

Detección del cáncer: ¿Pruebas válidas?¹

por

MIGUEL BARRERAS ALCONCHEL
(IES Matarraña, Valderrobres)

En un artículo de la revista *Investigación y Ciencia* de marzo de 2012, «Detección del cáncer. Estadística de los falsos positivos», el prestigioso divulgador matemático John Allen Paulos sorprendía cuestionando la validez de las pruebas para detectar distintos cánceres (Leer artículo en *cancer.xls[articulo]*). Y ponía un ejemplo numérico, concreto, que se va a analizar aquí, con ayuda de una hoja de cálculo Excel.

Supongamos que un millón de personas se somete a las pruebas para detectar un determinado cáncer y que la prevalencia de esa enfermedad es de un 0,4% (2 por cada 500). Es decir, suponemos que, aproximadamente, habrá unas 4.000 personas enfermas en ese millón. Supongamos también que esa prueba da falsos positivos en un 1% de los casos; esto es, 1% de las personas sanas dará positivo (se deja para el lector la reflexión de los corolarios psicológicos y físicos que el paciente sano puede llegar a sufrir). Por fin, pongamos también que esta prueba arroja un porcentaje del 0,5 de falsos negativos.

En el libro Excel *cancer.xls [cancer]* se plantea y resuelve el problema.

El formato de las celdas A7, C7, E7 lo hemos cambiado a *Porcentaje, 2 decimales*.

Vamos a aplicar el teorema de Bayes, pero elaborando de forma sencilla una tabla de doble entrada, para prescindir de las fórmulas clásicas que suelen confundir a los no matemáticos. El proceso es el siguiente:

	A	B	C	D	E
6	afección		falsos +		falsos -
7	0,40%		1,00%		0,50%

	A	B	C	D
9		C	no C	
10	= "+"			
11	= "-"			
12		=D12*A7		1000000

	A	B	C	D
9		C	no C	
10	= "+"			
11	= "-"			
12		=D12*A7	=D12-B12	1000000

	A	B	C	D
9		C	no C	
10	= "+"	=B12-B11		
11	= "-"	=B12*E7		
12		=D12*A7	=D12-B12	1000000

	A	B	C	D
9		C	no C	
10	= "+"	=B12-B11	=C12*C7	=B10+C10
11	= "-"	=B12*E7	=C12-C10	=B11+C11
12		=D12*A7	=D12-B12	1000000

1.º el rojo, 2.º el verde

Rojo / verde / azul / amarillo

Se llega, así, a la tabla completa, con las probabilidades condicionadas siguientes:

	C	no C	
+	3.980	9.960	13.940
-	20	986.040	986.060
	4.000	996.000	1.000.000

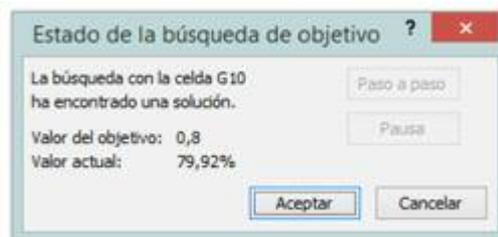
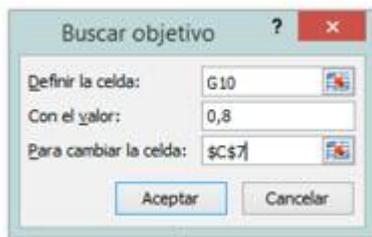
totales		condicionadas	
P(C)=	0,004	p(C/+)=	0,286
p(no C)=	0,996	p(C/-)=	0,000
p(+)=	0,014	p(+/C)=	0,995
p(-)=	0,986	p(+/no C)=	0,010

p(no C/+)=	0,714
p(no C/-)=	1,000
p(-/C)=	0,005
p(-/no C)=	0,990

Por tanto, entre todas las personas que dan positivo (13.940), solo 3.980 tienen, efectivamente, cáncer. Esto supone una probabilidad de 3.980 / 13.940, es decir 0,286 (el 28,6 %), porcentaje que sorprende a Paulos. Y también que 9.960 personas sanas den positivo.

Es claro que la cuestión puede solventarse, fácil, con lápiz y papel, pero la hoja de cálculo nos permite analizar todas las situaciones imaginables, sin cuentas adicionales, simplemente cambiando el valor de las celdas donde aparecen el porcentaje de incidencia de la enfermedad y los de falsos positivos y negativos de la prueba. [ver *cancer.xls*[fe]]. Pero, es más: podemos resolver inmediatamente problemas inversos del tipo siguiente: ¿Qué cota de falsos positivos podríamos permitirnos para que p(C/+) subiera al 80 %, por ejemplo?

1. Seleccionamos la celda C10.
2. *Datos / Herramientas de datos / Análisis Y si... / Buscar objetivo*. Completamos así:



falsos +
0,10%

3. La cota de falsos positivos no puede superar el 0,10 %.

Para acabar, supongamos que no se puede bajar la cota de falsos positivos y es del 1 %. Si una persona ha dado positivo, ¿se debería repetir la prueba? Parece razonable que sí. Si da otra vez positivo (esto es, ha dado dos positivos en dos pruebas), ¿cuál es la probabilidad de que tenga, efectivamente, cáncer? La solución está en *cancer.xls*[doble positivo]. Ahora es p(C/++)=0,98 (B7/D7=3.960,1/4.059,7), un resultado bastante aceptable.

	A	B	C	D	E
3	afección		falsos +		falsos -
4	0,004		0,01		0,005
6		C	no C		
7	"++"	=B11*(1-E4)^2	=C11*C4*C4	=B7+C7	
8	"+-"	=B11*(1-E4)*E4	=C11*C4*(1-C4)	=B8+C8	
9	"-+"	=E4*(1-E4)*B11	=C8	=B9+C9	
10	"--"	=B11*E4*E4	=C11*(1-C4)^2	=B10+C10	
11		=D11*A4	=D11-B11	1000000	

	C	no C	
++	3.960,1	99,6	4.059,7
+-	19,9	9.860,4	9.880,3
-+	19,9	9.860,4	9.880,3
--	0,1	976.179,6	976.179,7
	4.000	996.000	1.000.000

1. El libro Excel que se trabaja aquí, *cancer.xls*, está en la web <<http://catedu.es/calendas/catexcel/probabilidad.htm>> o directamente <<http://catedu.es/calendas/cancer.rar>>.