

Manual de Epidemiología

para Ciencias de la Salud



Alfredo Sánchez Alberca
asalber@ceu.es
<https://aprendeconalf.es>

Índice de contenidos

Prefacio	4
Licencia	4
1 Introducción a la Epidemiología	6
1.1 ¿Qué es la Epidemiología?	6
1.1.1 Algunos descubrimientos históricos	7
1.2 El lenguaje de la Epidemiología	7
1.3 Índices epidemiológicos	9
2 Probabilidad y odds	10
2.1 El concepto de probabilidad	10
2.1.1 Algunas propiedades de la probabilidad	11
2.1.2 Interpretación de una probabilidad	11
2.2 El concepto de odds	12
2.2.1 Interpretación de un odds	13
2.2.2 Conversión de odds en probabilidades	13
3 Medidas de frecuencia de una enfermedad	14
3.1 Prevalencia	14
3.2 Incidencia o riesgo absoluto	14
3.3 Prevalencia frente a incidencia	15
3.4 Algunas consideraciones en el caso de la COVID	16
4 Medidas de asociación	17
4.1 Riesgo relativo	17
4.1.1 Interpretación del riesgo relativo	18
4.2 Odds ratio	18
4.2.1 Interpretación del odds ratio	20
4.3 Riesgo relativo frente a odds ratio	20
4.4 Aplicación a la COVID	21
5 Tests diagnósticos	22
5.1 Sensibilidad y especificidad de un test	22
5.1.1 Cuándo usar un test más sensible o más específico	24
5.2 Valores predictivos de un test	25
5.2.1 Interpretación de los valores predictivos	26

6	La curva ROC	27
6.1	Definición de la curva ROC	27
6.1.1	Interpretación de la curva ROC	27
6.2	Área debajo de la curva ROC (AUC)	28
6.3	Aplicaciones a la COVID	29
	Apéndices	30
	Referencias	30
	Recursos generales	30
	Datos epidemiológicos	30
	Aplicaciones interactivas	30
	Otros manuales del autor	30

Prefacio

¡Bienvenida/os al manual de Epidemiología!

Este libro es una introducción a los conceptos básicos de la Epidemiología para alumnos de grados en Ciencias de la Salud.

La pandemia de la COVID-19 puso en boca de todos un montón de términos técnicos de la Epidemiología (prevalencia, incidencia, riesgo relativo, sensibilidad, falsos positivos, etc.) que, con demasiada frecuencia, se utilizan de manera errónea, incluso por los propios medios de comunicación, generando confusión en la población no experta.

El objetivo de este manual es explicar los principales índices epidemiológicos utilizados en el control de las enfermedades e ilustrar su uso con ejemplos de aplicación. Como todos ellos se basan en el cálculo de probabilidades, el libro comienza repasando el concepto de probabilidad y sus principales propiedades, para después presentar

- las medidas de frecuencia de una enfermedad (prevalencia e incidencia),
- las medidas de asociación entre un factor de exposición y una enfermedad (riesgo relativo y odds ratio), y
- las medidas de fiabilidad de los tests diagnósticos (sensibilidad, especificidad, valores predictivos y curva ROC).

Este libro se complementa con los siguientes recursos:

- [Manual de Estadística](#)
- [Colección de problemas resueltos](#)
- [Prácticas de Estadística con R](#)
- [Aplicación para el estudio de test diagnósticos](#)

Licencia

Esta obra está bajo una licencia Reconocimiento – No comercial – Compartir bajo la misma licencia 3.0 España de Creative Commons. Para ver una copia de esta licencia, visite <https://creativecommons.org/licenses/by-nc-sa/3.0/es/>.

Con esta licencia eres libre de:

- Copiar, distribuir y mostrar este trabajo.
- Realizar modificaciones de este trabajo.

Bajo las siguientes condiciones:

- **Reconocimiento** Debe reconocer los créditos de la obra de la manera especificada por el autor o el licenciador (pero no de una manera que sugiera que tiene su apoyo o apoyan el uso que hace de su obra).
- **No comercial** No puede utilizar esta obra para fines comerciales.
- **Compartir bajo la misma licencia** Si altera o transforma esta obra, o genera una obra derivada, sólo puede distribuir la obra generada bajo una licencia idéntica a ésta.

Al reutilizar o distribuir la obra, tiene que dejar bien claro los términos de la licencia de esta obra.

Estas condiciones pueden no aplicarse si se obtiene el permiso del titular de los derechos de autor.

Nada en esta licencia menoscaba o restringe los derechos morales del autor.

1 Introducción a la Epidemiología

1.1. ¿Qué es la Epidemiología?

La palabra *epidemiología* viene del griego *epi* (sobre), *demos* (gente) y *logos* (estudio), es decir, el estudio de lo que le ocurre a las poblaciones.

Definición 1.1 (Epidemiología). En el ámbito de la salud pública, la *Epidemiología* es la rama de la Medicina que se encarga del estudio de la distribución y las causas de los eventos relacionados con la salud (normalmente enfermedades) en las poblaciones, y de la aplicación de este estudio al control de los problemas de salud pública.

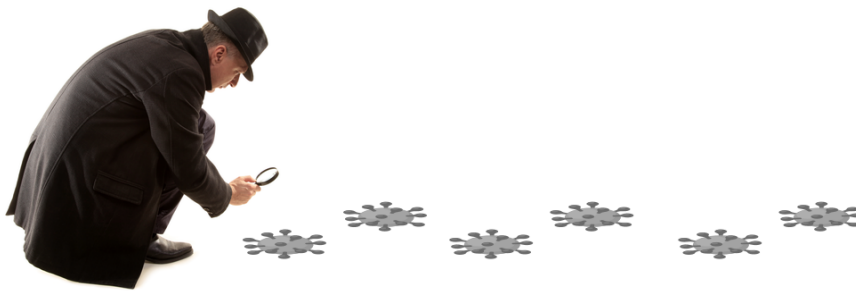


Figura 1.1: El epidemiólogo trabaja como un detective que busca las causas de una enfermedad

Obsérvese que, a diferencia de la Medicina clínica, que se ocupa del diagnóstico y el tratamiento de un paciente concreto, la Epidemiología estudia siempre *poblaciones*. Sus dos preguntas fundamentales son:

- ¿Cuánta enfermedad hay y cómo se distribuye? (*Epidemiología descriptiva*)
- ¿Qué factores la causan o la previenen? (*Epidemiología analítica*)

Debido a la epidemia provocada por el coronavirus, la Epidemiología se ha convertido en una de las ramas de la Medicina que más interés despiertan.

Sin embargo, mucho antes de la COVID, la Epidemiología ya había servido en otros momentos históricos para resolver o controlar algunos de los problemas de salud pública más serios que ha enfrentado la humanidad.



Figura 1.2: Fernando Simón, director del Centro de Coordinación de Alertas y Emergencias Sanitarias durante la pandemia de COVID-19

1.1.1. Algunos descubrimientos históricos

- **1854:** John Snow determina que la causa de la epidemia de cólera que asolaba Londres era que el agua estaba contaminada con heces.
- **1898:** Ronald Ross averigua que el transmisor de la malaria es el mosquito *Anopheles*.
- **1950:** Se descubre que fumar es el principal factor de riesgo del cáncer de pulmón.
- **1954:** Se valida la primera vacuna contra la poliomielitis (la de Jonas Salk).
- **1970:** Se observa que el ejercicio físico y una dieta sana reducen el riesgo de sufrir un infarto.
- **1983:** Robert Gallo, Luc Montagnier y Françoise Barré-Sinoussi identifican el virus que causa el SIDA. Poco después se observó que el riesgo de contraer el VIH aumentaba con ciertas prácticas sexuales y con el consumo de algunos tipos de drogas.
- **2020:** Y llegó la COVID...

1.2. El lenguaje de la Epidemiología

En estos tiempos de pandemia un montón de términos técnicos de la Epidemiología se han convertido en lugares comunes gracias a los medios de comunicación.

Sin embargo, muchos de estos términos se utilizan de manera errónea, incluso por los propios medios de comunicación, y generan confusión en la población no experta. En este manual se explican los principales conceptos epidemiológicos usados en el control de enfermedades como la COVID y se ilustra su uso con ejemplos de aplicación.



Figura 1.3: Términos epidemiológicos habituales en los medios de comunicación

1.3. Índices epidemiológicos

Los índices epidemiológicos que se estudian en este manual pueden agruparse en tres bloques:

Medidas de frecuencia de una enfermedad (Capítulo 3) Miden cuánta enfermedad hay en una población.

- Prevalencia.
- Incidencia o riesgo absoluto.

Medidas de asociación (Capítulo 4) Miden la relación entre la exposición a un factor y la aparición de la enfermedad.

- Riesgo y odds.
- Riesgo relativo y odds ratio.

Medidas de fiabilidad de los tests diagnósticos (Capítulo 5 y Capítulo 6) Miden la capacidad de un test para diagnosticar o descartar la enfermedad.

- Sensibilidad y especificidad.
- Valores predictivos.
- Curva ROC y área bajo la curva.

Todos estos índices están basados en el cálculo de probabilidades, por lo que en el siguiente capítulo se introduce el concepto de probabilidad y sus principales propiedades.

2 Probabilidad y odds

Todos los índices epidemiológicos que se estudian en este manual son, en el fondo, probabilidades o cocientes de probabilidades. Por eso, antes de definirlos, conviene repasar el concepto de probabilidad, sus propiedades y su interpretación, así como una forma alternativa de medir la incertidumbre: el *odds*.

2.1. El concepto de probabilidad

A lo largo de la historia ha habido diferentes intentos de definir matemáticamente el concepto de probabilidad. Quizá la más conocida, y la primera que se enseña en las escuelas, es la definición clásica de Laplace.

Definición 2.1 (Probabilidad: definición clásica de Laplace). Dado un experimento aleatorio con espacio muestral Ω en el que todos los casos posibles son equiprobables, la *probabilidad* de un suceso E es

$$P(E) = \frac{|E|}{|\Omega|} = \frac{\text{Casos favorables a } E}{\text{Casos posibles}}$$

Ejemplo 2.1. Al tirar un dado equilibrado, la probabilidad de sacar un número par $E = \{2, 4, 6\}$ es

$$P(E) = \frac{3}{6} = 0.5.$$

Sin embargo, esta definición tiene serios inconvenientes ya que, para poder usarla, todos los casos posibles del experimento deben tener la misma probabilidad de ocurrir (*equiprobabilidad*), y esto no suele ocurrir en la vida real (por ejemplo, no todos los grupos sanguíneos tienen la misma probabilidad de ocurrir).

Por este motivo, en el ámbito de las Ciencias de la Salud es mucho más común utilizar la definición de probabilidad basada en el cálculo de frecuencias.

Definición 2.2 (Probabilidad: definición frecuentista). La *probabilidad* de un suceso E se aproxima mediante su frecuencia relativa en una muestra de tamaño n ,

$$P(E) \approx f_E = \frac{n_E}{n} = \frac{\text{Frecuencia absoluta del suceso}}{\text{Tamaño muestral}}$$

Ejemplo 2.2. Se ha aplicado un tratamiento a 100 personas y se han curado 75. Entonces, la probabilidad de curación del tratamiento es

$$P(E) = \frac{75}{100} = 0.75 \Rightarrow 75 \%.$$

 Advertencia

La definición frecuentista no permite calcular el valor exacto de la probabilidad de un suceso, sino tan solo una aproximación, que será mejor cuanto mayor sea el tamaño de la muestra.

2.1.1. Algunas propiedades de la probabilidad

- Una probabilidad es un número real entre 0 y 1:

$$0 \leq P(E) \leq 1.$$

- La suma de las probabilidades de todos los casos posibles es 1:

$$P(\Omega) = P(e_1) + P(e_2) + \dots + P(e_n) = 1.$$

- La probabilidad de que ocurra lo contrario de un suceso es 1 menos la probabilidad del suceso:

$$P(\overline{E}) = 1 - P(E).$$

De este modo, cuanto más probable es que ocurra un suceso, menos probable es que ocurra su contrario, y viceversa.

2.1.2. Interpretación de una probabilidad

La probabilidad mide la verosimilitud de un suceso. De manera informal, se puede decir que la probabilidad mide la creencia o la confianza que tenemos en la ocurrencia de un suceso.

Interpretación

- $P(E) = 0 \Rightarrow$ Mínima verosimilitud (el suceso no ocurre nunca).
- $P(E) = 0.5 \Rightarrow$ Verosimilitud media (máxima incertidumbre).
- $P(E) = 1 \Rightarrow$ Máxima verosimilitud (el suceso ocurre siempre).



Figura 2.1: Escala de interpretación de una probabilidad

Aunque el concepto de probabilidad es el más extendido en las aplicaciones que requieren cuantificar la incertidumbre sobre la ocurrencia de un suceso, existen otras formas de cuantificar esa incertidumbre, como por ejemplo el *odds*.

2.2. El concepto de odds

Definición 2.3 (Odds). El *odds* de un suceso E es el cociente entre el número de casos en los que ocurre E y el número de casos en los que no ocurre,

$$O(E) = \frac{\text{N.º casos con } E}{\text{N.º casos sin } E} = \frac{P(E)}{P(\bar{E})}$$

Ejemplo 2.3. Se ha aplicado un tratamiento a 100 personas y se han curado 75. Entonces, el odds de curación del tratamiento es

$$O(E) = \frac{75}{25} = 3,$$

es decir, por cada persona que no se cura hay 3 que se curan.

Advertencia

A diferencia de una probabilidad, un odds puede ser mayor que 1. De hecho, no tiene cota superior.

2.2.1. Interpretación de un odds

Los odds también permiten cuantificar la verosimilitud de un suceso..., pero en una escala diferente, ya que se trata de una razón de probabilidades.

i Interpretación

- $O(E) = 0 \Rightarrow$ Mínima verosimilitud.
- $O(E) = 1 \Rightarrow$ Verosimilitud media (máxima incertidumbre).
- $O(E) = \infty \Rightarrow$ Máxima verosimilitud.

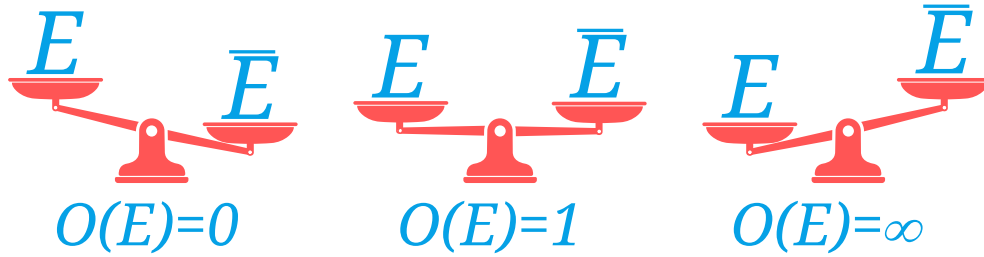


Figura 2.2: Escala de interpretación de un odds

2.2.2. Conversión de odds en probabilidades

Como el odds es una función de la probabilidad, es posible convertir un odds en una probabilidad aplicando la siguiente fórmula:

$$\frac{O(E)}{1 + O(E)} = \frac{\frac{P(E)}{P(\bar{E})}}{1 + \frac{P(E)}{P(\bar{E})}} = \frac{\frac{P(E)}{P(\bar{E})}}{\frac{P(\bar{E}) + P(E)}{P(\bar{E})}} = P(E).$$

Ejemplo 2.4. Se ha aplicado un tratamiento a 100 personas y se han curado 75, de manera que

$$O(E) = \frac{75}{25} = 3 \Rightarrow P(E) = \frac{3}{1 + 3} = 0.75.$$

Y, del mismo modo, se puede pasar de una probabilidad a un odds mediante la fórmula

$$O(E) = \frac{P(E)}{1 - P(E)}.$$

3 Medidas de frecuencia de una enfermedad

La primera tarea de la Epidemiología es cuantificar cuánta enfermedad hay en una población. Para ello se utilizan dos medidas de frecuencia que se confunden a menudo, pero que responden a preguntas distintas: la *prevalencia* y la *incidencia*.

3.1. Prevalencia

Definición 3.1 (Prevalencia). La *prevalencia* de una enfermedad E es la proporción de personas que tienen la enfermedad en un momento concreto,

$$\text{Prevalencia}(E) = \frac{\text{N.º individuos afectados por } E}{\text{Tamaño poblacional}}$$

Ejemplo 3.1. En una muestra de 1000 personas, 150 tenían gripe. La prevalencia de la gripe es aproximadamente

$$\frac{150}{1000} = 0.15 \Rightarrow 15 \%.$$

Obsérvese que en el numerador de la prevalencia se cuentan *todos* los casos existentes en ese instante, tanto los nuevos como los antiguos.

3.2. Incidencia o riesgo absoluto

Definición 3.2 (Incidencia). La *incidencia* o *riesgo absoluto* de una enfermedad E es la proporción de nuevos casos durante un periodo determinado (por día, por semana, por mes, etc.),

$$R(E) = \frac{\text{N.º nuevos casos con } E \text{ en el periodo}}{\text{Tamaño de la población en riesgo al comienzo del periodo}}$$

Ejemplo 3.2. Al comienzo del año se tomó una muestra de 1000 personas sin gripe y al finalizar el año 80 habían tenido gripe. La incidencia de la gripe ese año fue

$$R(E) = \frac{80}{1000} = 0.08 \Rightarrow 8\%.$$

! Importante

En la incidencia, el denominador no es el tamaño de la población, sino el tamaño de la *población en riesgo*, es decir, las personas que al comienzo del periodo podían enfermar (no se cuentan ni los ya enfermos ni los inmunizados).

3.3. Prevalencia frente a incidencia

Tabla 3.1: Diferencias entre la prevalencia y la incidencia

	Tiempo	Casos	Tipo de estudio
Prevalencia	Puntual	Nuevos y existentes	Transversal
Incidencia	Periodo	Solo nuevos	Longitudinal

- La prevalencia muestra el número de personas afectadas (la *carga* de la enfermedad) y es útil para planificar los recursos sanitarios necesarios.
- La incidencia muestra la evolución de la enfermedad y es más útil para detectar brotes y estudiar sus causas.
- La incidencia depende sobre todo de la contagiosidad de la enfermedad, mientras que la prevalencia depende también de la duración de la enfermedad y de lo agresiva que sea.

Ejemplo 3.3. Una enfermedad crónica como la diabetes o la artrosis tiene una incidencia prácticamente constante, al depender fundamentalmente de la edad y no ser contagiosa, y una prevalencia alta, ya que no existe cura para ella y las personas tampoco mueren a causa de ella, sino que viven con ella el resto de su vida.

Por otro lado, una enfermedad como el ébola tiene una incidencia pequeña, al no ser muy contagiosa, y también una prevalencia pequeña, al ser una enfermedad mortal, ya que casi todas las personas que se contagian acaban muriendo.

Finalmente, una enfermedad como la COVID tiene una incidencia muy alta, al ser muy contagiosa, y una prevalencia parecida a la incidencia, ya que la enfermedad suele acabar en un periodo de dos semanas desde el contagio (salvo en los casos que necesitan hospitalización).

3.4. Algunas consideraciones en el caso de la COVID

La incidencia de la COVID se suele dar sobre un periodo de dos semanas (14 días), aunque no siempre, y por cada 100 000 habitantes.

Los datos son poco precisos y subestiman el riesgo de la COVID:

- Muchos asintomáticos no son detectados.
- La detección de casos se realiza mediante tests diagnósticos que tienen un margen de error (falsos positivos y falsos negativos), como se verá en el Capítulo 5.
- Se calcula dividiendo por el tamaño de la población (nuevos casos por cada 100 000 habitantes), pero habría que dividir por el tamaño de la población en riesgo (sin contar los ya infectados o inmunizados).

Los datos oficiales pueden consultarse en la página del [Ministerio de Sanidad](#).

Tanto la prevalencia como la incidencia permiten estudiar la magnitud y la evolución de una enfermedad, pero no permiten analizar sus posibles causas. Para ello es necesario comparar los riesgos de distintos grupos de población, que es lo que se verá en el siguiente capítulo.

4 Medidas de asociación

Tanto la prevalencia como la incidencia permiten estudiar la magnitud y la evolución de una enfermedad, pero no permiten analizar sus posibles causas. Cuando se quiere investigar si la exposición a un determinado factor influye en el desarrollo de una enfermedad hay que comparar los riesgos de dos grupos:

- **Grupo tratamiento T :** individuos expuestos al factor.
- **Grupo control C :** individuos no expuestos al factor.

Los resultados del estudio se resumen en una tabla de contingencia como la siguiente, donde E es el suceso de padecer la enfermedad y $\bar{E}$ el de no padecerla.

Tabla 4.1: Tabla de contingencia de un estudio epidemiológico

	E	$\bar{E}$
Grupo tratamiento T	a	b
Grupo control C	c	d

A partir de esta tabla se pueden definir dos medidas de asociación entre el factor y la enfermedad: el *riesgo relativo* y el *odds ratio*.

4.1. Riesgo relativo

Definición 4.1 (Riesgo relativo). El *riesgo relativo* de un suceso E es el cociente entre el riesgo del grupo tratamiento y el riesgo del grupo control,

$$RR(E) = \frac{\text{Riesgo grupo tratamiento}}{\text{Riesgo grupo control}} = \frac{R_T(E)}{R_C(E)} = \frac{a/(a+b)}{c/(c+d)}$$

Ejemplo 4.1. Para estudiar la eficacia de una vacuna contra la gripe se ha tomado una muestra de 500 personas vacunadas y otra de 500 no vacunadas, obteniéndose los siguientes resultados.

	Gripe G	No gripe $\bar{G}$
Vacunados T	20	480
No vacunados C	80	420

El riesgo relativo de padecer gripe es

$$RR(G) = \frac{20/(20 + 480)}{80/(80 + 420)} = 0.25,$$

es decir, el riesgo de padecer gripe en los vacunados es la cuarta parte del riesgo en los no vacunados.

4.1.1. Interpretación del riesgo relativo

Interpretación

- $RR = 1 \Rightarrow$ No hay asociación entre el suceso y la exposición al factor.
- $RR < 1 \Rightarrow$ La exposición al factor disminuye el riesgo del suceso (factor de protección).
- $RR > 1 \Rightarrow$ La exposición al factor aumenta el riesgo del suceso (factor de riesgo).

4.2. Odds ratio

Del mismo modo que se pueden comparar los riesgos de los grupos tratamiento y control, se pueden comparar también sus odds.

Definición 4.2 (Odds ratio). El *odds ratio* o *razón de odds* de un suceso E es el cociente entre el odds del grupo tratamiento y el odds del grupo control,

$$OR(E) = \frac{\text{Odds grupo tratamiento}}{\text{Odds grupo control}} = \frac{O_T(E)}{O_C(E)} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

Ejemplo 4.2. Con los datos del ejemplo anterior,

	Gripe G	No gripe $\bar{G}$
Vacunados T	20	480
No vacunados C	80	420



Figura 4.1: Escala de interpretación del riesgo relativo

el odds ratio de padecer gripe es

$$OR(G) = \frac{20/480}{80/420} = 0.22.$$

4.2.1. Interpretación del odds ratio

Interpretación

- $OR = 1 \Rightarrow$ No existe asociación entre el suceso y la exposición al factor.
- $OR < 1 \Rightarrow$ La exposición al factor disminuye el riesgo del suceso.
- $OR > 1 \Rightarrow$ La exposición al factor aumenta el riesgo del suceso.

4.3. Riesgo relativo frente a odds ratio

El riesgo relativo es una comparación de probabilidades, pero depende de la incidencia de la enfermedad.

La interpretación del odds ratio es más enrevesada porque es contrafactual, ya que indica cuántas veces es más frecuente el suceso en el grupo tratamiento en comparación con el control, asumiendo que en el grupo control es tan frecuente que ocurra el suceso como que no. Su ventaja es que no depende de la incidencia de la enfermedad, lo que lo hace imprescindible en los estudios de casos y controles, donde el investigador fija de antemano el número de enfermos y sanos de la muestra.

Ejemplo 4.3. Para estudiar la asociación entre fumar y el cáncer de pulmón se han tomado dos muestras, la segunda con el doble de pacientes sanos que la primera.

Muestra 1	Cáncer E	No cáncer $\bar{E}$
Fumadores T	60	80
No fumadores C	40	320

$$RR(E) = \frac{60/(60 + 80)}{40/(40 + 320)} = 3.86,$$

$$OR(E) = \frac{60/80}{40/320} = 6.$$

Muestra 2	Cáncer E	No cáncer $\bar{E}$
Fumadores T	60	160
No fumadores C	40	640

$$RR(E) = \frac{60/(60 + 160)}{40/(40 + 640)} = 4.64,$$

$$OR(E) = \frac{60/160}{40/640} = 6.$$

Obsérvese que el riesgo relativo cambia al cambiar la proporción de sanos de la muestra, mientras que el odds ratio se mantiene constante.

4.4. Aplicación a la COVID

Durante la pandemia se publicaron multitud de estudios que utilizaban estas medidas de asociación para identificar los factores de riesgo y de protección frente a la COVID:

- La edad aumenta la gravedad.
- El riesgo de infección depende del grupo sanguíneo.
- El déficit de vitamina D aumenta el riesgo de infección.
- Las personas con demencia tienen mayor riesgo de infectarse.
- El remdesivir acelera la recuperación.

5 Tests diagn3sticos

Otra aplicaci3n de la Epidemiolog3a basada en el c3lculo de probabilidades son los *tests diagn3sticos*.

Definici3n 5.1 (Test diagn3stico). Un *test diagn3stico* es una prueba usada para diagnosticar una enfermedad o descartarla.

Normalmente producen dos resultados: positivo (+), a favor de la enfermedad, y negativo (−), en contra de ella. Al comparar el resultado del test con el verdadero estado del paciente se obtienen cuatro posibilidades:

Tabla 5.1: Resultados posibles de un test diagn3stico

Test	E	$\bar{E}$
Positivo +	Verdadero positivo (VP)	Falso positivo (FP)
Negativo −	Falso negativo (FN)	Verdadero negativo (VN)

Un test perfecto ser3a aquel que no produce ni falsos positivos ni falsos negativos, pero en la pr3ctica ning3n test lo consigue. Por eso es necesario medir su fiabilidad.

5.1. Sensibilidad y especificidad de un test

La fiabilidad *a priori* de un test diagn3stico, es decir, antes de aplicarlo, depende de las siguientes probabilidades.

Definici3n 5.2 (Sensibilidad). La *sensibilidad* de un test diagn3stico es la proporci3n de resultados positivos del test entre las personas con la enfermedad,

$$P(+|E) = \frac{VP}{VP + FN}$$

Definición 5.3 (Especificidad). La *especificidad* de un test diagnóstico es la proporción de resultados negativos del test entre las personas sin la enfermedad,

$$P(-|\overline{E}) = \frac{VN}{VN + FP}$$

Ejemplo 5.1. Un test de antígenos para detectar el SARS-CoV-2 tiene una sensibilidad del 70 % y una especificidad del 95 %. Esto significa que:

- Si aplicamos el test a 100 enfermos, dará 70 positivos y 30 negativos.
- Si aplicamos el test a 100 sanos, dará 95 negativos y 5 positivos.

Ahora bien, la utilidad del test depende también de la prevalencia de la enfermedad en la población a la que se aplica.

Ejemplo 5.2. Utilizando el test del ejemplo anterior en una población de 1000 personas y suponiendo una prevalencia del 1 %, se tiene

Test	E	$\overline{E}$
Positivo +	7	50
Negativo −	3	940

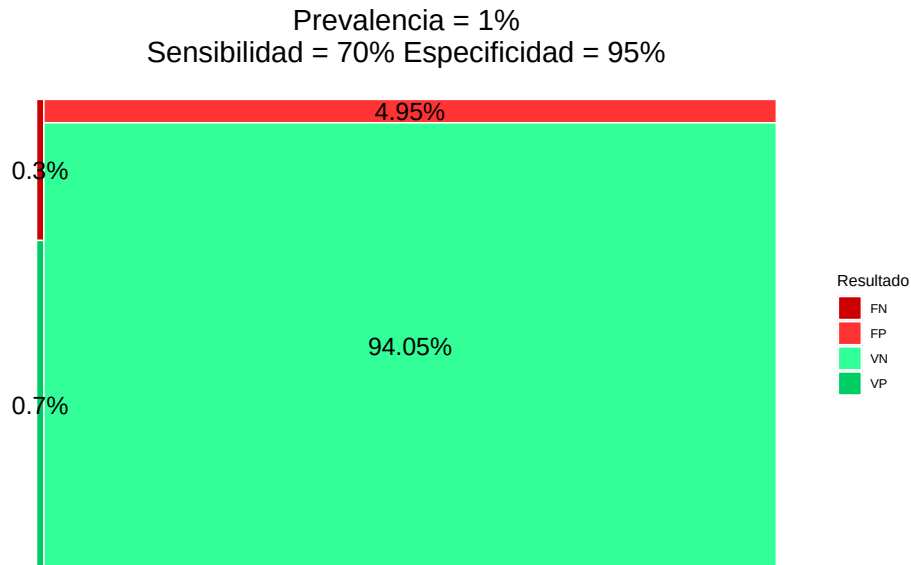


Figura 5.1: Resultados de un test diagnóstico con una sensibilidad del 70 %, una especificidad del 95 % y una prevalencia del 1 %

Mientras que, si la prevalencia es del 10 %, se tiene

Test	E	$\bar{E}$
Positivo +	70	45
Negativo –	30	855

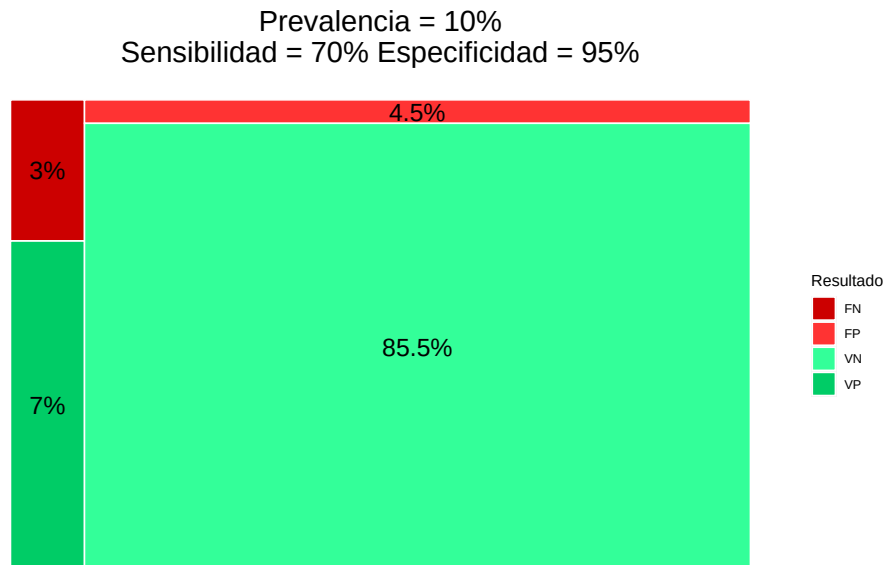


Figura 5.2: Resultados de un test diagnóstico con una sensibilidad del 70 %, una especificidad del 95 % y una prevalencia del 10 %

Obsérvese que, aunque la sensibilidad y la especificidad del test son las mismas en ambos casos, con una prevalencia del 1 % la mayoría de los positivos del test son falsos positivos, mientras que con una prevalencia del 10 % la mayoría son verdaderos positivos.

💡 Aplicación interactiva

Para ver los resultados de un test diagnóstico en función de la prevalencia, la sensibilidad y la especificidad se puede utilizar esta [aplicación para tests diagnósticos](#).

5.1.1. Cuándo usar un test más sensible o más específico

Una mayor sensibilidad aumenta el número de verdaderos positivos y disminuye el número de falsos negativos, mientras que una mayor especificidad aumenta el número de verdaderos negativos y disminuye el número de falsos positivos.

Por tanto, utilizaremos un test **más sensible** cuando:

- La enfermedad es grave o muy contagiosa y es importante detectarla.
- La enfermedad es curable.
- Los falsos positivos no provocan traumas serios.

Y utilizaremos un test **más específico** cuando:

- La enfermedad es importante pero difícil o imposible de curar.
- Los falsos positivos pueden provocar traumas serios.
- El tratamiento de los falsos positivos puede tener graves consecuencias.

Tanto la sensibilidad como la especificidad son indicadores de la fiabilidad de un test *a priori*, es decir, antes de aplicarlo. Una vez que el test se ha aplicado y se conoce su resultado, a la hora de diagnosticar la enfermedad o descartarla es mejor utilizar los *valores predictivos*.

5.2. Valores predictivos de un test

Definición 5.4 (Valor predictivo positivo). El *valor predictivo positivo* de un test diagnóstico es la proporción de personas con la enfermedad entre las personas con resultado positivo en el test,

$$VPP = P(E|+) = \frac{VP}{VP + FP}$$

Definición 5.5 (Valor predictivo negativo). El *valor predictivo negativo* de un test diagnóstico es la proporción de personas sin la enfermedad entre las personas con resultado negativo en el test,

$$VPN = P(\bar{E}|-) = \frac{VN}{VN + FN}$$

Ejemplo 5.3. Siguiendo con el ejemplo anterior y suponiendo una prevalencia del 1 %, se tiene

Test	E	$\bar{E}$
Positivo +	7	50
Negativo -	3	940

$$VPP = \frac{7}{7 + 50} = 0.123, \quad VPN = \frac{940}{3 + 940} = 0.997.$$

Mientras que, si la prevalencia es del 10 %, se tiene

Test	E	$\overline{E}$
Positivo +	70	45
Negativo −	30	855

$$VPP = \frac{70}{70 + 45} = 0.609, \quad VPN = \frac{855}{30 + 855} = 0.966.$$

! Importante

A diferencia de la sensibilidad y la especificidad, que son propiedades del test, los valores predictivos dependen de la prevalencia de la enfermedad en la población estudiada. Por eso un mismo test puede ser muy útil en una población con prevalencia alta e inútil en una población con prevalencia baja.

5.2.1. Interpretación de los valores predictivos

i Interpretación

$$\begin{aligned} VPP > 0.5 &\Rightarrow \text{Diagnosticar la enfermedad,} \\ VPN > 0.5 &\Rightarrow \text{Descartar la enfermedad.} \end{aligned}$$

En el ejemplo anterior, con una prevalencia del 1 % un resultado positivo no permite diagnosticar la enfermedad ($VPP = 0.123$), mientras que con una prevalencia del 10 % sí ($VPP = 0.609$). En ambos casos, en cambio, un resultado negativo permite descartarla con bastante seguridad.

6 La curva ROC

En los tests diagnósticos basados en la medición de una variable cuantitativa (como, por ejemplo, los tests de antígenos para la COVID) la sensibilidad y la especificidad dependen del umbral fijado para dar un positivo.

Al bajar el umbral aumenta la sensibilidad, pero disminuye la especificidad, y al subirlo ocurre lo contrario, de manera que no es posible resumir la fiabilidad del test con un único par de valores. Para evaluar la fiabilidad de estos tests se suele utilizar la *curva ROC*.

6.1. Definición de la curva ROC

Definición 6.1 (Curva ROC). La *curva ROC* (del inglés *Receiver Operating Characteristic*) de un test diagnóstico es la curva que resulta de representar la razón de verdaderos positivos (sensibilidad) frente a la razón de falsos positivos ($1 - \text{especificidad}$) para los diferentes umbrales de positivo del test.

6.1.1. Interpretación de la curva ROC

Interpretación

- Cada punto de la curva corresponde a un umbral para el positivo.
- El mejor test es el que se sitúa en la esquina superior izquierda del espacio (sensibilidad 1 y especificidad 1).
- La diagonal representa un test con un diagnóstico aleatorio.

Así, la curva ROC sirve también para seleccionar el umbral más adecuado en función de si interesa priorizar la sensibilidad o la especificidad, tal y como se explicó en el Capítulo 5.

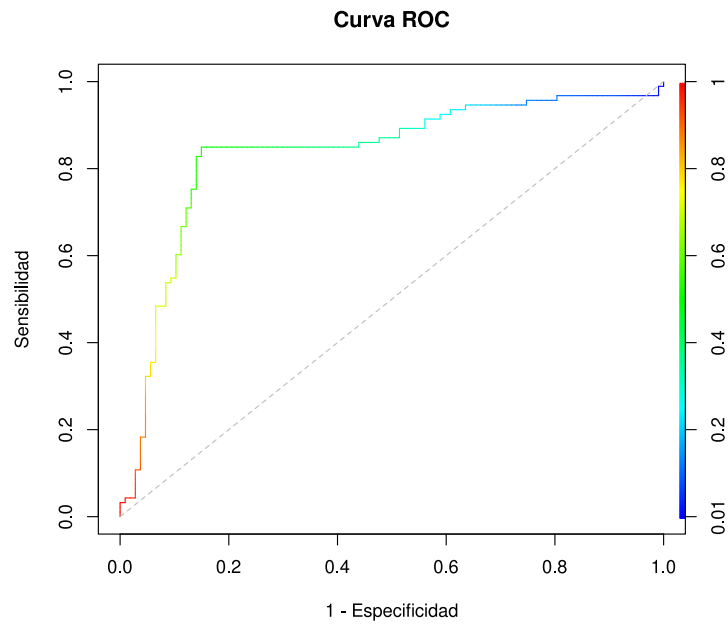


Figura 6.1: Curva ROC de un test diagnóstico

6.2. Área debajo de la curva ROC (AUC)

Para evaluar la fiabilidad de un test diagnóstico independientemente del umbral de positivos se suele medir el área bajo la curva ROC, también conocida como *AUC* (*area under the curve*).

Definición 6.2 (Área bajo la curva). El *área bajo la curva ROC* o *AUC* es la probabilidad de que, tomados al azar un individuo enfermo y otro sano, el test asigne al enfermo un valor más indicativo de la enfermedad que al sano.

Según el valor del AUC, se tiene:

Tabla 6.1: Interpretación del área bajo la curva ROC

AUC	Fiabilidad del test
0.5	Diagnóstico aleatorio
[0.5, 0.6)	Test malo
[0.6, 0.75)	Test regular
[0.75, 0.9)	Test bueno
[0.9, 0.97)	Test muy bueno
[0.97, 1)	Test excelente

6.3. Aplicaciones a la COVID

- [Fiabilidad del diagnóstico por PCR.](#)
- [Fiabilidad del diagnóstico por el test de antígenos.](#)
- [Comparativa de tests.](#)

Los tests de antígenos son más rápidos que las PCR, pero son menos fiables.

Por un lado, son menos sensibles que una prueba de PCR, debido a que se requiere una mayor cantidad de virus en las mucosas nasales o bucales para que se muestre un resultado positivo. Eso limita su efectividad cuando las personas llevan poco tiempo infectadas y el virus está empezando a reproducirse.

Por otro lado, también son menos específicos que la PCR y, por tanto, producen más falsos positivos.

Referencias

Recursos generales

- [Acción matemática contra el coronavirus.](#)
- [R Epidemic Consortium \(RECON\).](#)
- [Analysis of epidemiological data using R and Epicalc.](#)
- [R resources about COVID-19.](#)

Datos epidemiológicos

- [Situación actual de la COVID-19 \(Ministerio de Sanidad\).](#)
- [Centro Nacional de Epidemiología \(Instituto de Salud Carlos III\).](#)
- [European Centre for Disease Prevention and Control \(ECDC\).](#)
- [World Health Organization \(WHO\).](#)

Aplicaciones interactivas

- [Aplicación para el estudio de tests diagnósticos.](#)

Otros manuales del autor

- [Manual de Estadística.](#)
- [Colección de problemas resueltos de Estadística.](#)
- [Prácticas de Estadística con R.](#)