

CUADERNO DE TRABAJO

# Elementos básicos para el procesamiento, el análisis y la interpretación de la información estadística en salud



Ana Ariovich

EDICIONES **UNGS**



Universidad  
Nacional de  
General  
Sarmiento

**CUADERNO DE TRABAJO**

---

**Elementos  
básicos para el  
procesamiento,  
el análisis y la  
interpretación de  
la información  
estadística  
en salud**

**CUADERNO DE TRABAJO**

---

**Elementos  
básicos para el  
procesamiento,  
el análisis y la  
interpretación de  
la información  
estadística  
en salud**

---

Ana Ariovich

---

EDICIONES **UNGS**



Universidad  
Nacional de  
General  
Sarmiento

Ariovich, Ana

Elementos básicos para el procesamiento, el análisis y la interpretación de la información estadística en salud / Ana Ariovich. - 1a ed. - Los Polvorines : Universidad Nacional de General Sarmiento, 2020.

64 p. ; 28 x 20 cm. - (Política, políticas y sociedad. Cuadernos de trabajo ; 3)

ISBN 978-987-630-450-4

1. Salud. 2. Estadísticas. I. Título.

CDD 610.727

© Universidad Nacional de General Sarmiento, 2020

J. M. Gutiérrez 1150, Los Polvorines (B1613GSX)

Prov. de Buenos Aires, Argentina

Tel.: (54 11) 4469-7507

ediciones@campus.ungs.edu.ar

ediciones.ungs.edu.ar

Diseño gráfico: Daniel Vidable - Ediciones UNGS

Diagramación: Eleonora Silva

Corrección: Andrea Gardey

Hecho el depósito que marca la Ley 11.723

Prohibida su reproducción total o parcial

Derechos reservados

# Índice

**PRESENTACIÓN** .....7

## **PARTE I**

### **De los datos al procesamiento y análisis de la información en salud**

#### **Módulo 1**

Las nociones básicas y los desafíos de la medición..... 11

#### **Módulo 2**

El análisis univariado.....23

#### **Módulo 3**

El análisis de más de una variable y la búsqueda de relaciones .....41

## **PARTE II**

### **Aportes de la estadística al campo de la salud**

#### **Módulo 4**

Las estadísticas de salud y los indicadores de uso frecuente .....53



# Presentación

Este cuaderno de trabajo busca contribuir al proceso de formación y capacitación de quienes desempeñan su práctica profesional dentro del ámbito de la salud, brindando un marco general para la utilización, el análisis y la interpretación de la información cuantitativa en salud. De manera más específica, persigue colaborar en el desarrollo de conocimientos y competencias vinculados a la estadística a ser aplicados al campo de la gestión de las políticas de salud en el territorio.

Destinada a la medición, el procesamiento y el análisis de los datos sociosanitarios, la primera parte del cuaderno contempla tres módulos que abordan nociones y herramientas básicas de la estadística y ejemplos ilustrativos. Además, sistematizan contenidos que orientan hacia la lectura y comprensión crítica de la información cuantitativa.

El primero de estos módulos da cuenta de las complejidades, los problemas y las decisiones que supone el proceso de

la medición de la evidencia empírica, a la vez que introduce a los lectores en los propósitos del método estadístico y en sus principales conceptos (población, unidad de análisis y muestra).

Centrado en el procesamiento y análisis univariado, el segundo módulo revisa representaciones (tabulares y gráficas) y técnicas de la estadística descriptiva (medidas resumen, de posición y de variabilidad); el siguiente módulo, en cambio, focaliza en la utilización de aquellas técnicas que posibilitan analizar simultáneamente dos variables a los fines de determinar posibles relaciones.

Los aportes más importantes de la estadística al campo de la salud se desarrollan en la segunda parte del cuaderno. En esta, un cuarto módulo presenta las estadísticas en salud y los indicadores de uso frecuente. Asimismo, explora las principales fuentes de información en el sector y refiere tanto a las potencialidades como a las limitaciones en la utilización de los datos secundarios disponibles.





# PARTE I

De los datos  
al procesamiento  
y análisis de la  
información en salud







# Módulo 1

## Las nociones básicas y los desafíos de la medición

## El método estadístico

Podemos comenzar pensando a la *estadística* como una disciplina con base matemática, que provee los métodos y los procedimientos para recolectar, procesar, analizar, presentar e interpretar un conjunto de datos cuantitativos sobre la realidad. La característica principal que presentan estos datos es que sus valores cambian de un elemento a otro (no son constantes), y reciben, por ello, el nombre de *variables*; si no hay variabilidad en la información a procesar y a analizar, no tiene sentido el uso de la estadística.

Siguiendo a Torres (2009), el método estadístico tiene como propósito estudiar las propiedades (o variables) que presentan objetos, personas o eventos y medirlos con la finalidad de

- describirlos (en intensidad, variabilidad, etcétera),
- compararlos y
- predecir o conocer su probabilidad de ocurrencia.

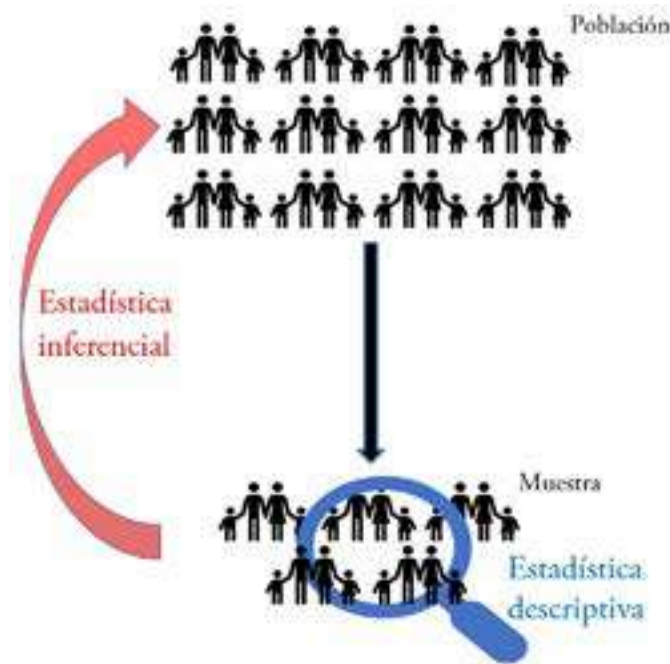
De acuerdo a estas finalidades, la estadística puede ser dividida en dos grandes ramas.

- *Estadística descriptiva*: tiene por objeto organizar y presentar un conjunto de datos para interpretar cómo estos se comportan, lo que permite describir de manera precisa la variable o las variables analizadas. En las técnicas descriptivas, los datos pueden ser resumidos numéricamente mediante estadísticos (por ejemplo, un promedio) o gráficamente.
- *Estadística inferencial*: utiliza los resultados estadísticos obtenidos en el análisis de una porción de la población (o muestra)<sup>1</sup> para realizar inferencias o generalizaciones acerca de determinados parámetros o valores poblacionales. La estadística inferencial se dedica, entonces, a sacar conclusiones sobre la población a partir de los datos de una muestra; recurre para este propósito a un conjunto de procedimientos estadísticos basados en la teoría de las probabilidades.

---

<sup>1</sup> Los conceptos de muestra y población son explicados más adelante.

Figura 1.1. Ramas de la estadística



Fuente: elaboración propia.

## ¿En qué consiste la medición de variables?

Antes de avanzar sobre la medición en estadística, debemos precisar un poco más el concepto de variable. Llamamos *variables* a aquellas características que presentan los individuos, los objetos o los fenómenos que integran una muestra o un universo que estamos interesados en estudiar. Estas características pueden asumir distintos valores o intensidades, y sus posibles valores reciben el nombre de *categorías* (por ejemplo, la variable *sexo* tiene como categorías “femenino” y “masculino”); las categorías tienen que ser exhaustivas y mutuamente excluyentes.

Si la variable presenta un atributo no cuantificable, se denomina *cualitativa*; inversamente, si la variable se expresa a través de valores numéricos y representa cantidades, se designa como *cuantitativa*.

En este marco, sostenemos que *la medición* es el proceso por el cual se asignan “número” u “otros símbolos” a las propiedades (o variables) que conforman un elemento, de acuerdo a ciertas reglas que especifican el procedimiento a utilizar (Babbie, 1996). No todas las variables admiten los mismos niveles de medición: las variables cuantitativas permiten la aplicación de distintas operaciones aritméticas en la medición,

mientras que las cualitativas solo posibilitan una clasificación o una jerarquización. Por este motivo, las variables se clasifican de acuerdo con sus escalas de medición.

- *Cualitativas nominales*: son aquellas que poseen el nivel más bajo de medición, ya que solo admiten una clasificación. Ejemplos: lugar de residencia, condición de actividad, sexo.
- *Cualitativas ordinales*: dejan establecer –además de la clasificación– un orden jerárquico. Ejemplos: estatus social (alto, medio, bajo), nivel socioeconómico (alto, medio, bajo).
- *Cuantitativas de intervalo*: se les puede asignar una clasificación, un orden jerárquico, números a sus categorías y realizar medidas aritméticas, como suma, resta, división y multiplicación. En estas variables, podemos conocer con precisión las distancias entre las categorías, que son siempre intervalos iguales. Ejemplos: años de edad, años de escolaridad.
- *Cuantitativas de razón*: presentan el nivel de medición más alto, ya que poseen todas las propiedades de las variables de intervalo, pero también se les puede atribuir un punto de origen verdadero de valor 0. Ejemplos: el peso de un objeto, la medición del tiempo (en que ese valor 0 no es arbitrario).

Las variables cuantitativas, a su vez, se distinguen según involucren “categorías continuas” (como la presión arterial o la estatura) o “categorías enteras sin valores intermedios” (como la cantidad de hijos de una mujer); las primeras reciben el nombre de *cuantitativas continuas*, mientras las segundas, de *cuantitativas discretas*.

Figura 1.2. Clasificación de variables



Fuente: elaboración propia.



*A tener en cuenta:*

Las variables cualitativas son tan comunes como las cuantitativas en los estudios que se desarrollan en el campo de la salud. Distinguir entre ambos tipos es importante, ya que las medidas, representaciones y cálculos asociados a cada una son diferentes (como veremos más adelante).

## Los problemas de la medición

Muchas veces la medición de las características que presentan las personas o los fenómenos de la realidad que queremos estudiar es indirecta y supone el pasaje de un nivel conceptual a un plano observacional. Es decir, en ciertas oportunidades, las variables que queremos analizar no son directamente observables, y su medición implica poner en práctica un proceso de operacionalización que nos permita arribar a indicadores empíricos (Babbie, 1996; Vieytes, 2004); este proceso involucra varios pasos concatenados:

- a. La definición de la variable en términos teóricos para precisar qué entendemos por esa variable; por este motivo sostenemos que la medición es un proceso que depende fundamentalmente de la teoría.
- b. La descripción de las distintas dimensiones que abarca según esa conceptualización.
- c. La selección de indicadores empíricos que permitan medir dichas dimensiones.

Para medir el “acceso a la atención”, por ejemplo, el primer paso es explicar conceptualmente qué entendemos por tal acceso. En este proceso, el marco teórico desde el que partimos tiene una relevancia determinante. Si nos situamos en la línea que considera la problemática un atributo de la oferta, la accesibilidad es definida en términos de *la capacidad del sistema para producir prestaciones acordes a las necesidades de salud de una población determinada*. Esta capacidad involucra dos dimensiones: una *geográfica*, que refiere a la distancia entre los servicios y usuarios, y otra *socioorganizativa*, que apunta a las características de la oferta que pueden facilitar o limitar el uso potencial de los servicios (Donabedian, 1988, 2003).

Contrariamente, si recuperamos la propuesta de Wallace, que *vincula el acceso a la atención tanto a la esfera de la oferta del sistema como a la demanda de la población* (Wallace y Enriquez-Haass, 2001), ahora el problema contempla como dimensiones a la disponibilidad, la accesibilidad y la aceptabilidad. Mientras la disponibilidad da cuenta de los recursos físicos y humanos existentes en el sistema sanitario para responder a las necesidades de una población determinada, la accesibilidad alude a

los medios (económicos y físicos) con que cuenta esa población para acceder a los servicios. Asimismo, las prestaciones deben ser aceptables para quienes las utilizan, esto es, responder a sus requisitos y valores.

En ambos casos, las dimensiones involucradas en el acceso a la atención son algo diferentes, al igual que los indicadores empíricos que debiéramos elegir para medirlas, como se ilustra en las figuras 1.3 y 1.4.

Figura 1.3. Esquema de operacionalización del “acceso a la atención” según propuesta de Donabedian (1988, 2003)



Fuente: elaboración propia sobre la base de las propuestas teóricas de Donabedian (1988; 2003).

Figura 1.4. Esquema de operacionalización del “acceso a la atención” según propuesta de Wallace



Fuente: elaboración propia sobre la base de la propuesta teórica de Wallace.



En suma, las variables presentan distintos grados de abstracción: las más complejas son las que conocemos como *conceptos* y las más empíricas son las que reciben el nombre de *indicadores*. Todas las conceptualizaciones constituyen construcciones lógicas que expresan, desde un determinado punto teórico, un fenómeno que no es simplemente observable en la realidad. Contrariamente, los indicadores son atributos directamente observables de un fenómeno.

Además, para arribar a resultados precisos y confiables, la *medición* debe ser *válida* y *fiable*:

- La *validez* refiere a que el procedimiento que se utiliza mida lo que realmente se pretende medir y no otra cosa. Aquí es clave la elección de indicadores empíricos adecuados.
- La *fiabilidad*, en cambio, alude a que la utilización repetida de ese procedimiento permita –en iguales circunstancias– obtener los mismos resultados.

Mientras que, en la validez, es fundamental la elección de indicadores empíricos adecuados; en la fiabilidad, cobra relevancia la utilización de buenos instrumentos de medición (Babbie, 1996).

## Otros conceptos básicos del método estadístico

Como vimos, para llevar adelante una investigación empírica es necesario –en muchas ocasiones– transformar nuestras *formulaciones teóricas* sobre las características que presentan las personas, los objetos o los fenómenos en *indicadores* directamente observables. Asimismo, es preciso elegir la *población* (o la porción de esta) en la que estudiaremos tales variables.

Una *población* es un conjunto de elementos (individuos, objetos, fenómenos o acontecimientos) que comparten algún o algunos atributos en los cuales el investigador está interesado. Por ejemplo, si el propósito es estudiar el hábito de fumar de los adolescentes de entre 12 y 17 años de la provincia de Buenos Aires, la población estará conformada por “todos los adolescentes de esta provincia que se encuentren entre esas edades”.

Cada uno de estos elementos de la población en los cuales se medirán o estudiarán las variables de interés recibe el nombre de *unidad de análisis*. Las unidades de análisis pueden ser tanto individuales como colectivas.

Cuadro 1.1. Ejemplos de poblaciones y unidades de análisis

| <b>Población</b>   | <b>Unidad de análisis</b>  |
|--|--|
| Mujeres embarazadas a término de la ciudad de San Miguel de Tucumán. | Cada mujer embarazada a término de la ciudad de San Miguel de Tucumán. |
| Archivo de historias clínicas de un hospital.                        | Cada historia clínica del hospital.                                    |
| Pacientes diabéticos que asisten al hospital.                        | Cada paciente que asiste al hospital con diagnóstico de diabetes.      |

Fuente: elaboración propia para el ejemplo.

Cuando la población a estudiar contiene un número de individuos o de elementos que hace difícil observarla y analizarla en su totalidad, ya sea por cuestiones de costos, recursos disponibles, tiempos o accesibilidad a los datos, podemos seleccionar una porción que denominaremos *muestra* para conocer algo sobre esa población.<sup>2</sup> La muestra será representativa, en la medida en que sus elementos sean seleccionados a través de algún criterio que garantice reproducir –lo más idénticamente posible– las características de la población de la cual fueron tomados.

La cantidad de elementos que integran la muestra constituye lo que se conoce como *tamaño muestral*, y los elementos utilizados para seleccionar la muestra, como *unidades de muestreo*.

### ¿Cómo deberíamos elegir la muestra sobre la que analizaremos a la población bajo estudio?

Si bien dentro de la estadística la construcción de una muestra es un complejo proceso que ha dado lugar al desarrollo de especialistas dentro de la propia disciplina, trataremos de introducir algunas cuestiones importantes para tener en cuenta en el diseño de una muestra.

El *tamaño de la muestra*: suele ser menor, cuanto mayor sea la frecuencia con que ocurre la característica, factor o fenómeno que se estudia en la población, más homogénea sea su distribución en esta y menor sea la precisión que se desea obtener en nuestros resultados (Baranger, 1992).

El *método de muestreo*: refiere a los criterios considerados por el investigador para seleccionar los elementos que integrarán su muestra. Básicamente, las muestras se clasifican en *no probabilísticas* y *probabilísticas*: en las primeras, la elección no es aleatoria

<sup>2</sup> Si una investigación obtiene información sobre la totalidad de la población bajo estudio, se denomina *censo*.

y sigue criterios específicos para los fines del estudio; en las segundas, media siempre el azar y todos los elementos tienen la misma probabilidad de ser seleccionados.<sup>3</sup>

Las técnicas de muestreo no probabilístico más usadas son las siguientes.

- *Incidental*: los elementos van siendo incorporados a medida que acuden al sitio donde se efectúan las mediciones, hasta alcanzar un tamaño muestral previamente establecido.
- *Intencional*: el investigador selecciona deliberadamente elementos que considera representativos o típicos de la población que quiere estudiar (exige un conocimiento importante de la población).
- *Por cuotas*: el investigador propone estratos, de acuerdo con las variables que considera relevantes, y se incorporan elementos hasta completar la cuota prefijada. Se utiliza mucho en las encuestas de opinión.

Los métodos de muestreo probabilístico que se utilizan pueden clasificarse, siguiendo a Baranger (1992), del siguiente modo.

- *Aleatorio simple*: se basa en la aleatoriedad del proceso de selección. Debe ser llevado a cabo a través de un procedimiento que garantice el azar como un sorteo entre todos los elementos de la población.
- *Sistemático*: consiste en seleccionar al azar la primera unidad y, a partir de allí, tomar las siguientes unidades a intervalos constantes (establecidos según el tamaño de la muestra que se desea obtener).
- *Estratificado*: de acuerdo con un criterio predeterminado, la población se divide en subgrupos (estratos) mutuamente excluyentes, en cada uno de los cuales están contenidas unidades que poseen características semejantes (cada estrato es homogéneo dentro de este). De cada estrato se obtiene una muestra aleatoria. El muestreo estratificado se aplica cuando la población bajo estudio es heterogénea y las características de los estratos difieren entre sí (exige un conocimiento importante de la población).
- *Por conglomerados*: consiste en dividir la población bajo estudio en subgrupos –llamados conglomerados– que tienen existencia en la realidad (tales como barrios, manzanas, escuelas, etcétera), y, posteriormente, se aplica en cada uno de ellos el método de muestreo aleatorio simple. A diferencia del muestreo estratificado, en el que los subgrupos son homogéneos en su interior, en el muestreo por conglomerados, resultan conjuntos heterogéneos.

<sup>3</sup> En este tipo de muestreo, se puede cuantificar la probabilidad de cometer un error en la elaboración de la muestra, utilizando una fórmula específica llamada *error estándar* ( $S_x$ ). La fórmula para calcularlo es  $S_x = SD/\sqrt{n}$ , en la que  $SD$  = desvío estándar de una muestra y  $n$  = tamaño de la muestra (Ritchey, 2001). Sobre la noción de  $SD$ , trabajaremos en el Módulo 2.

- o *Polietápico*: implica realizar el muestreo en varias etapas sucesivas, utilizando iguales o distintos procedimientos en las diferentes fases.

*A tener en cuenta:*

Cuando buscamos extender los resultados del estudio al conjunto de la población o universo, conviene trabajar con una muestra probabilística. Este tipo de muestra reproduce, en pequeña escala, las características de la población con la finalidad de hacer inferencias estadísticas sobre esta (Vieytes, 2004).

## Bibliografía

- Babbie, Earl (1996). “Conceptualización y medición”. En *Manual para la práctica de la investigación social*, pp. 147-175. España: Desclée de Brouwer.
- Baranger, Denis (1992). “Nociones de muestreo”. En *Construcción y análisis de datos. Una introducción al uso de las técnicas cuantitativas en la investigación social*, pp. 43-60. Posadas: Editorial Universitaria.
- Donabedian, Avedis (1988). *Los espacios de la salud: aspectos fundamentales de la organización de la atención médica*. México D. F.: Fondo de Cultura Económica.
- (2003). *An introduction to quality assurance in health care*. Nueva York: Oxford University.
- Ritchey, Ferris (2001). “Uso de la teoría de la probabilidad para producir distribuciones muestrales”. En *Estadística para las ciencias sociales. El potencial de la imaginación estadística*, pp. 193-217. México, D. F.: McGraw-Hill Interamericana.
- Torres, Silvana (2009). “El método estadístico”. En *Estadísticas en salud I. Estadística descriptiva*, pp. 10-15. Facultad de Medicina, Universidad Nacional del Tucumán.
- Vieytes, Ruth (2004). “Unidad de análisis y técnicas de muestreo”; “El proceso de medición”. En *Metodología de la investigación en organizaciones, mercado y sociedad. Epistemología y técnicas*, pp. 389-459. Buenos Aires: Editorial de las Ciencias.
- Wallace, Steven y Enriquez-Haass, Vilma (2001). “Disponibilidad, accesibilidad y aceptabilidad en el sistema de atención médica en vías de cambios para los adultos mayores de Estados Unidos”. *Revista Panamericana de Salud Pública*, vol. 10, n° 1, pp.18-28.









## Módulo 2

# El análisis univariado

Como señalamos en el módulo anterior, uno de los principales propósitos de la estadística es describir un grupo de datos para comprender cómo se comportan las variables.<sup>1</sup> El análisis univariado consiste en describir de manera independiente cada una de las variables estudiadas, por eso centra el procesamiento y el análisis exclusivamente en una de las variables de interés.

Una forma de hacerlo es mostrar las *frecuencias absolutas* que se observan en las distintas categorías de la variable considerada; es decir, indicar el número de casos que hay en cada uno de los valores que asume esa variable. También se pueden presentar las *frecuencias relativas* (o porcentajes), relacionando las cifras parciales que exhibe cada categoría con el total de los datos observados para la variable. Cabe señalar que el conjunto de las frecuencias de todas las categorías de la variable se denomina *distribución de frecuencias*, y una buena manera de representarla es a través de una tabla.

Así, si queremos exponer la distribución de frecuencia de la variable “edad de los egresos hospitalarios” en la provincia de Buenos Aires para el año 2013, se debe contabilizar la cantidad de egresos registrados en cada una de las categorías definidas para esta variable. Como se ilustra en el cuadro 2.1., la variable supone como categorías diferentes tramos etarios (que van desde menores de 1 año hasta 75 años y más) y la categoría “edad ignorada”.

Cuadro 2.1. Representación tabular de la edad de los egresos hospitalarios en la provincia de Buenos Aires, 2013

| Edad de los egresos hospitalarios | Frecuencias absolutas | Frecuencias relativas |
|-----------------------------------|-----------------------|-----------------------|
| 1 año                             | 57.901                | 7,72                  |
| 1 a 4                             | 38.414                | 5,12                  |
| 5 a 9                             | 22.259                | 2,97                  |
| 10 a 14                           | 21.621                | 2,88                  |
| 15 a 24                           | 140.702               | 18,77                 |
| 25 a 34                           | 110.445               | 14,73                 |
| 35 a 44                           | 72.439                | 9,66                  |
| 45 a 54                           | 59.386                | 7,92                  |
| 55 a 64                           | 69.099                | 9,22                  |
| 65 a 74                           | 62.544                | 8,34                  |
| 75 y mas                          | 73.194                | 9,76                  |
| Edad ignorada                     | 21.769                | 2,91                  |
| Total                             | 749.773               | 100                   |

Fuente: elaboración propia sobre la base de los datos de la Dirección de Información Sistemática de la Subsecretaría de Gestión y Contralor del Conocimiento, Redes y Tecnologías Sanitarias, Ministerio de Salud de la Provincia de Buenos Aires, 2013.

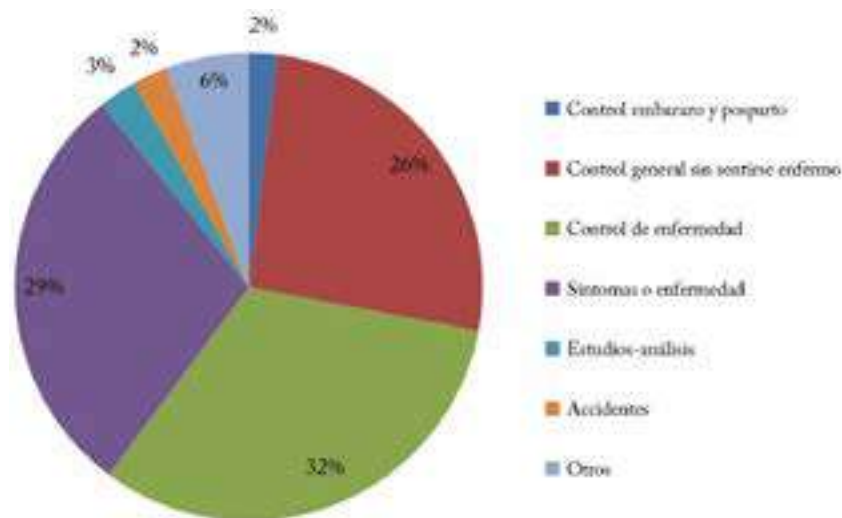
<sup>1</sup> Recordemos que, en el primer módulo, hemos referido a otros propósitos importantes de la estadística, como la comparación de datos o la inferencia de los valores que pueden asumir las variables en una población a partir de los resultados obtenidos a través de datos muestrales.



También los gráficos permiten representar una distribución de frecuencias. Para las variables cualitativas y para las variables cuantitativas discretas, podemos utilizar gráficos de tortas o de barras. Las variables cuantitativas continuas, en cambio, debemos graficarlas a través de gráficos de superficie, como los histogramas o los polígonos de frecuencias. Detallamos a continuación estos distintos tipos de gráficos.

*Gráficos de tortas:* la frecuencia expresada en % de cada categoría de la variable se representa mediante un sector del círculo proporcional a dicha frecuencia.

Figura 2.1. Gráfico de distribución de los motivos de consulta al médico en la Argentina, 2010

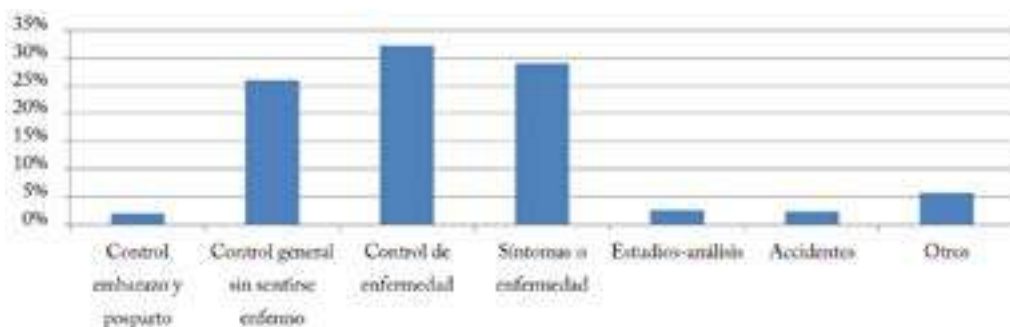


\*Población consultante con al menos 1 consulta: 15.237.030

Fuente: elaboración propia sobre la base de los datos de la Encuesta de utilización y gastos en servicios de salud 2010, Sistema Estadístico de Salud, Dirección de Estadística e Información en Salud (DEIS) del Ministerio de Salud de la Nación.

*Gráficos de barras:* el eje vertical se llama ordenada o eje “y”, en él se colocan las frecuencias (en números absolutos o en %). El eje horizontal se llama abscisa o eje “X”, en él se colocan las categorías de la variable.

Figura 2.2. Gráfico de distribución de los motivos de consulta al médico en la Argentina, 2010

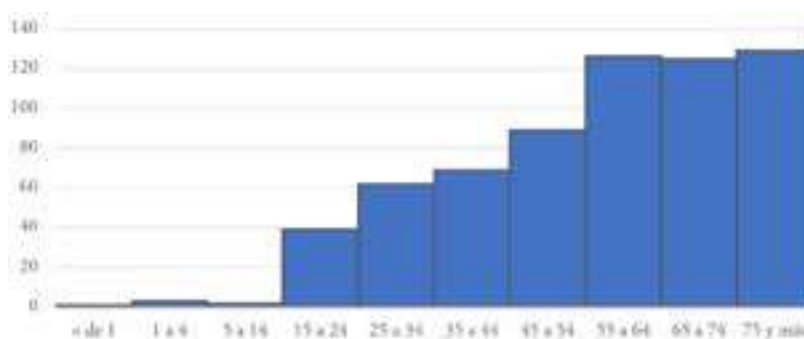


\*Población consultante con al menos 1 consulta: 15.237.030

Fuente: elaboración propia con base en los datos de la Encuesta de utilización y gastos en servicios de salud 2010, Sistema Estadístico de Salud, Dirección de Estadística e Información en Salud (DEIS) del Ministerio de Salud de la Nación.

*Histogramas:* es un gráfico formado por rectángulos que van unidos entre sí, debido a las características de continuidad de la escala. Cada rectángulo expresa la frecuencia para cada intervalo de la variable.

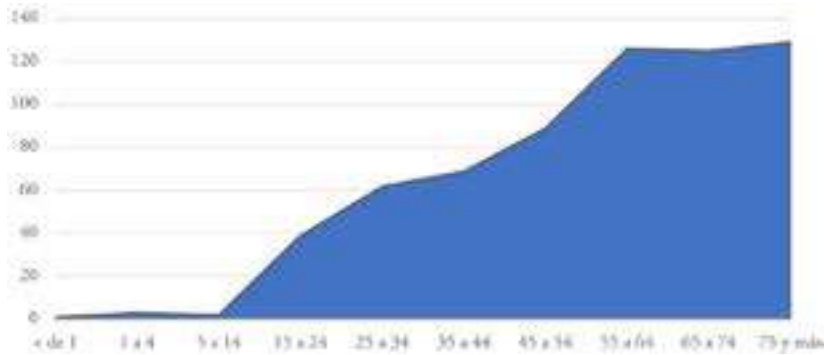
Figura 2.3. Gráfico de distribución de la edad de las defunciones por tuberculosis en la Argentina, 2016



Fuente: elaboración propia sobre la base de los datos del Sistema Estadístico de Salud, Dirección de Estadística e Información en Salud (DEIS) del Ministerio de Salud de la Nación, 2016.

*Polígonos de frecuencias:* asignan la frecuencia al punto medio en cada intervalo de la variable.

Figura 2.4. Gráfico de distribución de la edad de las defunciones por tuberculosis en la Argentina, 2016



Fuente: elaboración propia sobre la base de los datos del Sistema Estadístico de Salud, Dirección de Estadística e Información en Salud (DEIS) del Ministerio de Salud de la Nación, 2016.

## Medidas descriptivas de resumen

Además de representar las distribuciones a través de tablas y gráficos, podemos calcular algunas medidas que nos permitan describir los datos de manera resumida y cuantificar correctamente el impacto de una determinada variable.

*Razón:* es un cociente entre dos cifras de diferente o similar naturaleza y expresa la relación de tamaño entre una y otra. Por ejemplo, en un poblado de 2500 habitantes, hay 625 viviendas, la razón de habitantes por vivienda es

- $2500/625=4$ , hay 4 habitantes por vivienda.

*Proporción:* es un cociente en el que las unidades de observación que figuran en el numerador están incluidas en el denominador; constituye una comparación cuantitativa entre la parte y el todo. Cuando se multiplica por 100, se denomina *porcentaje*. Por ejemplo, si en una población de 25.000 habitantes se encontraron 1500 personas sin viviendas, la proporción de personas sin vivienda (en esa población) es

- $1500/25.000=0.06$  o el 6% de esa población carecía de viviendas.

*Tasa:* es similar a una proporción, con la diferencia de que las tasas llevan incorporados los conceptos de tiempo y lugar. Toman todos los casos de un evento, pertenecientes a una población total, en un lugar y período determinados; además, se multiplican por una constante (10 o múltiplos de 10) para facilitar la comparación de tasas de poblaciones diferentes (aunque estas sean de diferente tamaño).

En general, representan la fuerza con que se produce un hecho determinado en una población dada.

Las tasas se clasifican de la siguiente manera.

- *Tasas generales brutas*: son aquellas que se constituyen relacionando la totalidad de un fenómeno o evento con el conjunto de la población en un período determinado. Por ejemplo, la tasa de mortalidad infantil (TMI) para la provincia de Buenos Aires en el año 2017 fue el cociente entre el número de defunciones de niños menores de 1 año de edad y el número de nacidos vivos en ese mismo año multiplicado (por mil).

$$\text{➤ } T_{MI} = \frac{2461}{259.795} \times 1000 = 9,47$$

La fuente es de elaboración propia, con base en los datos del Sistema Estadístico de Salud, Dirección de Estadística e Información en Salud (DEIS) del Ministerio de Salud de la Nación, 2017.

- *Tasas específicas*: son aquellas que son calculadas en un grupo particular de la población, determinado por alguna característica (edad, sexo, ocupación, etcétera), en un período determinado. Por ejemplo, la tasa específica de la mortalidad femenina es el cociente entre el número de defunciones femeninas y número total de población de ese mismo sexo en ese año (por mil).

$$\text{➤ } T_{MF} = \frac{67.940}{8.678.079} \times 1000 = 7,82$$

La fuente es de elaboración propia, con base en datos del Sistema Estadístico de Salud, Dirección de Estadística e Información en Salud (DEIS) del Ministerio de Salud de la Nación, 2017.

## Medidas de tendencia central

Las medidas de tendencia central tienen como objeto ubicar el centro de distribución. Las más frecuentes son las siguientes:

- moda,
- promedio o media aritmética y
- mediana.

*Moda*: es el valor de la variable que se presenta mayor número de veces. Se identifica visualmente y se simboliza  $M_o$ . Existen distribuciones bimodales. Se puede utilizar para describir variables con escalas nominales, ordinales, intervalares o de razón.

*Promedio o media aritmética*: es el cociente que se obtiene al dividir la suma de los valores que asume una variable por el número de casos observados y se simboliza  $\bar{x}$ . La fórmula correspondiente a este cálculo es  $\bar{x} = (\sum x_i)/n$ , en la que

$x_i$  = cada uno de los valores de la variable,

$\sum$  = suma de esos valores, y

$n$  = número total de casos que contempla la variable medida.

Cuando los datos se encuentran agrupados, es decir que, para cada valor de la variable considerada, hay distintas frecuencias, la fórmula es  $\bar{x} = (\sum (x_i * f))/n$ ; y, en los casos en que las observaciones están agrupadas en intervalos (por ejemplo, intervalos de edades), la fórmula es  $\bar{x} = (\sum (p_m * f))/n$ , en la que  $p_m$  es el punto medio del valor del intervalo. La media aritmética puede calcularse únicamente en los niveles de intervalo o de razón.

*Mediana*: es la categoría que divide a los datos en dos partes iguales y se simboliza  $ME$ ; es decir, es el valor de la distribución que deja a ambos lados el mismo número de observaciones. Para calcularla, se ordenan los datos siguiendo un orden creciente (de menor a mayor) y, si el número de observaciones es impar, la mediana corresponde al valor situado en el medio; si el número de observaciones es par, la mediana corresponde al promedio de los dos valores centrales. No es posible utilizar esta medida en variables que solo admiten un sistema clasificatorio, como las nominales.

Cuando los datos están agrupados, el procedimiento es más complejo y comprende varios pasos:<sup>2</sup>

1. Armar las frecuencias absolutas acumuladas.
2. Calcular  $(n+1)/2$  = posición de la mediana.
3. Buscar en qué frecuencia absoluta acumulada está contenida esa posición de la mediana.
4. Identificar la categoría de la variable a la que corresponde la frecuencia absoluta acumulada en que se ubica la posición de la mediana.

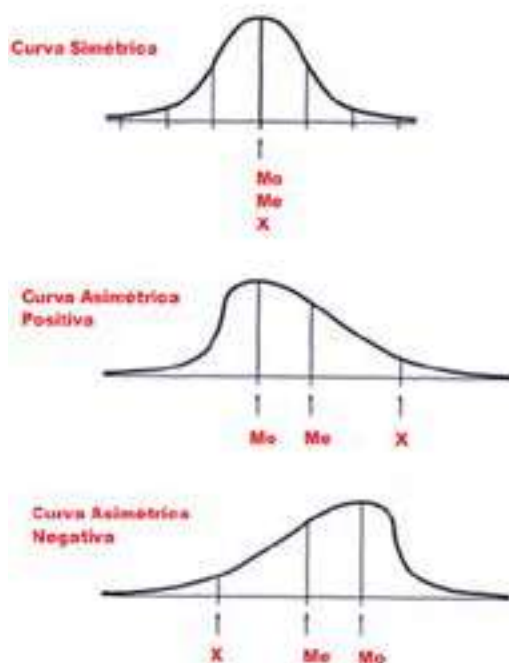
<sup>2</sup> La aplicación del procedimiento para calcular la mediana, en una variable con datos agrupados, se detalla en los ejemplos que se desarrollan a continuación de la presentación de las medidas de tendencia central.

En los casos en los que la media, la mediana y la moda de una variable coinciden entre sí y la distribución de los valores tiene forma de campana, decimos que estamos frente a una *distribución simétrica o normal* (recibe este nombre, ya que muchas variables en la vida real suelen mostrar una curva de este tipo, como, por ejemplo, la talla de las personas de una misma edad).

Otras variables pueden presentar distribuciones sesgadas, ya que muchos de sus elementos presentan valores sumamente altos o bajos. Las *asimétricas positivas*, o *sesgadas positivamente*, son aquellas distribuciones en las que la mediana y la moda se ubiquen por debajo de la media (o hacia la izquierda de esta), ya que muchas de sus puntuaciones son muy bajas. En los países latinoamericanos, las variables económicas, como los ingresos familiares, frecuentemente muestran distribuciones asimétricas positivas, debido a que gran parte de la población suele tener ingresos desfavorables y solo una pequeña porción, ingresos elevados.

La *asimetría es de tipo negativa o sesgada negativamente*, en cambio, en los casos en que la mayor cantidad de datos de la distribución presentan valores que se aglomeran por encima de la media (o hacia la derecha de esta), debido a que son significativamente altos.

Figura 2.5. Diferentes curvas de distribuciones de frecuencias



Fuente: elaboración propia.

Asimismo, cuando trabajamos con muestras muy homogéneas, las *distribuciones* son más bien apuntadas y reciben el nombre de *leptocúrticas*, muestran mucha concentración de los datos alrededor de la media. Es decir, si en una muestra las familias estudiadas tienen salarios muy similares (no hay ni muy ricos ni muy pobres), todos los habitantes tendrán rentas parecidas y próximas al valor de la media (García Ferrando, 2006).

Figura 2.6. Curva de distribución de frecuencias poco heterogénea



Fuente: elaboración propia.

Las curvas, entonces, suelen ser muy útiles para visualizar rápidamente cómo se reparten los distintos valores de una variable en una población o en una muestra, ya que representan la frecuencia absoluta o relativa con que se observan los distintos valores de una determinada variable. Nos facilitan información rápida sobre la tendencia central de los datos (¿dónde se sitúa la mayoría?) y la dispersión (¿están concentrados o hay muchas diferencias?).

Veamos ahora en el cuadro 2.2 cómo se aplica la fórmula para calcular la media ( $\bar{x}$ ) de una distribución. Supongamos que estamos interesados en conocer la edad promedio de un grupo de madres adolescentes que se atienden en el Centro de Salud Piaggi de San Fernando, cuyas edades son 13, 14, 15, 16 y 18 años, respectivamente:

Cuadro 2.2. Aplicación de la fórmula de la media ( $\bar{x}$ )

| Categorías de la variable<br>“edad de las madres<br>adolescentes” ( $x_i$ ) | $(\sum x_i)$ | <b>n</b> | $(\sum x_i)/n$ |
|---|--------------|----------|----------------|
| 13  | 76           | 5        | 15,2           |
| 14  |              |          |                |
| 15  |              |          |                |
| 16  |              |          |                |
| 18  |              |          |                |

$\bar{X}=15,2$

Fuente: elaboración propia para el ejemplo.

Y si queremos calcular la mediana en este caso, el procedimiento es muy sencillo, ya que sabemos que, una vez ordenadas de mayor a menor las categorías, la mediana corresponde al valor central que en este caso es 5:  $ME=15$ .

Nuestro ejemplo se complejiza, ya que nos interesa también medir la media de la “edad de las madres adolescentes” de los cuatro centros de salud que se encuentran próximos a la Escuela Secundaria Básica N° 4 de San Fernando. Se contabilizan un total de 30 casos, con distintas frecuencias para cada categoría o valor que asume esta variable; veamos cómo realizar el cálculo en esta situación en la que los datos de la variable se encuentran agrupados:

Cuadro 2.3. Aplicación de la fórmula de la media para datos agrupados

| Categorías de la variable<br>“edad de las madres<br>adolescentes” ( $x_i$ ) | Frecuencias (f) | $x_i \cdot f$                  | $(\sum (x_i \cdot f))/n$ |
|---|-----------------|--------------------------------|--------------------------|
| 13  | 1               | 13                             | 16,2                     |
| 14  | 3               | 42                             |                          |
| 15  | 4               | 60                             |                          |
| 16  | 9               | 144                            |                          |
| 17  | 7               | 119                            |                          |
| 18  | 6               | 108                            |                          |
|   | <b>N = 30</b>   | <b><math>\sum = 486</math></b> |                          |

$M_o= 16$

$\bar{X}= 16,2$

Fuente: elaboración propia para el ejemplo.

Además de conocer la edad modal (16) y la edad promedio (16,2) de estas 30 adolescentes que están transitando un embarazo, nos interesa estimar cuál es la edad



que divide al grupo de madres en dos partes iguales para saber cuál es tramo etario en que se encuentran las más jóvenes. Con tal propósito calcularemos la mediana.

Cuadro 2.4. Aplicación de la fórmula de la mediana para datos agrupados

| Categorías de la variable “edad de las madres adolescentes” (X <sub>i</sub> ) | Frecuencias (f) | Frecuencias absolutas acumuladas | n+1/2 (posición de la ME) |
|---|-----------------|----------------------------------|---------------------------|
| 13  | 1               | 1                                | 15,5                      |
| 14  | 3               | 4                                |                           |
| 15  | 4               | 8                                |                           |
| 16  | 9               | 17                               |                           |
| 17  | 7               | 24                               |                           |
| 18  | 6               | 30                               |                           |
|   | n= 30           |                                  |                           |

1. Buscar este valor en las frecuencias absolutas acumuladas.
2. Identificar la categoría de la variable que corresponde a la ME.
3. La edad que divide al grupo en dos partes iguales es 16.

Fuente: elaboración propia para el ejemplo.

Podemos concluir que las madres más jóvenes se encuentran en el grupo etario que va desde los 14 hasta los 16 años, ya que el valor que divide al grupo en dos partes es 16.

*A tener en cuenta:*

La media es una medida muy útil en variables de intervalo o razón, en la que influyen de modo proporcional a su valor y a su frecuencia todas y cada una de las categorías que asume una variable; por eso, su cálculo siempre se encuentra distorsionado por un sesgo en la curva de distribución, ya que se encuentra condicionado por los valores extremos. La mediana, en cambio, es sensible al tamaño de la muestra y en ella las puntuaciones no influyen por su valor, sino por el lugar que ocupan en la distribución (Ritchey, 2001).

### Otras medidas de posición

Existen otras medidas de posición que proporcionan información acerca de las características de un conjunto de datos, pero que no son de tendencia central, como *los cuartiles, los deciles y los percentiles*; mientras que los primeros dividen a la distribución en 4 partes, los segundos lo hacen en 10, y los terceros, en 100. Dentro de este grupo, los cuartiles (Q<sub>1</sub>, Q<sub>2</sub> y Q<sub>3</sub>) son las medidas que se utilizan más frecuentemente y refieren a los tres valores de la variable que dividen al conjunto

de los datos en cuatro partes iguales. Para calcularlos, el procedimiento es muy similar al de la mediana: se ordenan los datos siguiendo un orden creciente y, si el número de estos es impar,  $Q_1$  corresponde al valor que ocupa la posición central de las observaciones que se encuentran por debajo de la mediana (si el número es par, corresponde al promedio de los dos valores centrales); contrariamente, para calcular  $Q_3$ , se selecciona el valor que se encuentra justo en el medio de las observaciones ubicadas por encima de la mediana.  $Q_2$  coincide siempre con el valor de la mediana.<sup>3</sup>

Cuando los datos están agrupados, comprende los siguientes pasos:

1. Armar las frecuencias absolutas acumuladas.
2. Comenzar calculando  $Q_1$ . La fórmula sugerida es  $(n+1/4)$  = posición de la  $Q_1$ .
3. Buscar en qué frecuencia absoluta acumulada está contenida esa posición de  $Q_1$ .
4. Identificar la categoría de la variable a la que corresponde la frecuencia absoluta acumulada en que se ubica la posición de la  $Q_1$ .
5. El cálculo de  $Q_2$  y  $Q_3$  requiere el mismo procedimiento, pero las fórmulas se reemplazan por  $2(n+1/4)$  y  $3(n+1/4)$ .

## Medidas de variabilidad

Hemos visto que las medidas de posición y tendencia central nos permiten ubicar la posición y el centro de una distribución, respectivamente. Sin embargo, no son suficientes para describir de manera completa un grupo de datos, ya que no dan cuenta de la variabilidad que podamos encontrar en ese conjunto. Las medidas de dispersión (o de variabilidad) informan cómo están repartidos los valores de una variable alrededor del promedio o media de la distribución, pero pueden calcularse únicamente en variables que presenten niveles de intervalo o de razón; entre estas medidas se encuentran la amplitud o rango, el desvío estándar y el coeficiente de variación.

*Amplitud o rango (R)*: es la distancia entre el valor máximo y mínimo de la distribución. Se calcula como una simple resta,  $R = XI \text{ máximo} - XI \text{ mínimo}$ . Aunque su valor es de interés, cabe resaltar que solo considera las dos observaciones extremas, sin tener en cuenta los restantes valores de la distribución. En muchas ocasiones, las observaciones no están uniformemente distribuidas entre el máximo y el mínimo, por lo que es necesario recurrir a otra medida para poder dar cuenta de una variabilidad promedio que muestran los datos.

<sup>3</sup> El cuarto cuartil ( $Q_4$ ) no se calcula porque corresponde al total de  $n$ .

*Desviación estándar (SD)*: representa la distancia media existente entre los valores de las observaciones y la media de la variable. Cuanto mayor es la desviación estándar, más lejos están las observaciones de la media; cuanto menor es, contrariamente, significa que más próximas están las observaciones de la media. Se mide en las mismas unidades que la variable, por lo que es una de las medidas de variabilidad más utilizada. Siguiendo a Llorca *et al.* (2011), la fórmula para calcular la desviación estándar en una muestra<sup>4</sup> es  $SD = \sqrt{[\sum(x_i - \bar{X})^2 / n - 1]}$ , en la que

$\sum$ = sumatoria,

$x_i$ = cada uno de los valores de la variable,

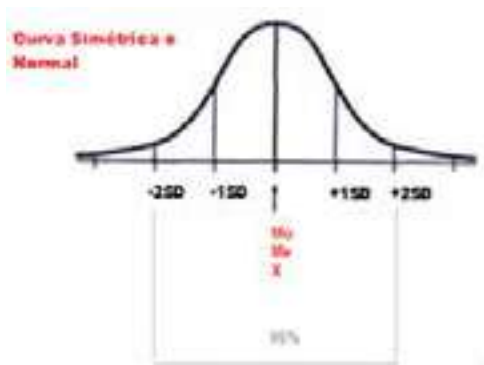
$\bar{X}$ = media, y

$n$ = número total de casos que contempla la variable medida.

Cuando los datos se encuentran agrupados, es decir que los valores de la variable asumen distintas frecuencias, la fórmula es  $SD = \sqrt{[\sum(x_i - \bar{X})^2 * f] / n - 1}$ ; y, en los casos en que las observaciones están agrupadas en intervalos (por ejemplo, intervalos de edades), la fórmula es  $SD = \sqrt{[\sum(pm - \bar{X})^2 * f] / n - 1}$ , en la que  $pm$  es el punto medio del valor del intervalo.

Cabe señalar que, en distribuciones simétricas o normales, se observa que, aproximadamente, el 68,26% de los datos de la variable se encuentra en un intervalo que va de 1 desvío estándar por debajo de la media a 1 desvío estándar por encima de la media ( $\bar{x} - 1SD$ ,  $\bar{x} + 1SD$ ). Asimismo, aproximadamente, el 95,44% de las observaciones se encuentra en el intervalo que va de 2 desvíos estándar por debajo de la media a 2 desvíos estándar por encima de la media ( $\bar{x} - 2SD$ ,  $\bar{x} + 2SD$ ). Esta premisa fundamenta la utilización de la curva normal como modelo de probabilidad en la estadística inferencial (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000).

Figura 2.7. Curva de distribución simétrica o normal



Fuente: elaboración propia.

<sup>4</sup> Cuando queremos calcular el desvío estándar en una población, la fórmula es  $SD = \sqrt{[\sum(x_i - \bar{x})^2 / n]}$ .

*Coefficiente de variación (cv)*: En algunas ocasiones, como cuando comparamos variables que no están expresadas en las mismas unidades, resulta necesario medir la dispersión en forma de porcentaje y se calcula un estadístico derivado de desviación estándar que se llama coeficiente de variación. Este coeficiente se obtiene como resultado de dividir el desvío estándar de la distribución por la media y se expresa como un porcentaje ( $cv = SD/x * 100$ ). Cabe señalar que la unidad original de medida es irrelevante para este coeficiente, ya que ella figura tanto en el numerador como en el denominador de la fracción.

Así, si entre un grupo de bebés el desvío estándar en la medición de su talla es 7 cm y la correspondiente media es 55 cm, el coeficiente de variación es el 12,73%. La medición del perímetro cefálico entre estos mismos bebés arroja un desvío estándar de 7 cm y una media de 39 cm, lo que da como resultado un coeficiente de variación, en este caso, de 17,95%. Los resultados obtenidos en torno al coeficiente de variación muestran que, aunque ambas mediciones tienen el mismo desvío estándar, los bebés presentan menor variabilidad respecto a la talla que al perímetro cefálico.

Retomemos el ejemplo de la medición de la “edad de las madres adolescentes” de los cuatro centros de salud que se encuentran próximos a la Escuela Secundaria Básica N° 4 del municipio de San Fernando.

Ahora, además de conocer el centro de la distribución ( $x = 16,2$ ), nos interesa saber cuán heterogéneo es este grupo de madres respecto a la edad en la que transitan su embarazo. Con este propósito, debemos recurrir a las medidas de dispersión, ya que nos permitirán saber cómo se encuentran repartidos los distintos valores que presenta esta variable respecto al promedio o media de la distribución. Los cuadros 2.5, 2.6 y 2.7 ilustran los cálculos de la amplitud o rango de la distribución, del desvío estándar respecto a la media y del coeficiente de variación, respectivamente.

Cuadro 2.5. Aplicación del cálculo de amplitud o rango (R)

| Categorías de la variable<br>“edad de las madres<br>adolescentes” (xi) | xi máximo - xi mínimo |
|--|-----------------------|
| 13   | 18 - 13               |
| 14   |                       |
| 15   |                       |
| 16   |                       |
| 17   |                       |
| 18   |                       |

R= 5

Fuente: elaboración propia para el ejemplo.

Cuadro 2.6. Aplicación de la fórmula del desvío estándar para datos agrupados

| Categorías de la variable "edad de las madres adolescentes" (xi) | Frecuencias (f) | xi - x | (xi - x) <sup>2</sup> | (xi - x) <sup>2</sup> * f | $\sqrt{ [ \sum (xi - x)^2 * f ] / n - 1 }$ |
|--|-----------------|--------|-----------------------|---------------------------|--|
| 13   | 1               | -3,2   | 10,24                 | 10,24                     | 1,37                                       |
| 14   | 3               | -2,2   | 4,84                  | 14,52                     |  |
| 15   | 4               | -1,2   | 1,44                  | 5,76                      |  |
| 16   | 9               | -0,2   | 0,04                  | 0,36                      |  |
| 17   | 7               | 0,8    | 0,64                  | 4,48                      |  |
| 18   | 6               | 1,8    | 3,24                  | 19,44                     |  |
|  | n= 30           |        |                       | $\Sigma = 54,8$           |  |

X= 16,2; sd=1,37

Fuente: elaboración propia para el ejemplo.

Cuadro 2.7. Aplicación del cálculo de coeficiente de variación (cv)

| sd/x      | (sd/x)*100 |
|-----------|------------|
| 1,37/16,2 | 8,46       |

CV= 8,46%

Fuente: elaboración propia para el ejemplo.

Podemos interpretar estos resultados y decir que la edad de embarazo entre las adolescentes que han sido medidas en estos 4 centros de salud muestra un rango de 5 años. Además, si bien la edad media de estas jóvenes que están embarazadas es de 16,2, encontramos un desvío estándar –en relación con esa edad promedio– de 1,37 años. Finalmente, el coeficiente de variación al interior del grupo respecto a la edad de embarazo es inferior al 9%.

*A tener en cuenta:*

Las medidas de dispersión serán mayores, cuanto más disgregados o esparcidos estén los datos en una distribución de frecuencias. Es decir, decimos que, cuanto más pequeñas sean las medidas de dispersión, más homogénea es esa distribución, ya que encontramos menor variación entre los valores; y, a la inversa, cuanto más grandes sean las medidas de dispersión, más heterogénea es la distribución de frecuencias de la variable considerada (García Ferrando, 2006).

Asimismo, mientras que las variables cualitativas se suelen describir a través de la utilización de porcentajes, las medidas de tendencia central (como la media y la

mediana) y las de dispersión (como el rango, el desvío estándar o el coeficiente de variación) son, frecuentemente, utilizadas para describir distribuciones de frecuencias de variables cuantitativas (Llorca *et al.*, 2011).

## Bibliografía

- Dirección de Estadística e Información en Salud, Proyecto Funciones Esenciales y Programas Priorizados de Salud Pública e Instituto de Investigaciones Gino Germani (2012). *Encuesta de utilización y gastos en servicios de salud. Argentina-2010. Primeros resultados*, serie 10, n° 21. Sistema Estadístico de Salud, Dirección de Estadística e Información en Salud (DEIS) del Ministerio de Salud de la Nación.
- García Ferrando, Manuel (2006). “Características de una distribución de frecuencias: tendencia central; dispersión y forma. La distribución normal”. En *Socioestadística. Introducción a la estadística en sociología*, pp. 85-118. Madrid: Alianza.
- Hernández Sampieri, Roberto; Fernández Collado, Carlos y Baptista Lucio, Pilar (2000). “Análisis de datos”. En *Metodología de la investigación*, pp. 341-423. México, D. F.: McGraw-Hill Interamericana.
- Llorca Díaz, Francisco Javier; Dierssen Sotos, Trinidad y Gómez Acebo, Inés (2011). “Introducción a los métodos estadísticos. Estadística Descriptiva. Probabilidad”. En Hernández-Aguado, Ildefonso; Gil de Miguel, Ángel; Delgado Rodríguez, Miguel; Bolúmar Montrull, Francisco; Benavides, Fernando; Porta Serra, Miguel; Álvarez-Dardet Díaz, Carlos; Vioque López, Jesús (dirs.) y Lumbreras Lacarra, Blanca (coord.), *Manual de epidemiología y salud pública: para grados en ciencias de la salud*, pp. 25-30. Argentina: Médica Panamericana.
- Ministerio de Salud y Organización Panamericana de la Salud/Organización Mundial de la Salud (2016). *Indicadores Básicos. Argentina 2016*. Argentina: Ministerio de Salud y la Organización Panamericana de la Salud.
- Ritchey, Ferris (2001). “Estimación de promedios”. En *Estadística para las ciencias sociales. El potencial de la imaginación estadística*, pp. 99-119. México D. F.: McGraw-Hill Interamericana.

### Páginas webs consultadas:

- Dirección de Información Sistematizada, Subsecretaría de Gestión y Contralor del Conocimiento y Redes Sanitarias, Ministerio de Salud de la Provincia de Buenos Aires:  
<http://www.ms.gba.gov.ar/estadisticas>.
- Dirección de Estadística e Información de Salud (DEIS) del Ministerio de Salud de la Nación:  
<http://deis.msal.gov.ar>.









## Módulo 3

# El análisis de más de una variable y la búsqueda de relaciones

Además de describir una tendencia o encontrar los patrones que muestran los datos respecto a una variable de interés (como vimos en el segundo módulo), frecuentemente, los investigadores estamos preocupados por explicar las variaciones que se producen en un determinado fenómeno, en función de ciertos factores o elementos causales. Por eso, solemos trabajar, en muchas oportunidades, con más de una variable simultáneamente y buscar posibles relaciones o asociaciones entre esas variables bajo estudio.

La variación que se trata de explicar recibe el nombre *variable dependiente*, ya que se supone que los valores que asumen sus categorías dependen de los valores que presentan las categorías de otras variables; estas últimas se llaman, contrariamente, *variables independientes*. La selección de las variables independientes está condicionada por la o las hipótesis de trabajo que orientan la investigación.

### **¿Cómo organizamos la información cuando queremos analizar dos o más variables de una población o una muestra?**

Cuando trabajamos con un conjunto de variables, debemos volcar la información obtenida en lo que llamamos *matriz de datos*. Por convención, en una matriz, cada fila corresponde a una unidad de análisis, y cada columna, a una variable; las celdas representan los valores o categorías que asume cada una de esas unidades de análisis para las distintas variables.

Para el procesamiento de los datos, necesitamos que todas las categorías de una matriz sean numéricas, por eso a las categorías de las variables cualitativas se les asigna siempre un código numérico. En el caso de la variable “lugar de residencia”, por ejemplo, los códigos podrían ser los siguientes: 1 Ciudad Autónoma de Buenos Aires; 2 Gran Buenos Aires.

Como puede apreciarse en el cuadro 3.1, la matriz de datos tiene el formato de una planilla de cálculos como la del programa Excel:

Cuadro 3.1. Ejemplo matriz de datos

| Unidades de análisis | Variables     |   |  |                                      |
|----------------------|---------------|---|--|--------------------------------------|
|                      | N° de médicos | N° de camas en establecimientos (todos los subsectores) | N° de establecimientos asistenciales (todos los subsectores) | Edad de esperanza de vida en mujeres |
| Buenos Aires         | 36.931        | 51.517  | 4.896  | 78,69                                |
| Catamarca            | 669           | 1.470   | 380  | 78,85                                |
| Córdoba              | 13.532        | 17.718  | 1.763  | 79,23                                |
| Corrientes           | 2.215         | 3.525   | 387  | 77,54                                |
| Chaco                | 1.732         | 3.537   | 519  | 76,41                                |
| Chubut               | 959           | 2.148   | 248  | 79,96                                |
| Entre Ríos           | 2.383         | 6.751   | 604  | 78,98                                |
| Formosa              | 719           | 1.680   | 271  | 76,98                                |
| Jujuy                | 1.145         | 2.823   | 394  | 78,08                                |
| La Pampa             | 695           | 1.179   | 206  | 79,51                                |
| La Rioja             | 775           | 1.260   | 281  | 78,57                                |
| Mendoza              | 4.478         | 4.142   | 676  | 79,36                                |
| Misiones             | 1.446         | 3.125   | 617  | 77,69                                |
| Neuquén              | 1.308         | 1.660   | 248  | 80,75                                |
| Río Negro            | 1.432         | 1.930   | 315  | 79,36                                |
| Salta                | 2.159         | 3.805   | 598  | 78,22                                |
| San Juan             | 1.551         | 1.742   | 301  | 77,95                                |
| San Luis             | 929           | 1.218   | 236  | 79,33                                |
| Santa Cruz           | 493           | 1.041   | 105  | 79,43                                |
| Santa Fe             | 9.469         | 11.011  | 1.869  | 79,13                                |
| Santiago del Estero  | 1.267         | 3.033   | 526  | 77,84                                |
| Tierra del Fuego     | 336           | 286   | 37   | 79,92                                |
| Tucumán              | 3.757         | 4.520   | 702  | 78,20                                |

Fuente: Elaboración propia sobre la base de los datos de los Indicadores Básicos Argentina, Ministerio de Salud de la Nación - OPS/OMS, 2016.

### La tabulación simultánea de variables

Las *tablas de contingencia* permiten representar el resultado del cruce de dos o más variables, para analizar las relaciones existentes entre variables dependientes e independientes; una tabulación simultánea comprende así un cierto número de celdas en las que se vuelcan las frecuencias obtenidas para cada combinación de categorías de las variables consideradas. En otros términos, cuando armamos una tabla de contingencia, partimos siempre de la hipótesis –más o menos explícita– sobre el efecto (o no) de una variable sobre otra. Hay variables que debido a su naturaleza

solo pueden funcionar como variables independientes (como el caso del sexo o la edad de las personas), pero hay muchas otras situaciones en las que los objetivos de investigación son los que definen estos roles entre las variables.

En este módulo, nos ocuparemos solamente de las *tablas bivariadas*. Este tipo de tabla es muy usual en la investigación en ciencias sociales, que se caracteriza por una utilización predominante de variables con niveles de medición nominal y ordinal (Baranger, 1992).<sup>1</sup>

Cabe destacar que –por convención–, en estas tablas, la variable dependiente suele ubicarse en las hileras, y la independiente, en las columnas. Asimismo, los totales de las columnas y las filas reciben el nombre de *frecuencias marginales*, las celdas al interior de la tabla se llaman *frecuencias condicionales*. Los requisitos que deben cubrir las tablas de contingencia son los siguientes:

- El título no solo debe reflejar la información que contiene la tabla, sino que debe expresar la relación entre variables postulada en la hipótesis de trabajo del investigador.
- Incluir un subtítulo para cada columna y subcolumna que se integre a la tabla.
- Indicar al final de cada columna el número total de casos que comprende; de manera optativa podemos sumar los totales en las hileras.

Cuadro 3.2. Periodicidad en las consultas obstétricas, según presencia de necesidades básicas insatisfechas (NBI)

| Periodicidad de las consultas obstétricas | Presencia de NBI |     |         |
|---|------------------|-----|---------|
|   | Sí               | No  | Totales |
| Baja                                      | 40               | 20  | 60      |
| Media                                     | 30               | 55  | 85      |
| Alta                                      | 15               | 35  | 50      |
| Totales                                   | 85               | 110 | 195     |

Fuente: elaboración propia para el ejemplo.

Para poder interpretar los valores obtenidos en una tabla bivariada, debemos comenzar calculando los porcentajes que corresponden a cada frecuencia. Recordemos que estos permiten apreciar la importancia relativa de las frecuencias observadas con relación a los totales que presenta cada variable. En el caso de las tablas de

<sup>1</sup> Cabe resaltar que también se pueden utilizar tablas de contingencia con variables definidas en niveles intervalares o de razón, pero, para estos niveles, la estadística cuenta con otras técnicas más precisas para medir la relación entre variables, como el *coeficiente de correlación  $r$  de Pearson* y la *regresión lineal* (Ritchev, 2001).



contingencia, los porcentajes tienen que ser estimados en el sentido de la variable independiente, ya que el foco del análisis está puesto en ver cómo esta incide (o no) en la distribución de frecuencias de la variable dependiente (García Ferrando, 2006).

Una vez computados estos porcentajes, estamos en condiciones de hacer interpretaciones comparando los porcentajes observados para cada variable. Esta comparación se hace en sentido inverso a como fueron originalmente estimados. Es decir, los porcentajes siempre se calculan en la dirección de la variable independiente (nuestro factor causal), pero la comparación se hace en sentido de la variable dependiente; este procedimiento se conoce como la *1ª regla de Zeisel*. Además, podemos analizar si hay asociación entre las variables de la tabla (en términos de comprobar la existencia de una relación). Un método muy sencillo es hacerlo a través de la *diferencia porcentual*, que se calcula haciendo una simple resta en el mismo sentido en que hemos comparado anteriormente los porcentajes (Baranger, 1992).<sup>2</sup>

Supongamos que nuestra hipótesis de trabajo es que “la presencia de necesidades básicas insatisfechas (NBI) en las mujeres condiciona la periodicidad con la que estas realizan sus consultas obstétricas”. Para comprobar esta hipótesis, debemos computar los porcentajes en la dirección de nuestra variable independiente (presencia de NBI), para luego hacer comparaciones o calcular las diferencias porcentuales en sentido contrario.

Cuadro 3.3. Periodicidad en las consultas obstétricas, según presencia de NBI (%)

| Periodicidad de las consultas obstétricas | Presencia de NBI |              | Diferencia porcentual |
|---|------------------|--------------|-----------------------|
|   | Sí               | No           |                       |
| Baja                                      | 47,1             | 18,2         | 28,9                  |
| Media                                     | 35,3             | 50           | -14,7                 |
| Alta                                      | 17,6             | 31,8         | -14,2                 |
| Totales                                   | 100<br>(85)      | 110<br>(110) |                       |

Fuente: elaboración propia para el ejemplo.

La primera lectura que podemos hacer es que, entre las mujeres que presentan NBI, casi un 50% de las consultas se realizan con una periodicidad baja, mientras que,

<sup>2</sup> Para saber si la asociación que hallamos entre las dos variables de nuestra muestra es estadísticamente significativa o tiene relevancia estadística, podemos recurrir al cálculo de un estadístico llamado cuadrado (en el caso que nuestras variables presenten un nivel de medición nominal). Es decir, para afirmar que la asociación encontrada entre las variables estudiadas representa una asociación realmente existente en el universo o población de la que tomamos nuestra muestra, necesitamos recurrir a otro tipo de estadísticos, como el chi cuadrada (Ritchey, 2001).

Cuando trabajamos con variables con niveles de medición de intervalo o razón, podemos utilizar estadísticos, como el *coeficiente de correlación r de Pearson* y la *regresión lineal*, para analizar la existencia de una relación estadística entre las *variables bajo estudio* (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000).

entre las mujeres sin NBI, ese porcentaje es inferior al 20% de los casos. Asimismo, el porcentaje de mujeres que no vive en condiciones de pobreza estructural y que realiza consultas con una periodicidad alta duplica prácticamente al que se observa en las mujeres que viven con NBI.

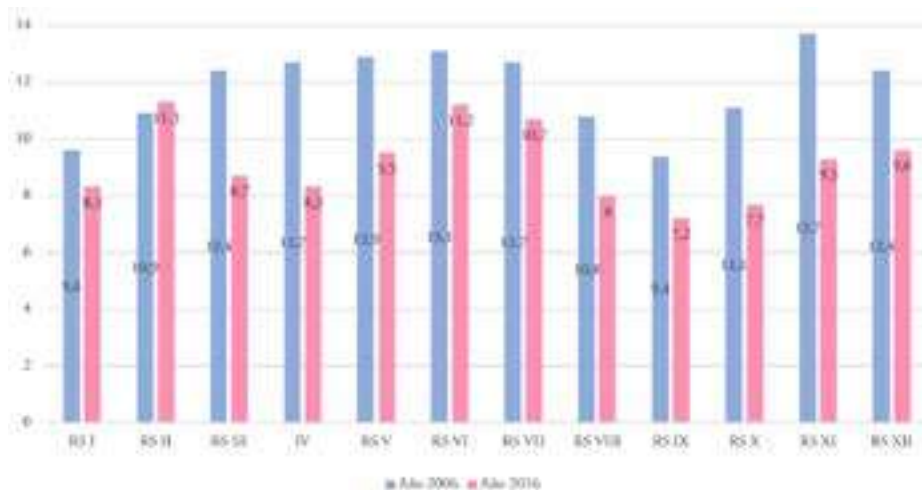
Además, la diferencia porcentual permite observar que la periodicidad clasificada como “baja” en las consultas obstétricas es casi un 30% más alta (28,9%) entre las mujeres que presentan NBI, a la vez que las periodicidades media y alta son un 14% inferior entre estas mujeres.

### ¿Cómo podemos graficar una tabulación conjunta de variables?

Al igual que como sucede con las distribuciones de frecuencias, existen distintas formas de exhibir gráficamente las tablas cuando hay una clasificación cruzada que presentan la relación entre 2 variables.

*Gráficos de barras agrupadas:* permiten desplegar la distribución conjunta de dos variables cualitativas. También pueden utilizarse para graficar variables cuantitativas discretas o continuas, cuando el número de valores (o intervalos de clases) considerado no es demasiado grande.

Figura 3.1. Gráfico de distribución de la tasa de mortalidad infantil en las regiones sanitarias de la provincia de Buenos Aires según período (2006 y 2016)

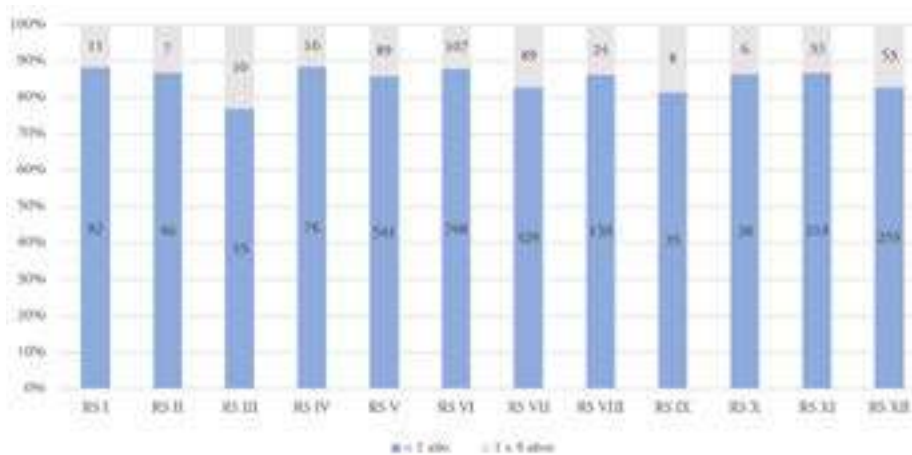


Fuente: Elaboración propia sobre la base de los datos de la Dirección de Información Sistemática de la Subsecretaría de Gestión y Contralor del Conocimiento, Redes y Tecnologías Sanitarias, Ministerio de Salud de la Provincia de Buenos Aires, 2016.

*Gráficos de barras proporcionales:* cada barra representa el 100% de una categoría de la variable independiente y se divide –proporcionalmente– en el número casos que

registra cada una de las categorías de la variable dependiente para esa categoría de la variable independiente. Resultan particularmente apropiados cuando el propósito es mostrar que una de las variables es un posible factor de riesgo para la otra. Como el gráfico de barras agrupadas, posibilita mostrar la relación entre variables cualitativas o cuantitativas (cuando el número de valores o intervalos de clases considerado es pequeño). En lugar de barras proporcionales, también podemos utilizar *barras segmentadas*, en las que cada barra representa la frecuencia absoluta para cada categoría de la variable independiente y, consecuentemente, se divide en la cantidad de casos que exhibe cada categoría de la variable dependiente.

Figura 3.2. Gráfico de distribución proporcional de las defunciones infantiles en las regiones sanitarias (rs) de la provincia de Buenos Aires según edad, año 2016



Fuente: Elaboración propia con base en los datos de la Dirección de Información Sistematizada de la Subsecretaría de Gestión y Contralor del Conocimiento, Redes y Tecnologías Sanitarias, Ministerio de Salud de la Provincia de Buenos Aires, 2016.

## Bibliografía

- Baranger, Denis (1992). “La matriz de datos”. En *Construcción y análisis de datos. Una introducción al uso de las técnicas cuantitativas en la investigación social*, pp. 17-42. Posadas: Editorial Universitaria.
- García Ferrando, Manuel (2006). “Estadística Descriptiva Bivariable: características de una asociación bivariable”. En *Socioestadística. Introducción a la estadística en sociología*, pp. 215-232. Madrid: Alianza.

- Hernández Sampieri, Roberto; Fernández Collado, Carlos y Baptista Lucio, Pilar (2000). “Análisis de datos”. En *Metodología de la investigación*, pp. 341-423. México, D. F.: McGraw-Hill Interamericana.
- Ministerio de Salud y Organización Panamericana de la Salud/Organización Mundial de la Salud (2016). *Indicadores Básicos. Argentina 2016*. Argentina: Ministerio de Salud y la Organización Panamericana de la Salud.
- Ritchey, Ferris (2001). “Correlación y regresión bivariadas. Parte 1: conceptos y cálculos”. En *Estadística para las ciencias sociales. El potencial de la imaginación estadística*, pp. 509-551. México D. F.: McGraw-Hill Interamericana.

Páginas web consultadas:

- Dirección de Información Sistematizada, Subsecretaría de Gestión y Contralor del Conocimiento y Redes Sanitarias, Ministerio de Salud de la Provincia de Buenos Aires:  
<http://www.ms.gba.gov.ar/estadisticas>.

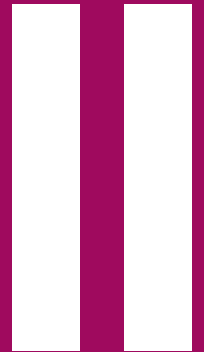




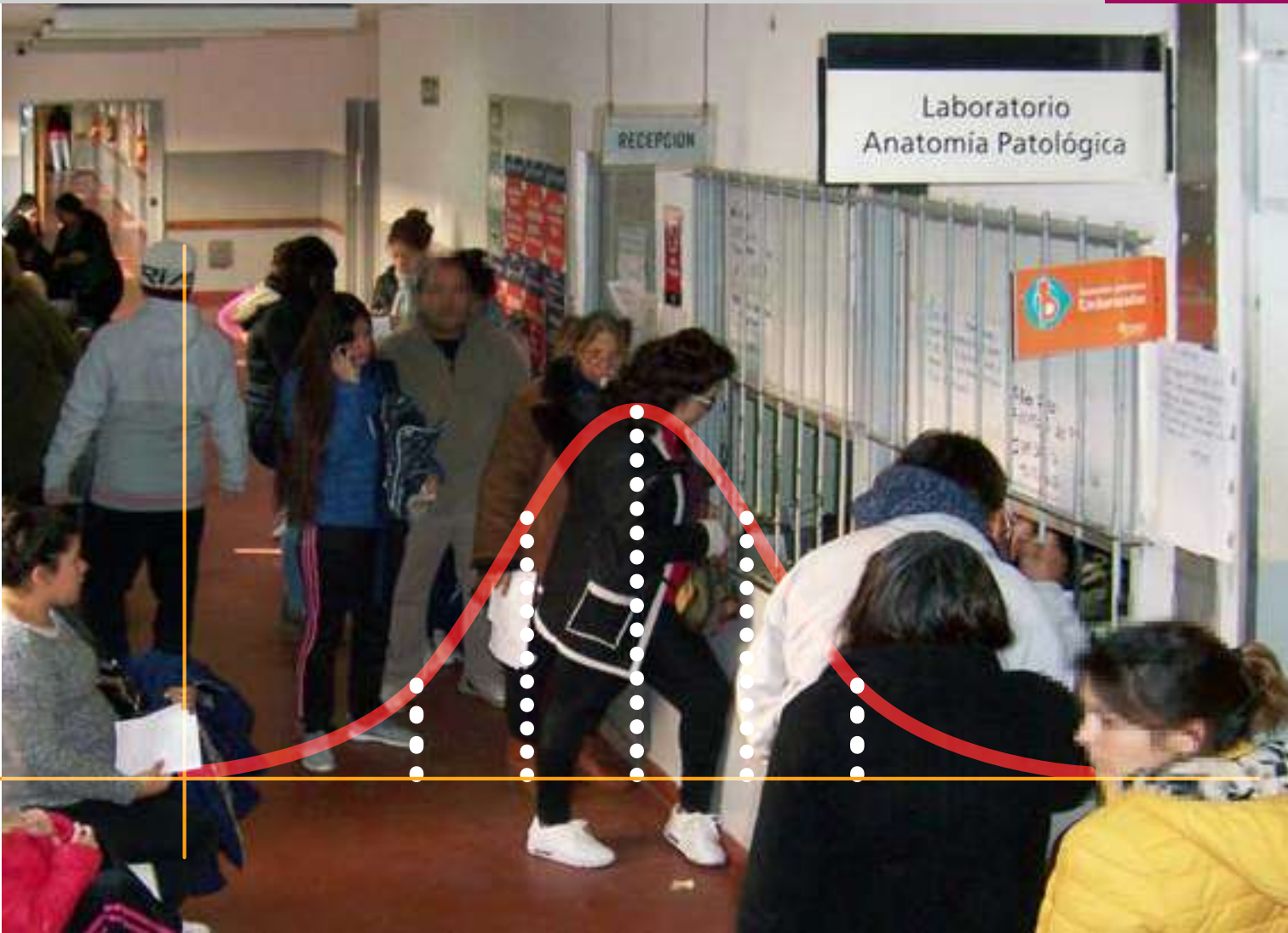


## **PARTE II**

Aportes de la  
estadística al campo  
de la salud







## Módulo 4

# Las estadísticas de salud y los indicadores de uso frecuente

Los conceptos básicos, los métodos y los procedimientos que hemos trabajado en la primera parte de este cuaderno son utilizados con bastante regularidad dentro del ámbito de la salud. Ya sea para realizar investigaciones específicas sobre la salud y enfermedad de conjuntos poblacionales, o para planificar y hacer intervenciones en salud pública –inclusive, para desarrollar la práctica profesional más cotidiana–, empleamos corrientemente técnicas y herramientas tanto de la estadística descriptiva como de la inferencial: utilizamos tasas para comparar el riesgo de morir o enfermar en dos poblaciones diferentes; buscamos los patrones que muestran determinadas enfermedades en una comunidad construyendo distribuciones de frecuencias y calculando las tendencias centrales o las desviaciones típicas; exploramos asociaciones entre factores de riesgo y ciertas patologías para programar actividades de prevención; estimamos valores o tendencias para distintas variables en la población, a partir de lo observado en un subgrupo de esta, etcétera.

Una de las contribuciones más importantes de la estadística en el campo de la salud refiere, no obstante, a la elaboración de sistemas de registros continuos que llamamos *estadísticas de salud* (Torres, 2009). La información procesada y volcada en estos registros permite no solo hacer diagnósticos, sino que es relevante para formular, implementar, evaluar o comparar acciones, programas o políticas de salud. Estos sistemas de estadísticas pueden clasificarse de la siguiente manera.

1. *Estadísticas vitales primarias*: proporcionan información referente a la dinámica de la población y son muy utilizadas para el cálculo de indicadores de gran importancia para el sector salud. Están basadas en el registro sistemático de los hechos vitales, es decir, aquellos eventos relacionados con el comienzo y final de la vida: nacimientos, defunciones y defunciones fetales (DEIS, 2016). En nuestro país, este registro se completa con información proveniente de los certificados médicos que notifican legalmente el nacimiento o defunción de una determinada persona y el lugar de ocurrencia (y, en el caso de las muertes, la causa inmediata de la defunción).

2. *Estadísticas epidemiológicas*: dentro de este grupo podemos encontrar aquellas que se vinculan con la morbilidad (presencia de enfermedades en la población), con la discapacidad (prevalencia, causa, tipo de discapacidad) o con los factores de riesgo y calidad de vida. Estas estadísticas surgen de distintas fuentes específicas, como el Sistema de Notificaciones Médicas Obligatorias, el registro de egresos hospitalarios y de motivos de consulta médica, y de relevamientos especiales de datos sobre determinadas enfermedades (Ortiz, Esandi y Bortman, 2004).

3. *Estadísticas de recursos para la salud*: contemplan información sobre los recursos humanos y físicos disponibles dentro del sector salud. Los datos reunidos surgen de

registros administrativos de personal, de recursos físicos, de insumos y de recursos financieros desarrollados por distintas instituciones del sector.

4. *Estadísticas de atención y acciones de salud*: son elaboradas a partir del registro que llevan adelante los establecimientos de salud, en torno a la realización de actividades de promoción, protección, recuperación y rehabilitación de la salud. Involucran también datos vinculados a las acciones de salud pública sobre el ambiente y la vigilancia de riesgos y enfermedades desarrolladas por los organismos oficiales de salud (como ministerios y secretarías).

5. *Estadísticas económicas sociales*: si bien están relacionadas con la salud, provienen de fuentes que no forman parte del sistema de salud. No obstante, son centrales tanto para el análisis epidemiológico (grupos de riesgo delimitados mediante variables sociales) como para identificar factores que influyen sobre las necesidades de salud de cada grupo social en los distintos territorios.

Las estadísticas de salud se construyen a través de la utilización de una serie de indicadores empíricos que son de uso frecuente y que permiten medir tanto las características de diversos eventos o fenómenos relacionados con la salud de la población como atributos que presentan los sistemas de salud. Recordemos que muchas de las características (o variables) que deseamos estudiar deben ser conceptualizadas con anticipación y definidas operacionalmente, de modo tal que puedan establecerse los componentes o dimensiones cuya intensidad o magnitud queremos medir por medio de algún indicador escogido.

Los *indicadores empíricos de salud* pueden ser entendidos, entonces, como expresiones numéricas que nos posibilitan cuantificar fenómenos de salud. En general, se encuentran estandarizados por autoridades nacionales o internacionales para posibilitar comparaciones entre poblaciones, regiones o a través del tiempo y estimar posibles riesgos epidemiológicos. Básicamente, pueden dividirse en dos grandes grupos: los directos y los indirectos (Torres, 2009).

Entre *los indicadores directos*, encontramos distintos tipos de tasas. De acuerdo con Vioque López, existen tasas destinadas a medir la frecuencia de una enfermedad o de un particular estado de salud en una población, así como la de sus posibles determinantes; es decir, la cantidad de personas expuestas a una enfermedad o afectadas por alguno de sus determinantes en un espacio y tiempo considerados (Vioque López y Navarrete Muñoz, 2011):

- Tasa de prevalencia puntual (número de casos existentes de una enfermedad –o de un determinante de una enfermedad– en un lugar y momento precisos sobre la población total de ese lugar y de ese momento, por 100) mide la proporción de individuos que presentan una enfermedad –o padecen alguno de sus determinantes– en una población de un momento muy concreto y permite estimar la probabilidad o riesgo de estar enfermo que tiene una persona de esa población.
- Tasa de prevalencia de período (número de casos antiguos + número de casos nuevos de una enfermedad –o de un determinante de una enfermedad–, en un lugar, durante un período, sobre la población total de ese lugar y de ese período, por 100) determina los casos existentes de una enfermedad –o de individuos que padecen alguno de sus determinantes– durante un período en el tiempo y permite estimar la probabilidad o riesgo de estar enfermo que tiene una persona de esa población.
- Tasa incidencia acumulada (casos nuevos que desarrollan una enfermedad –o un determinante de una enfermedad–, en un lugar y durante un período determinado, sobre la población libre de enfermedad, y en riesgo de enfermar al comienzo del período considerado, multiplicado por 100). Permite estimar la probabilidad o riesgo de enfermar que tiene una persona de esa población. La duración de ese período de observación influye en los resultados de la incidencia, pues, si este aumenta, ocurrirán más casos.

Otras tasas, en cambio, se orientan a cuantificar eventos vinculados al ciclo de vida de las personas (algunas de las cuales ya hemos mencionado en el segundo módulo del cuaderno):

- Tasa bruta de mortalidad (total de muertes en un período y lugar sobre la población de ese período y lugar, por 1000).
- Tasa de mortalidad por causa (muertes por causa “n” en un período y lugar sobre la población de ese período y lugar, por 100.000).
- Tasa específica de mortalidad por edad (muertos en una edad “n” en un período y lugar sobre la población con edad “n” de ese período y lugar, por 100.000).
- Tasa específica de mortalidad por sexo (muertos del sexo “n” en un período y lugar sobre la población de ese sexo “n” en un período y lugar, por 1000).
- Tasa de mortalidad materna (muertes por embarazo, parto y puerperio en un período y lugar sobre nacidos vivos en ese período y lugar, por 10.000).
- Tasa de mortalidad infantil (muertes de menores de 1 año en un período y lugar sobre nacidos vivos de ese período y lugar, por 1000).



- Tasa de mortalidad neonatal (muertes de menores de 28 días en un período y lugar sobre nacidos vivos de ese período y lugar, por 1000).
- Tasa de mortalidad posneonatal (defunciones de niños entre 28 días y 1 año en un período y lugar sobre nacidos vivos de ese período y lugar, por 1000).
- Tasa bruta de natalidad (nacidos vivos en un período y lugar sobre la población de ese período y lugar, por 1000).

Los *indicadores indirectos*, a su vez, pueden agruparse en tres grandes categorías:

*a) Indicadores de condiciones socioeconómicas.* Tratan de mostrar la correlación que existe entre el desarrollo económico y social alcanzado por una comunidad y el nivel de salud que muestra esta (por ejemplo, calorías consumidas por día per cápita o consumo per cápita de leche, sus derivados y carnes).

*b) Indicadores de condiciones ambientales* (por ejemplo, proporción de población con agua corriente o con cloacas, o porcentaje de hogares con hacinamiento crítico).

*c) Indicadores de recursos y de actividades de salud.* Dentro de este colectivo, podemos recuperar los indicadores de estructura, proceso y resultado propuestos por Donabedian (1988) para evaluar la calidad de la atención sanitaria:

- Los *indicadores de estructura* reflejan información cuantitativa sobre la calidad de la infraestructura de los servicios de salud, su nivel de organización, los tipos de servicios disponibles para la población y los recursos con que se cuenta. Dentro de estos están aquellos que miden la accesibilidad geográfica, la estructura física del área hospitalaria, las características y estructura de cada servicio, los recursos humanos (números y calificación), los recursos materiales (equipamiento médico y no médico, materiales) y las actividades asistenciales, docentes e investigativas.
- Los *indicadores de proceso* miden la calidad de la actividad llevada a cabo dentro del sistema para abordar los procesos de atención de la población; suelen incluir información sobre producción de los servicios. Algunos ejemplos pueden ser las consultas por horas médicas, visitas domiciliarias por horas médicas, egresos por servicios, porcentaje ocupacional de camas por servicios, giro de cama por servicio, promedio días de estadía por cama y servicio.
- Los *indicadores basados en resultados* computan el nivel de éxito alcanzado en los procesos de la atención de la salud de una población, ofrecen oportunidades para valorar la efectividad y la eficiencia del sistema en su conjunto,

tanto con relación a los proveedores de los servicios asistenciales como de las tecnologías utilizadas. Muchas veces, la manera de medir este impacto o el resultado de la atención es a través de los cambios producidos en los niveles de salud, enfermedad, invalidez o muerte que expresa una determinada comunidad y que pueden atribuirse –indirectamente– al desarrollo de las acciones llevadas a cabo por los servicios de salud; en estos casos, se utilizan las tasas de morbilidad o de mortalidad.

*A tener en cuenta:*

Mientras que las tasas de prevalencia buscan medir el número de casos de una enfermedad en una población para un momento o lapso temporal dados, las de incidencia se orientan a estimar el número de casos nuevos que se producen durante un período determinado, en una población especificada. Se trata de dos medidas muy utilizadas para precisar la frecuencia con que ocurre una enfermedad y diseñar intervenciones oportunas. La relación entre ellas puede variar de una enfermedad a otra: hay algunas que presentan alta prevalencia y baja incidencia, como la diabetes, y otras de baja prevalencia y alta incidencia, como la gripe.

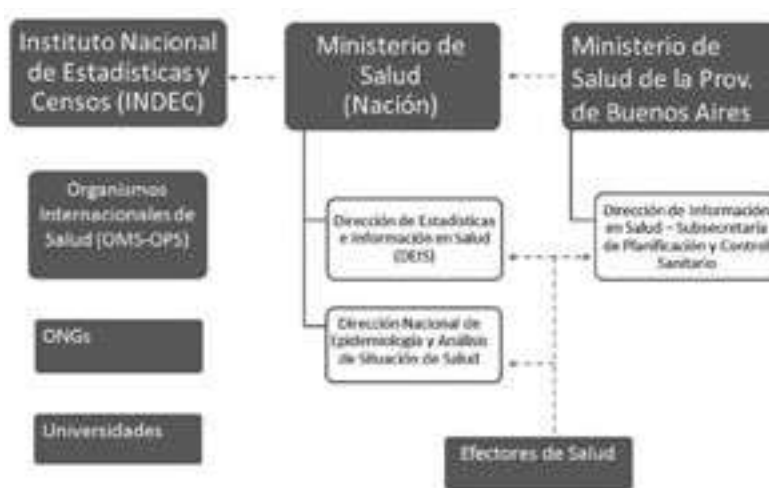
Otras tasas consideradas indicadores fundamentales para la planificación y programación de actividades en salud son las que cuantifican la mortalidad infantil, en virtud de que son altamente sensibles para registrar la situación sanitaria y social de una determinada población. Existen un conjunto de factores –interrelacionados entre sí– que influyen en el nivel que alcanzan estas tasas: biológicos, demográficos, socioeconómicos, culturales, ambientales y relacionados con el desempeño del sistema de salud (Tobar, 2006). No obstante, la influencia de estos factores difiere según la edad de defunción de los menores de un año. Así, en la tasa de mortalidad neonatal, prevalecen aquellos vinculados con las condiciones congénitas (salud de la madre) o con la atención de la salud (control del embarazo, atención del parto y control y atención del niño durante los primeros días de vida). En la mortalidad posneonatal, contrariamente, tienen mayor peso las condiciones ambientales y socioeconómicas sobre la salud del niño (DEIS, 2016).

## **La utilización de los datos disponibles en salud**

Es preciso comenzar distinguiendo entre datos recabados por el propio investigador en el marco de los objetivos propuestos en su investigación en curso (*fuentes primarias*) y datos previamente recolectados por otros investigadores, organismos e instituciones sociales diversas (*fuentes secundarias*).

En la Argentina, constituyen ejemplos de datos secundarios en salud aquellos que se encuentran reunidos en el Sistema Estadístico de Salud (SES),<sup>1</sup> en el Sistema Integrado de Información Sanitaria Argentino (SISA)<sup>2</sup> y en el Sistema Nacional de Vigilancia de la Salud (SNVS)<sup>3</sup>. La figura que se presenta a continuación muestra organismos oficiales e instituciones del país y de la provincia de Buenos Aires que producen regularmente estadísticas para alimentar a estos sistemas de información sanitaria.

Figura 4.1. Organismos productores de información secundaria de salud en la Argentina y en la provincia de Buenos Aires, 2018



Fuente: elaboración propia.

Algunas de las bases disponibles son interactivas y posibilitan al usuario crear sus propias tablas o gráficos, seleccionando los indicadores que le resulten de interés. Dentro del sector salud, la página web de la Organización Panamericana de la Salud (OPS)<sup>4</sup> proporciona un buen ejemplo de este tipo de herramientas.

Para utilizar la información proveniente de fuentes secundarias, el investigador deberá evaluar la adecuación de esta información como estrategia metodológica para su propio estudio: ¿Puedo medir lo que deseo estudiar recurriendo a datos secundarios? Paralelamente, es importante reconstruir el proceso en el que se generaron estos datos y dar cuenta de tres dimensiones:

<sup>1</sup> Sistema Estadístico de Salud (SES): <http://www.deis.msal.gov.ar/index.php/ses-sistema-estadistico-de-salud/>.

<sup>2</sup> Sistema Integrado de Información Sanitaria Argentino (SISA): <https://sisa.msal.gov.ar/sisa/>.

<sup>3</sup> Sistema Nacional de Vigilancia de la Salud (SNVS): <https://www.argentina.gob.ar/salud/epidemiologia/vigilancia>.

<sup>4</sup> Página web de la Organización Panamericana de la Salud: <http://www.paho.org/data/index.php/es/>.

- El *proceso de elaboración* de los datos: ¿cómo se construyeron los datos?
- El *período y las etapas de recolección* de la evidencia empírica.
- El *autor o la institución* que generó esta información.

Estas tres dimensiones pueden ser subsumidas en aquello que autores como González Rodríguez (2000) identifican como problemas de *fiabilidad* y de *validez* de la información secundaria.

Los *problemas de la fiabilidad* refieren a lo siguiente.

- La *disponibilidad y acceso a los datos*: ¿dónde se encuentran los datos que necesito?, ¿son accesibles?, ¿cuándo dispondría de ellos?, ¿cómo se publican, en bases de microdatos o en tabulados?, ¿qué información complementaria necesito?
- La *credibilidad de la fuente*: ¿por quién fueron recogidos los datos?, ¿es un organismo confiable?
- La *credibilidad de los datos*: ¿cómo se construyó el dato?, ¿se informa la metodología? (la definición de la unidad de análisis, el momento de recolección, el tipo de muestra, el nivel de respuesta, la cobertura, el destinatario de la información, el objetivo de la recogida de esa información, etcétera), ¿se definen los conceptos implicados?

El *problema de la validez* de los datos secundarios vuelve sobre la concordancia entre los conceptos teóricos y los indicadores que los expresan antes revisada:

- *Validez de constructo*: correspondencia entre el concepto abstracto y la medida concreta. Coherencia entre mis definiciones teóricas y las de la fuente.
- *Validez de contenido*: cobertura de las diferentes dimensiones del concepto.

*A tener en cuenta:*

La existencia de bases de datos con información sanitaria –sea de proceso, de estructura o de resultados– facilita el desarrollo de análisis comparativos dentro del sector, realizados a nivel nacional, regional e internacional, o entre distintos períodos temporales. De acuerdo con Ragin (2007), este tipo de análisis es recomendado para estudiar diversidades y rasgos similares entre grupos o poblaciones –en este caso vinculados a la salud–, indagando sobre aquellos factores determinantes tanto de las tendencias o de los patrones comunes como de las heterogeneidades observadas.

## Bibliografía

- Dirección de Estadística e Información en Salud (DEIS) (2016). Estadísticas vitales. Información básica Argentina - Año 2016. Serie 5 Número 60. Buenos Aires: Secretaría de Políticas, Regulación e Institutos Dirección de Estadística e Información en Salud, Ministerio de Salud de la Nación.
- Donabedian, Avedis (1988). *Los espacios de la salud: aspectos fundamentales de la organización de la atención médica*. México D. F.: Fondo de Cultura Económica.
- González Rodríguez, Benjamín (2000). “La utilización de los datos disponibles”. En García Ferrando, Manuel; Alvira Martín, Francisco e Ibáñez, Jesús (comps.), *El análisis de la realidad social. Métodos y técnicas de investigación*, pp. 103-133. Madrid: Alianza Universidad.
- Ortiz, Zulma; Esandi, María Eugenia y Bortman, Marcelo (2004). “Cuantificación de los Problemas de Salud”. En *Epidemiología Básica y Vigilancia de la Salud*, pp. 7-85. Programa VIGI+A (Ministerio de Salud-Banco Mundial).
- Ragin, Charles (2007). “El uso de los métodos comparativos para estudiar la diversidad”. En *La construcción de la investigación social. Introducción a los métodos y su diversidad*, pp. 177-212. Bogotá: Universidad de Los Andes-Sage Publications.
- Tobar, Federico (2006). “¿Qué indica la tasa de mortalidad infantil?”. *Revista ISALUD*, diciembre de 2006.
- Torres, Silvana (2009). “Conceptos Básicos”; “Indicadores de uso frecuente en salud”. En *Estadísticas en salud I. Estadística descriptiva*, pp. 4-9; 46-52. Facultad de Medicina, Universidad Nacional del Tucumán.
- Vioque López, Jesús y Navarrete Muñoz, Eva María (2011). “Medidas de frecuencia en epidemiología”. En Hernández-Aguado, Ildelfonso; Gil de Miguel, Ángel; Delgado Rodríguez, Miguel; Bolúmar Montrull, Francisco; Benavides, Fernando; Porta Serra, Miguel; Álvarez-Dardet Díaz, Carlos; Vioque López, Jesús (dirs.) y Lumbreras Lacarra, Blanca (coord.), *Manual de epidemiología y salud pública: para grados en ciencias de la salud*, pp. 51-54. Argentina: Médica Panamericana.





El principal propósito de este cuaderno de trabajo es acercar la *estadística* a quienes desempeñan su práctica profesional dentro del ámbito de la salud y, específicamente, a quienes abordan la gestión de las políticas de salud en el territorio.

Con el objetivo de brindar herramientas sencillas que permitan tanto capacitarse como ejercitarse en los procesos de medición, procesamiento y análisis de los datos sociosanitarios, el texto ofrece un marco general en el cual las nociones y las técnicas básicas de la estadística se articulan con ejemplos ilustrativos.

El cuaderno también busca sistematizar algunos contenidos que son orientadores en la lectura y la interpretación crítica de la información cuantitativa disponible dentro del sector, y que es frecuentemente utilizada en los espacios de trabajo, reflexionando en torno a las potencialidades y a las limitaciones que supone el uso de estos datos secundarios.