Chapter 09 / Capítulo 09

Evidence-based biomedicine: methodology for research, standardization, and scientific procedural aspects (English Edition)
ISBN: 978-9915-704-43-2
DOI: 10.62486/978-9915-704-43-2.Ch09
Pages: 52-56

©2025 The authors. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY) 4.0 License.

Statistics / Estadística

Statistics is a scientific discipline focused on the systematization, collection, organization, and presentation of data related to phenomena characterized by variability or uncertainty. Its purpose is to study these phenomena methodically to infer the laws that govern them and, from there, make predictions, take decisions, or reach conclusions. In general, statistical analyses are categorized into two types: Descriptive Studies, which focus on summarizing data, and Inferential Studies, which allow projections, decisions, and conclusions to be drawn from the data. The fundamental steps in a statistical study include:

1. Formulation of hypotheses.
2. Descriptive analysis of the sample collected.
3. Application of inference techniques to draw conclusions about the population under study.

9.1. Basic conceptualization

Before beginning a statistical analysis, it is essential to understand some key concepts. These include the population, which is the total set of interest for the analysis. Often, the population is so large that it is impractical or too costly to access all of its members. For this reason, a sample is used, which is a representative subset of the population. The information we collect from each individual in the sample is described through variables, which can be of two types:

1. Qualitative: These variables are not numerically measurable and refer to:
   - Attributes: These are nominal variables that are divided into different categories, such as dichotomous (male/female, single/married, exposed/unexposed) or polytomous (blood type).
   - Ordinal: Variables where categories can be ordered hierarchically (e.g., patient pain level: severe, moderate, mild, none; patient functional grade: I, II, III, IV).

2. Quantitative: These are variables that can be measured and are classified as:
   - Discrete: Those whose domain is a countable set of values and do not allow intermediate values between two consecutive values (e.g., number of children, number of surgical procedures).
   - Continuous: Their domain is continuous, and they can take infinite values between two points (height, weight, blood pressure).

9.2. Numerical description: frequency distribution

A frequency distribution is a tool that organizes and presents observations in an orderly manner.

For ungrouped frequency distributions, the following elements are defined: the values of the variable (xi), the frequency of each value (ni), the proportion of each value (fi), and the cumulative proportion up to a given value (Fi).

In the case of grouped frequency distributions, which are useful when the variable has a large number of different values, the data is organized into intervals. A common difficulty when creating a table with grouped data is that the upper limit of one interval may coincide with the

lower limit of the next. To solve this, it is assumed that each interval is closed at its lower end and open at its upper end. This data organization makes it easier to identify patterns in the responses.

L_{i-1}, L_i	n_i	f_i	F_i
$[L_0, L_1[$	n_1	f_1	F_1
$[L_1, L_2[$	n_2	f_2	F_2
...	...	...	...
$[L_{n-1}, L_n[$	n_n	f_n	F_n

Figure 9.1. Frequency distribution using confidence intervals

9.3. Graphical representations

In addition to the statistics already mentioned, we can summarize data using graphical representations, whose forms vary depending on the type of variable. For qualitative variables, common graphical representations are:

- Bar chart: In this chart, the categories of the variable are represented on the vertical axis (ordinates) and the relative frequencies on the horizontal axis (abscissas). The heights of the bars are proportional to the frequencies, whether absolute or relative.
- Pie chart: A circle is divided into sections, each representing a category of the variable. The area of each section is proportional to the frequency of the category it represents.

For quantitative variables, particularly continuous ones, the following are used:

- Histograms: In a histogram, the rectangles are adjacent, reflecting the continuous nature of the variable. The midpoint of each interval serves as a reference for each rectangle, representing the distribution of the data in continuous intervals.

9.4. Descriptive analysis

As mentioned above, descriptive analysis is useful for summarizing data. This summary can be done using frequency tables or graphical representations. To summarize the data, we have certain statistics (numerical values derived from the sample), which are classified into two categories:

1. Central tendency statistics: These provide information about the typical or average values of a data set. Among the central tendency statistics we will discuss are:
    - Mean: Calculated as the sum of the values divided by the total number of observations. It is susceptible to the influence of extreme values.
    - Median: This is the value that divides the data set in half, leaving 50 % of the observations on each side. It is not affected by extreme values.
    - Mode: This represents the value or values that appear most frequently. Unlike the mean and median, there can be more than one mode, resulting in unimodal or

multimodal distributions.

2. Dispersion statistics: These indicate how dispersed or concentrated the data is around the central value. The greater the dispersion, the further the values are from the center. They include:
    - Variance: Measures the average of the squared deviations of the observations from the mean. Sensitive to extreme values and difficult to interpret because its units are quadratic.
    - Standard deviation: This is the square root of the variance, expressed in the same units as the original data, and commonly used to describe the variability of data in research.
    - Pearson's coefficient of variation: A measure of relative dispersion that makes it easier to compare the dispersion between different distributions. It is calculated as the ratio between the standard deviation and the mean, and is generally expressed as a percentage. For example, if the mean is 10 and the standard deviation is 5, the standard deviation represents 50 % of the mean.

3. Shape statistics:
    - Skewness: Refers to how distributions can vary between being symmetric or asymmetric. In a distribution with positive skewness, the right tail is longer, resulting in a positive skewness coefficient. Conversely, in a distribution with negative skewness, the left tail is longer, and the coefficient will be negative.

In perfectly symmetric distributions, the asymmetry coefficient is zero. In such distributions, the mean, median, and, if unique, the mode coincides.

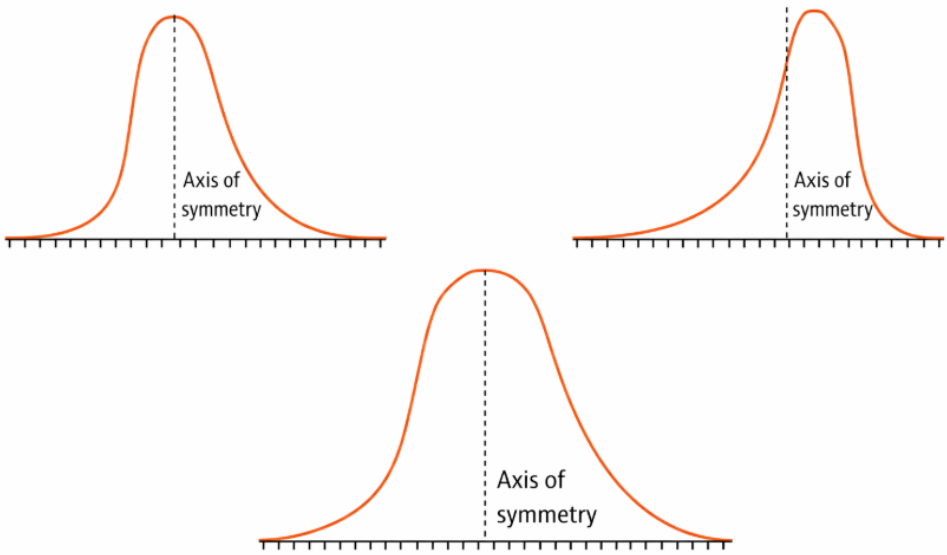

Figure 9.2. Asymmetry and symmetry of distributions

- Kurtosis: This statistic evaluates how peaked or flat a distribution is compared to a normal distribution. Depending on its value, the distribution can be classified as platykurtic (kurtosis less than 0), mesokurtic (kurtosis equal to 0), or leptokurtic (kurtosis greater than 0).

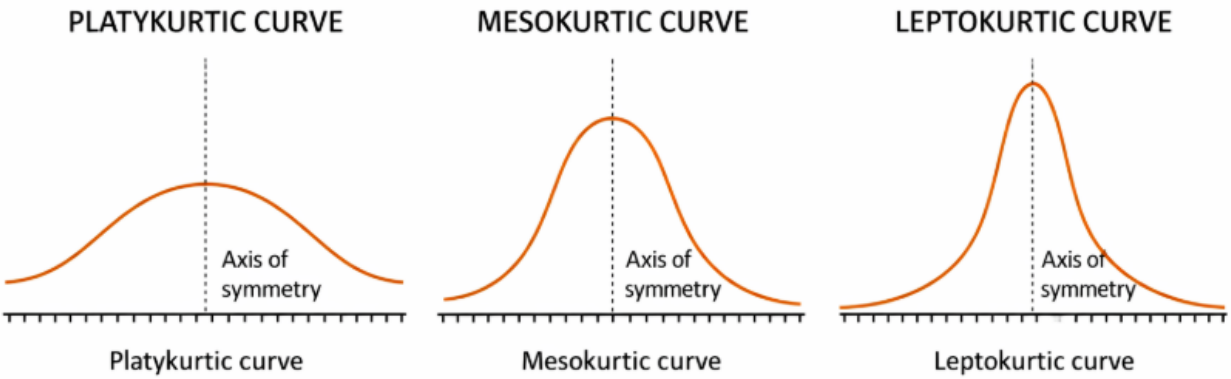

Figure 9.3. Distribution note

4. Position statistics:
   - Quartiles: These are three values that divide a data set, ordered from lowest to highest or vice versa, into four equal parts. Each part represents 25 % of the data.
   - Deciles: These are nine values that divide a data set, also ordered in ascending or descending order, into ten equal segments. Each segment contains 10 % of the data.
   - Percentiles: These are ninety-nine values that organize a data set, ordered from lowest to highest or vice versa, into one hundred equal parts. Each part covers 1 % of the data.