideration of whether or not a variable is distributed in a similar way to the so-called normal or Gaussian curve. There are several criteria that support the decision to

be made, such as the use of indicators such as asymmetry and kurtosis, the result of statistical tests of significance about the goodness of fit (chi-square, Kolmogorov-Smirnov, Shapiro) or a graphical distribution (Q-Q).

If it is concluded that the variable meets the corresponding requirements, the mean and its corresponding measure of dispersion, the standard deviation, can be used, which is considered to be more robust to subsequently carry out certain types of information analysis techniques. Otherwise, it is appropriate to use the median and other percentiles, considering that the median corresponds to the 50th percentile. Or some variants, such as quartiles, tertiles, deciles or quintiles.

Mode is only a curiosity, in the sense that it indicates which value was repeated the most times, but there is not always a data that meets that criterion, if all the data are different from each other, or there can be more than one mode.

The range speaks of the distance between the maximum and minimum values that were found in a series. The so-called interquartile or interquartile range concentrates 50% of the values in a series, that is, those that are located between the 25th and 75th percentile.

The last stage of statistical processing corresponds to the analysis, in which, in a complementary way, it is possible to resort, on the one hand, to the estimation of confidence intervals for the point estimators calculated in the sample (proportions, averages, measures of association), in order to extrapolate the results to the population, expressing the probability that the values at that level are within a certain range, and, on the other hand, the application of significance tests to evaluate the participation of chance in the results obtained. The selection of techniques depends, among other things, on the type of variables analyzed, as has been insisted throughout this work, the objective of the study (to identify differences between groups or associations between variables), and the number of groups studied.

Acknowledgments

None.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Villa-Romero A, Moreno-Altamirano L, García-de la Torre GS. *Epidemiology and public health statistics*. Mexico: McGraw-Hill; 2011.
2. Martínez-González MA, Sánchez-Villegas A, Toledo-Atucha EA, et al. *Friendly Biostatistics*. 3rd ed. Barcelona: Elsevier; 2014.
3. Dawson GF. *Easy Interpretation of Biostatistics: The Connection Between Evidence and Medical Decisions*. Barcelona: Elsevier; 2009.
4. Henquin R. *Epidemiology and Statistics for Beginners*. Buenos Aires: Corpus Editorial; 2113.
5. Pagano M, Gauvreau K. *Fundamentals of Biostatistics*. Mexico: International Thompson Editores; 2001.
6. Daniel W. *Biostatistics: Basis for the Analysis of Health Sciences*. 4th ed. Mexico: Editorial Limusa; 2002.
7. Dawson B, Trapp RG. *Medical Biostatistics*. 4th ed. Mexico: Modern Manual; 2005.