

Módulo 4: Análisis de datos

Unidad 2: Análisis de datos básicos

1. Análisis descriptivo.
2. Interpretación de información.

Objetivos de aprendizaje

Al final de esta sesión, usted estará en capacidad de:

- Describir los diferentes tipos de variables.
- Reconocer y calcular las medidas de frecuencia de los eventos de interés en salud pública.
- Describir e interpretar los datos resumidos usando un proceso sistemático.
- Explicar los posibles motivos de la variación en la notificación de casos de eventos de interés en salud pública informados.

Introducción

El análisis de datos resulta fundamental en el propósito de generar información clave para la toma de decisiones, contribuyendo a la formulación de políticas basadas en la evidencia, así como a la comprensión de fenómenos intervinientes en los procesos de salud enfermedad y la afectación de los determinantes de la salud, entre otros.

En la vigilancia en salud pública, el análisis involucra principalmente un proceso de descripción y comparación de datos con relación a características y atributos de tiempo, lugar y persona, así como entre los diferentes niveles organizativos del sistema de salud y tiene el propósito de (1):

- Establecer las tendencias de la enfermedad a fin de detectar y anticipar la ocurrencia de cambios en su comportamiento.
- Sugerir los factores asociados con el posible incremento o descenso de casos y/o defunciones e identificar los grupos sujetos a mayor riesgo.
- Identificar las áreas geográficas que requieren medidas de control ante brotes y epidemias.

Contenido

1. Análisis descriptivo

Por medio de la estadística descriptiva, podemos clasificar, organizar y resumir los datos que obtenemos, logrando un resumen completo y una visión general del fenómeno de interés en salud pública.

Una medida (o escala) es la comparación de un dato que se hace con un referente y que busca ser entendida. Una variable es cualquier característica o atributo que puede asumir valores diferentes, necesitamos dos tipos diferentes de mediciones debido a que los distintos tipos de variables se analizan de manera diferente. Existen

#OrgullosamenteINS

 @INS Colombia  @insaludColombia  @INS Colombia

varias maneras de caracterizar las variables, una simple es distinguir entre variables cualitativas o categóricas y variables cuantitativas o continuas.

Variable cuantitativa: también llamada variable numérica. "Cuantitativo" proviene de la misma palabra que cantidad, que significa cuánto o cuántos. Las variables cuantitativas tienen valores numéricos y representan medidas. Por ejemplo, la edad, la altura y el número de hijos son todas variables cuantitativas.

Variable cualitativa: también llamada variable nominal o categórica. "Cualitativo" proviene de la misma palabra que "cualidades", que significa características de algo. Las variables cualitativas son descriptivas, con valores de categoría o etiqueta. Por ejemplo, presencia de enfermedad (sí o no), el sexo o el distrito.

Los distintos tipos de variables se analizan de manera diferentes.

Tipo de variable	Herramienta para resumir
Cuantitativa	(Medidas de tendencia central): Media, mediana, moda
- Continua	(Medida de dispersión): Rango
- Discreta	
Cualitativa	(Medidas de frecuencia de la enfermedad) o
- Ordinal	Razones, proporciones, distribuciones de frecuencia, tasas
- Nominal	

La manera más frecuente para resumir los datos de la vigilancia en salud pública es a través de las variables tiempo, lugar y persona. Describiendo los eventos en salud mediante grupos usando estas variables se pueden revelar patrones de la enfermedad no aparentes de otras maneras en los datos de vigilancia, esto puede apoyar al personal de salud pública en la prevención y control de enfermedades.

Tiempo

La distribución de los casos en el tiempo permite el establecimiento de hipótesis acerca del comportamiento de una enfermedad.

- Fecha de inicio de síntomas
- Fecha de hospitalización
- Fecha de notificación

Lugar

Los datos de la vigilancia también pueden ser analizados o comparados según el lugar en que ocurrieron. Un buen apoyo es la descripción gráfica de las notificaciones (mapeo) según espacios y población, especialmente a nivel local. El uso de sistemas de información geográfica no solamente puede mejorar la descripción gráfica de los eventos bajo vigilancia con relación a la variable lugar, sino también el análisis geoespacial de dichos eventos y la identificación de conglomerados y brotes. Se debe tratar de localizar el lugar en el que se originó la enfermedad, así como el lugar en el que se encontraba el paciente al momento de detección de la enfermedad.

- Lugar de residencia

#OrgullosamenteINS

 @INS Colombia
  @insaludColombia
  @INS Colombia

- Lugar de trabajo
- Lugares visitados

Persona

El análisis de los datos de vigilancia por las características de las personas afectadas es valioso para identificar los grupos de riesgo. La mayoría de los sistemas de vigilancia proporcionan información por edad y sexo. Otras variables utilizadas o que pueden estar disponibles son:

- nacionalidad
- nivel de inmunidad
- nutrición
- estilos de vida
- escolaridad
- área de trabajo
- hospitalización
- factores de riesgo
- nivel socioeconómico.

Los factores o características que se pueden usar para distinguir entre las personas enfermas o no, son las características personales, sus actividades y las condiciones de vida.

Medidas de frecuencia

El paso inicial del análisis de datos de la vigilancia es medir la frecuencia de los eventos de salud con el fin de hacer comparaciones entre distintas poblaciones o en la misma población a través del tiempo. No obstante, dado que el número absoluto de eventos depende en gran medida del tamaño de la población objeto de la vigilancia, estas comparaciones no se pueden realizar utilizando cifras de frecuencia absoluta (o número absoluto de eventos) (2).

Las medidas de frecuencia más usadas en epidemiología se refieren a la medición de la mortalidad y la enfermedad en una población. La mortalidad es útil para estudiar enfermedades que provocan la muerte, especialmente cuando su letalidad es importante. Las medidas de frecuencia pueden usar distintas escalas, casi todas matemáticas y sin duda son las más empleadas en la epidemiología para la labor descriptiva:

- Los números absolutos o conteo de casos
- Las proporciones
- Las tasas
- Las razones

Números absolutos o conteo de casos

Los números absolutos, recuentos o conteos pueden ser bastante útiles. Son útiles para brindar un panorama de la carga de la enfermedad, para la planificación y evaluación de los servicios de salud.

Figura 1. Conteo de casos: número de muertes en el mundo por causas seleccionadas, 2000 y 2015

#OrgullosamenteINS

 @INS Colombia  @insaludColombia  @INS Colombia

Enfermedades	2000	2015
Enfermedad isquémica del Corazón	6883	8756
Accidente Cerebro Vascular	5407	6241
Infección vía respiratoria inferior	3408	3190
Enfermedad pulmonar obstructiva	2953	3170
Cáncer: Tráquea/bronquios/pulmón	1255	1695
Diabetes mellitus	958	1586
Enfermedad diarreica	2177	1389
Tuberculosis	1667	1373

Fuente: Fuente: WHO. *Global Health Observatory. Top 10 causes of death. 2017*

Esta tabla muestra datos de la Organización Mundial de la Salud (OMS) sobre 12 de las principales causas de muerte a nivel mundial entre los años 2000 y 2015. Son solo recuentos y números simples que representan el número de muertes atribuidas a cada causa (por ejemplo: el número de muertes atribuido a enfermedad cardíaca, accidente cerebrovascular, etc.). Las dos principales causas de muerte tanto en el año 2000 como en el año 2015 fueron las enfermedades no transmisibles: las enfermedades cardíacas y los accidentes cerebrovasculares. El número de muertes para las categorías de enfermedades infecciosas (infecciones de las vías respiratorias inferiores, enfermedades diarreicas, tuberculosis y VIH / SIDA) disminuyó entre el 2000 y el 2015.

Proporciones

Las proporciones son medidas que expresan la frecuencia con la que ocurre un evento en relación con la población total en la cual éste puede ocurrir, es decir describe una parte en comparación con el todo, por ejemplo, qué parte de la población tiene una característica específica. Esta medida se calcula dividiendo el número de eventos ocurridos entre la población en la que ocurrieron. Como cada elemento de la población puede contribuir únicamente con un evento es lógico que al ser el numerador (el volumen de eventos) una parte del denominador (población en la que se presentaron los eventos) aquel nunca pueda ser más grande que éste. A menudo las proporciones se expresan en forma de porcentaje, y en tal caso los resultados oscilan entre cero y 100.

Ejemplo: Entre 10.000 adultos inscritos en una encuesta de presión arterial (PA), a 570 personas se les diagnosticó hipertensión arterial (definida como una medición de la PA diastólica > 95 mm Hg).

¿Qué proporción de los participantes de la encuesta tenían hipertensión?

Respuesta: $570 / 10.000 = 0.057 = 5.7\%$

Siguiendo el ejemplo de las causas globales de mortalidad que se mostró anteriormente; la columna de porcentajes (en texto verde) es la proporción de cada causa de muerte como porcentaje de las muertes por

#OrgullosamenteINS

 @INS Colombia
  @insaludColombia
  @INS Colombia

TODAS LAS CAUSAS en los años 2000 y 2015, entonces en el año 2000, el VIH / SIDA causó el 2.8% de las muertes en todo el mundo, pero el porcentaje cayó al 1.9% en 2015.

Figura 2. Proporción causas mundiales de mortalidad, 2000 y 2015.

Causas de mortalidad	n*	%	n*	%
Enfermedad isquémica del Corazón	6883	13.2	8756	15.5
Accidente Cerebro Vascular	5407	10.4	6241	11.1
Infección vía respiratoria inferior	3408	6.5	319	5.7
Enfermedad pulmonar obstructiva	2953	5.7	317	5.6
Cáncer: Tráquea/bronquios/pulmón	1244	2.4	1695	3.0
Diabetes mellitus	958	1.8	1586	2.8
Enfermedad diarreica	2177	4.2	1389	2.5
Tuberculosis	1667	3.2	1373	2.4
Accidentes de tránsito	1118	2.1	1342	2.4
Cirrosis hepática	905	1.7	1162	2.1
Nefropatía	709	1.4	1129	2.0
HIV/Sida	1463	2.8	106	1.9
Todas las causas	52135	100	56441	100

Fuente: Fuente: WHO. *Global Health Observatory. Top 10 causes of death. 2017*

Razones

Las razones pueden definirse como magnitudes que expresan la relación existente entre dos eventos en una misma población, o un solo evento en dos poblaciones. Las razones son la comparación de dos valores calculados al dividir un número por otro número. Una razón corresponde a una fracción de la forma a/b en donde tanto a y b , se refieren a la frecuencia en que ocurre algún evento y el numerador no está incluido en el denominador.

Los dos números pueden ser partes diferentes de un todo, por ejemplo, la razón entre hombres y mujeres.

Ejemplo: En 2014, el País A registró 202.143 niños nacidos vivos y 193.141 niñas nacidas vivas. Calcule la razón de niñas y niños nacidos vivos.

La razón es 202.143 dividido por 193,141. ¿Podemos simplificar?

Al dividir y luego multiplicar por 100, obtenemos una razón de 104,6 (aproximadamente 105) nacimientos de niños por cada 100 nacimientos de niñas.

O podemos decir que, por cada 100 nacimientos de niños, solo hay 95.4 nacimientos de niñas. Ambas razones son correctas y son matemáticamente lo mismo. La que elija dependerá de usted y de qué quiera comunicar.

La manera correcta de expresar el resultado del cálculo de una razón es señalar el número de elementos del numerador que existen por cada elemento del denominador.

Tasas

#OrgullosamenteINS

 @INS Colombia
  @insaludColombia
  @INS Colombia

Las tasas expresan la dinámica de un suceso en una población a lo largo del tiempo. Se pueden definir como la magnitud del cambio de una variable (enfermedad o muerte) por unidad de cambio de otra (usualmente el tiempo) en relación con el tamaño de la población que se encuentra en riesgo de experimentar el suceso.

En las tasas, el numerador expresa el número de eventos ocurridos durante un periodo en un número determinado de sujetos observados. A diferencia de una proporción el denominador de una tasa no expresa el número de sujetos en observación sino el tiempo durante el cual tales sujetos estuvieron en riesgo de sufrir el evento. La unidad de medida empleada se conoce como tiempo-persona de seguimiento. Por ejemplo, la observación de 100 individuos libres del evento durante un año corresponde a 100 años-persona de seguimiento; de manera similar, 10 sujetos observados durante diez años corresponden a 100 años-persona. Dado que el periodo entre el inicio de la observación y el momento en que aparece un evento puede variar de un individuo a otro, el denominador de la tasa se estima a partir de la suma de los periodos de todos los individuos. Las unidades de tiempo pueden ser horas, días, meses o años, dependiendo de la naturaleza del evento que se estudia.

El cálculo de tasas se realiza dividiendo el total de eventos ocurridos en un periodo dado en una población entre el tiempo-persona total (es decir, la suma de los periodos individuales libres de la enfermedad) en el que los sujetos estuvieron en riesgo de presentar el evento. Las tasas se expresan multiplicando el resultado obtenido por una potencia de 10, con el fin de permitir rápidamente su comparación con otras tasas. número de eventos ocurridos en una población en un periodo t sumatoria de los periodos durante los cuales los sujetos de la población libres del evento estuvieron expuestos al riesgo de presentarlo en el mismo periodo.

Ejemplo:

- la tasa de mortalidad infantil se considera uno de los indicadores mas importantes en las políticas nacionales. Este se calcula de la siguiente forma:

$$\text{Tasa de mortalidad infantil} = \frac{\# \text{ número de muertes en niños menores de 1 año}}{\# \text{ número de nacidos vivos}} \times 1000$$

Indicador

Un indicador es una medida que permite ser comparada como un referente que estima la realidad de una situación determinada (3). Cuando se estudia la salud y la población, las situaciones de referencia de donde se pueden obtener indicadores son:

- La muerte
- La enfermedad
- La discapacidad de la población
- Los factores de riesgo de la población
- Aspectos intervinientes en la salud poblacional tales como prevalencia de felicidad, calidad de vida, desarrollo humano.

2. Interpretación de información.

#OrgullosamenteINS

 @INS Colombia  @insaludColombia  @INS Colombia

La interpretación de los hallazgos del análisis sirve para la generación de hipótesis, para lo cual debe tenerse en consideración una serie de posibles explicaciones alternativas. Factores tales como el aumento de la población, la migración, la introducción de nuevos métodos diagnósticos, el mejoramiento de los sistemas de notificación, el cambio en la definición de casos, la aparición de nuevos y efectivos tratamientos y la posibilidad de problemas con la validez de los datos de vigilancia, por subregistro, sesgos o duplicación de notificaciones pueden producir resultados espurios o falsos.

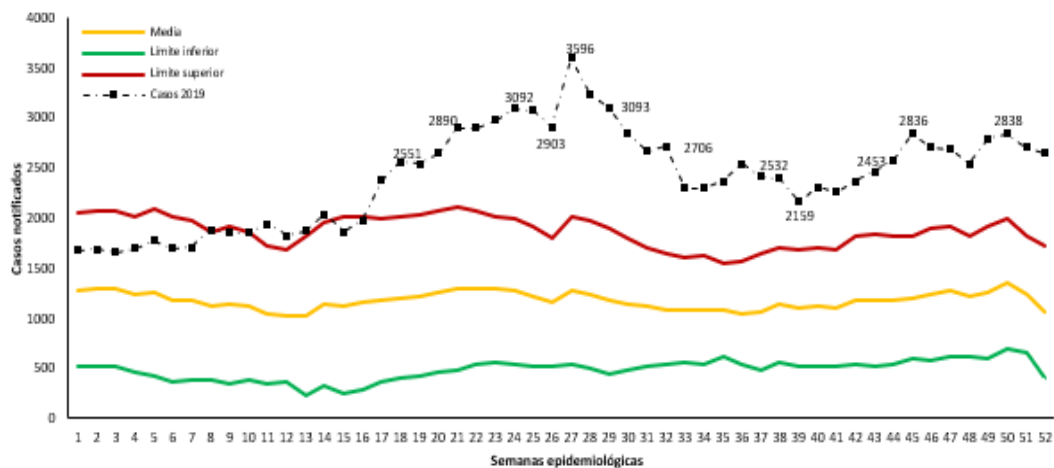
La interpretación debe seguir un proceso sistemático. Pueden verse como pasos o como métodos diferentes de interpretar los datos:

- Explicar las mediciones epidemiológicas y estadísticas en lenguaje sencillo. Esto implica el uso de palabras comunes para explicar qué sucede y colocar la información en contexto.

Ejemplo. “La incidencia estimada de VIH en Colombia en el 2019 fue de 22,1 por 100 000 habitantes y la prevalencia es de 0,2 %. La razón hombre mujer fue de 4,4; el promedio de edad fue de 33,2 años, la mediana fue de 30 años y el rango de edad es de 0 a 85 años”. Lo que se podría decir de manera sencilla es: “en el 2019 por cada 100 000 habitantes hubo 22 personas a las que se les diagnóstico VIH y en el país por cada 100 habitantes hay menos de una persona que vive con VIH o Sida. Por cada 4 hombres con VIH/Sida hay una mujer. La mitad de las personas que vive con VIH en el país tiene menos de 30 años y la edad de las personas que vive con VIH en Colombia oscila entre 0 y 85 años”. (4)

- Comparar lo que se observó con lo que espera o comparar observaciones con un nivel que esté establecido por el gobierno, por la Organización Mundial de la Salud o por otros; aquí es donde se utilizan los umbrales.

Figura 1. Canal endémico de dengue, Colombia, 2019



Fuente: Instituto Nacional de Salud

En este gráfico observamos el canal endémico para dengue en Colombia en el año 2019, elaborado con las medias geométricas de tasas históricas. Por debajo del límite inferior en este caso la línea verde se habla de decremento en el número de casos, entre la línea verde y la línea amarilla, que es la media geométrica el evento

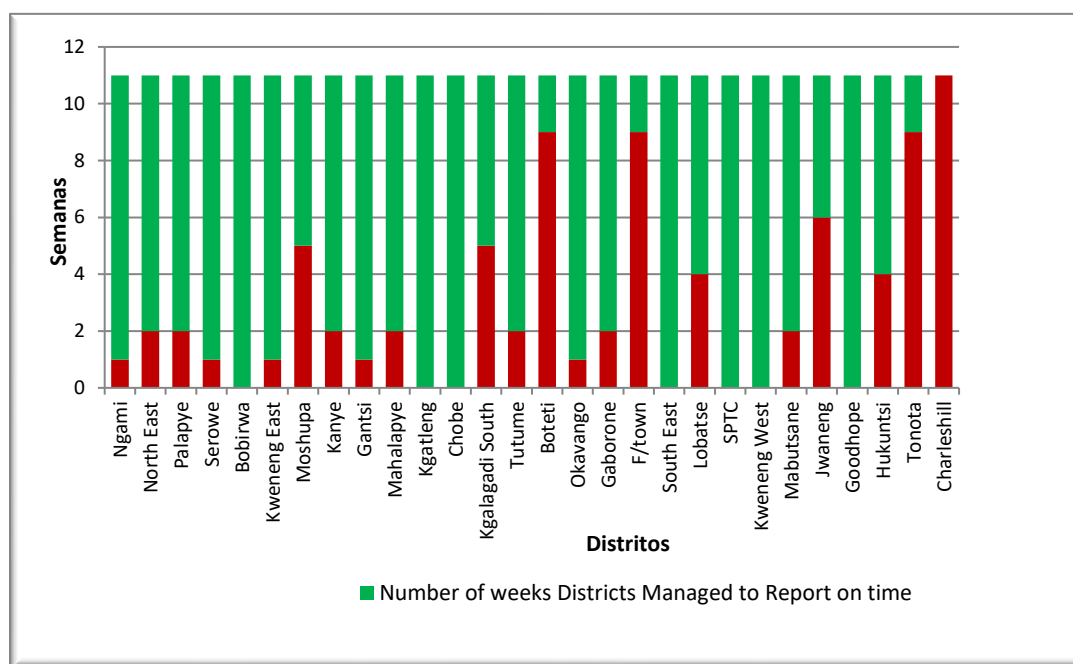
#OrgullosamenteINS

 @INS Colombia
  @insaludColombia
  @INS Colombia

se encuentra dentro de la esperado; sobre la línea geométrica y el límite superior, la línea roja se habla de zona de alerta y sobre el límite superior podemos hablar de epidemia. En el ejemplo podemos observar que desde el inicio del año el país se encontraba en alerta y desde la semana epidemiológica 8 el país entró en epidemia por dengue, el mayor número de casos se presentó en la Semana Epidemiológica 27, hay un descenso hasta la semana 39, pero vuelve a aumentar manteniéndose la epidemia hasta el final del 2019.

- c. Tener en cuenta la calidad de los datos. Haga la pregunta: ¿en qué medida son correctos los datos? ¿Todos los centros informaron esta semana? ¿Hay muchos errores?

Grafico 4. Número de semanas que el Distrito gestionó y no informó a tiempo, Botsuana, Semanas 1 a 11, 2015.



Fuente: Botswana IDSW Weekly Report, 2014

Este gráfico muestra la cantidad de semanas que cada distrito informó o no informó a tiempo; el color verde significa a tiempo y rojo significa que no se informó a tiempo. Entonces, observemos Goodhope, completamente verde, informaron a tiempo cada semana. Observe Charleshill, completamente rojo, jamás informaron a tiempo.

Por lo tanto, algo por considerar cuando se interpretan datos es la calidad de informar no solo la puntualidad sino también la precisión. Además, es posible que quiera saber algo acerca de cómo se recopilan los datos. ¿Estos se informan por si mismos? ¿Existe una definición de caso estándar? ¿La definición de caso cambió recientemente? ¿Los casos están basados en informes clínicos? En caso de que sea así, podría haber variaciones en el discernimiento entre los especialistas clínicos. ¿Qué sucede si algunas áreas tienen confirmación de laboratorio y otras no? Todas estas situaciones podrían afectar cómo reaccionan ustedes a los datos.

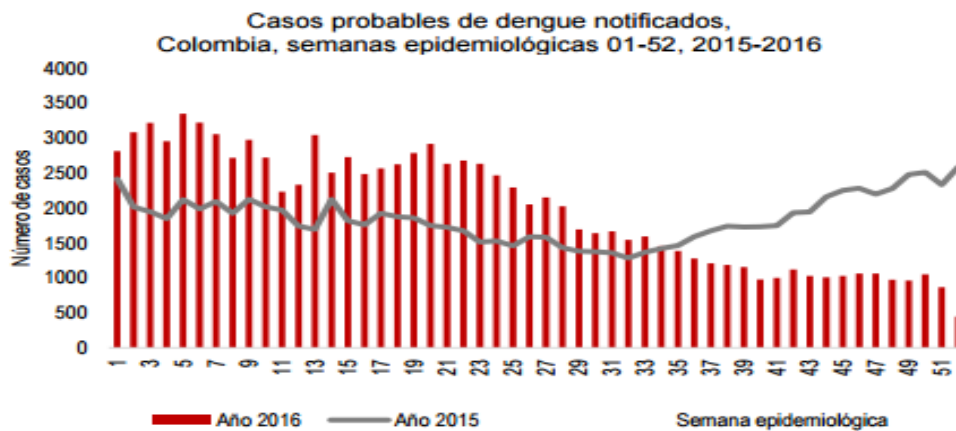
#OrgullosamenteINS

 @INS Colombia
  @insaludColombia
  @INS Colombia

d. Considere las posibles explicaciones por el posible incremento en los casos. ¿El incremento representa un verdadero brote o epidemia o un mejor informe o existe alguna otra explicación?

- Incremento verdadero en la ocurrencia de la enfermedad
- Brote / epidémico
- Patrón estacional
- Incremento repentino en el tamaño de la población
- Cambio en los procedimientos de informe o el sistema de vigilancia
- Cambio en la definición del caso
- Incremento o mejora en pruebas de laboratorio o procedimientos diagnósticos
- Mayor conciencia de la enfermedad (público o proveedores)
- Mayor acceso a la atención médica
- Nueva clínica, notificador o proveedor de atención médica
- Error de diagnóstico o laboratorio
- Informes en lote

e. Hacer inferencias acerca de la ocurrencia de la enfermedad a partir de los datos del resumen. Presentar conclusiones acerca de la aparición de la enfermedad a partir de los datos del resumen. Esto implica observar el gráfico, el cuadro o el mapa y determinar qué le sugiere. De las posibles explicaciones, ¿cuál es la más probable?



Fuente: Sivigila, Instituto Nacional de Salud, Colombia, 2015-2016

En esta gráfica se observa los casos de dengue en los dos años, lo que podría sugerir que hay epidemia de dengue en Colombia por las primeras 34 semanas del 2016.

Las mediciones epidemiológicas y estadísticas son formas de entender y describir lo que está pasando en la salud de una comunidad. Usamos palabras comunes para explicar qué está ocurriendo y ponemos la información en contexto, como si estuviéramos contando una historia.

#OrgullosamenteINS

 @INS Colombia  @insaludColombia  @INS Colombia

Finalmente, después de mirar todo esto, podemos contar cómo se está comportando la enfermedad y hacer un resumen. Así es como los expertos en salud pública nos ayudan a entender lo que está sucediendo y cómo mantenernos saludables.

Bibliografía

1. Módulos de principios de epidemiología para el control de enfermedades (MOPECE). Segunda Edición Revisada Unidad 4: Vigilancia en salud pública [Internet]. [citado 10 de marzo de 2023]. Disponible en: <https://www.paho.org/col/dmdocuments/MOPECE4.pdf>
2. Moreno-Altamirano A y col. Principales medidas en epidemiología. Rev Salud Pública de México / vol.42, no.4, julio-agosto de 2000.
3. Guillermo Restrepo Ch. Juan Carlos González Q. Libro de Biometría Comunitaria. Fundación Universitaria Juan N Corpas. Colombia 2010.
4. Fondo Colombiano de Enfermedades de Alto Costo, Cuenta de Alto Costo (CAC). Situación del VIH en Colombia 2019; Bogotá D.C. 2020.

#OrgullosamenteINS

 @INS Colombia  @insaludColombia  @INS Colombia