

# Bioestadística

## Tema 4: Probabilidad (recordatorio)

- ¿Cuál es la probabilidad de aprobar Bioestadística?
- ¿Cuál es la probabilidad de no encontrarme un atasco en la N 30 cuando voy a clase?
- Todos los días nos hacemos preguntas sobre probabilidad e incluso los que hayáis visto poco de la materia en cursos anteriores, tenéis una idea intuitiva lo suficientemente correcta para lo que necesitamos de ella en este curso.
- En este tema vamos a:
  - Recordar qué entendemos por probabilidad.
  - Recordar algunas reglas de cálculo.
  - Ver cómo aparecen las probabilidades en CC. Salud.
  - Aplicarlo a algunos conceptos nuevos de interés en CC. Salud.
    - [Pruebas diagnósticas.](#)

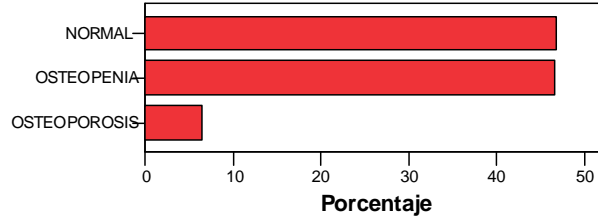
## Nociones de probabilidad

- **Frecuentista** (objetiva): Probabilidad de un suceso es la frecuencia relativa (%) de veces que ocurriría el **suceso** al realizar un experimento repetidas veces.

CLASIFICACION OMS

	Frecuencia	Porcentaje
Válidos		
NORMAL	469	46,9%
OSTEOPENIA	467	46,7%
OSTEOPOROSIS	64	6,4%
Total	1000	100,0

CLASIFICACION OMS

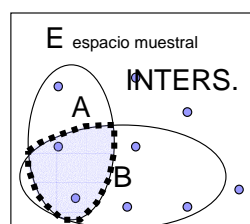
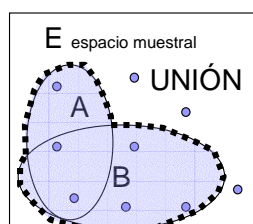
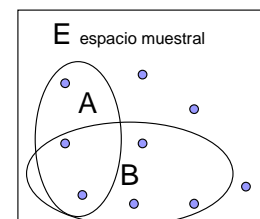
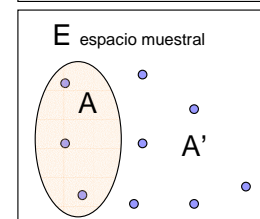
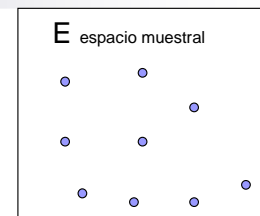


- **Subjetiva** (bayesiana): Grado de certeza que se posee sobre un **suceso**. Es personal.

En ambos tipos de definiciones aparece el concepto de **suceso**. Vamos a recordar qué son y algunas operaciones que se pueden realizar con sucesos.

## Sucesos

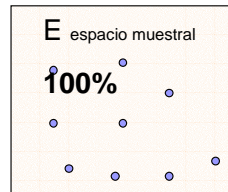
- Cuando se realiza un experimento aleatorio diversos resultados son posibles. El conjunto de todos los resultados posibles se llama **espacio muestral** (E).
- Se llama **suceso** a un subconjunto de dichos resultados.
- Se llama **suceso contrario** (complementario) de un suceso A,  $A'$ , al formado por los elementos que no están en A
- Se llama **suceso unión** de A y B,  $A \cup B$ , al formado por los resultados experimentales que están en A o en B (incluyendo los que están en ambos).
- Se llama **suceso intersección** de A y B,  $A \cap B$  o simplemente  $AB$ , al formado por los elementos que están en A y B



## Definición de probabilidad

- Se llama **probabilidad** a cualquier función,  $P$ , que asigna a cada suceso  $A$  un valor numérico  $P(A)$ , verificando las siguientes reglas (axiomas)

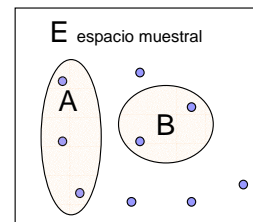
- $P(E)=1$



- $0 \leq P(A) \leq 1$

- $P(A \cup B) = P(A) + P(B)$  si  $A \cap B = \emptyset$

- $\emptyset$  es el conjunto vacío.



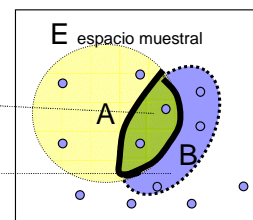
- Podéis imaginar la probabilidad de un subconjunto como el tamaño relativo con respecto al total (suceso seguro)

## Probabilidad condicionada

- Se llama **probabilidad de  $A$  condicionada a  $B$** , o **probabilidad de  $A$  sabiendo que pasa  $B$** :

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

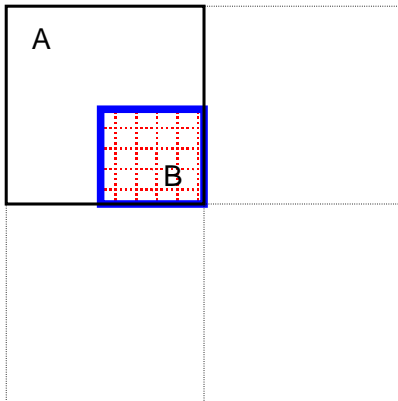
"tamaño" de uno respecto al otro



### Error frecuentíiiiiisimo:

- No confundáis probabilidad condicionada con intersección.
- En ambos medimos efectivamente la intersección, **pero...**
  - En  $P(A \cap B)$  con respecto a  $P(E)=1$
  - En  $P(A|B)$  con respecto a  $P(B)$

## Intuir la probabilidad condicionada

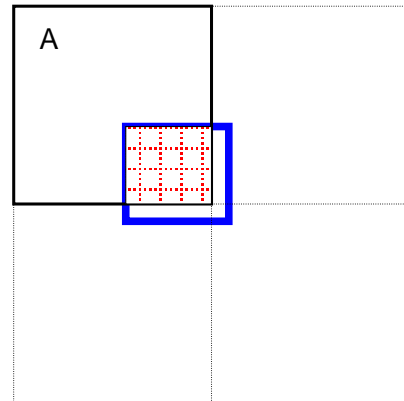


$$\begin{aligned} P(A) &= 0,25 \\ P(B) &= 0,10 \\ P(A \cap B) &= 0,10 \end{aligned}$$

¿Probabilidad de A sabiendo que ha pasado B?

$$P(A|B)=1$$

Bioestadística. U. Málaga.



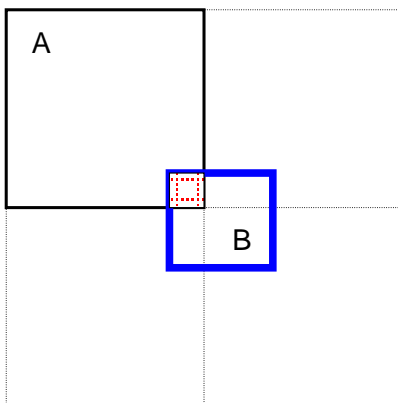
$$\begin{aligned} P(A) &= 0,25 \\ P(B) &= 0,10 \\ P(A \cap B) &= 0,08 \end{aligned}$$

$$P(A|B)=0,8$$

Tema 4: Probabilidad

7

## Intuir la probabilidad condicionada

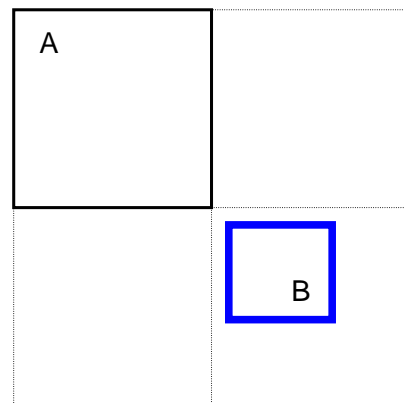


$$\begin{aligned} P(A) &= 0,25 \\ P(B) &= 0,10 \\ P(A \cap B) &= 0,005 \end{aligned}$$

¿Probabilidad de A sabiendo que ha pasado B?

$$P(A|B)=0,05$$

Bioestadística. U. Málaga.



$$\begin{aligned} P(A) &= 0,25 \\ P(B) &= 0,10 \\ P(A \cap B) &= 0 \end{aligned}$$

$$P(A|B)=0$$

Tema 4: Probabilidad

8

## Algunas reglas de cálculo prácticas

- Cualquier problema de probabilidad puede resolverse en teoría mediante aplicación de los axiomas. Sin embargo, **es más cómodo conocer algunas reglas de cálculo:**

- $P(A') = 1 - P(A)$

- $P(A \cup B) = P(A) + P(B) - P(AB)$

- $P(AB) = P(A) P(B|A)$   
 $= P(B) P(A|B)$

- Prob. de que pasen A y B es la prob. de A y que también pase B sabiendo que pasó A.

## Independencia de sucesos

- **Dos sucesos son independientes** si el que ocurra uno, no añade información sobre el otro.

- **A es independiente de B**

- $\Leftrightarrow P(A|B) = P(A)$

- $\Leftrightarrow P(AB) = P(A) P(B)$

## Ejemplo (I)

		Recuento		Total
		MENOPAUSIA		
		NO	SI	
CLASIFICACION	NORMAL	189	280	469
OMS	OSTEOPENIA	108	359	467
	OSTEOPOROSIS	6	58	64
Total		303	697	1000

- Se ha repetido en **1000** ocasiones el experimento de elegir a una mujer de una población muy grande. El resultado está en la tabla.

□ ¿Cuál es la probabilidad de que una mujer tenga osteoporosis?

- $P(\text{Osteoporosis}) = 64/1000 = 0,064 = 6,4\%$ 
  - Noción frecuentista de probabilidad

## Ejemplo (II)

		Recuento		Total
		MENOPAUSIA		
		NO	SI	
CLASIFICACION	NORMAL	189	280	469
OMS	OSTEOPENIA	108	359	467
	OSTEOPOROSIS	6	58	64
Total		303	697	1000

- ¿Probabilidad de tener osteopenia u osteoporosis?
  - $P(\text{Osteopenia} \cup \text{Osteoporosis}) = 467/1000 + 64/1000 = 0,531$ 
    - Son sucesos disjuntos
    - $\text{Osteopenia} \cap \text{Osteoporosis} = \emptyset$
- ¿Probabilidad de tener osteoporosis o menopausia?
  - $P(\text{Osteoporosis} \cup \text{Menopausia}) = 64/1000 + 697/1000 - 58/1000 = 0,703$ 
    - No son sucesos disjuntos
- ¿Probabilidad de una mujer normal? (entiéndase...)
  - $P(\text{Normal}) = 469/1000 = 0,469$
  - $P(\text{Normal}) = 1 - P(\text{Normal}') = 1 - P(\text{Osteopenia} \cup \text{Osteoporosis}) = 1 - 0,531 = 0,469$

### Ejemplo (III)

		MENOPAUSIA		Total
		NO	SI	
CLASIFICACION	NORMAL	189	280	469
OMS	OSTEOPENIA	108	359	467
	OSTEOPOROSIS	6	58	64
Total		303	697	1000

■ Si es menopáusica... ¿probabilidad de osteoporosis?

□  $P(\text{Osteoporosis}|\text{Menopausia}) = 58/697 = 0,098$

■ ¿Probabilidad de menopausia y osteoporosis?

□  $P(\text{Menop} \cap \text{Osteoporosis}) = 58/1000 = 0,058$

■ Otra forma:

$$P(\text{Menop} \cap \text{Osteoporosis}) = P(\text{Menop}) \times P(\text{Osteoporosis} | \text{Menop}) =$$

$$= \frac{697}{1000} \times \frac{58}{697} = 58/1000 = 0,058$$

### Ejemplo (IV)

		MENOPAUSIA		Total
		NO	SI	
CLASIFICACION	NORMAL	189	280	469
OMS	OSTEOPENIA	108	359	467
	OSTEOPOROSIS	6	58	64
Total		303	697	1000

■ ¿Son independientes menopausia y osteoporosis?

□ Una forma de hacerlo

■  $P(\text{Osteoporosis}) = 64/1000 = 0,064$

■  $P(\text{Osteoporosis}|\text{Menopausia}) = 58/697 = 0,098$

□ La probabilidad de tener osteoporosis es mayor si ha pasado la menopausia. Añade información extra. ¡No son independientes!

□ ¿Otra forma?

■  $P(\text{Menop} \cap \text{Osteoporosis}) = 58/1000 = 0,058$

■  $P(\text{Menop}) P(\text{Osteoporosis}) = (697/1000) \times (64/1000) = 0,045$

□ La probabilidad de la intersección no es el producto de probabilidades. No son independientes.

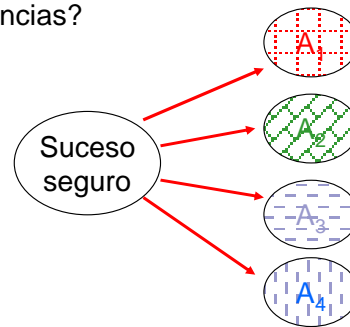
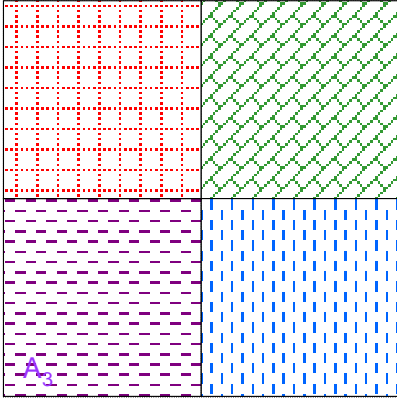
## Sistema exhaustivo y excluyente de sucesos

Son una colección de sucesos

$A_1, A_2, A_3, A_4, \dots$

Tales que la unión de todos ellos forman el espacio muestral, y sus intersecciones son disjuntas.

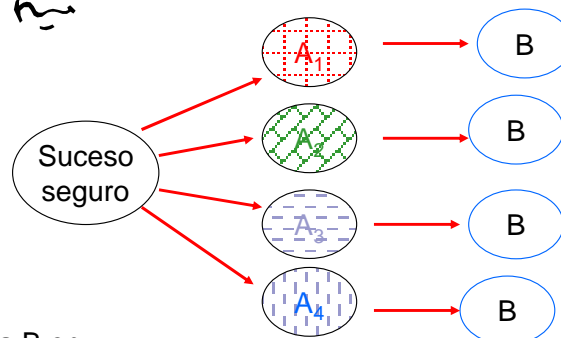
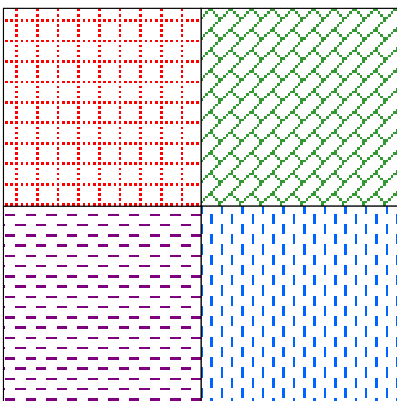
¿Recordáis cómo formar intervalos en tablas de frecuencias?



## Divide y vencerás

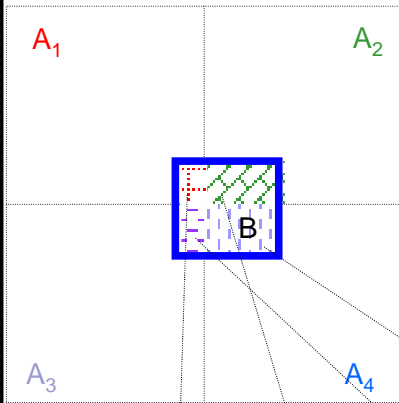
Todo suceso B, puede ser **descompuesto** en componentes de dicho sistema.

$$B = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup (B \cap A_4)$$



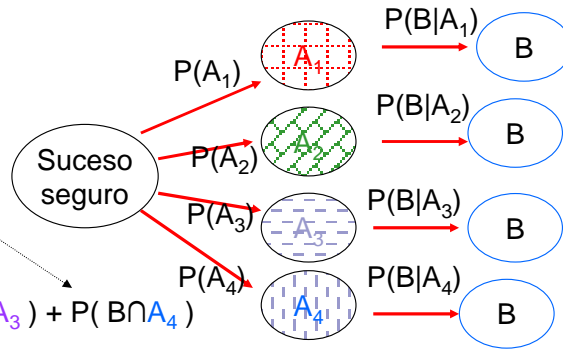
Nos permite descomponer el problema B en **subproblemas más simples**. Creedme . Funciona.

## Teorema de la probabilidad total



Si conocemos la probabilidad de B en cada uno de los componentes de un sistema exhaustivo y excluyente de sucesos, entonces...

... podemos calcular la probabilidad de B.



$$P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) + P(B \cap A_4)$$

$$= P(A_1) P(B|A_1) + P(A_2) P(B|A_2) + \dots$$

**Ejemplo (I):** En este aula el 70% de los alumnos son mujeres. De ellas el 10% son fumadoras. De los hombres, son fumadores el 20%.

### T. Prob. Total.

Hombres y mujeres forman un sist. Exh. Excl. de sucesos

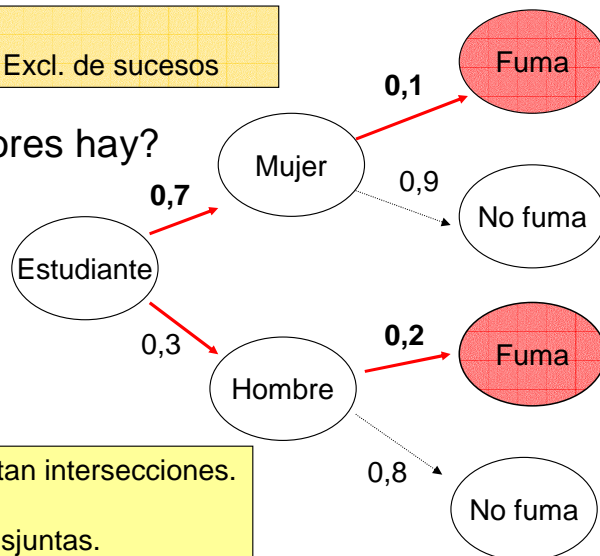
■ ¿Qué porcentaje de fumadores hay?

$$P(F) = P(M \cap F) + P(H \cap F)$$

$$= P(M)P(F|M) + P(H)P(F|H)$$

$$= 0,7 \times 0,1 + 0,3 \times 0,2$$

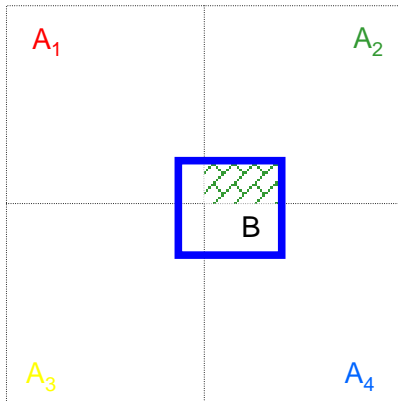
$$= 0,13 = 13\%$$



• Los caminos a través de nodos representan intersecciones.

• Las bifurcaciones representan uniones disjuntas.

## Teorema de Bayes



Si conocemos la probabilidad de B en cada uno de los componentes de un sistema exhaustivo y excluyente de sucesos, entonces...

...si ocurre B, podemos calcular la probabilidad (*a posteriori*) de ocurrencia de cada \$A\_i\$.

$$P(A_i | B) = \frac{P(B | A_i) P(A_i)}{P(B)}$$

donde \$P(B)\$ se puede calcular usando el teorema de la probabilidad total:

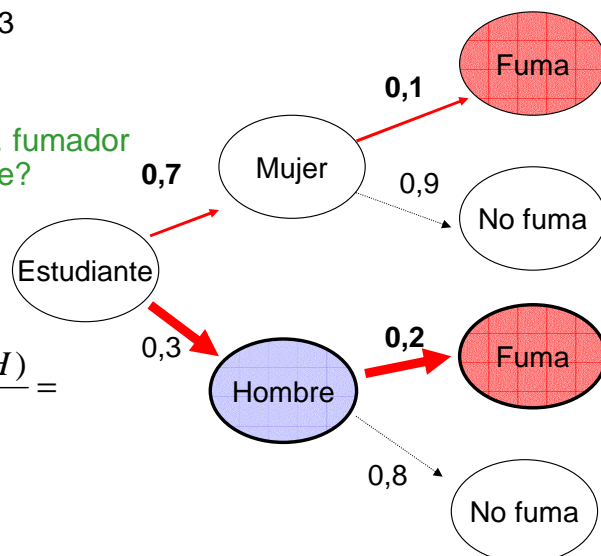
$$P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) + P(B \cap A_4)$$

$$= P(B | A_1) P(A_1) + P(B | A_2) P(A_2) + \dots$$

**Ejemplo (II):** En este aula el 70% de los alumnos son mujeres. De ellas el 10% son fumadoras. De los varones, son fumadores el 20%.

- ¿Qué porcentaje de fumadores hay?
  - $P(F) = 0,7 \times 0,1 + 0,3 \times 0,2 = 0,13$ 
    - (Resuelto antes)

- Se elige a un individuo al azar y es... fumador  
¿Probabilidad de que sea un hombre?



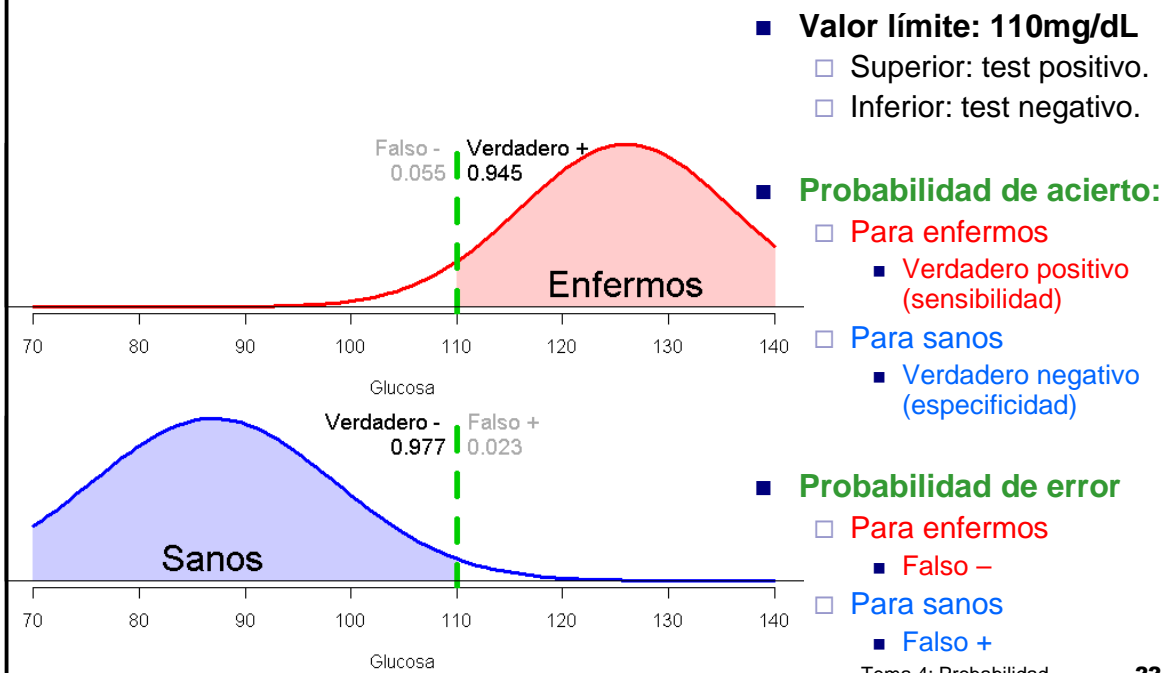
$$P(H | F) = \frac{P(H \cap F)}{P(F)} = \frac{P(H) \cdot P(F | H)}{P(F)}$$

$$= \frac{0,3 \times 0,2}{0,13} = 0,46$$

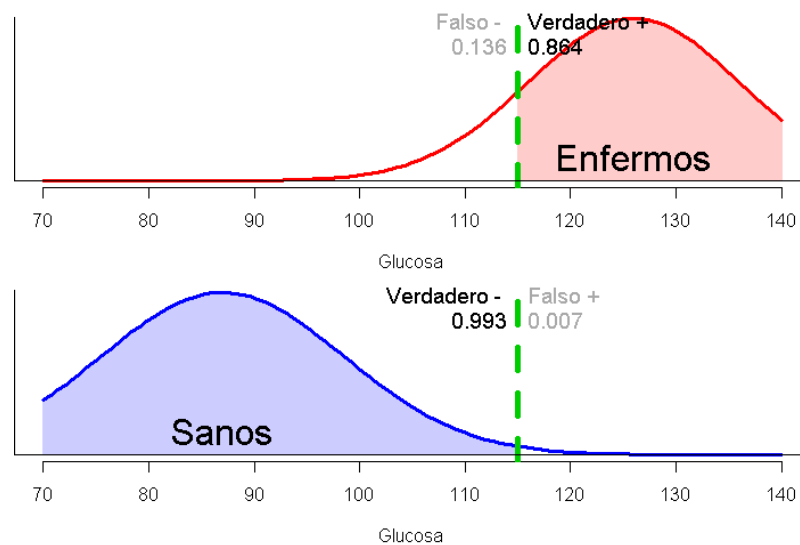
## Ejemplo de prueba diagnósticas: Diabetes

- Los carbohidratos ingeridos terminan como glucosa en la sangre. El exceso se transforma en glucógeno y se almacena en hígado y músculos. Este se transforma entre comidas de nuevo en glucosa según necesidades.
- La principal hormona que regula su concentración es la insulina. La diabetes provoca su deficiencia o bien la insensibilidad del organismo a su presencia. Es una enfermedad muy común que afecta al 2% de la población (**prevalencia**)
- Una prueba común para diagnosticar la diabetes, consiste en medir el nivel de glucosa. En individuos sanos suele variar entre 64 y 110mg/dL.
  - El cambio de color de un indicador al contacto con la orina suele usarse como indicador (**resultado del test positivo**)
- Valores por encima de 110 mg/dL se asocian con un posible estado pre-diabético.
  - Pero no es seguro. Otras causas podrían ser: hipertiroidismo, cancer de páncreas, pancreatitis, atracón reciente de comida...
- Supongamos que los enfermos de diabetes, tienen un valor medio de 126mg/dL.

## Funcionamiento de la prueba diagnóstica de glucemia



## ¿Cómo definir el punto de corte de la prueba diagnóstica?

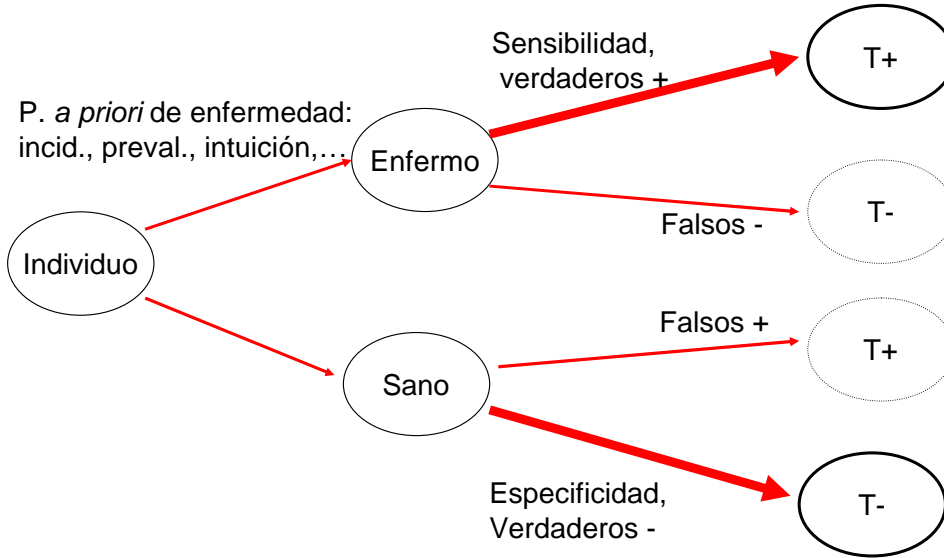


No es simple. **No es posible aumentar sensibilidad y especificidad al mismo tiempo.** Hay que elegir una solución de compromiso: Aceptable sensibilidad y especificidad.

Una **prueba diagnóstica** ayuda a mejorar una estimación de la probabilidad de que un individuo presente una enfermedad.

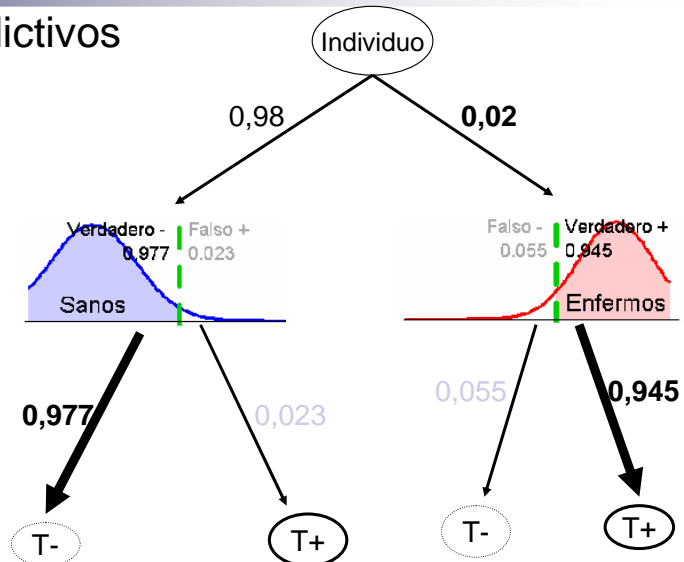
- En principio tenemos una **idea subjetiva** de  $P(\text{Enfermo})$ . Nos ayudamos de...
  - **Incidencia**: Porcentaje de nuevos casos de la enfermedad en la población.
  - **Prevalencia**: Porcentaje de la población que presenta una enfermedad.
- Para confirmar la sospecha, usamos una prueba diagnóstica. Ha sido evaluada con anterioridad sobre dos grupos de individuos: sanos y enfermos. Así **de modo frecuentista** se ha estimado:
  - $P(+ | \text{Enfermo}) = \text{Sensibilidad}$  (verdaderos +) = Tasa de acierto sobre enfermos.
  - $P(- | \text{Sano}) = \text{Especificidad}$  (verdaderos -) = Tasa de acierto sobre sanos.
- A partir de lo anterior y usando el **teorema de Bayes**, podemos calcular las probabilidades *a posteriori* (en función de los resultados del test): **Índices predictivos**
  - $P(\text{Enfermo} | +) = \text{Índice predictivo positivo}$
  - $P(\text{Sano} | -) = \text{Índice predictivo negativo}$

## Pruebas diagnósticas: aplicación T. Bayes.



## Ejemplo: Índices predictivos

- La diabetes afecta al 2% de los individuos.
- La presencia de glucosuria se usa como indicador de diabetes.
- Su sensibilidad es de 0,945.
- La especificidad de 0,977.
- Calcular los índices predictivos.



$$P(\text{Sano} | T-) = \frac{P(\text{Sano} \cap T-)}{P(\text{Sano} \cap T-) + P(\text{Enf} \cap T-)} = \frac{0,98 \cdot 0,977}{0,98 \cdot 0,977 + 0,02 \cdot 0,055} = 0,999$$

$$P(\text{Enf} | T+) = \frac{P(\text{Enf} \cap T+)}{P(\text{Enf} \cap T+) + P(\text{Sano} \cap T+)} = \frac{0,02 \cdot 0,945}{0,02 \cdot 0,945 + 0,98 \cdot 0,023} = 0,456$$

## Observaciones

- En el ejemplo anterior, al llegar un individuo a la consulta tenemos una idea *a priori* sobre la probabilidad de que tenga una enfermedad.
- A continuación se le pasa una **prueba diagnóstica** que nos aportará nueva información: Presenta glucosuria o no.
- En función del resultado tenemos una nueva idea (*a posteriori*) sobre la probabilidad de que esté enfermo.
  - Nuestra opinión a priori ha sido modificada por el resultado de un experimento.



-¿Qué probabilidad tengo de estar enfermo?

- En principio un 2%. Le haremos unas pruebas.



- Presenta glucosuria. La probabilidad ahora es del 45,6%.

## ¿Qué hemos visto?

- Álgebra de sucesos
  - Unión, intersección, complemento
- Probabilidad
  - Nociones
    - Frecuentista
    - Subjetiva o Bayesiana
  - Axiomas
  - Probabilidad condicionada
  - Reglas de cálculo
    - Complementario, Unión, Intersección
  - Independencia de sucesos
  - Sistema exhaustivo y excluyente de sucesos
    - Teorema probabilidad total.
    - Teorema de Bayes
      - Pruebas diagnósticas
        - *A priori*: Incidencia, prevalencia.
        - Eficacia de la prueba: Sensibilidad, especificidad.
        - *A posteriori*: Índices predictivos.

