

Organización Panamericana de la Salud
Organización Mundial de la Salud

**Módulo de Principios de
Epidemiología para el Control de
Enfermedades (MOPECE)**
Segunda Edición Revisada
**Medición de las condiciones de salud y
enfermedad en la población**

3



**Organización
Panamericana
de la Salud**

Oficina Regional de la
Organización Mundial de la Salud

Organización Panamericana de la Salud
Oficina Sanitaria Panamericana, Oficina Regional de la
Organización Mundial de la Salud

Módulos de principios de epidemiología para el control de enfermedades (MOPECE)

Segunda Edición Revisada

Unidad 3: Medición de las condiciones de salud y
enfermedad en la población



Organización Panamericana de la Salud
Oficina Sanitaria Panamericana
Oficina Regional de la Organización Mundial de la Salud
525 Twenty-third Street, N.W. Washington, D.C. 20037, E.U.A

Catalogación por la Biblioteca de la OPS:

Organización Panamericana de la Salud.

Módulos de Principios de Epidemiología para el Control de Enfermedades, segunda edición. Washington D.C.: OPS, © 2002, 92 p. –(Serie PALTEX N° para Técnicos Medios y Auxiliares N° 24).

ISBN 92 75 32407 7

I. Título

1. EPIDEMIOLOGÍA–principios

3. SALUD PÚBLICA

II. (serie)

2. CONTROL–enfermedades

4. REGION DE LAS AMERICAS

Este Módulo de capacitación están especialmente destinado a los profesionales de salud de América Latina y se publica dentro del Programa Ampliado de Libros de Texto y Materiales de Instrucción (PALTEX) de la Organización Panamericana de la Salud, organismo internacional constituido por los países de las Américas, para la promoción de la salud de sus habitantes y de la Fundación Panamericana para la Salud y Educación. Se deja constancia de que este programa está siendo ejecutado con la cooperación financiera del Banco Interamericano de Desarrollo.

ISBN 92 75 32407 7

© Organización Panamericana de la Salud, 2011

Segunda Edición Revisada

Las publicaciones de la Organización Panamericana de la Salud están acogidas a la protección prevista por las disposiciones sobre reproducción de originales del Protocolo 2 de la Convención Universal sobre Derecho de Autor. Reservados todos los Derechos.

Las denominaciones empleadas en esta publicación y la forma en que aparecen presentados los datos que contiene no implican, por parte de la Secretaría de la Organización Panamericana de la Salud, juicio alguno sobre la condición jurídica de países, territorios, ciudades o zonas, o de sus autoridades, ni respecto del trazado de sus fronteras o límites.

La mención de determinadas sociedades mercantiles o del nombre comercial de ciertos productos no implica que la Organización Panamericana de la Salud los apruebe o recomiende con preferencia a otros análogos.

De las opiniones expresadas en la presente publicación responden únicamente los autores.

Créditos

Módulos de Principios de Epidemiología para el Control de Enfermedades, segunda edición, fue elaborado por el Programa Especial de Análisis de Salud de la Oficina Central de la Organización Panamericana de la Salud (Washington DC, EUA) en 2001.

Editores:

Carlos Castillo-Salgado
Oscar J Mujica
Enrique Loyola
Jaume Canela

Revisores técnicos:

Gabriela Fernández
Enrique Vázquez
Patricia Gassibe
Soledad Velázquez
Edgar Navarro
Patricia Arbeláez
Mayra Cartín
Eduardo Velasco

Revisión editorial:

Lucila Pacheco

Se agradece especialmente la colaboración de: Gilberto Ayala, Julio Alberto Armero, Xiomara Badilla, Itza Barahona de Mosca, Herbert Caballero, Marco Tulio Carranza, Rocío Cuevas, Thais Dos Santos, Carlos Flores, Modesta Haughton, José Federico Hernández, Marlo Libel, Miguel Machuca, Alfredo Moltó, José Moya, Carlos Muñoz, Maritza Ortega, Alberto Paredes, Rosalía Quinteros, Mirta Roses, Patricia Ruiz, Gloria Tewres, Guadalupe Verdejo, Reinaldo Viveros, así como a múltiples epidemiólogos de la Región de las Américas, por su participación y recomendaciones sugeridas durante el proceso de prueba de materiales.

Colaboración para la presente Segunda Edición Revisada:

José Moya, Oscar J Mujica, Steven K Ault, Jacobo Finkelman, Fátima Marinho, Diego Victoria.

Tapa, Proyecto Gráfico y Diagramación:

All Type Assessoria Editorial Ltda.

Indice

Contenidos y objetivos	5
Medición de salud y enfermedad en la población	6
Tipos de datos y su tabulación	10
Presentación gráfica de datos.	15
Medidas de resumen de una distribución	20
Medidas de tendencia central	20
Medidas de dispersión	25
Medidas de frecuencia	31
Prevalencia e incidencia	32
Comparación de la frecuencia de enfermedad y medidas de asociación	58
Medidas de asociación	59
Comparación de dos Proporciones: La Prueba Chi Cuadrado	61
Medidas de la fuerza de asociación	65
Riesgo Relativo	65
Apéndice estadístico - epidemiológico	72
Comparación de dos Promedios: La Prueba Z.	72
Intervalos de Confianza para promedios y proporciones	73
Razón de Posibilidades (<i>Odds Ratio</i>).	77
Medidas de impacto potencial	81
Riesgos atribuibles	81
Fracciones atribuibles	82
Estandarización de tasas	86
Referencias bibliográficas	91

Contenidos y objetivos

Esta Unidad presenta las medidas comunes de cuantificación de las condiciones de salud y enfermedad, sus aplicaciones y limitaciones; el ordenamiento y presentación tabular y gráfica de datos epidemiológicos; las medidas de resumen y de asociación. Esta Unidad se enfoca en los métodos para la presentación y análisis de datos cuantitativos.

Los objetivos de la presente Unidad son:

- Describir las formas de medición de la salud y sus aplicaciones.
- Identificar las formas y usos de la presentación tabular y gráfica de datos.
- Describir, calcular, interpretar y aplicar las medidas de resumen de datos.
- Describir, calcular, interpretar y aplicar las medidas de asociación.

Medición de salud y enfermedad en la población

Existen diversas formas de medir la salud, dependiendo de cuál es su definición; una definición amplia mediría el nivel de salud y bienestar, la capacidad funcional, la presencia y causas de enfermedad y muerte y la expectativa de vida de las poblaciones (Donaldson, 1989). Existen distintas medidas e indicadores de bienestar (social o económico) en salud y se han desarrollado ciertos índices de “salud positiva” (Alleyne, 1998), tanto con fines operacionales, como para investigación y promoción de condiciones saludables, en dimensiones tales como la salud mental, autoestima, satisfacción con el trabajo, ejercicio físico, etc. La recolección de datos y la estimación de indicadores tienen como fin generar, en forma sistemática, evidencia que permita identificar patrones y tendencias que ayuden a emprender acciones de protección y promoción de la salud y de prevención y control de la enfermedad en la población.

Entre las formas más útiles y comunes de medir las condiciones generales de salud de la población destacan los censos nacionales, que se llevan a cabo decenalmente en varios países. Los censos proporcionan el conteo periódico de la población y varias de sus características, cuyo análisis permite hacer estimaciones y proyecciones.

Para permitir las comparaciones a lo largo del tiempo en una misma población o bien entre poblaciones diferentes, se requiere de procedimientos de medición estandarizados.

Medición: Es el procedimiento de aplicar una escala estándar a una variable o a un conjunto de valores.

La medición del estado de salud requiere sistemas armonizadores y unificados como la Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud (CIE), en su Décima Revisión, cuyos XXI capítulos inician con ciertas enfermedades infecciosas y parasitarias (A00-B99) y culminan con factores que influyen en el estado de salud y contacto con los servicios de salud (Z00-Z99).

Los indicadores de salud miden de la población distintos aspectos relacionados con la función o discapacidad, la ocurrencia de enfermedad o muerte, o bien aspectos relacionados con los recursos y desempeño de los servicios de salud.

Los indicadores de salud funcional tratan de medir el impacto de los problemas de salud en la vida diaria, como por ejemplo la capacidad para llevar a cabo actividades cotidianas, lesiones y accidentes en el hogar y el lugar de trabajo, y años de vida libres de discapacidad. Los datos se obtienen generalmente a través de encuestas y registros de discapacidades. Los índices de calidad de vida incluyen variables de función tales como

la actividad física, la presencia de dolor, el nivel de sueño, de energía, o el aislamiento social.

Los indicadores de morbilidad miden la frecuencia de problemas de salud específicos tales como infecciones, cánceres, accidentes en el trabajo, etc. Las fuentes de datos suelen ser registros de hospitales y servicios de salud, notificación de enfermedades bajo vigilancia y encuestas de seroprevalencia y de autoreporte de enfermedad, entre otros. Cabe mencionar que las enfermedades crónicas, por su larga evolución, requieren de monitoreo de etapas clínicas, por lo que es preferible contar con registros de enfermedad (cáncer, defectos congénitos) (Newcomer, 1997).

Los indicadores de mortalidad general o por causas específicas permiten comparar el nivel general de salud e identificar causas de mortalidad relevantes como accidentes, tabaquismo, etc. El registro de la mortalidad requiere de la certificación de la muerte, para lo cual se usa el Certificado Médico de Defunción. La mortalidad se presenta comúnmente como números crudos, proporciones, o tasas por edad, sexo y causas específicas.

Además de la medición del estado de salud, también es necesario medir el desempeño de los servicios de salud. Tradicionalmente esta medición se ha enfocado a insumos y servicios; en la actualidad se considera preferible medir los procesos y funciones de los servicios de salud (Turnock, 1997).

Conjuntamente con los indicadores mencionados, la medición en salud requiere de la disponibilidad de datos sobre características relevantes de la población (variables), tales como su tamaño, composición, estilos de vida, clases sociales, eventos de enfermedad, nacimientos y muertes.

Los datos para la medición en salud provienen de diversas fuentes, por lo que deben de tomarse en cuenta aspectos relacionados con la validez, calidad, integridad y cobertura de los datos mismos y sus fuentes. Los datos, cuantitativos o cualitativos, que se obtienen y registran de los servicios de salud y las estadísticas vitales representan la “materia prima” para el trabajo epidemiológico. Si los datos son incompletos o inconsistentes, se obtendrán medidas sesgadas o inexactas, sin importar la sofisticación del análisis epidemiológico, y las intervenciones derivadas de su uso no serán efectivas. La deficiente cobertura de los servicios en amplios sectores de población en varios países, limita la generación de información útil y necesaria para resolver los problemas de salud que afectan de manera específica a sus comunidades. Aún cuando los datos estén disponibles y sean confiables, su utilización para la gestión en salud puede ser insuficiente.

Con el propósito de responder a la necesidad de contar con un conjunto de datos validados, estandarizados y consistentes de los países de las Américas, la Organización Panamericana de la Salud (OPS) trabaja desde 1995 en la Iniciativa Regional de Datos

Básicos de Salud. Se incluye en esta fuente una serie histórica de 117 indicadores demográficos, socioeconómicos, de morbilidad y mortalidad y de recursos, acceso y cobertura de servicios de salud, de los 48 Estados y Territorios de la Región.

Una vez que se cuenta con los datos e indicadores de salud, una de las dificultades presente en los servicios de salud tiene que ver con las limitaciones para el manejo correcto de la información numérica, su análisis e interpretación, funciones que requieren el uso de los principios de la epidemiología y la bioestadística. Resulta paradójico que en el nivel en que ocurren los problemas y en donde se solucionan, los procedimientos y técnicas para la obtención, medición, procesamiento, análisis, interpretación y uso de datos aún no estén plenamente desarrollados.

Para la correcta toma de decisiones en todos los niveles de los servicios de salud, basada en información pertinente, es necesaria la capacitación permanente del equipo local de salud y de sus redes en la recolección, manejo, análisis e interpretación de datos epidemiológicos.

La cuantificación de los problemas de salud en la población requiere de procedimientos y técnicas estadísticas diversas, algunas de ellas de relativa complejidad. Dadas las características multifactoriales de los problemas de salud, las técnicas cualitativas son también valiosas para aproximarse al conocimiento de los determinantes de salud. Es por ello que existe la necesidad de incorporar, en forma dialéctica, métodos y técnicas cuantitativas y cualitativas que permitan estudiar los diversos componentes de los objetos de estudio.

En el análisis cuantitativo el empleo de programas computacionales facilita el manejo y análisis de datos, pero no se deben sobrestimar sus alcances y aplicaciones. Su utilidad es mayor cuando se establecen redes de colaboración y sistemas de información en salud, que permiten el manejo eficiente de grandes bases de datos y generan información oportuna y útil para la toma de decisiones. Un programa computarizado reduce notablemente el tiempo de cálculo, procesamiento y análisis de los datos, pero es el trabajo humano el que aporta resultados racionales y válidos para el desarrollo de los objetivos de salud pública. Existen dos paquetes de programas de cómputo diseñados específicamente para salud que facilitan el almacenamiento, proceso y análisis de información epidemiológica: Epi-Info producido por el Centro de Prevención y Control de Enfermedades de los Estados Unidos (CDC) y la Organización Mundial de la Salud (OMS) y el Epidat de la OPS y la Xunta de Galicia, España. Lejos de competir entre sí, estos paquetes de programas, de gran uso y de libre distribución, ofrecen procesos y rutinas de manejo y análisis epidemiológico de datos que son complementarios.

En un sentido amplio, podemos considerar que el quehacer de la salud pública parte de constatar una realidad de salud no deseable en una población y apunta a conseguir un cambio social, deliberado y sostenible en dicha población. En ese sentido, y desde

un punto de vista metodológico, la epidemiología como toda ciencia tiene exigencia de método desde una perspectiva estadística. El **enfoque epidemiológico**, consiste básicamente en: i) la *observación* de los fenómenos de salud y enfermedad en la población; ii) la *cuantificación* de éstos en frecuencias y distribuciones; iii) el *análisis* de las frecuencias y distribuciones de salud y de sus determinantes; y, iv) la *definición de cursos de acción* apropiados. Este proceso cíclico de observar–cuantificar–comparar–proponer sirve también para evaluar la efectividad y el impacto de las intervenciones en salud, para construir nuevos modelos que describan y expliquen las observaciones y para utilizarlos en la predicción de nuevos fenómenos.

En resumen, en todo este proceso, los procedimientos y técnicas de cuantificación son de gran relevancia y la capacitación del equipo local de salud en estos aspectos del enfoque epidemiológico es, en consecuencia, fundamental.

Tipos de datos y su tabulación

La cuantificación del estado de salud y patrones de enfermedad en la población, requiere de métodos y técnicas que permitan recolectar datos en forma objetiva y eficiente, convertir los datos en información para facilitar su comparación y simplificar su interpretación, y transformar la información en conocimiento relevante para las acciones de control y prevención. Para conocer los grupos de población que presentan mayor número de casos, los lugares con mayor incidencia o prevalencia de determinadas enfermedades y el momento en que ocurren los eventos en salud, se deben aplicar sistemáticamente algunos procedimientos básicos de manejo de datos o variables.

Uno de los primeros pasos en el proceso de medición del estado de salud en la población es la definición de las variables que lo representan o caracterizan.

Variable: cualquier característica o atributo que puede asumir valores diferentes.

Las variables pueden ser de dos tipos, cualitativas y cuantitativas. Denominamos variables cualitativas a aquellas que son atributos o propiedades. Las variables cuantitativas son aquellas en las que el atributo se mide numéricamente y a su vez se pueden clasificar en discretas y continuas. Las **variables discretas o discontinuas** asumen valores que son siempre números enteros; por ejemplo, el número de hijos de una pareja, el número de dientes con caries, el número de camas de hospital, el número de hemáties por campo, el pulso, etc., que sólo pueden tomar valores de un conjunto finito. Las **variables continuas** pueden tomar tantos valores como permita la precisión del instrumento de medición; por ejemplo, el peso al nacer de un bebé de 2.500 gramos podemos medirlo con mayor precisión, como 2.496,75 gramos, si nuestra báscula lo permite.

Las variables también pueden clasificarse según el nivel o tipo de medición que podamos aplicarles. Así, se pueden distinguir cuatro niveles de medición de las variables: nominal, ordinal, de intervalo y proporcional o de razón. Una **variable nominal** tiene categorías a las que se les asignan nombres que no tienen ningún orden entre ellos; por ejemplo, el sexo. La categoría “hombre” no tiene ninguna relación de orden sobre la categoría “mujer”. Las variables nominales no tienen que ser necesariamente dicotómicas (dos categorías) sino que pueden tener varias categorías, como por ejemplo el estado civil (soltero, casado, divorciado, viudo, unión libre) o el grupo sanguíneo según el sistema ABO (A, B, AB y O). El hecho de cambiar el orden no tiene ninguna implicación en el análisis de los datos.

Ahora supongamos que se nos pregunta sobre la calidad de un curso que acabamos de realizar y se nos ofrecen las siguientes opciones de respuesta: *muy malo*, *malo*, *regular*, *bueno* y *excelente*. Esta clasificación tiene un orden: *excelente* es mejor que *bueno*, *bueno* que *regular* y así sucesivamente; sin embargo, la “distancia” que hay entre *excelente* y

bueno no tiene porque ser la misma que entre *mal* y *muy malo*. Estamos ante una **variable ordinal**, que se define como aquella cuyas categorías tienen un orden, aunque las diferencias entre ellas pueden no ser iguales. Otros ejemplos de variables ordinales son los estadios de un cáncer (I, II, III y IV) o los resultados de un cultivo de laboratorio (-, +, ++, +++).

El siguiente nivel de medición de variables es el interválico. Una **variable de intervalo** tiene distancias iguales entre sus valores y una característica fundamental: *el cero es arbitrario*. El ejemplo típico de variable de intervalo es la temperatura corporal. Existe la misma diferencia entre 37°C y 39°C que entre 38°C y 40°C (o sea, 2°C). Sin embargo, no podemos decir que una temperatura de 60°C sea ‘tres veces más caliente’ que una de 20°C. Tampoco podemos concluir que un individuo con un coeficiente de inteligencia de 120 es el doble de inteligente que otro con coeficiente 60.

Por último, si la variable de intervalo tuviese un punto de origen que es el valor *cero significativo*, estaríamos hablando de una **variable proporcional o de razón**. Ésta tiene intervalos iguales entre valores y punto de origen cero. El peso y la altura son ejemplos típicos de variables proporcionales, 80 Kg. es el doble que 40Kg. y hay la misma diferencia entre 50 y 35Kg. que entre 105 y 90 Kg. En este nivel se puede sumar, restar, multiplicar y dividir.

En el siguiente esquema se sintetiza la relación entre los distintos niveles de medición de las variables.

Tipo de variable	Valores
Nominal	categorías con nombre convencional
Ordinal	como las nominales, + categorías ordenadas
De Intervalo	como las ordinales, + intervalos iguales
Proporcional o de Razón	como las de intervalo, + cero significativo

Adaptado de Norman y Streiner, 1996.

Los datos sobre casos de enfermedades atendidas o notificadas por el centro o servicio de salud pueden provenir de un listado de nombres, edades, sexo, etc., del cual se puede obtener el número total o frecuencia de casos. Si se busca agrupar los casos según algunas características de los mismos, la tarea de identificar los grupos poblacionales con mayores problemas se simplifica. Por ejemplo, entre las características de la **persona**, se pueden agrupar los casos según su edad, sexo o etnia. A partir de este listado de datos se puede, además, identificar los casos que han ocurrido en determinado **lugar** y comparar la frecuencia de la enfermedad en distintos lugares o áreas geográficas. Por último, se puede examinar el número de casos según el momento en que ocurrieron o fueron notificados. La distribución de los casos en el **tiempo** se puede agrupar en días, semanas, meses o cualquier otro período de tiempo que se considere adecuado. Esto permite saber

en qué momento se presentó el mayor número de casos, cuándo empezó a aumentar y cuándo a disminuir. Es necesario recordar que la **frecuencia** es el número de veces que se repite un valor de la misma variable.

Los datos agrupados según determinadas características (edad, sexo, residencia, clase social, etc.) pueden presentarse en cuadros y/o gráficos; esto facilita los cálculos y la comparación e interpretación de los resultados como se puede ver en el siguiente Cuadro

Cuadro 3.1 Distribución de los casos de gastroenteritis según edad. Centro de Salud A, mayo de 2000.

edad (años)	Casos (N°)	Porcentaje %
< 1	63	36,4
1	55	31,8
2	25	14,4
3	14	8,1
4	5	2,9
5 y más	11	6,4
Total	173	100,0

Ejemplo para la elaboración de un cuadro:

El siguiente listado corresponde a las edades en años de 120 personas afectadas por malaria durante el verano de 2001, en una isla del Caribe

27	32	58	44	24	32	29	50
28	36	38	48	38	47	29	39
40	37	35	36	36	36	42	45
32	36	48	42	46	35	32	54
30	29	36	44	30	36	27	37
35	33	38	63	37	53	35	46
36	27	34	32	46	38	43	20
25	41	27	53	40	31	47	43
29	49	50	34	47	36	38	24
30	51	43	46	38	49	47	30
29	33	54	40	28	63	36	41
46	51	49	37	41	37	39	38
30	35	36	34	43	43	37	55
29	44	38	42	43	35	42	50
35	47	32	54	41	41	35	40

La edad mayor es 63 años y la menor es 20 años. Si agrupamos estas edades en clases, por intervalos de 4 años, tendremos 11 grupos de edad, a los que procedemos a asignar sus respectivas frecuencias (“paloteo”), como se muestra a continuación:

20 – 23 años	(un caso)
24 – 27 años	(siete casos)
28 – 31 años	(14)
32 – 35 años	(19)
36 – 39 años	(27)
40 – 43 años	(19)
44 – 47 años	(14)
48 – 51 años	(10)
52 – 55 años	(6)
56 – 59 años	(1)
60 – 63 años	(2)

Finalmente, preparamos la presentación tabular de esta información, es decir, el cuadro propiamente dicho, al que se denomina tabla o cuadro resumen de frecuencias, útil porque presenta la distribución de la variable que estamos observando.

Distribución: es el resumen completo de las frecuencias de los valores o categorías de la medición realizada. La distribución muestra cuántos o qué proporción del grupo se encuentra en un determinado valor o rango de valores dentro de todos los posibles que la medida cuantitativa puede tener.

En nuestro ejemplo, la distribución del número de casos de la enfermedad según grupos de edad, se presenta en el Cuadro 3.2. La distribución del número de casos (frecuencia absoluta) se acompaña del porcentaje de cada grupo (frecuencia relativa simple) y del porcentaje acumulado (frecuencia relativa acumulada), que suelen aportar información adicional útil. Por ejemplo, el 15,8% de los casos correspondió a personas de 32 a 35 años de edad y cerca de un tercio de los casos (34,1%) tenía menos de 36 años.

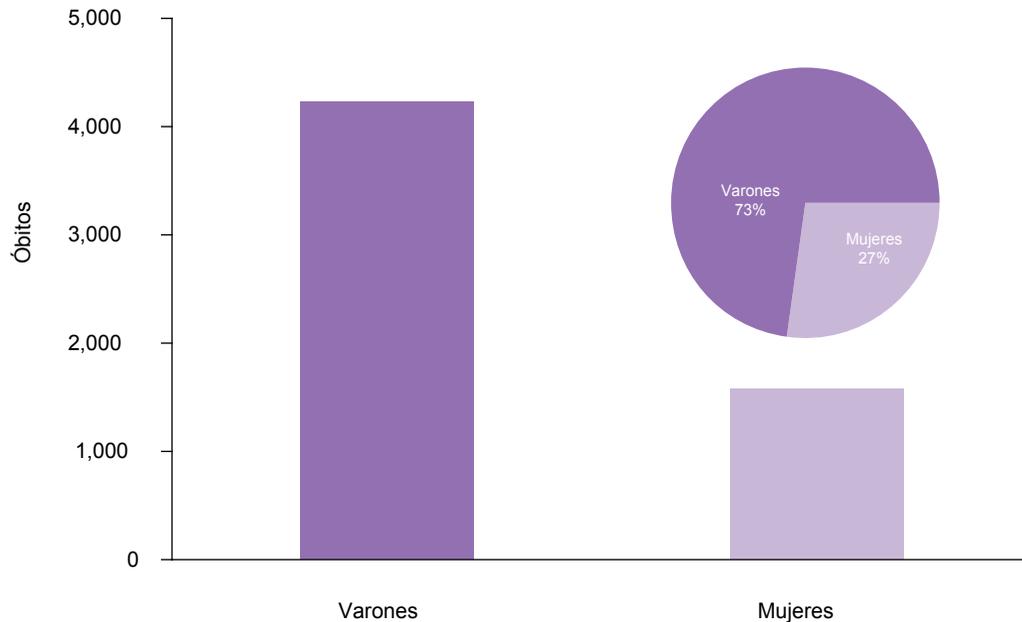
Cuadro 3.2 Distribución de casos de malaria por grupos de edad. Isla del Caribe, verano de 2001

grupos de edad (en años)	número de casos (frecuencia absoluta)	Porcentaje (%)	Porcentaje acumulado (%)
20 - 23	1	0,8	0,8
24 - 27	7	5,8	6,6
28 - 31	14	11,7	18,3
32 - 35	19	15,8	34,1
36 - 39	27	22,5	56,6
40 - 43	19	15,8	72,4
44 - 47	14	11,7	84,1
48 - 51	10	8,4	92,5
52 - 55	6	5,0	97,5
56 - 59	1	0,8	98,3
60 - 63	2	1,7	100,0
Total	120	100,0	

Presentación gráfica de datos

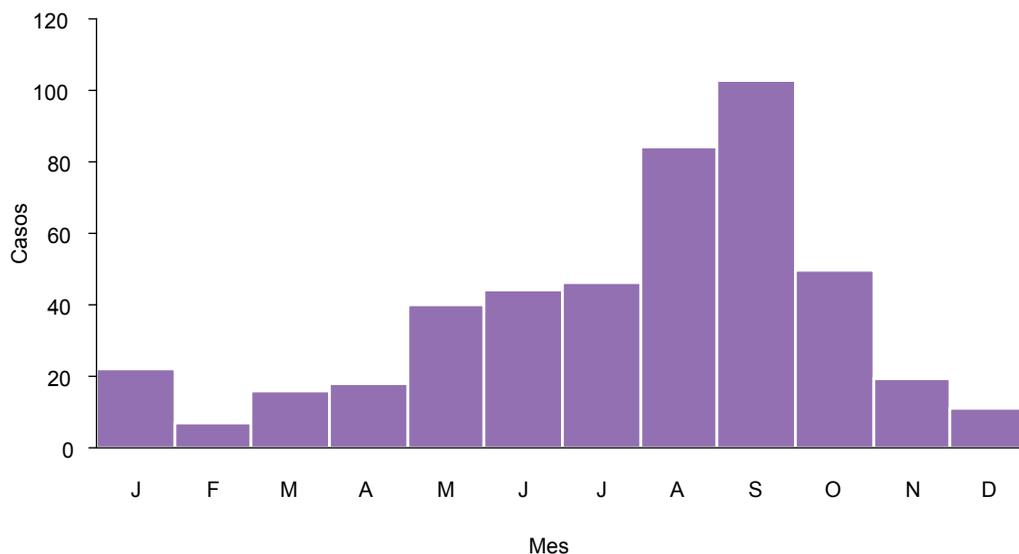
La distribución de variables cualitativas así como las cuantitativas discretas se suele representar gráficamente por medio de diagramas de barras o bien por gráficos de sectores, ya sea como frecuencias absolutas o relativas, como se muestra a continuación:

Gráfico 3.1 Distribución de muertes por suicidio según sexo. Lugar X, 1995-2000



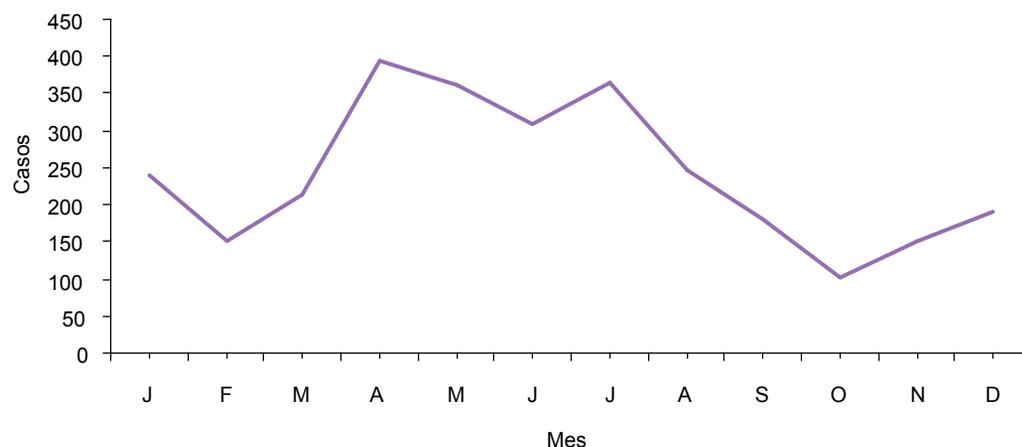
Las variables cuantitativas continuas se representan gráficamente por medio de histogramas y polígonos de frecuencia. Aunque parecidos a los diagramas de barras, en los histogramas las barras se disponen en forma adyacente, precisamente para ilustrar la continuidad y distribución de la variable representada. En el eje de las “x” se ubica la variable continua y en el eje de las “y” se representa la frecuencia.

Gráfico 3.2 Giardiasis. Casos notificados por mes, Lugar X, 2000
(total de casos notificados = 460)

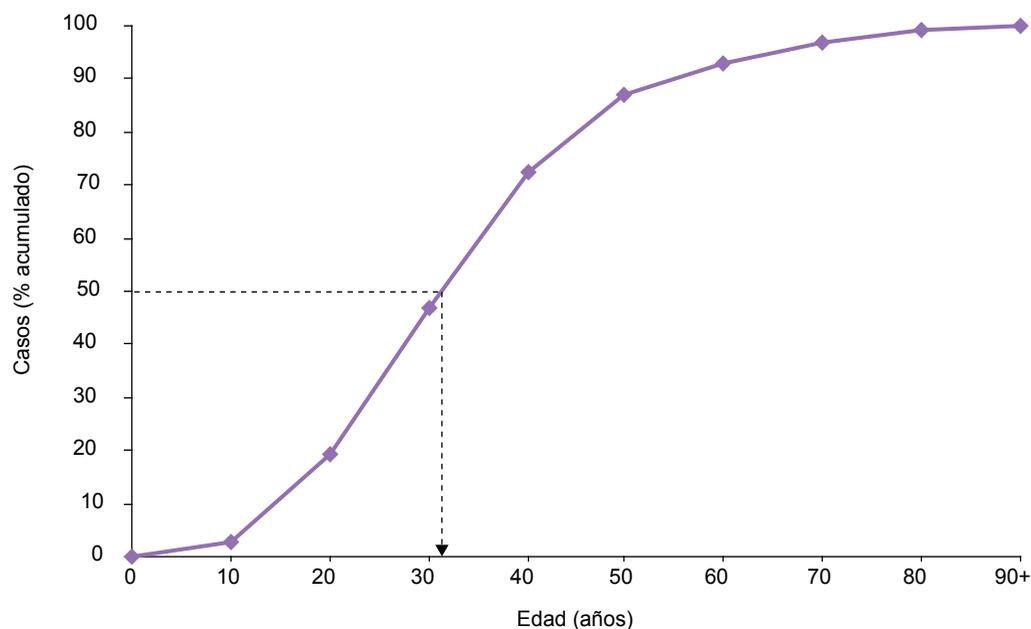


En el Gráfico 3.2, las categorías (meses) de la variable tiempo, una variable continua, se representan en el eje horizontal y el número de casos de giardiasis notificados en el eje vertical. La altura de cada barra representa la frecuencia absoluta de casos (puede también ser la frecuencia relativa) en cada una de las categorías de la variable tiempo, llamadas también intervalos de clase. Note que los intervalos de clase pueden ser de igual tamaño.

El polígono de frecuencias también permite graficar la distribución de una variable y se construye uniendo con líneas rectas los puntos medios del extremo superior de cada barra de un histograma (Gráfico 3.3). Es particularmente útil para visualizar la forma y simetría de una distribución de datos y para presentar simultáneamente dos o más distribuciones.

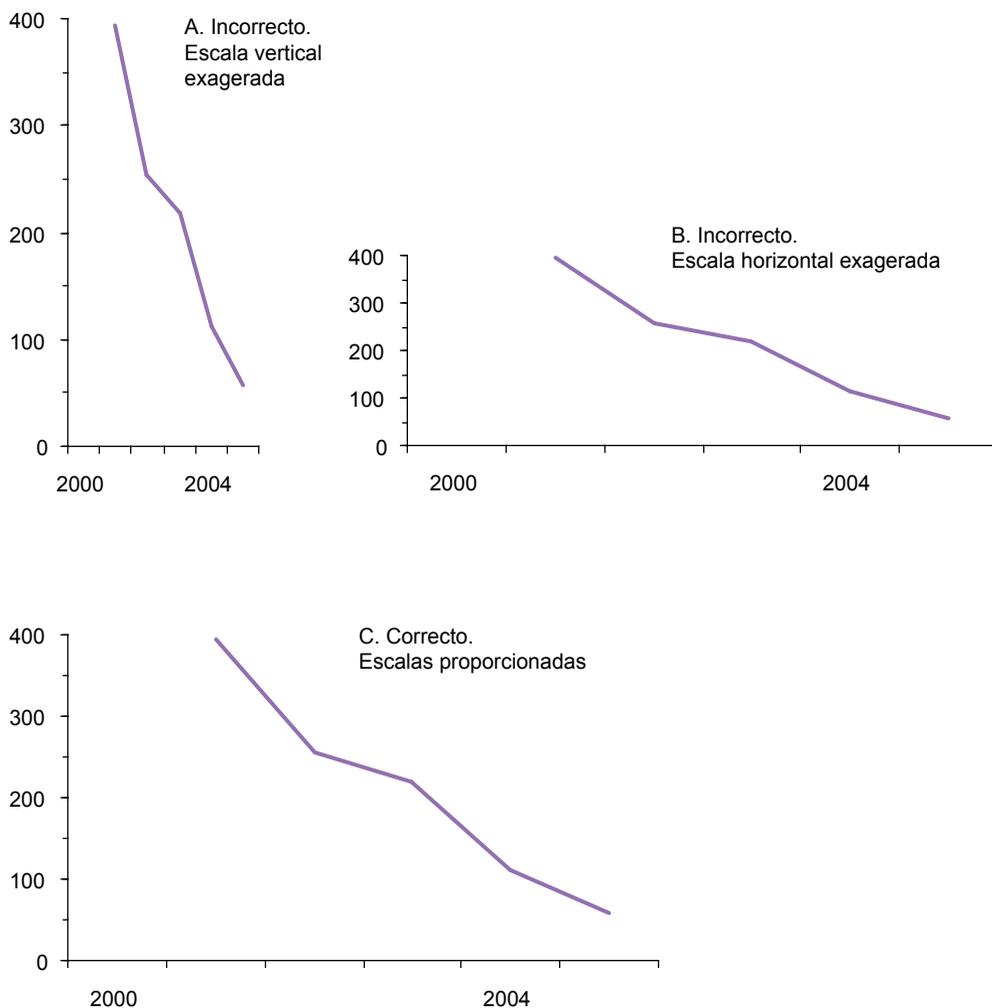
Gráfico 3.3 Casos de dengue por mes de inicio. Lugar Y, 2000

Una variante del polígono de frecuencias es la llamada “**ojiva porcentual**”, que es un polígono de frecuencias relativas acumuladas. Cada punto de este polígono representa el porcentaje acumulado de casos en cada intervalo de clase y, por tanto, va de cero a 100%. La ojiva porcentual permite identificar de manera gráfica el valor correspondiente a la mediana (i.e., 50%) de la distribución (Gráfico 3.4).

Gráfico 3.4 Casos de tuberculosis por edad. Lugar Z, 2000

Es importante mantener la proporción de las escalas del gráfico, ya que de otra manera éste puede dar impresiones erróneas (Gráfico 3.5, A, B, C).

Gráfico 3.5 Efecto de las dimensiones de las escalas en los gráficos



Aunque no existe una regla explícita sobre la proporcionalidad entre las escalas de un gráfico, se recomienda que la razón entre la escala horizontal respecto de la vertical se aproxime a 1,6:1 (la clásica “razón de oro”). Algunas veces el rango 1,2 a 2,2 se da como referencia apropiada para la razón entre el eje horizontal respecto del eje vertical.

El siguiente esquema resume los tipos básicos de gráficos más apropiados según cada tipo de variable.

Tipo de variable	Tipo de gráfico
Nominal	diagrama de barras gráfico de sectores
Ordinal	diagrama de barras (*) gráfico de sectores
De Intervalo	diagrama de barras (*) histograma gráfico de sectores (**) polígonos de frecuencias (simples y acumuladas)
Proporcional o de Razón	diagrama de barras (*) histograma gráfico de sectores (**) polígonos de frecuencias (simples y acumuladas)

(*) manteniendo el orden de las categorías

(**) en variables discretas o categorizadas

Una forma de presentación gráfica del comportamiento histórico de una enfermedad y los límites de variabilidad esperados sobre su ocurrencia en el futuro se denomina **corredor o canal endémico**. Esta herramienta de la vigilancia (Unidad 4) utiliza polígonos de frecuencia y permite visualizar el comportamiento secular (i.e., en el tiempo) de las enfermedades en un territorio determinado y orientar la decisión sobre la necesidad de desencadenar acciones de prevención y control.

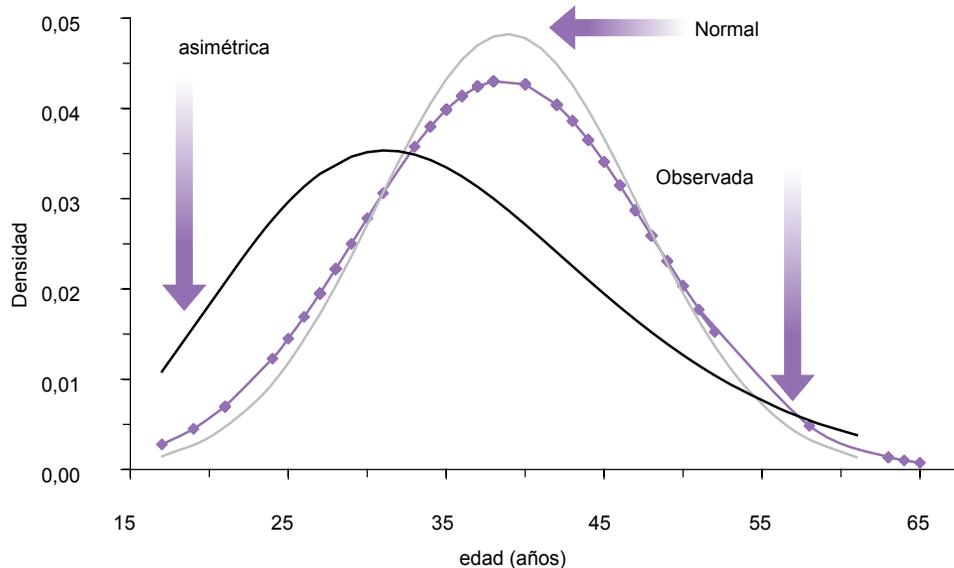
Medidas de resumen de una distribución

Muchas veces es necesario utilizar un valor resumen que represente la serie de valores en su conjunto, es decir, su distribución. Para los datos de variables cualitativas, la proporción o porcentaje, la razón y las tasas son unas típicas medidas de resumen. Para los datos de variables cuantitativas, sin embargo, hay medidas que resumen su tendencia hacia un valor medio (medidas de tendencia central) y otras que resumen su grado de variabilidad (medidas de dispersión). Cada una proporciona información complementaria y útil para el análisis epidemiológico.

Medidas de tendencia central

Las medidas de tendencia central de los datos son la moda, la mediana, y la media o promedio. La selección de las medidas depende del tipo de datos y propósitos. Los valores de muchas variables biológicas, como la talla, se distribuyen de manera simétrica. Otras variables, como la mortalidad y la tasa de ataque en una epidemia, tienen distribución asimétrica. Para distribuciones normales (ejemplo de distribución simétrica), la media, la mediana y la moda son idénticas. Para distribuciones asimétricas, la mediana representa mejor al conjunto de datos, aunque la media tiene mejores propiedades para el análisis estadístico y pruebas de significancia. El Gráfico 3.6 ilustra estas distribuciones, usando la serie observada de datos de la página 12 (listado de edades).

Gráfico 3.6 Distribuciones normal, observada y asimétrica de la edad



Para indicar que, por ejemplo, la rubéola tiene un período de incubación de 15 a 21 días (generalmente 18 días), fue necesario observar una gran cantidad de casos, desde el primer contacto de una persona sana con un enfermo hasta la aparición de los signos y síntomas y medir este período en días. La diferencia entre el límite inferior (15 días) y el superior (21 días) se conoce como rango y se considera, junto con la desviación estándar y la varianza, una medida de dispersión de los datos, como se verá más adelante. El Cuadro 3.3 presenta el período de incubación de la rubéola, en días, observado en 11 niños (a esta forma de listar los datos se le denomina Serie Simple de datos).

Cuadro 3.3 Período de incubación de la rubéola observado en 11 niños.

niño 1	19 días
niño 2	16 días
niño 3	37 días
niño 4	15 días
niño 5	16 días
niño 6	32 días
niño 7	15 días
niño 8	16 días
niño 9	20 días
niño 10	16 días
niño 11	15 días

El valor más frecuente, o sea el que más se repite, se denomina modo o moda. En el ejemplo, observamos que lo más frecuente fueron períodos de incubación de 16 días (4 veces). Este valor puede ser utilizado para representar el período de incubación en este grupo; i.e., “el período de incubación de rubéola más frecuente es de 16 días”.

Si ordenamos los valores en forma ascendente, como por ejemplo:

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
15	15	15	16	16	16	16	19	20	32	37

Observaremos que el sexto valor, o sea el que se encuentra en medio de la serie, también es de 16 días. Este valor que ocupa la posición central de una serie ascendente o descendente se denomina mediana y divide la serie en 50% de las observaciones arriba y 50% abajo de ella. Esta medida también es útil y puede ser empleada para representar la tendencia central del conjunto de datos, sobre todo cuando no es conveniente usar la media debido a la presencia de valores extremos o cuando la distribución es asimétrica, como en una curva epidémica.

Si en vez de un número impar de observaciones tuviéramos un número par de valores, como por ejemplo:

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
15	15	15	16	17	18	19	20	32	37

Se puede constatar que en esta serie no hay un valor central. Para calcular la mediana se suman los dos valores centrales (en el caso, 17 y 18) y se divide el resultado entre dos:

$$\frac{17 + 18}{2} = \frac{35}{2} = 17,5 \text{ días}$$

La **media o promedio aritmético** (\bar{x}) también es muy útil y se obtiene sumando los valores de todas las observaciones y dividiendo el resultado entre el número de observaciones. En nuestro ejemplo, es la suma de todos los valores de los períodos de incubación dividida entre el número de niños observados.

$$\frac{15 + 15 + 15 + 16 + 16 + 16 + 16 + 19 + 20 + 32 + 37}{11} = \frac{217}{11} = 19,7 \text{ días}$$

El promedio de 19,7 días es mayor que los valores del modo y de la mediana ya que, como toma en cuenta los valores de todos los casos, se ve afectado por la influencia de los casos con 32 y 37 días de incubación, que son valores extremos.

En muchas ocasiones los datos están disponibles como distribución de frecuencias, en cuyo caso la serie se presentaría de la manera siguiente:

período de incubación (días)	Casos (número)
15	3
16	4
19	1
20	1
32	1
37	1

El cálculo de la media de estos datos se haría así:

$$3(15)+4(16)+1(19)+1(20)+1(32)+1(37) = 45+64+19+20+32+37 = 217/11 = 19,7 \text{ días}$$

Las medidas de tendencia central son de gran utilidad también para comparar grupos de valores. Por ejemplo, de las personas que participaron en un paseo, un grupo se enfermó después de la comida y otro grupo no presentó ningún síntoma. Las edades de las personas en los dos grupos fueron las siguientes:

- enfermos: 8, 12, 17, 7, 9, 11, 6, 3 y 13
- sanos: 19, 33, 7, 26, 21, 36, 33 y 24

Los promedios aritméticos calculados fueron:

$$\text{enfermos} = \frac{8 + 12 + 17 + 7 + 9 + 11 + 6 + 3 + 13}{9} = 10 \text{ años}$$

$$\text{sanos} = \frac{19 + 33 + 7 + 26 + 21 + 36 + 33 + 24}{8} = 25 \text{ años}$$

Por lo tanto la enfermedad afectó más a los niños que a los adultos. Los niños eran más susceptibles o se expusieron más (consumieron mayor cantidad del alimento contaminado).

Series agrupadas

Cuando se tienen muchos datos se requiere agruparlos, para ello se construyen intervalos, que pueden contener igual o diferente número de unidades, y a ellos se asignan los datos observados. Tenemos entonces una serie de datos agrupados como en el Cuadro 3.4a.

Cuadro 3.4a Casos de suicidio. Lugar X, 1995-2000

edad (años)	Casos (<i>f</i>)
10 a 14	37
15 a 19	176
20 a 29	693
30 a 39	659
40 a 49	784
50 a 59	1.103
60 a 69	1.005
Total	4.457

Cuando nuestros datos sobre una variable continua están agrupados (categorizados) también podemos calcular una media y mediana aproximadas. Para estimar la media se debe construir una columna con los puntos medios (*x*) de cada intervalo de clase

de la variable y otra (fx) resultado de multiplicar el valor de cada punto medio (x) por el número de casos (f) del intervalo correspondiente (Cuadro 3.4b). La suma de estos productos ($\sum fx$) dividida entre la suma de casos ($\sum f$) nos da una aceptable aproximación a la media.

Para calcular el punto medio (x) de cada intervalo de clase de la variable se obtiene la media del intervalo, esto es se suman el límite inferior y superior del intervalo y se divide entre dos, por ejemplo en el cuadro 3.4a el primer intervalo es de 10 a 14 años ($10+14.9/2 = 12.5$), note que se utiliza 14.9 por ser el límite superior real del intervalo. Los puntos medios de los siguientes intervalos de clase se calculan de manera similar.

Cuadro 3.4b Casos de suicidio. Lugar X, 1995-2000

edad (años)	Casos (f)	punto medio (x)	fx	Casos acumulados
10 a 14	37	12,5	462,5	37
15 a 19	176	17,5	3.080,0	213
20 a 29	693	25,0	17.325,0	906
30 a 39	659	35,0	23.065,0	1.565
40 a 49	784	45,0	35.280,0	2.349
50 a 59	1.103	55,0	60.665,0	3.452
60 a 69	1.005	65,0	65.325,0	4.457
Total	$\sum f$ 4.457		$\sum fx$ 205.202,5	

En este ejemplo, la aproximación a la media sería:

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{205.202,5}{4.457} \approx 46,0 \text{ años}$$

Mientras que si utilizáramos en el cálculo los 4.457 casos, uno a uno, obtendríamos una media de 45,7 años de edad.

También podemos estimar la mediana a partir de datos agrupados en un cuadro de frecuencias. Para ello tendremos que construir una columna de “casos acumulados”, como la que se presenta en el Cuadro 3.4b, que se obtiene a partir de la columna de “casos”. La primera celda corresponde a los casos de 10 a 14 años (37), la segunda se obtiene sumando a esos 37 los casos de 15 a 19 años (176) y así sucesivamente, hasta completar la última celda, cuyo valor tiene que coincidir con el total de casos (4.457). Una vez construida la columna de frecuencias acumuladas podemos aproximar la mediana de la edad mediante el siguiente proceso de cálculo: primero, localizar el intervalo de clase que contiene la posición de la mediana (PM); es decir:

$$PM = \frac{(\sum f) - 1}{2} = \frac{4.457 - 1}{2} \approx 2.228$$

En la columna de casos acumulados el caso N° 2.228 está situado en el intervalo de edad de 40 a 49 años.

Después de obtener la posición de la mediana, se estima la mediana por interpolación; es decir:

$$mediana = Li + \left[\frac{PM - fIPM}{fPM} \right] (AIPM)$$

en donde:

- Li = Límite inferior del intervalo de la posición de la mediana.
- PM = Posición de la mediana.
- $fIPM$ = Frecuencia acumulada del intervalo anterior a la posición de la mediana.
- fPM = Frecuencia de la posición de la mediana.
- $AIPM$ = Amplitud del intervalo de la posición de la mediana.

$$mediana = 40,0 + \left[\frac{2.228 - 1.565}{784} \right] (10) = 40,0 + 8,45 \approx 48,45 \text{ años}$$

en donde:

- 2.228 = caso situado en el punto medio de la serie
- 1.565 = casos acumulados en el intervalo anterior al que contiene el caso 2.228
- 784 = casos del intervalo que contiene la mediana
- 10 = amplitud del intervalo

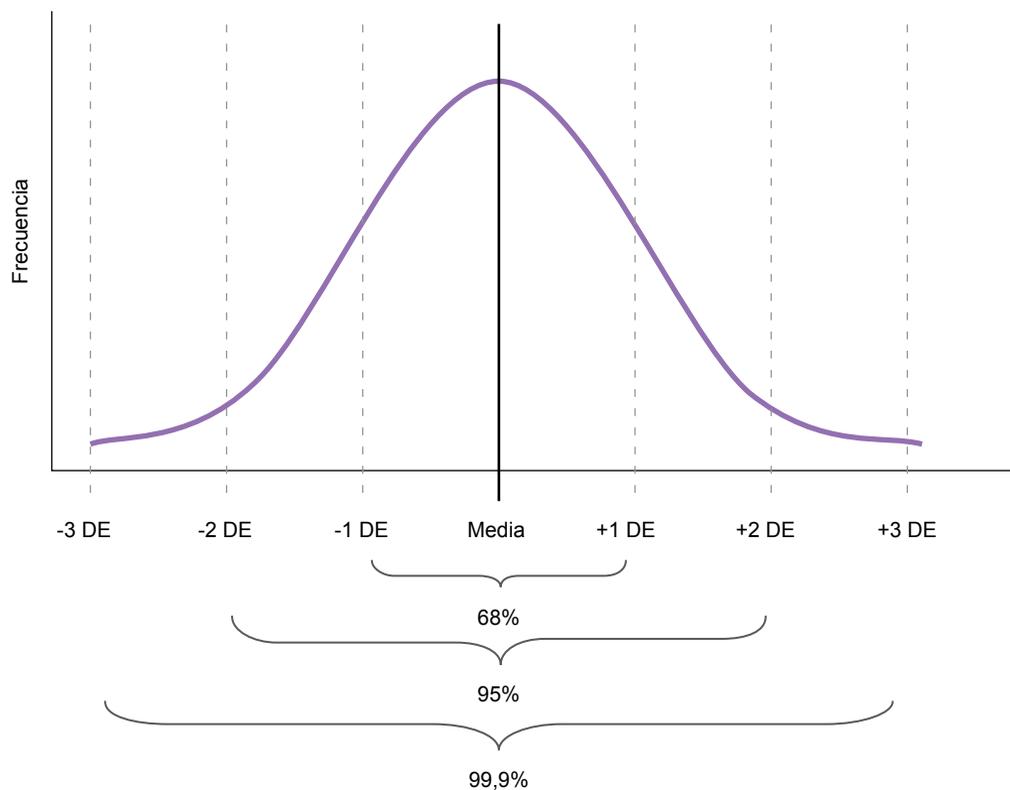
Si calculásemos la mediana de edad de esta serie a partir de los 4.457 casos de suicidio *individualmente*, ésta sería 48 años de edad.

Medidas de dispersión

Para las variables cuantitativas continuas las medidas de dispersión más usadas son básicamente tres: el rango o amplitud, la varianza y la desviación estándar. Estas medidas representan la dispersión o variabilidad de los datos continuos. El **rango** o amplitud es la diferencia entre el valor máximo y el valor mínimo de una serie de datos. La **varianza** (s^2), que mide la desviación promedio de los valores individuales con respecto a la media, es el cociente entre la suma de los cuadrados de la diferencia entre cada valor y el

promedio, y el número de valores observados (menos 1). La **desviación estándar** (DE) es la raíz cuadrada de la varianza. La desviación estándar junto con la media permiten describir la distribución de la variable. Si la variable se distribuye normalmente, entonces el 68% de sus valores estará dentro de ± 1 desviación estándar de la media, 95% dentro de ± 2 y 99,9% dentro de ± 3 DE (Gráfico 3.7). (Es necesario notar que se trata de aproximaciones, ya que, por ejemplo, el área bajo la curva que abarca 95% no es exactamente 2 desviaciones, sino 1.96).

Gráfico 3.7 Distribución normal



Veamos un ejemplo. Volviendo al Cuadro 3.3, el rango del período de incubación de la rubéola, con base en las 11 observaciones, es de 22 días (i.e., 37-15 días). La varianza sería:

$$\text{varianza} = \frac{(15-19,7)^2 + (15-19,7)^2 + \dots + (16-19,7)^2 + \dots + (32-19,7)^2 + (37-19,7)^2}{11-1}$$

$$\text{varianza} = \frac{572,19}{10} = 57,219 \text{ días}^2$$

y la desviación estándar sería:

$$\text{desviación estándar (DE)} = \sqrt{57,219} = 7,56 \text{ días}$$

En resumen, los estimadores de los parámetros de la distribución del período de incubación de la rubéola en los 11 niños del ejemplo serían:

$$\text{media } (\bar{x}) = 19,7 \text{ días}$$

$$\text{desviación estándar (DE)} = 7,6 \text{ días}$$

Cuando nuestros datos sobre una variable continua están agrupados también podemos calcular la varianza y correspondiente desviación estándar aproximadas. Para ello, al igual que para estimar la media en datos agrupados, se parte de la columna con los puntos medios (x) de cada intervalo de clase. Luego generamos tres columnas; una con las diferencias entre el punto medio de cada intervalo de clase y la media de nuestros datos ($x - \bar{x}$) llamada desviación; otra con esta desviación elevada al cuadrado ($(x - \bar{x})^2$) llamada desviación cuadrática y, finalmente, otra columna ($f(x - \bar{x})^2$), resultado de multiplicar la desviación cuadrática por el número de casos (f) del intervalo correspondiente (Cuadro 3.4c). La suma de todos estos productos ($\sum f(x - \bar{x})^2$) dividida entre la suma de casos ($\sum f$) nos da una razonable aproximación a la varianza y extrayendo su raíz cuadrada obtendremos un estimado aceptable de la desviación estándar de nuestros datos agrupados.

Cuadro 3.4c Casos de suicidio. Lugar X, 1995-2000

edad (años)	Casos (f)	punto medio (x)	desviación ($x - \bar{x}$)	desviación cuadrática ($(x - \bar{x})^2$)	producto $f(x - \bar{x})^2$
10 a 14	37	12,5	-33,5	1.124,97	41.623,71
15 a 19	176	17,5	-28,5	814,56	143.362,57
20 a 29	693	25,0	- 21,0	442,70	306.792,87
30 a 39	659	35,0	- 11,0	121,89	80.327,22
40 a 49	784	45,0	- 1,0	1,08	848,79
50 a 59	1.103	55,0	9,0	80,27	88.540,76
60 a 69	1.005	65,0	19,0	359,46	361.260,03
Total	$\Sigma=4.457$				$\Sigma=1.022.755,94$

media (\bar{x}) = 46,0 años

En este ejemplo, la aproximación a la varianza (s^2) sería:

$$s^2 = \frac{\sum f(x - \bar{x})^2}{\sum f} = \frac{1.022.763,25}{4.457} \approx 229,47$$

y la correspondiente desviación estándar (D.E., o “S”) para datos agrupados sería:

$$D.E. = \sqrt{229,47} \approx 15,1 \text{ años}$$

mientras que si utilizáramos en el cálculo los 4.457 casos, uno a uno, obtendríamos una desviación estándar de 15,3 años de edad.

Finalmente, otra forma útil de representar la dispersión de la distribución de una serie de datos es usando **cuantiles**, que son los valores que ocupan una determinada posición en función de la cantidad de partes iguales en que se ha dividido una serie ordenada de datos. Si dividimos nuestra serie en 100 partes iguales, hablamos de *percentiles*; si la dividimos en 10 partes iguales, *deciles*; en cinco partes, *quintiles*; y en cuatro, *cuartiles*. Como ejemplo, el Cuadro 3.5 presenta nueve percentiles de edad correspondientes a la serie de casos de malaria del Cuadro 3.2: el percentil 25 de esta distribución es 33 años, que equivale a decir que 25% de los casos tenían edad menor o igual a 33 años (y, en consecuencia, 75% de los casos eran mayores de 33 años). Nótese que la mediana de la distribución corresponde al percentil 50 (o cuartil 2).

Cuadro 3.5 Percentiles y sus valores en una distribución de casos.

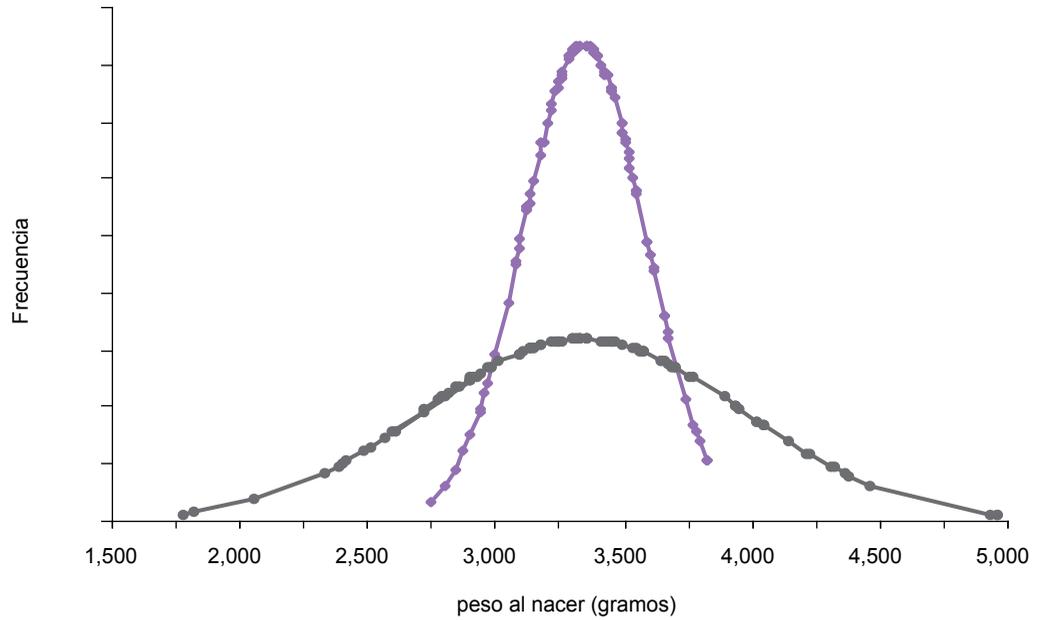
Percentiles	Edad
1%	24
5%	27
10%	29
25%	33
50%	38
75%	44
90%	50
95%	54
99%	63

Como hemos visto, el promedio y la desviación estándar definen la **distribución normal** y, por ello, se les llaman sus *parámetros*. El promedio, como medida resumen de tendencia central de los datos, es un indicador de la precisión de las observaciones. Por su parte, la desviación estándar, como medida resumen de la dispersión de los datos, es un indicador de la *variación* de las observaciones. Estos dos conceptos, precisión y variación, son de gran importancia para documentar la *incertidumbre* con la que observamos los fenómenos en la población y constituyen los principios básicos del proceso de inferencia estadística, cuyo uso nos permite derivar conclusiones acerca de toda la población observando solamente una muestra de la misma. El Gráfico 3.8 ilustra esquemáticamente dos situaciones posibles y frecuentes con relación a la distribución del peso al nacer en la población.

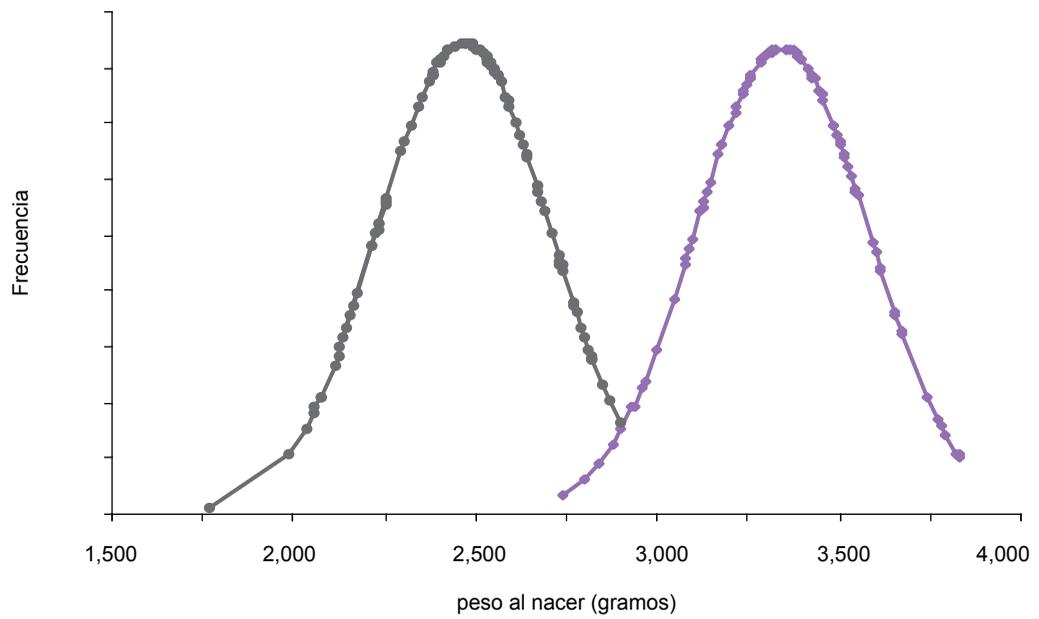
Una vez que se ha revisado la forma de presentación de los datos, las medidas de tendencia central y de dispersión de los mismos, se continúa con el estudio de las distintas distribuciones o modelos teóricos. El más conocido de ellos es la llamada Curva Normal (Gráfico 3.7) que acabamos de introducir. Existen otros modelos teóricos que tienen aplicación en Epidemiología y Salud Pública, pero que van más allá de los objetivos de este taller. Mencionaremos que la Distribución Normal queda definida por 4 elementos característicos:

1. Tiene un eje de simetría.
2. La media aritmética, la mediana y la moda coinciden en el mismo valor por el cual pasa el eje de simetría.
3. La distancia entre el eje de simetría y los puntos de inflexión de la curva equivalen a la desviación estándar.
4. Es asintótica al eje de las “x” (abscisas), es decir nunca lo cruza.

Gráfico 3.8 Curvas de distribución normal del peso al nacer



A. misma precisión, diferente variación (igual promedio, distinta desviación estándar)



B. misma variación, diferente precisión (diferente promedio, igual desviación estándar)

Medidas de frecuencia

Como revisamos en la Unidad 2 del MOPECE, el enfoque epidemiológico parte de la **observación** de hechos en la población. Estos hechos son la presencia de enfermedad, la exposición a determinados factores, u otros atributos o eventos de interés. Son “de interés” porque consideramos que, eventualmente, actuando sobre ellos podremos modificar en forma positiva la realidad observada. El segundo aspecto de este enfoque epidemiológico es la **cuantificación**, es decir, asignar números a los hechos y para ello nos valemos de procedimientos estadísticos básicos como los descritos hasta ahora en esta Unidad. Así, medir la **frecuencia** u *ocurrencia* de tales hechos en la población es una tarea fundamental. Para medir la frecuencia de enfermedad en la población, el epidemiólogo recurre a tres conceptos muy importantes: probabilidad, riesgo y tasa.

En su acepción más sencilla, la **probabilidad** es una medida de ocurrencia de un hecho incierto, o sea, el número de veces que se espera ocurra un evento a futuro. El **riesgo a la salud** es una medida que refleja la *probabilidad* de que se produzca un efecto o daño a la salud dentro de un periodo de tiempo establecido. La **tasa** es una medida de la rapidez de cambio de un estado a otro estado (v.g., de sano a enfermo) y, por tanto, también expresa un *riesgo*. Sin embargo, a diferencia del riesgo, la tasa involucra la experiencia de exposición de la población en unidades efectivas de tiempo-persona

Probabilidad: Número de eventos que ocurren dentro de un número de eventos posibles.

Riesgo a la salud: probabilidad de experimentar un efecto adverso o daño en un tiempo determinado.

Tasa: medida de la rapidez de cambio de un fenómeno dinámico por unidad de población y de tiempo (tiempo-persona de exposición).

La probabilidad es medible o cuantificable y su valor numérico se sitúa entre 0 y 1 (siendo 1 la certeza absoluta). Por conveniencia se puede expresar como un número decimal, una fracción o un porcentaje.

El riesgo es también medible y se obtiene a partir de un cociente o fracción. En sentido estricto, al igual que la probabilidad, su valor se situaría entre 0 y 1. Se suele expresar como un valor de fracción multiplicado por una constante.

La tasa es la medida clásica de la epidemiología. Tiene 3 componentes básicos:

- un *numerador*, que corresponde al número de individuos que experimenta el evento de interés (v.g., muerte, enfermedad, nacimiento, ingreso hospitalario);
- un *denominador*, que corresponde al número total de individuos en la población expuestos o en riesgo de presentar el evento; y,
- un *período de tiempo específico*, durante el cual se observa la frecuencia del evento de interés y la población que ha estado expuesta efectivamente.

De ahí que la tasa pueda asumir un valor de 0 a infinito y se expresa en número de eventos que ocurren en un tiempo-persona de exposición, que es la característica fundamental de la tasa.

Como se puede apreciar, la tasa combina en una sola expresión las tres dimensiones básicas del análisis epidemiológico: persona, lugar y tiempo. Aplicada correctamente, la tasa no solamente es una medida de frecuencia de enfermedad en la población, sino una medida del riesgo de enfermar en la población. De esta forma, al emplear tasas podríamos comparar la frecuencia y riesgo de enfermedad en diferentes poblaciones, diferentes subgrupos de la misma población o diferentes periodos de tiempo en la misma población.

Prevalencia e incidencia

Las clásicas medidas de frecuencia de enfermedad son dos: prevalencia e incidencia, que veremos en esta Unidad. Más específicamente, revisaremos la **incidencia acumulada**, y la **tasa de incidencia**. Aunque de uso común, el término “tasa de prevalencia” debe evitarse, pues no representa una tasa en tanto no es una medida dinámica; una denominación más apropiada es **proporción de prevalencia**, que también revisaremos.

Medidas de morbilidad

En la cuantificación de las condiciones de salud y enfermedad en la población, resulta imprescindible el uso de indicadores que faciliten su descripción y análisis. Se considera que estos estimadores son *convencionales*, pues el consenso científico-técnico y la práctica avalan su utilidad como herramientas relevantes para la interpretación de los perfiles de enfermedad en la población, con fines comparativos.

Para que puedan responder a las necesidades de las poblaciones en forma adecuada, los servicios de salud deben efectuar sistemáticamente mediciones con el objeto de precisar la frecuencia de las condiciones de salud relevantes en la comunidad. Por ejemplo, reconocer que existen 700 personas en la comunidad con diagnóstico de lepra es una información esencial para organizar los recursos existentes y obtener, desde otro nivel del sistema si fuera preciso, el apoyo adicional requerido para la atención de todos los

enfermos. Así, el recuento de los casos de una enfermedad es una medida de gran importancia que sirve para orientar los servicios a la magnitud de los recursos necesarios.

El recuento de nacimientos permite estimar la cantidad de niños menores de 1 año de edad que existen en una comunidad en un año determinado. Esta medida puede ser usada, por ejemplo, para proyectar la cantidad de vacunas necesarias para inmunizar a dicha población en ese año. Esta información podrá tener mayor precisión si, además, se conoce también el número de defunciones en menores de 1 año ocurridas en esa misma comunidad en el año considerado. Es decir, en una población y tiempo específicos, el número de niños a vacunarse puede ser operacionalmente definido como el número de niños sobrevivientes al primer año de vida y se puede estimar por la diferencia entre el número de nacidos vivos y número de defunciones en menores de 1 año de edad.

En consecuencia, la **enumeración o recuento** de casos de enfermedad (así como de nacimientos y defunciones, junto con estimados censales de la población y sus características), constituyen los datos básicos que permiten a los servicios de salud obtener un mejor conocimiento sobre las condiciones de salud y enfermedad en las poblaciones y, por tanto, desempeñarse más eficientemente.

Por otra parte, es prácticamente imposible conocer la magnitud *real* de la **morbilidad** en una población debido a múltiples y complejos factores, que van desde la diversidad de la percepción cultural sobre salud y enfermedad y la presencia de casos subclínicos (tal como hemos revisado en la sección sobre historia natural de la enfermedad, en el Módulo 2) hasta la inaccesibilidad a los servicios de salud y la falta de confianza en la asistencia médica. No obstante, es la *disponibilidad de registros de buena calidad* lo que hace posible obtener estimaciones epidemiológicamente válidas, que nos acercan al conocimiento de la realidad y nos permiten modificarla positivamente.

Consideremos, como ejemplo, un área geográfica determinada en un país de América Latina. En la población se observaron 60 casos de tuberculosis en el año 1995. Un nuevo recuento en el 2000 demostró la existencia de 80 casos en la población.

Cuadro 3.6a Casos de tuberculosis en una comunidad latinoamericana por unidad de tiempo. 1995-2000

año	Nº de casos
1995	60
2000	80

¿Cuál sería la mejor explicación de la diferencia observada entre los dos años?. Los hechos que podrían explicar la diferencia observada se pueden resumir como sigue:

- la atención a los enfermos y las demás medidas de control fueron inadecuadas y por ello se produjo un aumento en el número de casos;
- las medidas de búsqueda activa y/o un cambio en la definición de caso de tuberculosis permitieron una mejor identificación de casos antes desconocidos, dando la impresión de aumento de la enfermedad en la población;
- ciertos factores socioeconómicos, independientes de las medidas de control, provocaron un aumento de la enfermedad (v.g., crisis económica-desempleo-desnutrición); y/o,
- hubo un aumento de la población local, por crecimiento natural o por inmigración neta (v.g., atracción por oferta de empleo en áreas de desarrollo industrial).

Cuadro 3.6b Casos de tuberculosis en una población latinoamericana por unidad de tiempo. 1995-2000

año	Nº de casos	Total población
1995	60	30.000
2000	80	50.000

Observamos que si bien hubo un aumento de 60 a 80 en el número de casos, también aumentó de 30.000 a 50.000 la población en el mismo periodo. Por tanto, lo que deseamos comparar es la diferencia entre 60 casos en 30.000 personas y 80 casos en 50.000. Es decir,

60 casos	vs	80 casos
30.000 habitantes		50.000 habitantes

Un cálculo sencillo nos permitirá la comparación más directa:

- en 1995: $60 / 30.000 = 0,0020$
- en 2000: $80 / 50.000 = 0,0016$

Con el fin de facilitar la comparación, expresando la proporción en números enteros y no decimales, se acostumbra multiplicar el resultado por 100, 1.000, 10.000 o 100.000, según nuestra conveniencia o siguiendo una convención preestablecida. En el ejemplo, si multiplicamos el resultado de la división entre casos y población por 10.000, obtendremos:

- en 1995: 20 casos por 10.000 habitantes
- en 2000: 16 casos por 10.000 habitantes

Ello nos permite indicar que hubo una disminución de la prevalencia de tuberculosis en ese período de tiempo y en esa comunidad latinoamericana.

Cuadro 3.6c Prevalencia de tuberculosis en una comunidad latinoamericana

1995	2000
20 (casos por 10.000 personas)	16 (casos por 10.000 personas)

El cálculo realizado fue el de la *proporción* de prevalencia de la tuberculosis (en el área específica, en los años 1995 y 2000).

Prevalencia: es la medida del número total de casos existentes, llamados casos prevalentes, de una enfermedad en un punto o periodo de tiempo y en una población determinados, sin distinguir si son o no casos nuevos. La prevalencia es un indicador de la magnitud de la presencia de una enfermedad u otro evento de salud en la población.

$$\text{Proporción de Prevalencia de la enf. A} = \frac{\text{N}^\circ \text{ de personas con enf. A en un periodo}}{\text{N}^\circ \text{ total de personas en el mismo periodo}} \times \text{factor}$$

Como ya se mencionó, la prevalencia no puede ser considerada una tasa, pues no toma en cuenta el inicio ni duración de la enfermedad. Su denominador no necesariamente corresponde a la población *en riesgo*, es decir, a aquella población expuesta y susceptible de donde surgen los casos. Sin embargo, la prevalencia es un indicador de gran importancia en salud pública, pues provee una medida del volumen o carga de enfermedad en una comunidad o población en un momento dado –información esencial en la planeación de los servicios de salud.

Desde el punto de vista epidemiológico, hay mayor interés en conocer específicamente cuántos casos *nuevos* de una enfermedad aparecen en una población durante un período de tiempo. Más concretamente, es importante conocer cuántos casos nuevos surgen de una población *que está en riesgo* de padecer una determinada enfermedad o daño a la salud; es decir, un indicador de la rapidez de cambio del proceso dinámico de salud y enfermedad en la población. La medida de ocurrencia de casos nuevos de enfermedad en una población en riesgo en un tiempo determinado se denomina incidencia. En general, la incidencia nos da una idea del riesgo promedio que tienen los individuos en la población de padecer la enfermedad, así como evaluar la eficacia de las acciones de control adoptadas.

Incidencia: es la medida del número de casos nuevos, llamados casos incidentes, de una enfermedad originados de una población en riesgo de padecerla, durante un periodo de tiempo determinado. La incidencia es un indicador de la velocidad de ocurrencia de una enfermedad u otro evento de salud en la población y, en consecuencia, es un estimador del riesgo absoluto de padecerla.

$$\text{incidencia la enf. B} = \frac{\text{número de casos nuevos de la enf. B en un período}}{\text{nº total de personas en riesgo al comienzo del mismo período}} \times \text{factor}$$

En los cálculos de incidencia y de prevalencia siempre es importante dejar bien claro a qué población y a qué momento o período de tiempo se refieren. Pueden relacionarse a la población entera de una región o a un grupo específico que estaría expuesto al problema. Así, el denominador de incidencia de cáncer de cuello de útero debería incluir únicamente a mujeres y el de incidencia de gonorrea a población sexualmente activa.

Por ejemplo, la incidencia de gastroenteritis, en la Provincia del Sur, durante el mes de diciembre de 2001 fue de 20 por mil en niños de 5 a 10 años. Es decir,

$$\begin{array}{l} \text{incidencia de} \\ \text{gastroenteritis en niños} \\ \text{de 5 a 10 años en la} \\ \text{Provincia del Sur en el mes} \\ \text{de diciembre de 2001} \end{array} = \frac{\begin{array}{l} \text{número de niños de 5 a 10 años} \\ \text{que desarrollaron gastroenteritis} \\ \text{en el mes de diciembre de 2001} \end{array}}{\begin{array}{l} \text{población de niños de 5 a 10 años} \\ \text{residentes en la Provincia del Sur} \\ \text{en el mes de diciembre de 2001} \end{array}} \times \text{factor}$$

Conviene precisar que las fórmulas sobre incidencia que acabamos de presentar en este ejemplo corresponden específicamente a lo que se denomina **incidencia acumulada**. Al calcularse como el cociente entre el número de casos nuevos y el tamaño de la población en riesgo en un periodo de tiempo, la incidencia acumulada asume que *todos* los individuos de la población en riesgo estuvieron *efectivamente* en riesgo de presentar la enfermedad durante *todo* el periodo de tiempo observado. Intuitivamente sabemos que esto raras veces ocurre así; en principio porque al momento de presentar la enfermedad, la persona deja de estar en riesgo (i.e., deja de “pertenecer al denominador”), pues se convierte en caso (i.e., “pasa al numerador”).

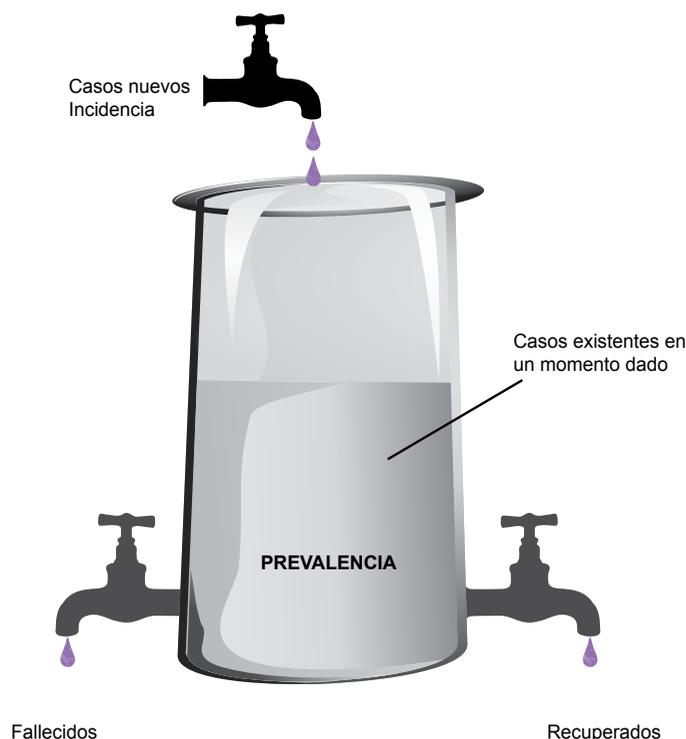
Una manera más precisa de representar el riesgo de pasar del estado sano al estado enfermo, o sea estimar la incidencia, la brinda la investigación epidemiológica de factores de riesgo causales en cohortes de población, como veremos en la Unidad 5. En tales condiciones es posible “seguir” a la población, midiendo con exactitud el tiempo que cada individuo estuvo en riesgo de enfermar y obtener el número total de “tiempo-persona” de observación (v.g., años-persona; días-persona). Si ahora dividimos el número de ca-

Los nuevos entre el número total de años-persona observados, obtendremos la **tasa de incidencia** propiamente dicha, llamada también *densidad de incidencia* o *fuerza de morbilidad* (Miettinen, 1975).

En términos prácticos, cuando calculamos la incidencia de una enfermedad en la población a menudo empleamos la incidencia acumulada, pues el cálculo de la tasa de incidencia propiamente dicha suele estar restringido al contexto de una investigación epidemiológica. Una forma de estimar la tasa de incidencia cuando no se cuenta con los años-persona consiste en tomar como denominador a la población a mitad de período.

Tanto la prevalencia como la incidencia son medidas de morbilidad (enfermedad) en la población, pero difieren en que la prevalencia mide el número de personas que tienen la enfermedad en un momento dado (i.e., su magnitud) y la incidencia mide los casos nuevos que se presentan en un período determinado de tiempo (i.e., su velocidad). Las relaciones entre incidencia y prevalencia pueden apreciarse en la siguiente Figura:

Figura 3.1 Relación entre incidencia y prevalencia



Note que si aumenta la incidencia (aparecen más casos nuevos) y el número de muertes y recuperados se mantiene sin cambio, aumentará la prevalencia. Si aumenta la mortalidad o más gente se recupera y la incidencia no cambia, la prevalencia disminuirá.

Supongamos que se introduce una nueva prueba que detecta la presencia de enfermedad tempranamente en el período subclínico; el resultado práctico será un aumento en la incidencia, en la duración de la enfermedad y también en la prevalencia. Por otra parte, si se introduce un medicamento que pospone o evita la mortalidad prematura pero no cura definitivamente, el resultado también será un aumento en la prevalencia.

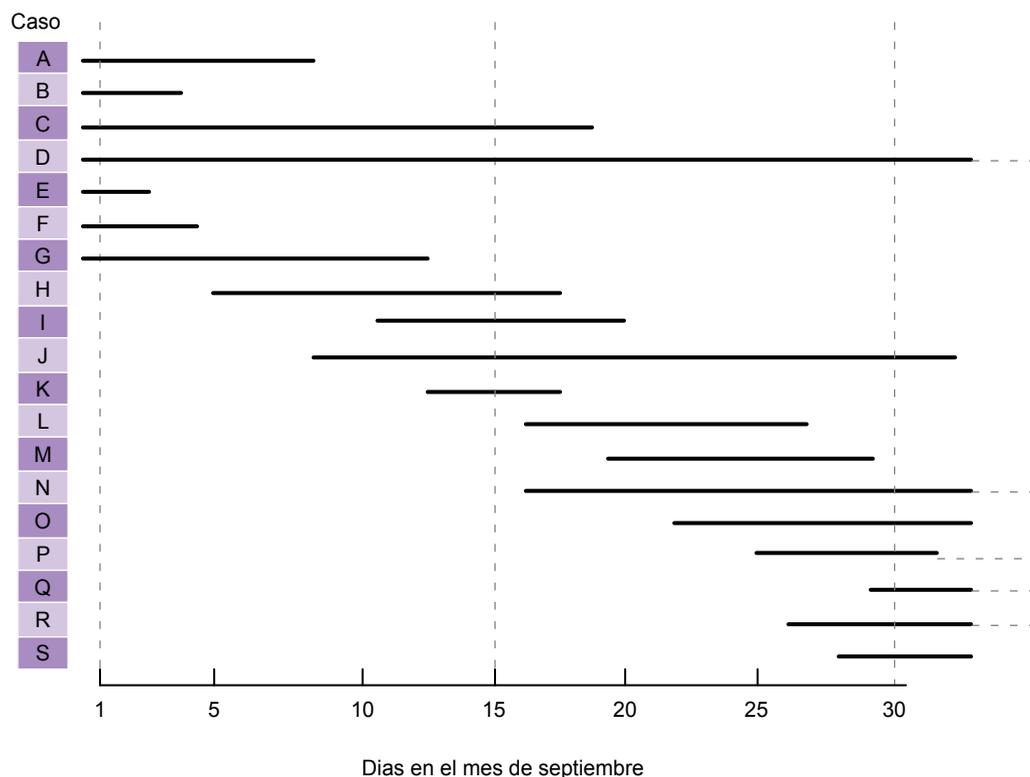
La duración de la enfermedad (D) también influye en esta dinámica y, en términos generales, se dice que en situación de equilibrio, la prevalencia (P) es el producto de la incidencia (I) por la duración de la enfermedad ($P=I \times D$). Esta dinámica entre incidencia, prevalencia y duración de la enfermedad tiene importantes implicaciones para el control de enfermedades en la población como veremos en el Módulo 6 del MOPECE.



Ejercicio 3.1

En la Figura 3.2, cada línea representa un caso de enfermedad respiratoria (neumonía) y la duración en días de cada caso, que se presentaron durante el mes de septiembre.

Figura 3.2 Casos de enfermedad respiratoria en el mes de septiembre



Conteste las siguientes preguntas:

1. ¿Cuál es el número de casos incidentes de la enfermedad en el mes de septiembre?
2. ¿Cuál es el número de casos prevalentes en el día 15 de septiembre?

Otras Medidas de morbilidad

Ahora bien, ¿Qué medida de incidencia se obtiene entonces cuando se investiga una epidemia?. El numerador es el número de casos nuevos de enfermedad y el denominador es el total de personas expuestas al factor de riesgo o agente causal. El tiempo generalmente se trata de manera implícita, ya que la mayoría de casos suelen ocurrir durante horas,

días o semanas, según la enfermedad de que se trate. La tasa de incidencia que se obtiene en una situación de brote o epidemia se denomina **tasa de ataque** de la enfermedad y se expresa usualmente como un porcentaje. Consideremos, como ejemplo, una situación en la que 96 personas fueron expuestas a un agente (v.g., *Pseudomonas aeruginosa* contaminando el equipo quirúrgico), de las cuales 26 se enfermaron en un período corto de tiempo. La tasa de ataque se obtiene de la siguiente manera:

$$\text{tasa de ataque} = \frac{26 \text{ enfermos}}{96 \text{ expuestos}} \times 100 = 27,1 \%$$

Existe también la llamada tasa de ataque secundario, que mide la contagiosidad de una enfermedad de transmisión persona-a-persona y, por tanto, es de utilidad para evaluar la efectividad de las medidas de control de un brote como veremos en el Módulo 5. La tasa de ataque secundario expresa el número de casos de una enfermedad que aparecen dentro del periodo de incubación entre los contactos susceptibles expuestos a un caso primario o índice, en relación con el número total de contactos susceptibles expuestos. Se calcula de la manera siguiente:

$$\text{tasa de ataque secundario} = \frac{\text{número de casos secundarios}}{\text{número de contactos susceptibles expuestos}} \times 100$$

Medidas de mortalidad

Otra forma importante de medir la ocurrencia de enfermedad en la población es a través de la **tasa de mortalidad**, un estimador del riesgo absoluto de morir.

$$\text{tasa general de mortalidad} = \frac{\text{número de defunciones}}{\text{población total}} \times 1.000$$

Las tasas de mortalidad pueden referirse a toda la población de un país o territorio o restringirse a una comunidad, institución o una muestra poblacional y pueden, también, calcularse para grupos específicos de población, según sexo, edad, grupos de enfermedades u otras características relevantes (en cuyo caso constituyen tasas específicas).

Existen algunos indicadores de mortalidad referidos a grupos de población específicos y a los que convencionalmente se les llama “tasas”. En particular, conviene recordar dos de ellos, de especial importancia en salud pública; se trata de las “tasas” de mortalidad infantil y de mortalidad materna.

$$\text{“tasa” de mortalidad infantil} = \frac{\text{número de defunciones en menores de 1 año de edad}}{\text{número de nacidos vivos}} \times 1.000$$

$$\text{“tasa” de mortalidad materna} = \frac{\text{número de muertes maternas}}{\text{número de nacidos vivos}} \times 100.000$$

Se puede constatar que para el cálculo de estas “tasas” se emplea el número de nacidos vivos como denominador y, por tanto en sentido estricto no es una tasa en tanto el denominador no es tiempo-persona. Por convención, se utiliza el número de nacidos vivos como denominador de estos indicadores porque se considera que es un dato más factible de obtener que los requeridos para construir las respectivas tasas, a saber: en el primer caso, el total de niños menores de un año que deben existir en la misma población y año donde se hizo el recuento de defunciones; en el segundo caso, el número total de mujeres que estuvieron embarazadas y que dieron a luz en la misma población y año.

En ocasiones se sospecha que una enfermedad determinada está causando un elevado o inusual número de muertes. Es de interés conocer cuántas de las personas enfermas mueren, es decir, la *proporción* de casos fatales entre el total de casos. En una situación epidémica, a este cálculo se le llama **tasa de letalidad** y es particularmente importante para evaluar la severidad de una epidemia:

$$\text{tasa de letalidad de la enfermedad A} = \frac{\text{número de defunciones por enfermedad A}}{\text{total de casos de enfermedad A}} \times 100$$

Mientras la mortalidad hace referencia a las defunciones entre la población total (sana o enferma), la letalidad sólo hace referencia a las defunciones entre la población enferma. Por ejemplo, si en un distrito de 30.000 habitantes ocurrieron 200 casos de tifoidea con 6 defunciones, en un año determinado, podemos afirmar que, en ese distrito y en ese año, la mortalidad por tifoidea fue 2 por 10.000 y su letalidad 3%; es decir:

$$\text{mortalidad por tifoidea} = \frac{6 \text{ muertes}}{300.000 \text{ habitantes}} \times 10.000 = 2 \times 10.000$$

$$\text{letalidad por tifoidea} = \frac{6 \text{ muertes}}{200 \text{ casos}} \times 100 = 3 \times 100$$

Distribución proporcional

La distribución proporcional es una forma sencilla de expresar la cantidad de casos o muertes según alguna característica de interés, como sexo, edad o causa específica, como porcentaje del total de casos o muertes observados. La distribución proporcional corresponde a una distribución de frecuencias relativas simples, como se vio al inicio de esta Unidad. A diferencia de las tasas, la distribución proporcional no mide el riesgo de

enfermar o morir, sino solamente indica cómo se distribuyen los casos o muertes entre las personas afectadas.

Veamos el siguiente ejemplo. Durante un brote de tifus exantemático ocurrido en una comunidad andina, se enfermaron 38 personas; 24 hombres y 14 mujeres. Se desconoce el número total de hombres y mujeres en dicha población.

Cuadro 3.7 Distribución proporcional de casos de tifus exantemático por sexo

Sexo	Número de casos	distribución proporcional
Masculino	24	63,2
Femenino	14	36,8
Total	38	100,0

Es decir,

$$\text{proporción de casos de sexo masculino: } \frac{24}{38} \times 100 = 63,2\%$$

$$\text{proporción de casos de sexo femenino: } \frac{14}{38} \times 100 = 36,8\%$$

Esta información nos permite afirmar que la enfermedad en dicha comunidad afectó más a los varones que a las mujeres (por ejemplo, “de cada 10 enfermos, al menos 6 fueron varones”; o, “existieron 1,7 casos en varones por cada caso en mujeres”). Sin embargo, esta información no nos permite afirmar que los varones tuvieran mayor riesgo de enfermar.

mortalidad proporcional

Cuando la distribución proporcional se refiere a datos de mortalidad, entonces hablamos específicamente de **mortalidad proporcional**, una medida de mortalidad útil para describir el perfil y las contribuciones de causas de muerte específicas a la mortalidad general de un lugar y periodo determinados; esto es:

$$\text{mortalidad proporcional} = \frac{\text{n}^\circ \text{ de muertes por causa específica en un año}}{\text{n}^\circ \text{ total de muertes en el mismo año}} \times 100$$

Las medidas de prevalencia, incidencia y mortalidad, tomadas en conjunto, permiten describir **perfiles de salud** de la población, útiles para el **análisis de la situación de salud**, la primera función esencial de la salud pública. Estas medidas básicas permiten

comparar la importancia relativa de riesgos, enfermedades y daños a la salud en las poblaciones entre diferentes períodos de tiempo y entre localidades distintas, facilitando la identificación de necesidades y prioridades en salud y orientando la gestión sanitaria. No obstante, cuando comparemos dos o más poblaciones debemos tener siempre presente que existen otras características propias de cada población que podrían explicar las diferencias que observamos. Por ejemplo, la diferencia entre las tasas de mortalidad de dos poblaciones puede estar distorsionada por las propias diferencias en la distribución de edad o la distribución de las muertes en distintas edades en cada población y, por tanto, llevarnos a conclusiones erróneas. Ello puede ser evitado con el uso de técnicas de *estandarización o ajuste de tasas* (los interesados en este tema pueden revisar el Apéndice Estadístico-Epidemiológico al final de este módulo).



Ejercicio 3.2

Los siguientes datos corresponden a dos localidades latinoamericanas, a las que denominaremos A y B, en un año determinado.

Cuadro 3.8 Población y mortalidad en dos localidades latinoamericanas

Datos básicos	Localidad	
	A	B
población total	10.320	76.311
número de nacidos vivos	850	1.226
población menor de 5 años	3.350	6.901
defunciones en menores de 1 año	105	10
defunciones en menores de 1 año por infecciones	32	4
defunciones en menores de 5 años	161	12
defunciones en menores de 5 años por diarrea	36	0

Calcule:

Indicadores de Salud	Localidad	
	A	B
a) la tasa de mortalidad en menores de 5 años (por 1.000)		
b) la tasa de mortalidad infantil (por 1.000)		
c) la tasa de mortalidad por diarrea en menores de 5 años (por 1.000)		
d) la tasa de mortalidad por infecciones en menores de 1 año (por 1.000)		
e) ¿Qué se observa al comparar las localidades A y B?	<hr/> <hr/> <hr/>	



Ejercicio 3.3

Pregunta 1 ¿Qué denominador emplearía para el cálculo de las siguientes medidas?:

a) la incidencia acumulada de dengue hemorrágico en el país “X” en 2000.

b) la incidencia acumulada de neumonía en los escolares de la Provincia “Y” en 2001.

Pregunta 2 En una isla del Caribe la letalidad por tifoidea es 5%. Durante el año 2000 ocurrieron 40 defunciones por esa enfermedad. ¿Cuántos casos de tifoidea ocurrieron en esa comunidad insular en dicho año?.

Pregunta 3 Una medida de incidencia usualmente expresada en porcentaje y que se refiere a poblaciones específicas, en períodos de tiempo limitados, como por ejemplo, epidemias, se conoce como:

- a) prevalencia
- b) tasa ajustada
- c) tasa de letalidad
- d) tasa de ataque
- e) tasa de mortalidad

Pregunta 4 Analice el siguiente cuadro y, de acuerdo con la información presentada, resuelva los puntos considerados a continuación:

Cuadro 3.9 Defunciones por grupo de edad y población

grupo de edad (años)	defunciones (número)	población
Menores de 1	285	12.681
1 – 4	251	49.002
5 – 24	274	201.820
25 – 44	408	116.538
45 – 64	576	51.356
65 y más	1.076	18.603
Total	2.870	450.000

a) la tasa de mortalidad (por 100.000) para el grupo de edad de 1 a 4 años

b) la tasa de mortalidad (por 100.000) para el grupo de edad de 65 y más años.

c) la tasa de mortalidad (por 100.000) para el grupo de 25 a 44 años.

d) Comente estos resultados con respecto a la magnitud de la mortalidad en los grupos de edad. Considere que estos resultados son típicos de cualquier población.

Pregunta 5 Veintiséis casos de tuberculosis fueron diagnosticados en Ciudad Alta entre el 1° de enero y el 30 de junio de 2000. El total de casos de tuberculosis activos al 30 de junio era 264. La población de Ciudad Alta era de 183.000 habitantes.

a) ¿Cuál fue la incidencia de tuberculosis en Ciudad Alta durante el período 1° de Enero al 30 de Junio?

- (a) 7,6 casos nuevos por 100.000 habitantes
- (b) 14,2 casos nuevos por 100.000 habitantes
- (c) 27,3 casos nuevos por 10.000 habitantes
- (d) 78,7 casos nuevos por 100.000 habitantes
- (e) 144,3 casos nuevos por 10.000 habitantes

b) ¿Cuál fue la prevalencia de tuberculosis en Ciudad Alta al 30 de junio de 2000?

- (a) 14,2 casos por 100.000 habitantes
- (b) 144,3 casos por 100.000 habitantes
- (c) 158,5 casos por 10.000 habitantes
- (d) 290,0 casos por 10.000 habitantes
- (e) 85,2 casos por 100.000 habitantes

c) Describa y explique las diferencias en los valores obtenidos sobre prevalencia e incidencia de tuberculosis encontrados e indique el uso adecuado de cada una de estas medidas epidemiológicas ¿Existe similitud de esta morbilidad por tuberculosis con la de su área de salud?

Pregunta 6 El tercer domingo de febrero de 2001 se celebró, en una comunidad de 462 habitantes, un baile de carnaval al que asistieron 287 personas. En las dos últimas semanas de marzo el centro de salud local atendió a 79 personas que consultaron por prurito intenso y erupción cutánea papulovesicular. Setenta y seis de ellas habían asistido al baile. De los 161 varones participantes de ese evento social, 53 se habían enfermado. El 68,3% de los asistentes al baile tenía entre 20 y 39 años de edad; en este grupo se encontraba el 82,9% de los enfermos. Calcule:

- a) la tasa de ataque en los participantes de la fiesta;
- b) la tasa de ataque en los hombres;
- c) la tasa de ataque en las mujeres;
- d) la tasa de ataque en el grupo de 20 a 39 años de edad.
- e) ¿Qué interpretación daría a estos resultados?

Pregunta 7 Examine el siguiente cuadro y resuelva lo considerado a continuación.

Cuadro 3.10 Número de casos por grupo de edad y población

Columna 1 edad (años)	Columna 2 casos (N°)	Columna 3 población	Columna 4	Columna 5
Menores de 5	3	48	14,3	6,3
5 – 19	6	17	28,6	35,3
20 – 39	5	23	23,8	21,7
40 y más	7	109	33,3	6,4
Total	21	197	100,0	10,7

a) los números en la Columna 4 representan:

- (a) la distribución proporcional de la población por edad
- (b) las tasas de ataque por edad
- (c) la distribución proporcional de los casos por edad
- (d) las tasas de mortalidad por edad
- (e) nada de lo anterior

b) los números en la Columna 5 representan:

- (a) la distribución proporcional de la población por edad
- (b) las tasas de ataque por edad
- (c) la distribución proporcional de los casos por edad
- (d) las tasas de mortalidad por edad
- (e) nada de lo anterior

c) Comente brevemente la información presentada en el Cuadro 3.10

Pregunta 8 El programa de atención integrada de enfermedades prevalentes en la infancia (AIEPI) de un centro de salud, informó que el tiempo de duración de todos los casos de neumonía en niños de 5 a 9 años de edad atendidos

en los últimos dos meses fue 9, 7, 11, 9, 8, 4, 6, 12, 6, 8, 8 y 5 días, respectivamente. Usando esta información, calcule:

a) la moda;

b) la mediana;

c) la media;

d) el rango;

e) la desviación estándar.

f) En el espacio provisto a continuación represente la distribución de la variable de interés empleando un gráfico apropiado.

g) Comente los resultados observados con respecto a esta distribución.



Ejercicio 3.4

El nivel intermedio del sistema de salud de Nicaragua, en el istmo centroamericano, comprende 17 Sistemas Locales de Atención Integral de Salud (SILAIS), que corresponden a los 15 departamentos y 2 regiones autónomas en que se divide geopolíticamente el país. A inicios de 1999 ocurrió un brote de rubéola de alcance nacional, poniéndose en operación un sistema de vigilancia activa intensificada que permitió establecer medidas oportunas para el control de la enfermedad. En este ejercicio se describe la situación observada en la SILAIS León, ubicada en la región occidental del país y que comprende 10 municipios. El Cuadro 3.11 contiene datos sobre los 130 casos confirmados de rubéola ocurridos en León entre las semanas epidemiológicas (S.E.) 4 y 33 de 1999. Los Cuadros 3.12a, 3.12b y 3.13 presentan datos adicionales relevantes para la caracterización de la situación epidemiológica.

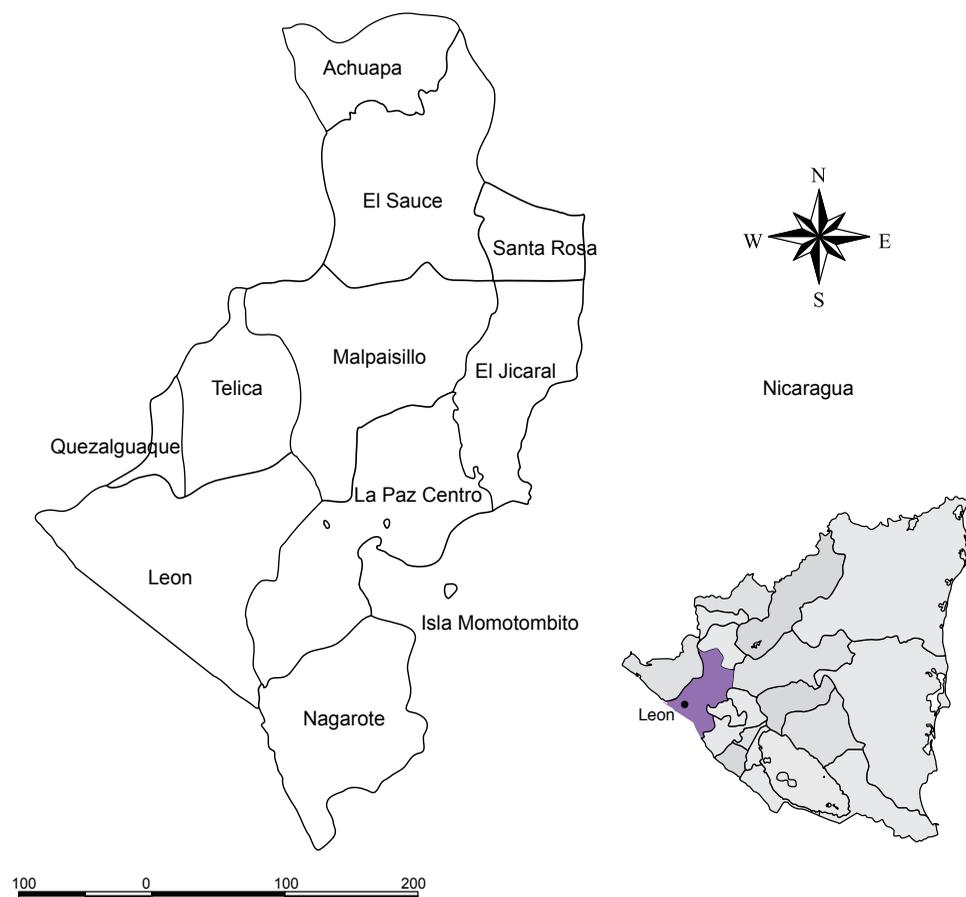
Pregunta 1 ¿Cuales fueron las características epidemiológicas más relevantes del brote de rubéola en la SILAIS León, Nicaragua, en 1999?. Caracterice la epidemia (describa las características de la epidemia) según:

- a) edad y sexo de los afectados;
- b) su distribución geográfica; y,
- c) su distribución en el tiempo.

Para ello:

- complete los Cuadros 3.12a, 3.12b y 3.13;
- calcule las tasas correspondientes usando los cuadros y mapa anexos; y,
- complete el Gráfico 3.9.

Figura 3.3 Mapa del SILAIS León, Nicaragua



Fuente: Programa Especial de Análisis de Salud (SHA) y Programa Ampliado de inmunizaciones (HVI), OPS

Cuadro 3.11 Ocurrencia de casos de rubéola; SILAIS León, 1999

N°	Sexo	edad (años)	Municipio	fecha de inicio	S.E.*
1	F	7	León	26-Jan	4
2	F	6	León	31-Jan	5
3	F	28	León	5-Feb	5
4	F	15	León	6-Feb	5
5	F	28	León	7-Feb	6
6	F	4	León	11-Feb	6
7	F	13	León	16-Feb	7
8	F	4	León	21-Feb	8
9	F	11	Nagarote	8-Mar	10
10	M	5	León	11-Mar	10
11	F	28	León	13-Mar	10
12	F	4	León	14-Mar	11
13	F	29	León	18-Mar	11
14	M	1	León	20-Mar	11
15	M	24	León	21-Mar	12
16	F	4	León	22-Mar	12
17	F	11	León	22-Mar	12
18	M	24	Quezalaguaque	23-Mar	12
19	F	9	León	25-Mar	12
20	F	3	León	29-Mar	13
21	M	4	León	29-Mar	13
22	F	8	León	4-Abr	14
23	M	3	León	6-Abr	14
24	F	34	Telica	6-Abr	14
25	M	7	León	8-Abr	14
26	F	20	Telica	8-Abr	14
27	F	6	León	9-Abr	14
28	F	7	León	9-Abr	14
29	F	10	León	10-Abr	14
30	M	13	León	11-Abr	15
31	F	9	León	12-Abr	15
32	M	7	León	13-Abr	15
33	F	6	León	14-Abr	15
34	M	12	Telica	14-Abr	15
35	F	10	León	16-Abr	15
36	F	24	León	16-Abr	15
37	F	4	León	17-Abr	15
38	M	5	León	17-Abr	15
39	F	3	Telica	18-Abr	16
40	M	3	Malpaisillo	18-Abr	16
41	M	7	León	19-Abr	16
42	F	12	Malpaisillo	19-Abr	16

N°	Sexo	edad (años)	Municipio	fecha de inicio	S.E.*
43	F	26	León	19-Abr	16
44	F	12	León	20-Abr	16
45	M	33	Malpaisillo	20-Abr	16
46	F	11	León	21-Abr	16
47	F	17	Achuapa	21-Abr	16
48	F	5	Telica	22-Abr	16
49	F	12	León	22-Abr	16
50	F	21	Malpaisillo	22-Abr	16
51	M	6	León	23-Abr	16
52	F	10	León	23-Abr	16
53	F	10	Malpaisillo	25-Abr	17
54	M	10	Telica	26-Abr	17
55	F	30	León	26-Abr	17
56	M	11	Malpaisillo	27-Abr	17
57	F	15	León	27-Abr	17
58	M	10	Malpaisillo	1-May	17
59	F	9	Achuapa	2-May	18
60	F	10	Achuapa	2-May	18
61	F	6	León	3-May	18
62	F	26	Telica	4-May	18
63	F	17	Malpaisillo	7-May	18
64	F	19	Achuapa	8-May	18
65	F	16	León	10-May	19
66	F	11	León	11-May	19
67	F	22	León	11-May	19
68	F	29	León	14-May	19
69	M	22	León	16-May	20
70	M	9	León	17-May	20
71	M	7	León	19-May	20
72	M	5	León	20-May	20
73	M	22	León	20-May	20
74	M	4	León	21-May	20
75	F	5	León	21-May	20
76	F	20	León	22-May	20
77	M	1	León	24-May	21
78	F	7	Nagarote	24-May	21
79	M	1	León	26-May	21
80	M	4	Malpaisillo	26-May	21
81	M	13	León	27-May	21
82	M	5	León	28-May	21
83	F	17	León	28-May	21
84	F	4	León	30-May	22
85	F	15	León	1-Jun	22
86	F	5	Nagarote	2-Jun	22
87	F	17	Telica	4-Jun	22

Módulos de principios de epidemiología para el control de enfermedades (MOPECE)

N°	Sexo	edad (años)	Municipio	fecha de inicio	S.E.*
88	M	4	León	5-Jun	22
89	F	20	León	6-Jun	23
90	F	5	León	7-Jun	23
91	M	8	León	10-Jun	23
92	F	10	León	11-Jun	23
93	F	10	León	12-Jun	23
94	F	9	León	15-Jun	24
95	F	5	León	16-Jun	24
96	F	31	León	18-Jun	24
97	F	36	León	18-Jun	24
98	M	8	El Jicaral	20-Jun	25
99	F	12	León	21-Jun	25
100	F	8	León	22-Jun	25
101	M	10	El Jicaral	22-Jun	25
102	M	10	León	22-Jun	25
103	F	6	León	24-Jun	25
104	M	13	El Jicaral	24-Jun	25
105	M	15	El Jicaral	25-Jun	25
106	F	15	Santa Rosa	29-Jun	26
107	F	4	León	30-Jun	26
108	F	8	León	1-Jul	26
109	M	7	Santa Rosa	2-Jul	26
110	F	9	León	3-Jul	26
111	F	10	Malpaisillo	3-Jul	26
112	F	11	Malpaisillo	5-Jul	27
113	F	7	Telica	7-Jul	27
114	F	14	León	7-Jul	27
115	F	9	Nagarote	11-Jul	28
116	M	28	León	11-Jul	28
117	M	14	León	12-Jul	28
118	F	30	El Sauce	12-Jul	28
119	F	30	León	13-Jul	28
120	F	4	León	15-Jul	28
121	F	6	La Paz Centro	17-Jul	28
122	F	15	León	26-Jul	30
123	M	6	León	27-Jul	30
124	F	13	León	29-Jul	30
125	F	6	El Jicaral	30-Jul	30
126	F	7	Malpaisillo	3-Ago	31
127	F	12	Nagarote	8-Ago	32
128	F	28	Malpaisillo	8-Ago	32
129	F	7	León	11-Ago	32
130	F	26	La Paz Centro	16-Ago	33

Cuadro 3.12a Rubéola en León, Nicaragua; 1999. Tasa de ataque (TA) por 100.000 habitantes por edad

edad (años)	Casos	población	TA
0 - 4		61.932	
5 - 9		52.555	
10 -14		46.940	
15 - 19		43.285	
20 - 24		34.373	
25 - 29		28.097	
30 - 34		23.419	
35 - 39		19.318	
40 y más		64.602	
Total		374.521	

Cuadro 3.12b Rubéola en León, Nicaragua; 1999. Tasa de ataque (TA) por 100.000 habitantes por edad y sexo

edad (años)	varones			mujeres		
	casos	población	TA	casos	población	TA
0 - 4		29.809			32.123	
5 - 9		25.539			27.016	
10 -14		22.886			24.054	
15 - 19		20.001			23.284	
20 - 24		15.770			18.603	
25 - 29		12.693			15.404	
30 - 34		10.385			13.034	
35 - 39		8.654			10.664	
40 y más		27.347			37.255	
Total		173.084			201.437	

Cuadro 3.13 Rubéola en León, Nicaragua; 1999. Tasa de Ataque (TA) por 100.000 habitantes por Municipio

Municipio	Casos	población	TA
Achuapa		14.681	
El Jicaral		11.174	
El Sauce		28.917	
La Paz Centro		30.627	
León		180.352	
Malpaisillo		32.025	
Nagarote		32.510	
Quezalguaque		8.710	
Santa Rosa		10.164	
Telica		25.361	
Total		374.521	

Gráfico 3.9 Rubéola en León, Nicaragua; 1999. Distribución de casos por fecha de inicio



□ = 1 caso

Pregunta 2 Establezca cuáles serían las recomendaciones epidemiológicas al equipo de salud del SILAIS de León.

Comparación de la frecuencia de enfermedad y medidas de asociación

Ahora que hemos revisado las principales medidas de frecuencia y distribución de los fenómenos en tiempo, espacio y persona, el siguiente paso del enfoque epidemiológico es la comparación de dichas medidas. Esta comparación es la estrategia básica del análisis epidemiológico y el paso fundamental para transformar los datos en información relevante.

En todo acto de comparación hay una intencionalidad analítica subyacente: encontrar igualdades o, alternativamente, encontrar diferencias. Esta capacidad de discriminar entre las observaciones que se hace de la realidad es un requisito para identificar las necesidades en salud de la población, establecer prioridades y, consecuentemente, desarrollar acciones en salud dirigidas a modificarlas positivamente. Este es el enfoque no-igualitario que pone en práctica la epidemiología.

Un aspecto central en este raciocinio analítico de la epidemiología es la generación de explicaciones tentativas, suposiciones, pronósticos o conjeturas verosímiles respecto a una relación causa-efecto que encierran incertidumbre; es decir, la **generación de hipótesis**. Una hipótesis es una suposición que se hace de una observación o reflexión, que lleva a predicciones refutables. Para la aplicación de pruebas estadísticas a los datos obtenidos, se requiere la elaboración de dos hipótesis: la hipótesis alterna o de investigación (HA) y la hipótesis nula o de no-diferencia (HO), sobre las cuales se enfocaran los resultados del análisis estadístico. La obtención de datos relevantes y la comparación racional de los mismos es la forma de contrastar nuestras hipótesis sobre la salud y la enfermedad en la población.

Medidas de asociación

En esta sección revisaremos los principios y métodos estadísticos básicos que ayudan a los equipos locales de salud a establecer la presencia de una **asociación** entre la exposición a un factor que se considera de riesgo y la ocurrencia de enfermedad en la población. En términos estadísticos, se dice que dos variables están asociadas cuando existe una relación de dependencia (usualmente de tipo matemático) entre ambas; es decir, el cambio en una de ellas necesariamente se acompaña del cambio en la otra (covariación).

Asociación: relación de dependencia estadística entre dos o más eventos, características u otras variables. Una asociación está presente si la probabilidad de ocurrencia de un evento depende de la ocurrencia de otro u otros.

En epidemiología, la aplicación del término ‘asociación’ siempre implica la intención de establecer una **relación de causa a efecto** entre una exposición y una enfermedad o evento en salud. Sin embargo, debe tenerse presente que una asociación puede ser fortuita o ‘espuria’ o puede ser producida por varias circunstancias y, por tanto, la presencia de una asociación estadística no necesariamente implica una relación causal.

Por otra parte, un **factor de riesgo** es un aspecto del comportamiento o estilo de vida personal, constitución genética o hereditaria o exposición ambiental que, con base en la evidencia epidemiológica disponible, se sabe que está asociado a condiciones relacionadas con la salud consideradas importantes de prevenir. Desde el punto de vista epidemiológico, lo más importante de un factor de riesgo es que sea identificable, cuantificable y, en lo posible, modificable *antes* de la ocurrencia del hecho que predice. Los factores de riesgo son indicadores o marcadores del riesgo de enfermar en la población aunque, debe tenerse presente que el hallazgo de un factor de riesgo no necesariamente implica que sea un factor causal.

Factor de riesgo: característica o circunstancia detectable en individuos o grupos, asociada con una probabilidad incrementada de experimentar un daño o efecto adverso a la salud. En general, un factor de riesgo es un atributo o exposición que incrementa la probabilidad de ocurrencia de una enfermedad u otro daño a la salud.

Los aspectos que hemos revisado en la primera mitad de esta Unidad corresponden al campo de la llamada **estadística descriptiva**: la *descripción* cuantitativa de la frecuencia y distribución de los fenómenos de salud y enfermedad observados en la población. La utilidad de presentar los datos resumidos en buenos cuadros y gráficos radica en permitir hacer *inferencias* sobre los determinantes de la situación de salud observada en la población y proceder con cierto nivel de confianza a intervenir sobre ellos.

La estadística es la ciencia matemática que proporciona a la epidemiología los elementos básicos para el tratamiento científico de la información numérica en salud. El análisis epidemiológico usa procedimientos estadísticos diseñados para estimar la probabilidad de que una conclusión acerca de la población de referencia, basada en el análisis de datos de una porción de ella (muestra) o en la comparación con otra población, sea correcta o verosímil; ésto corresponde a la llamada **estadística inferencial**.

En la práctica, para explorar una posible asociación entre exposición y enfermedad se requiere de tres elementos: dos grupos comparables de la población, una medida de la variable exposición para cada grupo y una medida de la variable enfermedad en cada grupo. En general, las variables epidemiológicas de exposición y de enfermedad son continuas o discretas y sus medidas se resumirán en **promedios** o en **proporciones**. La situación más común en los servicios de salud es la comparación de dos proporciones. Por ejemplo, para evaluar la asociación entre ingreso económico y tuberculosis, podríamos comparar dos proporciones: incidencia de tuberculosis entre los pobres y los ricos.

Las medidas de asociación estadística se basan en las llamadas **pruebas de significancia** (los aspectos teóricos relacionados con sus bases conceptuales escapan a los propósitos de este material. Para una breve introducción a ellos vea el Apéndice al final de este módulo). El propósito de estas pruebas es determinar si la presencia de un factor de riesgo evaluado está efectivamente relacionada con la frecuencia de la enfermedad. En dichas condiciones se espera que la prevalencia de exposición a dicho factor sea razonablemente más alta entre los que han enfermado o sufrido un daño a la salud que en aquellos aparentemente sanos.

La bioestadística posee una gran variedad de pruebas de significancia y otros recursos analíticos de potencial utilidad para la práctica epidemiológica. Por otra parte, el desarrollo de la informática y las posibilidades tecnológicas están facilitando el acceso a un número cada vez mayor de programas estadísticos de computadora. Dos paquetes populares en epidemiología de campo son el Epi-Info del CDC y la OMS y el Epidat de la OPS y la Xunta de Galicia, como se mencionó al inicio de este módulo. Es un hecho que el nivel de *exposición* a esta tecnología está aumentando y, por ello, un conocimiento básico del enfoque estadístico y epidemiológico en los equipos locales de salud es cada vez más necesario para asegurar el uso racional y eficiente de tales recursos tecnológicos. Precisamente para familiarizar a los equipos locales de salud con las aplicaciones de la bioestadística, el MOPECE presenta a continuación una de las pruebas de significancia estadística de mayor utilidad para la práctica epidemiológica de campo.

Comparación de dos Proporciones: La Prueba Chi Cuadrado

Para ejemplificar el uso de esta prueba estadística, nos referiremos a continuación a una situación en la que se evalúa si existe una asociación estadísticamente significativa entre ciertos tipos de ocupación y el riesgo de contraer malaria en una población.

En una localidad rural de 760 habitantes, situada en una zona malárica, se observó que, en el último año, la incidencia acumulada de malaria en campesinos fue 88,2 por mil, mientras que en la población no campesina fue 55,8 por mil. De acuerdo al censo local más reciente, en la comunidad hay 204 campesinos. Interesa saber si la ocupación campesino se asocia a la malaria. Los datos que resumen la situación descrita son:

	con malaria	sin malaria			tasa por mil	
campesino	18	186	204	$I_c =$	18/204	88,2
no-campesino	31	525	556	$I_{nc} =$	31/556	55,8
	49	711	760			

Desde el punto de vista estadístico, interesa conocer si hay diferencia significativa entre las dos medidas de incidencia acumulada. En este caso se puede aplicar una prueba de significación estadística llamada Chi Cuadrado (χ^2), cuya fórmula es:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

En la fórmula, “O” se refiere al valor observado en una celda y “E” a su valor esperado. La fórmula trabaja exclusivamente con las celdas centrales de un cuadro, en este caso una tabla 2x2, y hace referencia a las frecuencias observadas en dichas celdas y las que se esperaría ocurrieran (valores esperados) si efectivamente no hubiera diferencia entre las proporciones que se comparan. Los valores esperados para cada celda central se obtienen multiplicando sus totales marginales correspondientes y dividiendo este producto por el gran total. Por ejemplo, en la primera celda central (campesino con malaria), el valor observado es 18 y el valor esperado sería: $(204 \times 49) \div 760 = 13,2$. El cuadro de resultados para cada celda quedaría así:

	con malaria	sin malaria	
campesino	13,2 18	190,8 186	204
no-campesino	35,8 31	520,2 525	556
	49	711	760

Reemplazando la formula,

$$x^2 = \frac{(18 - 13,2)^2}{13,2} + \frac{(186 - 190,8)^2}{190,8} + \frac{(31 - 35,8)^2}{35,8} + \frac{(525 - 520,2)^2}{520,2} = 2,56$$

El valor de chi cuadrado calculado de esta manera (observado) se compara con un valor tabulado (esperado) tomado de la distribución de probabilidades teóricas. Este valor teórico corresponde al que se esperaría encontrar si los resultados observados ocurrieran puramente por azar. A este valor teórico se le llama *valor crítico*: si el valor observado es mayor que el valor crítico se concluye que la diferencia observada no es debida al azar y se dice que es *estadísticamente significativa*. El valor crítico indica el *nivel de significancia* de la prueba, que expresa la probabilidad de que la diferencia observada haya ocurrido por azar (dado que, en realidad, no existan diferencias). Usualmente esta probabilidad se fija en 5% y se denota como $p < 0,05$. El complemento de esta probabilidad se llama *nivel de confianza*, en general, 95%.

Para un nivel de confianza de 95%, el valor crítico del Chi Cuadrado (de acuerdo a una tabla de distribución teórica) es 3.84, que corresponde al llamado chi cuadrado con un grado de libertad, específico para tablas 2x2. Los grados de libertad de una tabla se refieren al número mínimo de frecuencias en las celdas centrales que se necesita conocer para poder completar los valores de las demás celdas, dados los valores de los totales marginales. En la situación analizada, el valor observado (2,56) no rebasa el valor crítico (3,84), por tanto se concluye entonces que concluye entonces que no se puede rechazar la hipótesis de no-diferencia (de nulidad). En consecuencia, se puede afirmar con 95% de confianza estadística que la ocupación campesino no está significativamente asociada a la presencia de malaria ($p > 0.05$).

La prueba de Chi cuadrado es ampliamente usada en epidemiología, especialmente en el análisis de tablas 2x2. Por ello, se ha desarrollado la siguiente fórmula alternativa simplificada:

$$x^2 = \frac{n(ad - bc)^2}{(a + c)(b + d)(a + b)(c + d)}$$

	enfermo	no enfermo	
expuesto	a	b	(a+b)
no expuesto	c	d	(c+d)
	(a+c)	(b+d)	n

que, aplicada a los datos de nuestro ejemplo, proporciona resultados similares:

$$x^2 = \frac{760[(18 \times 525) - (186 \times 31)]^2}{49 \times 711 \times 204 \times 556} = 2,61 \quad (p > 0,05)$$

La decisión sobre el nivel de confianza seleccionado depende de la situación que se analice; por las implicaciones que una decisión tenga, se querrá tener un nivel de confianza mayor o menor. Para servir a los distintos propósitos, existen varias alternativas. El Cuadro 3.14 presenta otros valores críticos de x^2 (chi cuadrado) para diferentes niveles de significancia.

Cuadro 3.14 Valores críticos de los estadígrafos x^2 y Z a distintos niveles de significancia y confianza estadísticas

nivel de significancia (p)	Valores Críticos		nivel de confianza
	Chi Cuadrado*	Prueba Z**	
0,001	10.83	3,29	99,999
0,01	6.63	2,58	99,99
0,05	3.84	1,96	95,0
0,10	2.71	1,64	90,0
0,20	1.64	1,28	80,0

* con un grado de libertad (sólo para tablas 2x2)

** para comparación de dos promedios (ver Apéndice)



Ejercicio 3.5

Durante el otoño de 1990, en una localidad de América del Norte, se reportó la inusual ocurrencia de una enfermedad caracterizada por fiebre, náuseas, edema, dificultad para respirar, taquicardia y, sobre todo, intensos dolores por contractura muscular, acompañada de una marcada elevación del número de eosinófilos, un tipo especial de glóbulos blancos que actúan sobre ciertos procesos alérgicos. En menos de un año se habían reportado más de 1.500 casos de Síndrome Eosinofilia-Mialgia. Los estudios epidemiológicos implicaron al consumo de triptofano, un importante aminoácido en la dieta humana, como responsable de la enfermedad; en especial asociado al consumo de un producto popular para combatir la depresión, el insomnio y el síndrome premenstrual.

Pregunta 1 Analice los siguientes datos y evalúe si existe una diferencia estadísticamente significativa en la incidencia de enfermedad entre los dos grupos estudiados según la exposición a triptofano. Emplee una prueba estadística apropiada. Utilice un nivel de significancia igual a 0.05. Interprete los resultados en términos de una hipótesis nula (de no-diferencia)

	Número	Incidencia (%)
expuestos	30	66,7
no-expuestos	36	22,2

Medidas de la fuerza de asociación

A diferencia de las llamadas pruebas de significancia estadística, útiles porque determinan la *presencia* de una asociación entre dos variables, la epidemiología propone el uso de dos medidas básicas que *cuantifican* la fuerza de esa asociación: el riesgo relativo y la OR (la OR proviene del inglés *odds ratio* que ha sido traducido como: razón de productos cruzados, razón de posibilidades u oportunidad relativa). En esta Unidad revisaremos los conceptos básicos sobre estas medidas de análisis epidemiológico y en la Unidad 5 las veremos en el contexto de la investigación epidemiológica de factores de riesgo de enfermedad, que es donde adquieren especial relevancia

Riesgo Relativo

Como vimos anteriormente, la incidencia de una enfermedad en una población y período determinados (incidencia acumulada y la tasa de incidencia) nos proporciona una medida del riesgo absoluto de padecer la enfermedad en esa población.

Riesgo absoluto: incidencia de enfermedad u otro evento de interés en la población o grupo poblacional; cuantifica la probabilidad de experimentar dicha enfermedad o evento.

La comparación de dos medidas de incidencia de una enfermedad, es decir, dos riesgos absolutos, permite detectar un posible **exceso de riesgo** en un grupo *con relación a otro*. En epidemiología es de particular interés comparar la ocurrencia de enfermedad entre un grupo expuesto a un factor considerado de riesgo respecto a otro no expuesto. Las pruebas estadísticas ayudan a detectar el exceso de riesgo entre ambos grupos, el **riesgo relativo** permite *cuantificar* la magnitud de tal exceso y mide la fuerza de la asociación entre exposición y enfermedad. Para calcular el riesgo relativo de una enfermedad con relación a una exposición, se requiere una medida del riesgo absoluto entre los expuestos a un factor de riesgo y una medida del riesgo absoluto entre los no expuestos a tal factor, es decir, la incidencia en expuestos y la incidencia en no expuestos. Conviene tener presente que nos referimos a la incidencia *de una enfermedad específica* y a la exposición *a un factor de riesgo específico*.

La epidemiología dispone de una serie de diseños de estudios para observar cuándo existe una asociación entre la exposición a un factor y el desarrollo subsecuente de una enfermedad. Entre estos diseños, los estudios de cohortes y caso-control no sólo demuestran si esa asociación existe sino cuán fuerte es. Los datos obtenidos a partir de estos estudios observacionales se disponen usualmente en una tabla 2x2, llamada así por la naturaleza dicotómica de las variables de exposición y de enfermedad, como la que se muestra a continuación:

Cuadro 3.15 Tabla 2x2 (Diseño de Cohortes)

	enfermo	no enfermo	
expuesto	a	b	a + b
no expuesto	c	d	c + d
	a + c	b + d	a + b + c + d

Convencionalmente, las columnas de la tabla 2x2 representan la presencia o ausencia de enfermedad y las filas la presencia o ausencia de exposición. En los estudios de cohortes se parte de dos grupos de sujetos *sin la enfermedad*, uno expuesto a un hipotético factor que se sospecha de riesgo y otro no expuesto al factor, y se observa el desarrollo posterior de la enfermedad en ambos grupos, durante un tiempo de seguimiento. De esta manera, en los estudios de cohortes la tabla 2x2 (Cuadro 3.15) tiene los siguientes componentes:

- a = expuestos al factor de riesgo que durante el seguimiento desarrollaron la enfermedad
- b = expuestos al factor de riesgo que durante el seguimiento no desarrollaron la enfermedad
- c = no expuestos al factor de riesgo que durante el seguimiento desarrollaron la enfermedad
- d = no expuestos al factor de riesgo que durante el seguimiento no desarrollaron la enfermedad
- a + c = total de sujetos que durante el seguimiento desarrollaron la enfermedad
- b + d = total de sujetos que durante el seguimiento no desarrollaron la enfermedad
- a + b = total de sujetos expuestos al factor de riesgo
- c + d = total de sujetos no expuestos al factor de riesgo

El **riesgo relativo** (RR), como medida de fuerza de asociación, se obtiene a partir de los estudios de cohortes, ya que su diseño nos permite calcular la incidencia de la enfermedad en ambos grupos. El riesgo relativo es una *razón de incidencias*, o sea el cociente entre la incidencia de enfermedad en los expuestos y la incidencia en los no expuestos al supuesto factor de riesgo. Es decir,

$$\text{riesgo relativo} = \frac{\text{incidencia en expuestos}}{\text{incidencia en no-expuestos}}$$

y, en la tabla 2x2, esto es:

$$RR = \frac{I_E}{I_{\bar{E}}} = \frac{a/a+b}{c/c+d}$$

Riesgo Relativo: razón entre el riesgo absoluto de enfermar o morir de aquellos con la exposición de interés y el riesgo absoluto de enfermar o morir de aquellos sin la exposición de interés.

Un RR igual a la unidad (RR=1) se interpreta como la ausencia de asociación entre exposición y enfermedad (el riesgo de enfermar es similar en ambos grupos); un RR mayor de 1 (RR>1) indica mayor riesgo en los expuestos (la exposición está asociada a la enfermedad y es un factor de riesgo); un RR menor de 1 (RR<1) indica menor riesgo en los expuestos (la exposición está asociada a la enfermedad y es un factor protector). La magnitud del RR cuantifica la *fuerza de asociación* entre la exposición y la enfermedad; así un RR igual a 3,5 expresa una asociación más fuerte entre exposición y enfermedad que, por ejemplo, un RR igual a 1,4; o un RR igual a 0,2 indica una asociación más fuerte que un RR igual a 0,7.

Consideremos un ejemplo clásico. Entre 1950 y 1952, los doctores Dawber, Meadors y Moore del Servicio de Salud Pública de los EE.UU. seleccionaron 5.127 varones y mujeres sanos, de 30 a 59 años, residentes de Framingham, Massachusetts, a quienes desde entonces se ha venido estudiando prospectivamente con el fin de observar la relación entre una serie de factores de riesgo y el desarrollo de enfermedad cardíaca coronaria. Como parte del Estudio del Corazón de Framingham, el Cuadro 3.16 presenta la situación observada al decimosexto año de seguimiento de un grupo de 1.112 varones, sanos y de 35 a 44 años de edad al inicio del estudio, con relación al desarrollo de enfermedad cardíaca coronaria según su exposición a tres factores de riesgo seleccionados.

Cuadro 3.16 Exposición a factores de riesgo y enfermedad cardíaca coronaria (ECC). Seguimiento de 16 años a varones de 35-44 años de edad. Framingham, EE.UU.

cohorte	hipertensión		cardiomegalia		tabaquismo	
	Total	ECC	Total	ECC	Total	ECC
expuestos	22	12	111	41	800	181
no-expuestos	1.090	206	1.001	177	312	37

Veamos el primer factor de riesgo estudiado: hipertensión (definida aquí como presión arterial sistólica igual o mayor a 180 mmHg). De acuerdo a los datos, de los 1.112 varones al inicio del estudio, 22 tenían hipertensión (estaban expuestos) y 1.090 no la tenían (no estaban expuestos). Dieciséis años después, 12 de los expuestos y 206 de los no expuestos (i.e., 218 casos) habían desarrollado enfermedad cardíaca coronaria (ECC). Esto quiere decir que el riesgo absoluto de enfermar con ECC entre quienes tienen hipertensión es:

incidencia en expuestos: $\frac{12}{22} \times 1.000 = 545,5$ por 1.000 expuestos

y el riesgo absoluto de enfermar con ECC entre quienes no tienen hipertensión es:

$$\text{incidencia en no expuestos: } \frac{206}{1.000} \times 1.000 = 189,0 \text{ por } 1.000 \text{ no expuestos}$$

Ahora podremos comparar ambos riesgos absolutos y determinar el exceso de riesgo como una razón de tasas. Así, el riesgo relativo será:

$$\text{riesgo relativo: } \frac{545,5}{189,0} = 2,89$$

esto es, los individuos hipertensos tienen 2,89 veces el riesgo de enfermar con ECC que el de los individuos no hipertensos. Dicho de otro modo, la exposición al factor de riesgo incrementa 1,89 veces el riesgo de desarrollar la enfermedad. Así, el riesgo relativo indica cuánto más riesgo tienen los expuestos *en relación* con los no expuestos.

Examinemos el siguiente ejemplo. En la última quincena de mayo de 1991 se presentó un intenso brote de cólera en tres caseríos aledaños de la selva amazónica (población 1.761 habitantes), que afectó a 125 personas y provocó la muerte de 7 de ellas. Durante la primera semana de junio se hizo una encuesta rápida en toda la población sobre su exposición a un conjunto de supuestos factores de riesgo. El Cuadro 3.17 presenta los resultados con relación al consumo de ciertos productos que, se sospechaba, podrían estar implicados en el brote.

Cuadro 3.17 Brote de cólera en tres caseríos rurales, selva amazónica; Junio 1991.

supuesto factor de riesgo	enfermaron (n=125)		no enfermaron (n=1.636)	
	expuestos	no expuestos	expuestos	no expuestos
agua no tratada	111	14	1.093	543
pescado crudo	7	118	14	1.622
pescado cocinado	17	108	198	1.438
arroz recalentado	47	78	522	1.114
tamal de arroz	24	101	272	1.364
fruta sin lavar	71	54	683	953

De lo que se trata es de determinar si existe alguna *asociación* entre la exposición a los supuestos factores de riesgo y la presencia de cólera en la población. Para ello procedemos a construir una tabla 2x2 para *cada uno* de los supuestos factores de riesgo, a saber:

Agua no tratada	Enfermo	No enfermo	Total
expuesto	111	1.093	1.204
No expuesto	14	543	557
Total	125	1.636	1.761

pescado crudo	Enfermo	No enfermo	Total
expuesto	7	14	21
No expuesto	118	1.622	1.740
Total	125	1.636	1.761

pescado cocinado	Enfermo	No enfermo	Total
expuesto	17	198	215
No expuesto	108	1.438	1.546
Total	125	1.636	1.761

arroz recalentado	Enfermo	No enfermo	Total
expuesto	47	522	569
No expuesto	78	1.114	1.192
Total	125	1.636	1.761

tamal de arroz	Enfermo	No enfermo	Total
expuesto	24	272	296
no expuesto	101	1.364	1.465
Total	125	1.636	1.761

fruta sin lavar	Enfermo	No enfermo	Total
expuesto	71	683	754
no expuesto	54	953	1.007
Total	125	1.636	1.761

Ahora podremos calcular más fácilmente las medidas de asociación. Asumiendo que este es un estudio de cohorte, por tanto, los riesgos relativos en cada exposición serían los siguientes:

$$RR_{\text{agua no tratada}} = \frac{\frac{111}{14}}{\frac{1.204}{557}} = \frac{0,092}{0,025} = 3,67$$

$$RR_{\text{arroz recalentado}} = \frac{\frac{47}{78}}{\frac{569}{1.192}} = \frac{0,083}{0,065} = 1,26$$

$$RR_{\text{pescado crudo}} = \frac{\frac{7}{118}}{\frac{21}{1.740}} = \frac{0,333}{0,068} = 4,92$$

$$RR_{\text{tamal de arroz}} = \frac{\frac{24}{101}}{\frac{296}{1.465}} = \frac{0,081}{0,069} = 1,18$$

$$RR_{\text{pescado cocinado}} = \frac{\frac{17}{108}}{\frac{215}{1.546}} = \frac{0,079}{0,070} = 1,13$$

$$RR_{\text{fruta sin lavar}} = \frac{\frac{71}{54}}{\frac{754}{1.007}} = \frac{0,094}{0,054} = 1,76$$

Parece claro que la exposición que se asocia con más fuerza al hecho de enfermar es el consumo de pescado crudo, así como el consumo de agua no tratada. El riesgo relativo de 4,92 nos indica que la probabilidad de desarrollar cólera fue 3.92 veces mayor en los sujetos que consumieron pescado crudo que en los que no lo consumieron. El riesgo de presentar cólera fue también casi tres veces mayor en aquellos que consumieron agua no tratada que en aquellos que no la consumieron. A la vista de los restantes riesgos relativos, esta evidencia apunta hacia una implicación causal de estos productos en el brote de cólera en estas comunidades amazónicas.



Ejercicio 3.6

A fin de cuantificar la asociación entre la exposición de mujeres gestantes a una serie de supuestos factores de riesgo y la presencia de bajo peso al nacer (BPN) en sus productos, entre mayo y septiembre de 1996 se realizó un estudio epidemiológico en la Provincia N de un país de América del Sur. Se definió como enfermo a toda madre que haya parido un niño/a con BPN (menos de 2.500 gramos) y como no enfermo a toda madre que haya parido un niño/a sin BPN. Se seleccionaron 1.556 enfermos y 16.910 no enfermos, registrados entre 1988 y 1995 en la base de datos del sistema informático perinatal disponible en dicha provincia. El Cuadro 3.18 presenta un extracto de los resultados de dicho estudio.

Cuadro 3.18 Prevalencia de exposición (%) a factores de riesgo de bajo peso al nacer. Estudio epidemiológico; Provincia N, América del Sur; 1988-1995.

factor de riesgo	enfermo (n=1.556)	no enfermo (n=16.910)
madre adolescente (menor de 17 años)	13,1	7,1
enfermedad de la placenta	24,4	7,2
sin atención prenatal	15,6	2,1
atención prenatal tardía (a partir del 5° mes)	56,8	31,0

- a) Disponga apropiadamente los datos en tablas 2x2 y cuantifique la correspondiente asociación entre exposición y enfermedad.

- b) Interprete y sintetice sus resultados.

Razón de Posibilidades (*Odds Ratio*)

Como acabamos de ver, para calcular el riesgo relativo necesitamos la incidencia de la enfermedad en expuestos y no expuestos y éstas se obtienen de un estudio de cohortes. Más frecuentemente, cuando necesitamos identificar asociación entre exposición y enfermedad sucede que nos encontramos ante una serie de sujetos que *ya presentaron la enfermedad*. En tal situación podemos recurrir a un estudio caso-control, en el cual se compara la *historia* de exposición de los enfermos con la de un grupo de sujetos similares, pero sanos, al que se llama ‘grupo control o testigo’. El diseño caso-control es muy versátil y popular en el campo y es uno de los estudios que nos permite obtener una medida denominada **razón de posibilidades** (razón de productos cruzados, razón de ventajas, razón de suertes, razón de momios, *odds ratio*), análoga al riesgo relativo, que estima el exceso de riesgo.

En los estudios caso-control se parte de dos grupos de sujetos, uno con la enfermedad y otro sin ella, y se investiga si habían estado previamente expuestos al factor de riesgo. Así, en los estudios caso-control la tabla 2x2 (Cuadro 3.20) tiene los siguientes componentes

Cuadro 3.19 Tabla 2x2 (Diseño Caso-Control)

	caso	control	
expuesto	a	b	a + b
no expuesto	c	d	c + d
	a + c	b + d	a + b + c + d

- a = enfermos (casos) que estuvieron expuestos al factor de riesgo
- b = no enfermos (controles) que estuvieron expuestos al factor de riesgo
- c = enfermos (casos) que no estuvieron expuestos al factor de riesgo
- d = no enfermos (controles) que no estuvieron expuestos al factor de riesgo

- a + c = total de sujetos enfermos (casos)
- b + d = total de sujetos no enfermos (controles)
- a + b = total de sujetos que estuvieron expuestos al factor de riesgo
- c + d = total de sujetos que no estuvieron expuestos al factor de riesgo

En su original acepción inglesa (*odds ratio*), la razón de posibilidades literalmente significa “razón de ‘odds’”. Para comprender mejor su significado, debemos revisar, una vez más, el término probabilidad. La definición frecuentista nos dice que la probabilidad de la ocurrencia de un evento, indica la frecuencia relativa límite con que dicho evento tendrá lugar a largo plazo, en pruebas repetidas en condiciones similares (Colton, 1975). El término ‘frecuencia relativa’ implica que el valor numérico de cualquier probabilidad se sitúa entre 0 y 1.

Como hemos visto a lo largo de esta Unidad, la probabilidad se utiliza a menudo para cuantificar la frecuencia esperada de ciertas características de interés, bajo condiciones de incertidumbre como, por ejemplo, el porcentaje de individuos en quienes se espera esté presente un riesgo o una enfermedad concretos. Esta misma información está contenida en otra medida, relacionada con la probabilidad, pero que se expresa de forma diferente. Se trata del odds. El odds (o 'ventaja') se define como la probabilidad de que ocurra un evento dividida entre la probabilidad de que no ocurra, es decir, el odds viene a ser una razón de probabilidades complementarias. Esto es,

$$\text{odds} = \frac{\text{probabilidad del evento}}{1 - \text{probabilidad del evento}}$$

Por ejemplo, es lo mismo decir que nuestro equipo de fútbol tiene una probabilidad de 80% de ganar su partido del domingo que decir que su odds es 4 a 1 para esa ocasión (Fletcher, 1998). O, como señala Last en su Diccionario, si 60 fumadores desarrollan tos crónica y 40 no, el odds entre estos 100 fumadores en favor de desarrollar tos es 60:40, o 1,5; en contraste, la probabilidad de que estos fumadores desarrollen tos es 60/100 o 0,6.

En un estudio caso-control se definen los odds en favor de la exposición al factor de riesgo. Así, en los casos, el odds de haber estado expuesto será:

$$\text{odds}_{\text{casos}} = \frac{a/a + c}{c/a + c} = \frac{a}{c}$$

y, en los controles, será:

$$\text{odds}_{\text{controles}} = \frac{b/b + d}{d/b + d} = \frac{b}{d}$$

De esta manera, la razón de posibilidades (OR) no es sino la razón o cociente entre los odds, en favor de la exposición de los casos y de los controles; es decir:

$$\text{OR} = \frac{a/a + c / c/a + c}{b/b + d / d/b + d} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Razón de posibilidades: es el cociente entre dos odds. Se define de forma diferente según la situación de que se trate. En un estudio caso-control, la razón de posibilidades (OR) es el cociente entre las ventajas (odds) a favor de la exposición en los casos (a/c) y los controles (b/d) (OR de exposición). En un estudio de cohortes o un estudio transversal, el OR es el cociente entre las ventajas (odds) a favor de la enfermedad en los expuestos (a/b) y los no expuestos (c/d) (OR de enfermedad). En ambos casos el cálculo queda reducido a ad/bc .

La razón de posibilidades de los estudios caso-control proporciona una medida que es conceptual y matemáticamente análoga al riesgo relativo de los estudios de cohortes. Desde un punto de vista más práctico, el OR, corresponde a la razón de productos cruzados en una tabla 2x2, como la presentada en esta Unidad y se calcula mediante la siguiente fórmula:

$$OR = \frac{a \times d}{b \times c}$$

Esta medida de fuerza de asociación tiene la misma interpretación que el riesgo relativo y en determinadas circunstancias (de baja frecuencia de la enfermedad) constituye una buena aproximación de éste. Así, un OR igual a 1 (OR=1) indica ausencia de asociación exposición-enfermedad; un OR mayor de 1 (OR>1) indica exposición de riesgo y un OR menor de 1 (OR<1) efecto protector.

Retomando nuestro ejemplo sobre el cólera en la Amazonía, asumamos que la situación corresponde a un diseño caso-control, como podría ser si se tratara de consultas en nuestro centro de salud. En tal situación, los casos serían todos los enfermos con cólera atendidos en el centro de salud y los controles, por ejemplo, todos los otros pacientes atendidos por otra causa. Aunque así ya no podemos calcular la incidencia, sí podemos medir la fuerza de asociación entre exposición y enfermedad por medio de la razón de posibilidades (OR), a saber:

$$OR_{\text{agua no tratada}} = \frac{111 \times 543}{1.093 \times 14} = \frac{60.273}{15.302} = 3,94 \quad OR_{\text{arroz recalentado}} = \frac{47 \times 1.114}{522 \times 78} = \frac{52.358}{40.716} = 1,29$$

$$OR_{\text{pescado crudo}} = \frac{7 \times 1.622}{14 \times 118} = \frac{11.354}{1.652} = 6,87 \quad OR_{\text{tamal de arroz}} = \frac{24 \times 1.364}{272 \times 101} = \frac{32.736}{27.472} = 1,19$$

$$OR_{\text{pescado cocinado}} = \frac{17 \times 1.438}{198 \times 108} = \frac{24.446}{21.384} = 1,14 \quad OR_{\text{fruta sin lavar}} = \frac{71 \times 953}{683 \times 54} = \frac{67.663}{36.882} = 1,83$$

Si comparamos la fuerza de asociación medida con el RR y con el OR apreciaremos que las diferencias observadas no cambian la conclusión acerca de las exposiciones que parecen estar causalmente implicadas en la propagación del cólera en la comunidad.

Conviene reiterar que tanto el riesgo relativo como la razón de posibilidades, miden el exceso de riesgo en los expuestos con respecto a los no expuestos a un determinado factor y ambos tienen como punto de referencia la unidad ($RR=1$ ó $OR=1$). Así, un RR u OR igual a 2, por ejemplo, no significa que los expuestos tengan dos veces más riesgo que los no expuestos, sino una vez más, es decir, los expuestos tienen el doble de riesgo de los no expuestos (100% más riesgo); un RR u OR igual a 1,5 significa que los expuestos tienen 0,5 veces más riesgo que los no expuestos (o sea, 50% de exceso de riesgo).

Apéndice estadístico - epidemiológico

En esta sección se ofrecen elementos adicionales para el análisis epidemiológico, los cuales pueden consultarse posteriormente. Estos incluyen pruebas de significancia estadística, estimación de intervalos de confianza, métodos para el análisis de impacto de factores de riesgo en la población y métodos para el control de factores que pueden distorsionar la comparación entre poblaciones.

Comparación de dos Promedios: La Prueba Z

Después de un brote de malaria, un centro de salud realiza un programa de tamizaje en el cual 150 frotis sanguíneos de niños de 1 a 4 años de edad son examinados para detectar la presencia de parásitos *Plasmodium falciparum*. Se encontraron 70 láminas positivas y el nivel promedio de hemoglobina en esos niños fue 10,6 g/dL, con una desviación estándar de 1,4 g/dL. El nivel promedio de hemoglobina en los 80 niños con láminas negativas fue 11,5 g/dL, con una desviación estándar de 1,3 g/dL. El centro de salud estaba interesado en saber si la infección por *P. falciparum* disminuye los niveles de hemoglobina en los niños de la comunidad. Los datos que resumen la situación descrita son:

	niños con malaria (grupo 1)	niños sin malaria (grupo 2)
tamaño del grupo (n)	70	80
promedio (\bar{x})	10,6	11,5
desviación estándar (DE)	1,4	1,3

Desde el punto de vista estadístico, el interés del centro de salud consiste en saber si existe o no diferencia entre los promedios de hemoglobina observados en los dos grupos de niños. En este caso podemos usar una prueba de significación estadística llamada Prueba Z, cuyo estadígrafo es:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{DE_1^2}{n_1} + \frac{DE_2^2}{n_2}}}$$

Según esta prueba, el valor crítico del estadígrafo Z para un nivel de significancia de 0.05 (5%) es 1,96. Si el valor de Z calculado es mayor que el Z crítico (1,96), se rechaza la hipótesis nula y se concluye que la diferencia observada es estadísticamente significativa con 95% de confianza ($p < 0,05$). Reemplazando,

$$z = \frac{10,6 - 11,5}{\sqrt{\frac{1,4^2}{70} + \frac{1,3^2}{80}}} \quad z = \frac{-0,9}{0,2216} = -4,06 \quad (p < 0,05)$$

El resultado es significativo al nivel 0,05 (5%), puesto que 4,06 es mayor que 1,96 (en esta prueba se considera el valor absoluto). En consecuencia, se puede afirmar con 95% de confianza que la infección por *P. falciparum* disminuye los niveles de hemoglobina en los niños afectados. Esta confianza es *estadística*, porque expresa la seguridad relativa con la que se toma una decisión *estadística*: decir que hay o no hay diferencia entre dos grupos observados, en función de la influencia del azar. Esta decisión, en estadística, corresponde a rechazar o aceptar la *hipótesis nula* o de no-diferencia. El Cuadro 3.14 presenta otros valores críticos de Z para diferentes niveles de significancia. Esta prueba se emplea sólo cuando ambas muestras son grandes (>30 en cada grupo) (de otro modo se suele aplicar la prueba *t* de Student, no tratada en este Módulo).

Intervalos de Confianza para promedios y proporciones

Un procedimiento alternativo o complementario de la inferencia estadística para estimar el grado de incertidumbre que rodea a los estimadores de los parámetros poblacionales es el cómputo de sus **intervalos de confianza**. Su comparación directa, incluso, puede ser tanto o más informativa que los valores *p* del nivel de significancia para decidir si una diferencia entre los estimadores (promedios o proporciones) es o no estadísticamente significativa.

En su forma general, un intervalo de confianza es simétrico respecto del estimador que contiene; es decir que se construye sumando y restando una misma cantidad al promedio o la proporción observada en la población estudiada. Esa cantidad se llama error de muestreo y corresponde al producto del error estándar del estimador y un valor crítico del estadígrafo correspondiente, que suele ser el valor de Z para un nivel de confianza de 95% (1,96). Así,

$$\text{intervalo de confianza} = \text{estimador} \pm \underbrace{Z_{95\%} \times \text{error estándar (E.S.)}}_{\text{error de muestreo}}$$

El error estándar es el mismo que se emplea en las pruebas de significancia estadística y, como hemos visto, es una expresión de la variación o variabilidad entre los individuos en las muestras de la población. Mas concretamente, el error estándar representa la desviación estándar de una distribución de *muestras* repetidas de la misma población. Imaginemos que en una misma población se repite 100 veces el mismo estudio para determinar el período de incubación promedio, en días, de la difteria. Cada estudio resumirá sus resultados en dos medidas: el promedio y la desviación estándar del período de incubación de la difteria. Así, tendremos 100 promedios y 100 desviaciones estándar de lo mismo. Si hacemos un gráfico de los 100 promedios obtenidos, veremos que éstos siguen una distribución normal y, por lo tanto, esta distribución de muestras tendrá también un promedio y una desviación estándar. Este promedio es un promedio de pro-

medios y esa desviación estándar es, precisamente, el **error estándar**; ambos representan los verdaderos parámetros poblacionales del período de incubación de la difteria. El error estándar es una medida de gran importancia en la teoría muestral; en la práctica se expresa como una relación entre la dispersión de los datos observados y el tamaño de la muestra estudiada.

El intervalo de confianza así construido indica el rango en el que, con una probabilidad conocida, el verdadero parámetro poblacional estudiado está contenido. Por ejemplo, si un estudio sobre el período de incubación de la difteria reportara que el promedio es 3,5 días y el Intervalo de Confianza al 95% (IC95%) es 2,4–6,1 días, lo que están diciendo los autores es: “si bien nuestro resultado puntual es 3,5 días en promedio, reconocemos que éste es un estimado y por lo tanto está sujeto a error aleatorio; no obstante, tenemos 95% de confianza que el verdadero tiempo de incubación de la difteria está entre 2,4 días y 6,1 días”. En otras palabras, se está diciendo que si se repitiera 100 veces el mismo estudio, en 95 de las veces el promedio obtenido estará entre 2,4 y 6,1 días, o sea el intervalo de confianza, pero en 5 de las veces podrá estar fuera de dicho intervalo. De esta manera, el intervalo de confianza cuantifica de manera explícita la inherente imprecisión de los datos.

IC95% de un promedio:

$$\bar{x} \pm 1,96 \cdot \left(\frac{DS}{\sqrt{n}} \right)$$

El IC95% del promedio de hemoglobina en el grupo de niños con malaria (pag. 76) será:

$$10,6 \pm 1,96 \left(\frac{1,4}{\sqrt{70}} \right) = 10,6 \pm 1,96 \times 0,1673 = 10,6 \pm 0,328$$

Límite inferior (Li) = 10,3g/dL; Límite superior (Ls) = 10,9g/dL.

El IC95% del promedio de hemoglobina en el grupo de niños sin malaria (pag. 76) será:

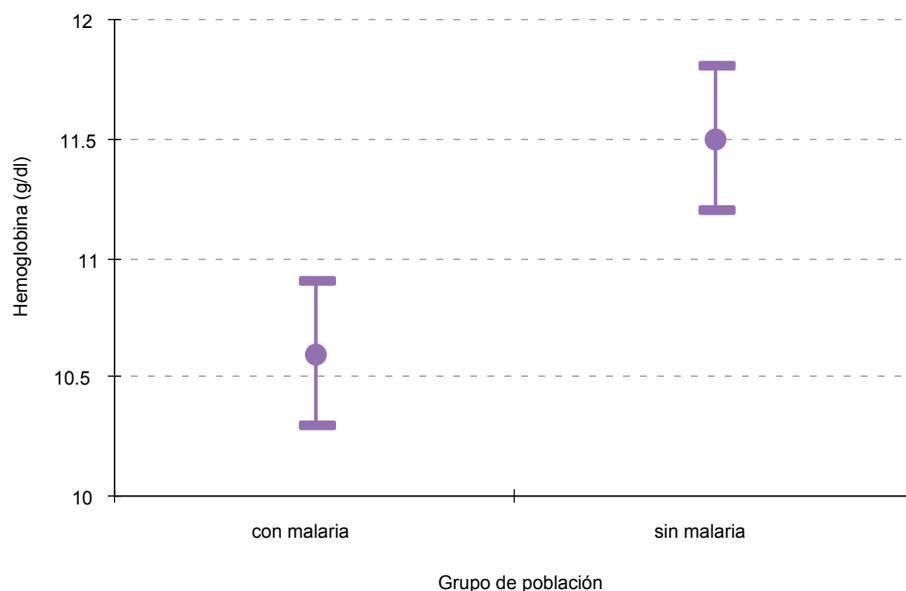
$$11,5 \pm 1,96 \cdot \left(\frac{1,3}{\sqrt{80}} \right) = 11,5 \pm 1,96 \times 0,1453 = 11,5 \pm 0,285$$

Li = 11,2g/dL; Ls = 11,8g/dL.

El Gráfico 3.10 compara visualmente ambos grupos de población en función de sus respectivos intervalos de confianza. El hecho de observar que sus rangos no se superponen entre sí (el límite superior de uno y el límite inferior de otro “no se tocan”) sugiere que

hay diferencia estadísticamente significativa entre ambos, en correspondencia con los resultados de la prueba de significancia empleada (Prueba Z).

Gráfico 3.10 Comparación de intervalos de confianza de dos promedios



IC95% de una proporción:

$$p \pm 1,96 \sqrt{\frac{p \cdot (1-p)}{n}}$$

El IC95% de confianza para la incidencia de malaria en campesinos (pag. 61) será:

$$0,0882 \pm 1,96 \sqrt{\frac{0,0882 \cdot (1-0,0882)}{204}} = 0,0882 \pm 1,96 \times 0,0199 = 0,0882 \pm 0,0389$$

$L_i = 0,0493$ (49,3 por mil); $L_s = 0,1272$ (127,2 por mil).

Obsérvese que la proporción p en la fórmula se expresa en tanto por uno (o sea, va de 0 a 1); los resultados se multiplican por 1.000 para expresarlos como tasas por mil.

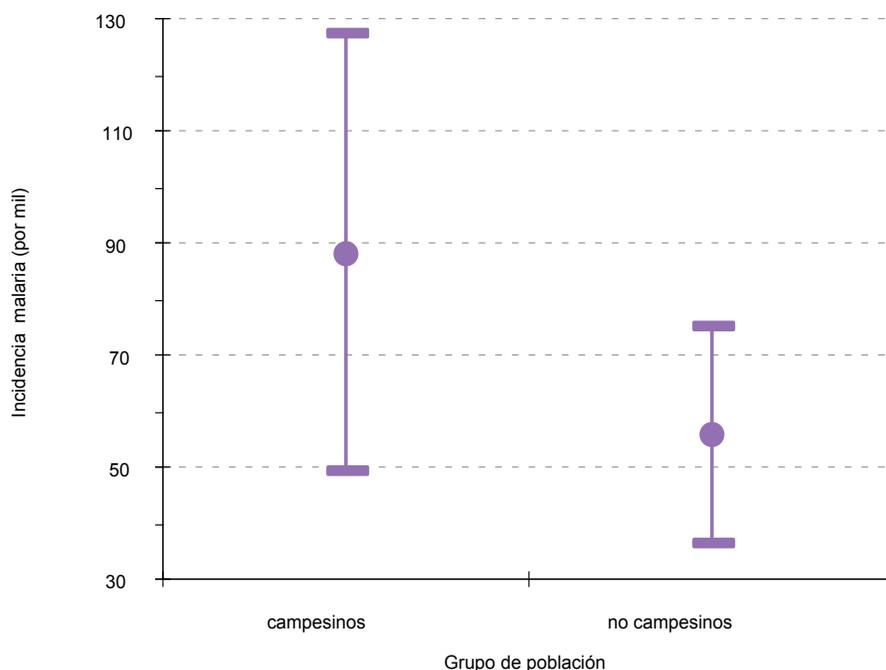
El IC95% de confianza para la incidencia de malaria en no campesinos (pag. 64) será:

$$0,0558 \pm 1,96 \sqrt{\frac{0,0558 \cdot (1-0,0558)}{556}} = 0,0558 \pm 1,96 \times 0,0097 = 0,0558 \pm 0,0191$$

$Li = 0,0367$ (36,7 por mil); $Ls = 0,0749$ (74,9 por mil).

El Gráfico 3.11 compara visualmente ambos grupos de población en función de sus respectivos intervalos de confianza. El hecho de observar que sus rangos se superponen entre sí (el límite superior de uno y el límite inferior de otro “se tocan”) sugiere que no hay diferencia estadísticamente significativa entre ambos, en correspondencia con los resultados de la prueba de significancia empleada (Chi cuadrado). Además, obsérvese la longitud del recorrido del intervalo de confianza para la incidencia en campesinos: la amplitud de un intervalo de confianza también informa sobre la precisión del estudio; los intervalos amplios indican pobre precisión muestral.

Gráfico 3.11 Comparación de intervalos de confianza de dos proporciones



La estimación de intervalos de confianza es un procedimiento básico para documentar el grado de precisión de nuestros resultados. Los métodos que hemos revisado son los más comunes y útiles en la práctica cotidiana de la epidemiología. Sin embargo, debe señalarse que existe una amplia variedad de técnicas para la estimación de intervalos de confianza de distintos parámetros poblacionales, como los intervalos de confianza de una diferencia de promedios, de una diferencia de proporciones; riesgos relativos, razones de posibilidades y otras importantes mediciones en salud. Aunque el desarrollo de estos cálculos no es parte de los objetivos del MOPECE, los principios básicos que ha descrito sobre la construcción de intervalos de confianza se aplican universalmente para tales casos.

Medidas de impacto potencial

Riesgos atribuibles

La comparación de dos medidas de incidencia acumulada, como se ha señalado, permite cuantificar el exceso de riesgo entre dos grupos con diferente nivel de exposición al factor de interés. Este exceso de riesgo se puede medir en dos direcciones: la razón de tasas, dividiéndolas para obtener el riesgo relativo, o la diferencia de tasas, restándolas, en cuyo caso obtendremos el llamado **riesgo atribuible en expuestos**, es decir:

riesgo atribuible en expuestos = incidencia en expuestos – incidencia en no expuestos

$$RA_E = I_E - I_{\bar{E}}$$

En el ejemplo de Framingham, el riesgo atribuible en expuestos (hipertensos) será:

riesgo atribuible en expuestos: $545,5 - 189,0 = 356,5$ por 1.000 expuestos

que es la tasa de enfermedad ECC en los expuestos que se considera debida o atribuible a la exposición, es decir, al hecho de ser hipertensos. En otras palabras, si los individuos expuestos dejaran de ser hipertensos (esto es, si se eliminase su exposición al factor de riesgo), su riesgo absoluto original (545,5 por 1.000) quedaría reducido solamente al riesgo absoluto de los no expuestos (189,0 por 1.000); este exceso de riesgo (356,5 por 1.000) de enfermar con ECC se atribuye a la hipertensión.

En forma análoga, el **riesgo atribuible en la población** (RAP), es la magnitud absoluta de incidencia de enfermedad que se atribuye a la exposición, ya no sólo en los expuestos sino *en el conjunto de la población*. El RAP es la diferencia entre la incidencia en la población (I_p) y la incidencia en los no expuestos; es decir:

riesgo atribuible en la población, $I_p - I_{\bar{E}}$

En el ejemplo de Framingham, podemos determinar que la incidencia de ECC en la población fue 196,0 por mil personas (218 casos en 1.112 personas). El RAP será:

riesgo atribuible en la población, RAP: $196,0 - 189,0 = 7,0$ por 1.000 habitantes

que es la tasa de enfermedad ECC en la población que se considera debida o atribuible a la hipertensión (la exposición). Es decir que, si no hubiera hipertensión en la población, el riesgo absoluto de enfermar con ECC sería 189,0 por mil personas en vez de 196,0 por mil: el exceso de riesgo, atribuible a la exposición, es sólo 7,0 por mil habitantes.

Fracciones atribuibles

A partir de los riesgos atribuibles podemos calcular otras dos medidas importantes, llamadas *medidas de impacto potencial*: la fracción atribuible en expuestos y la fracción atribuible en la población. La **fracción atribuible en expuestos** es simplemente el riesgo atribuible en expuestos expresado porcentualmente, es decir, la proporción de la incidencia en expuestos que se considera debida a la exposición al factor de riesgo. Esto es,

$$\text{fracción atribuible en expuestos: } \frac{I_E - I_{\bar{E}}}{I_E} \times 100$$

En el ejemplo de Framingham sobre hipertensión y ECC, la fracción atribuible en expuestos –llamada también riesgo atribuible porcentual, RA%, será:

$$\text{fracción atribuible en expuestos (RA%): } \frac{545,5 - 189,0}{545,5} \times 100 = 65,4\%$$

que quiere decir que el 65% del riesgo absoluto en expuestos es debido a la exposición al factor de riesgo, es decir, al hecho de ser hipertensos (el riesgo atribuible en expuestos, 356,5 por 1.000, equivale al 65,4% del riesgo absoluto en expuestos, 545,5 por 1.000).

Finalmente, la **fracción atribuible en la población** o *fracción etiológica* es una medida que cuantifica el exceso de riesgo de enfermar, que se atribuye a la exposición ya no sólo en los expuestos, sino *en el conjunto de la población*. Esta medida es simplemente el riesgo atribuible en la población expresado porcentualmente (RAP%). De esta manera, la fracción atribuible en la población queda expresada como:

$$\text{fracción atribuible en la población, RAP%: } \frac{I_P - I_{\bar{E}}}{I_P} \times 100$$

La fracción atribuible en la población permite identificar la importancia relativa de la exposición a un determinado factor de riesgo en la población, pues expresa la magnitud en que se reduciría el riesgo absoluto de enfermar en el conjunto de la población (es decir, la incidencia de la enfermedad en la población) si se eliminara dicha exposición. Por ello, se considera que el RAP% es una medida de gran trascendencia en salud pública.

En el ejemplo de Framingham sobre hipertensión y ECC, la fracción etiológica o riesgo atribuible poblacional porcentual, RAP%, será:

$$\text{fracción atribuible en la población, RAP%: } \frac{196,0 - 189,0}{196,0} \times 100 = 3,6\%$$

lo que quiere decir que la hipertensión da cuenta únicamente del 3.6% del riesgo absoluto de enfermar con ECC, i.e., la incidencia, en la población estudiada. En otras palabras, si se tuviera éxito en eliminar la hipertensión en toda la población, se habría conseguido reducir únicamente en 3.6% el riesgo de desarrollar enfermedad cardíaca coronaria en esa población. Esto, como veremos luego, tiene obvias implicaciones en la priorización e implementación de medidas de control de alcance poblacional en salud.



Ejercicio 3.7

Complete la información del cuadro siguiente con relación al impacto potencial asociado a los otros dos factores de riesgo de ECC presentados en el Cuadro 3.16a:

factor de riesgo	cardiomegalia (medida por el incremento de la sombra cardiaca a los rayos X)	tabaquismo (hábito de fumar cigarrillos presente)
Medida		
riesgo relativo		
riesgo atribuible en expuestos		
riesgo atribuible en la población		
fracción atribuible en expuestos		
fracción atribuible en la población		

Hagamos ahora un resumen de nuestras observaciones sobre la situación descrita en el ejemplo de Framingham. Hemos estado interesados en ver cuál es el riesgo de desarrollar enfermedad cardiaca coronaria en una población en función de la exposición a tres factores de riesgo: hipertensión, cardiomegalia y tabaquismo. Hemos visto que el factor con mayor riesgo relativo es la hipertensión (RR=2,9) y el de menor el tabaquismo (RR=1,9)

pero, por otro lado, la fracción etiológica más alta es la del tabaquismo (RAP%=39,5%) y la más baja la de hipertensión (RAP%=3,6%). En otras palabras, aunque claramente tiene más riesgo de desarrollar ECC un individuo hipertenso que uno fumador, a nivel poblacional resultaría más recomendable aplicar una estrategia de reducción del tabaquismo, que una de reducción de la hipertensión. ¿Por qué?

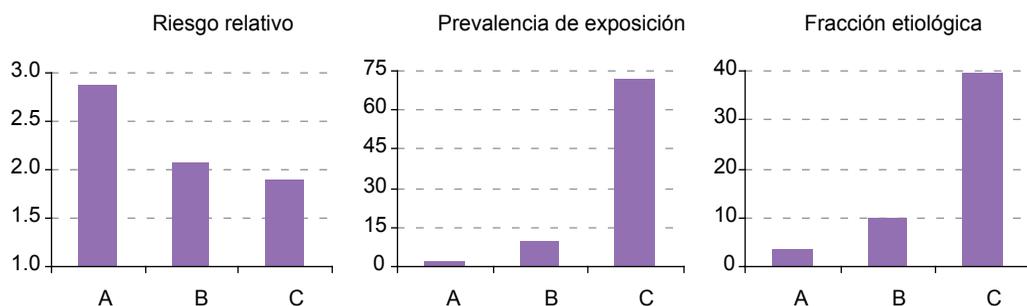
Esta situación, relativamente común, ilustra el contraste entre riesgo individual y riesgo poblacional. Esta situación se explica por las diferencias en la **prevalencia de exposición en la población** a los distintos factores de riesgo. Con los datos del Cuadro 3.21 podemos inferir que la prevalencia de hipertensión en la población fue 2%, mientras que la de tabaquismo fue 72%. Así, el RAP% es sensible no sólo a la magnitud del exceso de riesgo entre expuestos y no expuestos, sino también a la magnitud de la exposición al riesgo en la población. En el ejemplo de Framingham, que sintetizamos en el Cuadro 3.21 y Gráfico 3.12 siguientes, si tuviéramos que decidir por una intervención de salud pública para disminuir el riesgo de ECC –una situación habitual cuando hay limitación de recursos– la intervención dirigida a disminuir la prevalencia de consumo de cigarrillos tendrá definitivamente mayor impacto en dicha población.

Cuadro 3.21 Exposición a factores de riesgo y enfermedad cardiaca coronaria (ECC). Seguimiento de 16 años a varones de 35-44 años de edad. Framingham, EE.UU.

	factor de riesgo inicial	riesgo relativo (RR)	prevalencia de exposición (PE)	fracción etiológica (RAP%)
A	hipertensión sistólica	2,9	2,0	3,6
B	cardiomegalia	2,1	10,0	9,8
C	tabaquismo	1,9	71,9	39,5

Adaptado de Kahn y Sempos.

Gráfico 3.12 Exposición a factores de riesgo y enfermedad cardiaca coronaria (ECC). Seguimiento de 16 años a varones de 35-44 años de edad. Framingham, EE.UU.



Otro ilustrativo ejemplo de la importancia de considerar el impacto potencial de las medidas de control en la comunidad a la hora de tomar decisiones se encuentra en el caso del cólera en la Amazonía, presentado en el Cuadro 3.17 (página 68). De acuerdo con la información disponible, podemos determinar la fracción atribuible en la población (fracción etiológica, RAP%) de la exposición a pescado crudo y a agua no tratada:

$$\text{RAP\%}_{\text{pescado crudo}} : \frac{\left(\frac{125}{1.761}\right) - \left(\frac{118}{1.740}\right)}{\frac{125}{1.761}} \times 100 = \frac{(0,071 - 0,068)}{0,071} \times 100 = 4,46\%$$

$$\text{RAP\%}_{\text{agua no tratada}} : \frac{\left(\frac{125}{1.761}\right) - \left(\frac{14}{557}\right)}{\frac{125}{1.761}} \times 100 = \frac{(0,071 - 0,025)}{0,071} \times 100 = 64,59\%$$

Esto nos permite saber que, aunque el consumo de pescado crudo está más fuertemente asociado a la presencia de cólera (RR=4,9), su impacto potencial en la población es muy bajo, por que la prevalencia de exposición (consumo de pescado crudo) en la población es muy baja (1,2%). Si elimináramos el consumo de pescado crudo en esa población, solamente conseguiríamos reducir cerca de 5% su incidencia de cólera. En cambio, el consumo de agua no tratada –una práctica altamente prevalente en dicha comunidad (68,4% consume agua no tratada)– tiene gran impacto potencial: eliminar este factor de riesgo reduciría 65% la incidencia de cólera en la población. Aunque, es obvio que las campañas de prevención del cólera en esa comunidad habrán de advertir del riesgo de consumir pescado crudo o fruta sin lavar, la inversión tendrá que dirigirse prioritariamente a evitar que su población consuma agua no tratada.

Estandarización de tasas

Con el fin de sintetizar la información disponible, frecuentemente las tasas se presentan para la población completa o para grandes categorías de la misma. A estas tasas se les llaman **tasas crudas**. Con estas tasas-resumen se suele comparar la mortalidad o la incidencia de una enfermedad entre dos áreas geográficas, dos grupos de población o dos momentos en el tiempo y ver si existen diferencias relevantes. Consideremos la siguiente situación:

grupos de edad (años)	PAÍS A			PAÍS B		
	defunciones	población	tasa por mil	defunciones	población	tasa por mil
Total	1.269.166	68.386.000	18,6	5.564.944	198.250.000	28,1

Los datos de mortalidad y población corresponden al mismo año calendario. El País A es un país escasamente industrializado; el País B es un país altamente industrializado. ¿Qué información comparativa se obtiene a partir de estos datos?. Básicamente, podríamos

sacar dos conclusiones: 1) que la tasa de mortalidad en el País B es más alta que en el País A; y, 2) que el riesgo de morir en el País B es 50% más alto que en el País A. Es decir que, en términos de mortalidad, el País B está en una situación más desfavorable que el País A. El País B, como se mencionó, un país altamente industrializado.

Las tasas pueden también ser presentadas en forma desagregada para varias categorías de población, definidas sobre la base de características relevantes a la comparación, tales como edad, sexo, etnia, ocupación o nivel de exposición a determinado factor de riesgo. A estas tasas se les denomina **tasas específicas**. Puesto que el riesgo de morir o de contraer la gran mayoría de enfermedades está relacionado, en general, con la edad y a menudo difiere entre los sexos, el análisis de la mortalidad o la incidencia de enfermedad en una población debe necesariamente hacer uso de las correspondientes tasas específicas. Regresemos ahora a la situación entre los países A y B:

grupos de edad (años)	PAÍS A			PAÍS B		
	defunciones	población	tasa por mil	defunciones	población	tasa por mil
Total	1.269.166	68.386.000	18,6	5.564.944	198.250.000	28,1
<15	317.308	19.831.740	16,0	94.169	24.781.250	3,8
15 – 44	338.100	35.218.790	9,6	380.430	79.256.250	4,8
45 – 64	270.261	10.941.760	24,7	1.223.875	61.501.250	19,9
65 y +	343.497	2.393.710	143,5	3.866.470	32.711.250	118,2

A partir de esta información es posible identificar al menos tres hechos relevantes: 1) el País A tiene tasas de mortalidad específicas por edad más altas que el País B en *todos* los grupos de edad considerados; 2) la estructura por edades difiere marcadamente entre las dos poblaciones: el país A concentra su población hacia edades tempranas, el país B hacia edades tardías; y, 3) hay una aparente contradicción entre lo que esta información refleja y lo que se concluye observando las tasas crudas de mortalidad de los dos países.

¿Cómo se explica esta aparente *confusión*?. Dado que, como ha sido mencionado, el riesgo de morir o enfermar está habitualmente asociado con la edad, las tasas crudas de mortalidad e incidencia dependen críticamente de la composición etárea de una población. Esto cobra mayor relevancia cuando el objetivo es comparar dos poblaciones. Puede ser, por tanto, inapropiado emplear tasas crudas para comparar dos poblaciones *distintas* a menos que tengan la misma estructura por edades. La diferencia de composición etárea (o sea, la variable edad) ejerce un *efecto confusor* en la comparación de tasas crudas de mortalidad por país. De hecho, una tasa cruda es básicamente un promedio ponderado de las tasas específicas por categoría, siendo los pesos las proporciones de población en cada categoría.

Para poder realizar una comparación libre de las distorsiones que pueden provocar las diferencias en la composición de la población se deben emplear **tasas estandarizadas**. Las tasas estandarizadas o ajustadas son tasas-resumen construidas estadísticamente para tomar en cuenta y remover el potencial efecto confusor de la variable edad u otra tercera variable, al comparar las tasas de mortalidad o incidencia de dos poblaciones diferentes.

El procedimiento básico para el ajuste de tasas (el llamado método directo) requiere contar con las tasas específicas por categoría de la variable a ajustar (por ejemplo, la edad) en ambas poblaciones a comparar. Es decir, se requiere dos conjuntos de tasas específicas por edad (uno por cada país) y una población estándar. La idea general es ver cuál sería el número total de defunciones que habría en cada uno de los dos países, si éstos tuvieran exactamente la misma estructura de edad (la población estándar) y ocurrieran las tasas específicas por edad realmente *observadas* en cada uno. En otras palabras, se trata de *ajustar* la estructura de mortalidad observada de cada país, a una estructura de edad única y estándar para los dos. El procedimiento incluye los siguientes cuatro pasos:

1. fijar la población estándar: se puede seleccionar una ya conocida o se puede construir una a partir de los datos; por ejemplo, sumando las poblaciones específicas por categoría de edad en cada grupo;

grupos etáreos	PAÍS A	PAÍS B	POBLACIÓN ESTÁNDAR (A + B)
Total	68.386.000	198.250.000	266.636.000
<15	19.831.740	24.781.250	44.612.990
15 – 44	35.218.790	79.256.250	114.475.040
45 – 64	10.941.760	61.501.250	72.443.010
65 y +	2.393.710	32.711.250	35.104.960

2. calcular el número esperado de defunciones en cada categoría de edad de la población estándar, aplicando las tasas específicas por edad observadas en cada una de las dos poblaciones a comparar;

grupos etáreos	POBLACIÓN ESTÁNDAR	PAÍS A		PAÍS B	
		tasa observada	defunciones esperadas	tasa observada	defunciones esperadas
Total	266.636.000				
<15	44.612.990	16,0	713.808	3,8	169.530
15 – 44	114.475.040	9,6	1.098.959	4,8	549.480
45 – 64	72.443.010	24,7	1.789.339	19,9	1.441.616
65 y +	35.104.960	143,5	5.037.556	118,2	4.149.407

por ejemplo, el número de defunciones esperadas en el grupo de menores de 15 años para el País A, se calcula por medio de una regla de tres simple: si ocurren 16 defunciones por cada mil personas, ¿cuántas defunciones ocurrirán en 44.612.990 personas?. Esto es,

número de defunciones esperadas en <15 años, País A: $\frac{44.612.990}{1.000} \times 16 = 713.808$

3. obtener el número total de defunciones esperadas en cada país, sumando los resultados correspondientes del paso anterior; y,

grupos etáreos	POBLACIÓN ESTÁNDAR	PAÍS A		PAÍS B	
		tasa observada	defunciones esperadas	tasa observada	defunciones esperadas
Total	266.636.000		8.639.663		6.310.033
<15	44.612.990	16,0	713.808	3,8	169.530
15 – 44	114.475.040	9,6	1.098.959	4,8	549.480
45 – 64	72.443.010	24,7	1.789.339	19,9	1.441.616
65 y +	35.104.960	143,5	5.037.556	118,2	4.149.407

4. calcular las respectivas tasas ajustadas por edad para cada población, dividiendo el número total de casos esperados obtenido en el paso previo, entre el total de la población estándar.

tasa ajustada de mortalidad, País A: $\frac{8.639.663}{266.636.000} \times 1.000 = 32,4$ por mil

tasa ajustada de mortalidad, País B: $\frac{6.310.033}{266.636.000} \times 1.000 = 23,7$ por mil

Comparemos una vez más los resultados crudos y ajustados:

tasa de mortalidad por mil	PAÍS A	PAÍS B
cruda	18,6	28,1
estandarizada	32,4	23,7

Luego de haber removido la distorsión producida por la diferencia en la estructura de edades, se cuenta con una medida-resumen válida para comparar la mortalidad entre los dos países: la tasa de mortalidad es cerca de 40% más alta en el País A que en el País B. Cabe recalcar que el uso de tasas estandarizadas sólo está indicado con fines comparativos; su construcción estadística está basada en la elección arbitraria de un estándar y, por ello, la magnitud de la cifra carece de valor intrínseco. Por último, la estandarización de tasas no suple las deficiencias en la calidad, cobertura ni registro de los datos.

Referencias bibliográficas

Alleyne GAO. Gente sana en entornos saludables. Informe Anual del Director OPAS 1998. Organización Panamericana de la Salud; Washington DC, 1999.

Alleyne GAO. Emerging diseases—What now?. *Emerging Infectious Diseases* 1998 July; 4(3):498-500.

Beaglehole R, Bonita R, Kjellström T. *Epidemiología básica*. Organización Panamericana de la Salud; Washington DC, 1994.

Bortman M. Factores de riesgo de bajo peso al nacer. *Revista Panamericana de Salud Pública/Pan American Journal of Public Health* 1998 May;3(5):314-21.

Castillo-Salgado C [Editor]. *Manual sobre el enfoque de riesgo en la atención maternoinfantil*. Segunda edición. Organización Panamericana de la Salud; Washington DC, 1999.

Centers for Disease Control. *Epidemiologic notes and reports: Eosinophilia-Myalgia Syndrome—New Mexico. Mortality and Morbidity Weekly Report* 1989 November 17,38(45):765-67.

Colton T. *Estadística en medicina*. Ediciones Díaz de Santos S.A.; Madrid, 1995.

Donaldson RJ, Donaldson LJ. *Medicina comunitaria*. Editorial Díaz de Santos; Madrid, 1989.

Fletcher RH, Fletcher SW, Wagner EH. *Epidemiología clínica: aspectos fundamentales*. Segunda edición. Masson, Williams & Wilkins España, S.A.; Madrid, 1998.

Freund JE, Williams FJ. *Dictionary/Outline of basic statistics*. Dover Publications, Inc.; New York, 1991.

Gordis L. *Epidemiology*. Second edition. W.B. Saunders Company; Philadelphia, 2000.

Gore SM, Altman DG. *Statistics in practice*. British Medical Association; London, 1982.

Kahn HA, Sempos CT. *Statistical methods in epidemiology*. Oxford University Press; New York, 1989.

Kirkwood BR. *Essentials of medical statistics*. Backwell Scientific Publications; Osney Mead, 1988.

Last J [Editor]. *A dictionary of epidemiology*. Third edition. International Epidemiological Association. Oxford University Press; New York, 1995.

Lwanga SK, Tye CY, Ayeni O. *Teaching health statistics. Lesson and seminar outlines*. Second edition. World Health Organization; Geneva, 1999.

Newcomer RJ, Benjamin AE. *Indicators of chronic health conditions. Monitoring community-level delivery systems*. The Johns Hopkins University Press; Baltimore, 1997.

Norell SE. *Workbook of epidemiology*. Oxford University Press, Inc.; New York, 1995.

Norman GR, Streiner DL. *Bioestadística*. Mosby/Doyma Libros, S.A.; Madrid, 1996.

Organización Panamericana de la Salud. *Las condiciones de salud en las Américas. Volumen I*. Edición de 1994. Organización Panamericana de la Salud; Washington DC, 1994.

Quick RE, Vargas R, Moreno D, et al. Epidemic cholera in the Amazon: the challenge of preventing deaths. *American Journal of Tropical Medicine and Hygiene* 1993;48(5):597-602.

Rosner B. *Fundamentals of Biostatistics*. Fifth Edition. Duxbury Thompson Learning; Pacific Grove, 2000.

Silva LC. *Cultura estadística e investigación científica en el campo de la salud: una mirada crítica*. Ediciones Díaz de Santos, S. A.; Madrid, 1997.

Tukey JW. *Exploratory Data Analysis*. Addison Wesley Publishing Co. 1977.

Turnock BJ. Good news and bad news for public health. *Journal of Public Health Management and Practice* 1997 May;3(3):x-xi.

Wallgren A, Wallgren B, Persson R, Jorner U, Haalad JA. *Graphing statistics & data. Creating better charts* SAGE Publications, Inc; Newbury Park, 1996.

ISBN: 92 75 32407 7



9 789275 324073



**Organización
Panamericana
de la Salud**



Oficina Regional de la
Organización Mundial de la Salud

