

¿Doctor, estoy bien?

PLANTEAMIENTO CONTEXTUAL

Aunque no os lo creáis, una de las cosas que hacemos los profesores de matemáticas es leer, leer mucho y leer de todo. Y entre las cosas que leemos, nos gusta leer qué escriben algunos de nuestros compañeros de profesión (aquellos que denominamos la comunidad científica).

*Hace poco leí un artículo de divulgación matemática que me llamó la atención y me pareció interesante para comentarlo con vosotros en clase. Su autor era Raúl Ibáñez, profesor del Departamento de Matemáticas de la UPV/EHU y colaborador de la Cátedra de Cultura Científica. El artículo en cuestión tenía por título “**Falsos positivos, o la importancia de comprender la información**” y he aquí como empezaba:*

“La primera vez que leí, que yo recuerde, sobre falsos positivos fue en el interesante libro **El hombre anumérico** (Tusquets, 1990) del matemático y divulgador estadounidense John Allen Paulos. La situación que presentaba era la siguiente:

Supongamos que existe una prueba médica para detectar si una persona padece una cierta enfermedad (por ejemplo, cáncer), cuya fiabilidad es del 98%, y que esa enfermedad la sufren 1 de cada 200 personas. Un día nos hacemos dicha prueba y nos da positiva... ¿Hasta qué punto nos debemos de preocupar? ¿Qué probabilidad existe de que tengamos dicha enfermedad?

La cuestión es ciertamente importante, puesto que de la comprensión de la misma dependen tanto aspectos relacionados con la comunicación entre el equipo médico y el paciente, el primero debe transmitir al paciente una gravedad en consonancia con la probabilidad mencionada, como aspectos médicos, es decir, la gestión médica de la situación.

Pero el problema planteado por el matemático en su libro, que no es otro que el de los falsos positivos, no solamente afecta a los análisis médicos, sino que aparece en diferentes situaciones de nuestra vida cotidiana, que van desde los tests de embarazo o los análisis de drogas, hasta los experimentos científicos.”

¿Qué os parece si, con nuestros conocimientos básicos de probabilidad de la Regla de Laplace, intentamos responder matemáticamente a la pregunta planteada en el problema?

ACTIVIDADES

A1. Después de leer atentamente la introducción anterior, vamos a realizar un debate dirigido en el aula sobre ciencia, medicina, diagnosis y prueba médica:

1. *¿La medicina es una ciencia? ¿Cuál es tu opinión al respecto? Presentad argumentos que sustenten vuestra afirmación.*
2. *¿En qué creéis que consiste una diagnosis médica?*
3. *¿Puede una diagnosis estar equivocada? ¿Por qué? ¿Habéis oído hablar alguna vez de “pedir una segunda opinión”? ¿En qué consiste?*
4. *¿Es posible que dos profesionales médicos puedan discrepar sobre datos científicos?*
5. *¿La ciencia se puede equivocar? ¿Cómo se tratan los errores científicamente?*
6. *Teniendo en cuenta la regla de Laplace, ¿cómo debe interpretarse la probabilidad de padecer una enfermedad? ¿Qué es un factor de riesgo?*
7. *¿Consideráis que una prueba médica ejecutada rigurosamente es totalmente fiable?*
8. *¿Pero puede medirse la fiabilidad de una prueba médica? ¿Cómo?*
9. *¿Qué relación, si es que hay alguna, puede establecerse entre riesgo y fiabilidad?*
10. *En base a la pregunta anterior, y según vuestro parecer, ¿si la fiabilidad de una prueba médica es muy elevada, cómo dirías que es la probabilidad de padecer la enfermedad que se diagnostica con esa prueba?*

Habiendo anotado las respuestas y reflexiones que os han parecido más interesantes en el debate, realizad una redacción de 250 palabras con vuestras propias conclusiones argumentadas.



A2. Imaginemos que se sobre una muestra de población de 10.000 personas se realizan 10.000 pruebas médicas (una por persona) como las que se indican por el problema citado en el fragmento de texto introductorio:

1. *En esta muestra en concreto, y asumiendo que por los datos de población existentes esta enfermedad es sufrida por 1 de cada 200 personas ¿qué puede afirmarse sobre el número de personas que padezcan la enfermedad?*
2. *¿A cuántas de las personas enfermas de la muestra les dará positiva (es decir, confirmará la enfermedad) la prueba médica realizada? ¿Y, para cuántas el resultado de la prueba será negativo (es decir, la prueba médica no confirmará la enfermedad aunque la padezcan)?*
3. *¿Cuántas personas sanas de la muestra serán diagnosticadas como enfermas por la prueba médica?*
4. *Construye una tabla de contingencia para recoger y presentar los resultados anteriores.*
5. *Si la prueba ha dado positiva ¿Cuál es la probabilidad de estar realmente enfermo? ¿Y de estar realmente sano?*
6. *Por el contrario, si la prueba ha dado negativa ¿Cuál es la probabilidad de estar realmente enfermo? ¿Y de estar realmente sano?*
7. *¿Cómo se modificarían los resultados anteriores en el caso de una enfermedad rara con una tasa de incidencia poblacional de 1 por cada 1.000 habitantes? (suponiendo que la prueba médica correspondiente tiene la misma fiabilidad que la anterior)*
8. *Anotad vuestras conclusiones al respecto para discutir las en clase.*

Tipos de diagnósticos		Enfermedad	
		ausente	presente
Prueba diagnóstica	negativa	Verdadero negativo (diagnóstico negativo enfermedad ausente)	Falso negativo (diagnóstico negativo enfermedad presente)
	positiva	Falso positivo (diagnóstico positivo enfermedad ausente)	Verdadero positivo (diagnóstico positivo enfermedad presente)

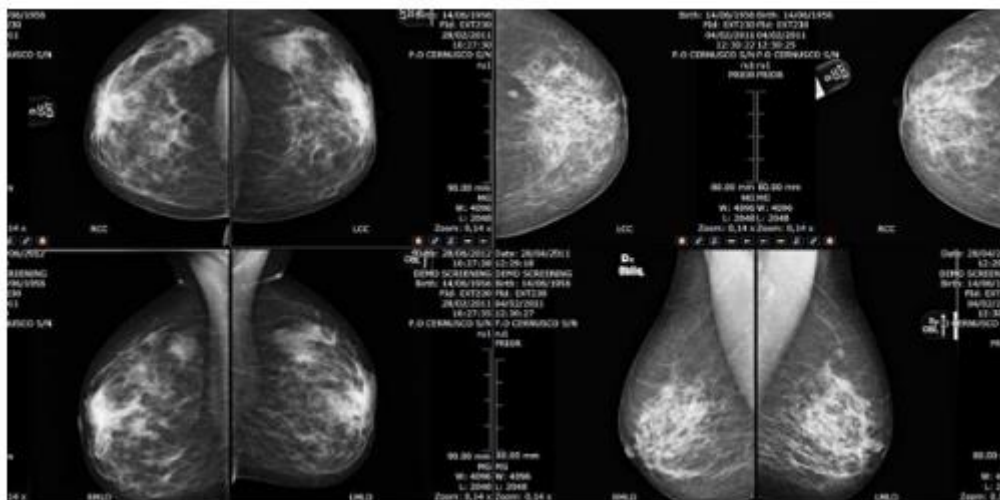
A3. En la actividad anterior hemos realizado unos cálculos de probabilidad asumiendo que la fiabilidad de la prueba es simétrica, es decir es la misma aplicada sobre las personas sanas que sobre las personas enfermas. Pero esto no es siempre así, como por ejemplo pasa en el caso de las pruebas de diagnóstico por la imagen tipo mamografía para detectar indicios de cáncer de mama.

En la siguiente tabla encontrarás la fiabilidad de la prueba de mamografía en función de si el bulto detectado es maligno (paciente enfermo) o benigno (paciente sano):

mamografía	masa maligna (cáncer)	masa benigna (no cáncer)
positiva	79,2%	9,6%
negativa	20,8%	90,4%

Si se estima que la incidencia del cáncer de mama entre la población femenina es de un 1% (es decir 1 de cada 100 mujeres), al tomar una muestra de población de 10.000 mujeres a las cuáles se les realiza una mamografía:

1. ¿Cuál es la probabilidad de que una mujer que ha dado positivo en la prueba tenga cáncer?
2. Anotad vuestras conclusiones al respecto para discutir las en clase.





MATERIALES I RECURSOS.

Para la realización de las actividades propuestas pueden usarse calculadoras científicas o bien ordenadores, en función de la disponibilidad de recursos por parte de los alumnos.

- Calculadora científica para cada alumno (modelo recomendado CASIO CLASSWIZ fx-82SPX Iberia, 570 o 991SPX)
- Hoja de cálculo (Excel Microsoft Office, Calc LibreOffice, o menú Hoja de Cálculo en CLASSWIZ 570 o 991 SPX)

Recursos y bibliografía complementaria:

- Ibáñez, Raúl (2015): *Falsos positivos, o la importancia de comprender la información*. Revista de Cultura Científica
<http://culturacientifica.com/2015/10/07/falsos-positivos-o-la-importancia-de-comprender-la-informacion/>
- Wikipedia: *Falso positivo* [https://es.wikipedia.org/wiki/Falso_positivo_\(medicina\)](https://es.wikipedia.org/wiki/Falso_positivo_(medicina))
- Cuevas Renaud, Corina (2010): *Validez y fiabilidad de las medidas de exposición y medición*.
<http://www.psicol.unam.mx/Investigacion2/pdf/SENSIBILIDAD%20Y%20ESPECIFICIDAD.pdf>

ORIENTACIONES DIDÁCTICAS.

Las actividades están diseñadas para trabajar el concepto de probabilidad a partir de la Regla de Laplace extrayendo información de tablas de contingencia. El trabajo didáctico también está pensado para introducir al alumnado en el concepto de probabilidad condicionada así como para desarrollar el sentido crítico en relación a las pruebas y mediciones científicas basadas en muestras estadísticas.

Pueden integrarse indistintamente en el curso de Matemáticas aplicadas de 3º y 4º ESO (bloque de estándares de probabilidad y estadística):

- *Distingue población y muestra justificando las diferencias en problemas contextualizados.*
- *Utiliza un vocabulario adecuado para describir, cuantificar y analizar situaciones relacionadas con el azar.*
- *Identifica y describe situaciones y fenómenos de carácter aleatorio, utilizando la terminología adecuada para describir sucesos.*
- *Calcula la probabilidad de sucesos con la regla de Laplace y utiliza, especialmente, diagramas de árbol o tablas de contingencia para el recuento de casos.*
- *Aplica técnicas de cálculo de probabilidades en la resolución de diferentes situaciones y problemas de la vida cotidiana.*
- *Resuelve problemas sencillos asociados a la probabilidad condicionada.*
- *Toma la decisión correcta teniendo en cuenta las probabilidades de las distintas opciones en situaciones de incertidumbre.*

Asimismo en Matemáticas de 1º de Bachillerato Científico, ubicándola en el tema de Probabilidad y, concretamente, en la unidad correspondiente a la probabilidad condicionada y el Teorema de Bayes.

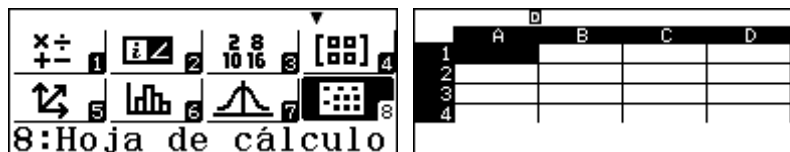
Con respecto al tiempo de dedicación a las actividades, se recomiendan tres sesiones (1h) de trabajo con los alumnos, una para cada actividad. En el planteamiento de las actividades resulta fundamental dedicar tiempo al debate en el aula precisamente para provocar que emerjan ideas preconcebidas y errores intuitivos en relación a la probabilidad.

Se recomienda trabajar las actividades conjuntamente con la asignatura de Biología, puesto que los contenidos de aplicación están directamente relacionados con esta materia.

ORIENTACIONES TÉCNICAS.

Para la realización de los cálculos propuestos en las actividades puede hacerse uso del menú de trabajo con hoja de cálculo disponible en la calculadora científica

CLASSWIZ fx-570SPX: **ON** **MENU** **8**



Para más información sobre la configuración de formatos, funciones y usos de la hoja cálculo consultar página 38 del manual técnico de la calculadora:

http://support.casio.com/storage/es/manual/pdf/ES/004/fx-570_991sp_x_es_-_guia_de_usuario_espanol.pdf

RESOLUCIÓN DE LAS ACTIVIDADES PROPUESTAS.

A2.

1. En esta muestra en concreto, y asumiendo que por los datos de población existentes esta enfermedad es sufrida por 1 de cada 200 personas ¿qué puede afirmarse sobre el número de personas que padezcan la enfermedad?

Puesto que la tasa de incidencia es en promedio de 1 sobre cada 200 personas, cabe esperar que sobre una muestra aleatoria de 10.000 personas puedan encontrarse 50 individuos enfermos.

Es decir atendiendo a la Regla de Laplace la probabilidad de estar enfermo viene dada por:

$$P(\text{enfermo}) = \frac{\text{casos favorables}}{\text{casos posibles}} = \frac{1}{200} = 0,005 = 0,5\%$$

The image shows three sequential calculator screens. The first screen shows the fraction 1/200 being entered. The second screen shows the result 0.005. The third screen shows the result 50, which is the product of 0.005 and 10,000.

Así pues sobre una muestra de 10.000 individuos, la frecuencia absoluta de casos de personas enfermas vendrá dada por:

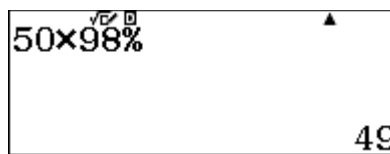
$$N(\text{enfermo}) = M \cdot P(\text{enfermo}) = 10.000 \cdot 0,005 = 50$$

Por lo tanto, si la selección de la muestra se ha hecho de forma científicamente correcta y sin sesgos sobre la población, se espera que de los 10.000 individuos hay 9.950 sanos y 50 que padecen la enfermedad. Pero hay que tener en cuenta que esa es una afirmación que se hace dentro de los parámetros de certeza estadística, por tanto no podemos garantizarlo con absoluta seguridad. Bien podría ser que hubiera más individuos enfermos o menos individuos enfermos que los inicialmente previstos de forma estadística.

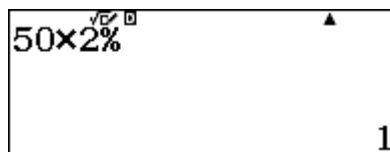
2. ¿A cuántas de las personas enfermas de la muestra les dará positiva (es decir, confirmará la enfermedad) la prueba médica realizada? ¿Y, para cuántas el resultado de la prueba será negativo (es decir, la prueba médica no confirmará la enfermedad aunque la padezcan)?

En el enunciado se nos informa que la prueba médica tiene una fiabilidad del 98% (se entiende que de forma simétrica, es decir tanto para individuos sanos como para enfermos). Por lo tanto, el número de personas enfermas con un resultado positivo vendrá dado por:

$$N(\text{enfermo } +) = N(\text{enfermo}) \cdot P(+)= 50 \cdot 98\% = 49$$



Y de forma equivalente el número de individuos enfermos con resultado negativo en la prueba será: $FN = N(\text{enfermo } -) = N(\text{enfermo}) \cdot P(-) = 50 \cdot 2\% = 1$

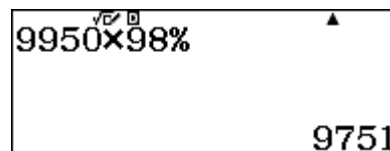


Así pues, 1 individuos enfermos sobre una muestra 10.000 ha sido diagnosticados falsamente como sano (**FN, falsos negativo**)

3. ¿Cuántas personas sanas de la muestra serán diagnosticadas como enfermas por la prueba médica?

Puesto que la fiabilidad de la prueba es simétrica, una persona sana puede dar negativo de padecer enfermedad con un 98% de probabilidad:

$$N(\text{sanos } -) = N(\text{sanos}) \cdot P(-) = 9.950 \cdot 98\% = 9.751$$



Por tanto, el número de individuos sanos falsamente diagnosticados serán aquellos con resultado positivo:

$$FP = N(\text{sanos } +) = N(\text{sanos}) \cdot P(+)= 9.950 \cdot 2\% = 199$$

Así pues, 199 individuos sanos sobre una muestra 10.000 han sido diagnosticados falsamente como enfermos (**FP, falsos positivos**).

4. Construye una tabla de contingencia para recoger y presentar los resultados anteriores.

Obtenidos los valores anteriores los presentamos en una tabla de contingencia de doble entrada (también llamada matriz de confusión):

	Sanos	Enfermos	TOTAL
Prueba +	199	49	248
Prueba -	9.751	1	9.752
TOTAL	9.950	50	10.000

5. Si la prueba ha dado positiva ¿Cuál es la probabilidad de estar realmente enfermo? ¿Y de estar realmente sano?

A partir de la tabla anterior y aplicando la regla de Laplace para el cálculo de probabilidad sobre la muestra dada:

$$P(\text{enfermo si } +) = \frac{\text{casos favorables enfermo}}{\text{casos posibles } +} = \frac{49}{248} = 0,197 \approx 20\%$$

$$P(\text{FP}) = P(\text{sano si } +) = \frac{\text{casos favorables sano}}{\text{casos posibles } +} = \frac{199}{248} = 0,80 \approx 80\%$$

$\frac{49}{248}$	$\frac{199}{248}$
0.1975806452	0.8024193548

6. Por el contrario, si la prueba ha dado negativa ¿Cuál es la probabilidad de estar realmente enfermo? ¿Y de estar realmente sano?

Nuevamente aplicando la regla de Laplace:

$$P(\text{FN}) = P(\text{enfermo si } -) = \frac{\text{casos favorables enfermo}}{\text{casos posibles } -} = \frac{1}{9.752} = 0,0001 \approx 0,01\%$$

$$P(\text{sano si } -) = \frac{\text{casos favorables sano}}{\text{casos posibles } -} = \frac{9.751}{9.752} = 0,9999 \approx 99,99\%$$

$\frac{1}{9752}$	$\frac{9751}{9752}$
0.00010254306	0.9998974569



7. ¿Cómo se modificarían los resultados anteriores en el caso de una enfermedad rara con una tasa de incidencia poblacional de 1 por cada 1.000 habitantes? (suponiendo que la prueba médica correspondiente tiene la misma fiabilidad que la anterior)

Los cálculos anteriores pueden sistematizarse y automatizarse mediante el uso de una hoja de cálculo (por ejemplo con EXCEL) en la cual se van introduciendo los diferentes parámetros que configuran la tabla de contingencia en función de los valores de muestra, de la tasa de incidencia de la enfermedad y de la fiabilidad de la prueba médica:

	A	B	C	D	E	F	G	H
1								
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +			
5		Fiabilidad sanos	0,98		Prueba -			
6		Fiabilidad enfermos	0,98		TOT		=C3*C4	
7								

	A	B	C	D	E	F	G	H
1								
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +			
5		Fiabilidad sanos	0,98		Prueba -			
6		Fiabilidad enfermos	0,98		TOT	=C3-G6	50	
7								

	A	B	C	D	E	F	G	H
1								
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +			
5		Fiabilidad sanos	0,98		Prueba -	=F6*C5		
6		Fiabilidad enfermos	0,98		TOT	9950	50	
7								

	A	B	C	D	E	F	G	H
1								
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +	=F6-F5		
5		Fiabilidad sanos	0,98		Prueba -	9751		
6		Fiabilidad enfermos	0,98		TOT	9950	50	
7								



	A	B	C	D	E	F	G	H
1								
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +	199	=G6*C6	
5		Fiabilidad sanos	0,98		Prueba -	9751		
6		Fiabilidad enfermos	0,98		TOT	9950	50	

	A	B	C	D	E	F	G	H
1								
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +	199	49	
5		Fiabilidad sanos	0,98		Prueba -	9751	=G6-G4	
6		Fiabilidad enfermos	0,98		TOT	9950	50	

	A	B	C	D	E	F	G	H
1								
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +	199	49	=F4+G4
5		Fiabilidad sanos	0,98		Prueba -	9751	1	
6		Fiabilidad enfermos	0,98		TOT	9950	50	

	A	B	C	D	E	F	G	H
1								
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +	199	49	248
5		Fiabilidad sanos	0,98		Prueba -	9751	1	=F5+G5
6		Fiabilidad enfermos	0,98		TOT	9950	50	

	A	B	C	D	E	F	G	H
1								
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +	199	49	248
5		Fiabilidad sanos	0,98		Prueba -	9751	1	9752
6		Fiabilidad enfermos	0,98		TOT	9950	50	=F6+G6



Una vez construida la tabla de contingencia de casos podemos calcular la tabla de probabilidades:

	A	B	C	D	E	F	G	H
1					TABLA DE CONTINGENCIA			
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +	199	49	248
5		Fiabilidad sanos	0,98		Prueba -	9751	1	9752
6		Fiabilidad enfermos	0,98		TOT	9950	50	10000
7								
8					TABLA DE PROBABILIDADES			
9								
10						Sanos	Enfermos	TOT
11					Prueba +			
12					Prueba -			
13					TOT			

	A	B	C	D	E	F	G	H
1					TABLA DE CONTINGENCIA			
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +	199	49	248
5		Fiabilidad sanos	0,98		Prueba -	9751	1	9752
6		Fiabilidad enfermos	0,98		TOT	9950	50	10000
7								
8					TABLA DE PROBABILIDADES			
9								
10						Sanos	Enfermos	TOT
11					Prueba +	=F4/H4		
12					Prueba -			
13					TOT			

	A	B	C	D	E	F	G	H
1					TABLA DE CONTINGENCIA			
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +	199	49	248
5		Fiabilidad sanos	0,98		Prueba -	9751	1	9752
6		Fiabilidad enfermos	0,98		TOT	9950	50	10000
7								
8					TABLA DE PROBABILIDADES			
9								
10						Sanos	Enfermos	TOT
11					Prueba +	80,2%	=G4/H4	
12					Prueba -			
13					TOT			



	A	B	C	D	E	F	G	H
1					TABLA DE CONTINGENCIA			
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +	199	49	248
5		Fiabilidad sanos	0,98		Prueba -	9751	1	9752
6		Fiabilidad enfermos	0,98		TOT	9950	50	10000
7								
8					TABLA DE PROBABILIDADES			
9								
10						Sanos	Enfermos	TOT
11					Prueba +	80,2%	19,8%	
12					Prueba -	=F5/H5		
13					TOT			

	A	B	C	D	E	F	G	H
1					TABLA DE CONTINGENCIA			
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +	199	49	248
5		Fiabilidad sanos	0,98		Prueba -	9751	1	9752
6		Fiabilidad enfermos	0,98		TOT	9950	50	10000
7								
8					TABLA DE PROBABILIDADES			
9								
10						Sanos	Enfermos	TOT
11					Prueba +	80,2%	19,8%	
12					Prueba -	99,99%	=G5/H5	
13					TOT			

	A	B	C	D	E	F	G	H
1					TABLA DE CONTINGENCIA			
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,005		Prueba +	199	49	248
5		Fiabilidad sanos	0,98		Prueba -	9751	1	9752
6		Fiabilidad enfermos	0,98		TOT	9950	50	10000
7								
8					TABLA DE PROBABILIDADES			
9								
10						Sanos	Enfermos	TOT
11					Prueba +	80,2%	19,8%	
12					Prueba -	99,99%	0,01%	
13					TOT	=F6/H6		



	A	B	C	D	E	F	G	H
1					TABLA DE CONTINGENCIA			
2								
3		<i>Muestra</i>	10000			Sanos	Enfermos	TOT
4		<i>Tasa de incidencia</i>	0,005		Prueba +	199	49	248
5		<i>Fiabilidad sanos</i>	0,98		Prueba -	9751	1	9752
6		<i>Fiabilidad enfermos</i>	0,98		TOT	9950	50	10000
7								
8					TABLA DE PROBABILIDADES			
9								
10						Sanos	Enfermos	TOT
11					Prueba +	80,2%	19,8%	
12					Prueba -	99,99%	0,01%	
13					TOT	99,5%	=G6/H6	

	A	B	C	D	E	F	G	H
1					TABLA DE CONTINGENCIA			
2								
3		<i>Muestra</i>	10000			Sanos	Enfermos	TOT
4		<i>Tasa de incidencia</i>	0,005		Prueba +	199	49	248
5		<i>Fiabilidad sanos</i>	0,98		Prueba -	9751	1	9752
6		<i>Fiabilidad enfermos</i>	0,98		TOT	9950	50	10000
7								
8					TABLA DE PROBABILIDADES			
9								
10						Sanos	Enfermos	TOT
11					Prueba +	80,2%		=F11+G11
12					Prueba -	99,99%	0,01%	
13					TOT	99,5%	0,5%	

Y finalmente tenemos la hoja de cálculo completada:

	A	B	C	D	E	F	G	H
1					TABLA DE CONTINGENCIA			
2								
3		<i>Muestra</i>	10000			Sanos	Enfermos	TOT
4		<i>Tasa de incidencia</i>	0,005		Prueba +	199	49	248
5		<i>Fiabilidad sanos</i>	0,98		Prueba -	9751	1	9752
6		<i>Fiabilidad enfermos</i>	0,98		TOT	9950	50	10000
7								
8					TABLA DE PROBABILIDADES			
9								
10						Sanos	Enfermos	TOT
11					Prueba +	80,2%	19,8%	100,0%
12					Prueba -	99,99%	0,01%	100,0%
13					TOT	99,5%	0,5%	100,0%

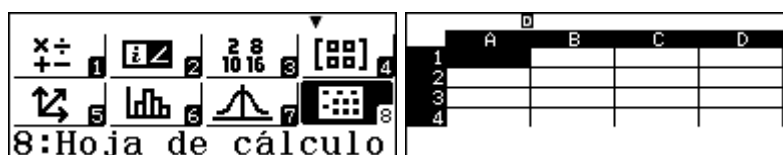
Ahora con sólo cambiar los datos iniciales de la hoja obtenemos los nuevos resultados:

	A	B	C	D	E	F	G	H
1					TABLA DE CONTINGENCIA			
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	=1/1000		Prueba +	199	49	248
5		Fiabilidad sanos	0,98		Prueba -	9751	1	9752
6		Fiabilidad enfermos	0,98		TOT	9950	50	10000
7								
8					TABLA DE PROBABILIDADES			
9								
10						Sanos	Enfermos	TOT
11					Prueba +	80,2%	19,8%	100,0%
12					Prueba -	99,99%	0,01%	100,0%
13					TOT	99,5%	0,5%	100,0%

	A	B	C	D	E	F	G	H
1					TABLA DE CONTINGENCIA			
2								
3		Muestra	10000			Sanos	Enfermos	TOT
4		Tasa de incidencia	0,001		Prueba +	199,8	9,8	209,6
5		Fiabilidad sanos	0,98		Prueba -	9790,2	0,2	9790,4
6		Fiabilidad enfermos	0,98		TOT	9990	10	10000
7								
8					TABLA DE PROBABILIDADES			
9								
10						Sanos	Enfermos	TOT
11					Prueba +	95,3%	4,7%	100,0%
12					Prueba -	100,00%	0,00%	100,0%
13					TOT	99,9%	0,1%	100,0%

Así que cuanto más rara es la enfermedad en la población (es decir con menor tasa de incidencia), para la misma fiabilidad de prueba, mayor es la probabilidad de los falsos positivos!

Las mismas tablas que hemos construido con EXCEL pueden introducirse en la hoja de cálculo de la calculadora:



D				
A	B	C	D	
2	10000			
3	5×10^3			
4	0.98			50
5	0.98			
0.98				
=A2×A3				
D				
A	B	C	D	
2	10000			
3	5×10^3			9751
4	0.98	9950		50
5	0.98			
=A2-D4				
=C4×A4				
D				
A	B	C	D	
2	10000		199	
3	5×10^3	9751		49
4	0.98	9950		50
5	0.98			
=C4-C3				
=D4×A5				
D				
A	B	C	D	E
2	10000		199	49
3	5×10^3	9751		1
4	0.98	9950		50
5	0.98			
=D4-D2				
=C2+D2				
D				
B	C	D	E	
2		199	49	248
3		9751	1	
4		9950	50	
5				
=C2+D2				
D				
B	C	D	E	
2		199	49	248
3		9751	1	9752
4		9950	50	10000
5				
=C2+D2				
Pegar: [=]				

1: Cortar y pegar
2: Copiar y pegar
3: Borrar todo
4: Recalcular

Y una vez hecha la tabla de contingencia, generamos la tabla de probabilidades condicionadas:

D				
B	C	D	E	
4	9950	50	10000	
5				
6	0.8024			
7				
=C2÷E2				
D				
B	C	D	E	
4	9950	50	10000	
5				
6	0.8024	0.1975		
7				
=D2÷E2				
D				
B	C	D	E	
4	9950	50	10000	
5				
6	0.8024	0.1975	1	
7				
=C6+D6				
D				
B	C	D	E	
5				
6	0.8024	0.1975	1	
7				
=C3÷E3				
D				
B	C	D	E	
5				
6	0.8024	0.1975	1	
7				
=D3÷E3				
D				
B	C	D	E	
6	0.8024	0.1975	1	
7				
=C7+D7				
D				
B	C	D	E	
6	0.8024	0.1975	1	
7				
=C4÷E4				
D				
B	C	D	E	
6	0.8024	0.1975	1	
7				
=D4÷E4				
D				
B	C	D	E	
6	0.8024	0.1975	1	
7				
=C8+D8				

Y ahora, de nuevo, con sólo cambiar el valor de la tasa de incidencia de la enfermedad los valores de la tabla se modificarán automáticamente:

D				
A	B	C	D	
2	10000		199	49
3	5×10^3		9751	1
4	0.98		9950	50
5	0.98			
=1÷10000				
D				
B	C	D	E	
2		199.8	9.8	209.6
3		9790.2	0.2	9790.4
4		9990	10	10000
5				
=1÷10000				
D				
B	C	D	E	
5				
6		0.9532	0.0467	1
7		0.9999	2×10^{-5}	1
8		0.999	1×10^{-3}	1

A3. Si se estima que la incidencia del cáncer de mama entre la población femenina es de un 1% (es decir 1 de cada 100 mujeres), al tomar una muestra de población de 10.000 mujeres a las cuáles se les realiza una mamografía:

mamografía	masa maligna (cáncer)	masa benigna (no cáncer)
positiva	79,2%	9,6%
negativa	20,8%	90,4%

1. ¿Cuál es la probabilidad de que una mujer que ha dado positivo en la prueba tenga cáncer?

Una vez realizada la actividad 2 y con la hoja de cálculo diseñada, tan sólo hay que modificar los datos e interpretar los resultados de las tablas:

	A	B	C	D	E	F	G	H
1					TABLA DE CONTINGENCIA			
2								
3		<i>Muestra</i>	10000			Sanos	Enfermos	TOT
4		<i>Tasa de incidencia</i>	0,01		Prueba +	950,4	79,2	1029,6
5		<i>Fiabilidad sanos</i>	0,904		Prueba -	8949,6	20,8	8970,4
6		<i>Fiabilidad enfermos</i>	0,792		TOT	9900	100	10000
7								
8					TABLA DE PROBABILIDADES			
9								
10						Sanos	Enfermos	TOT
11					Prueba +	92,3%	7,7%	100,0%
12					Prueba -	99,77%	0,23%	100,0%
13					TOT	99,0%	1,0%	100,0%

	A	B	C	D
2	10000		950,4	79,2
3	0,01		8949,6	20,8
4	0,904		9900	100
5	0,792			

10000

	B	C	D	E
2		950,4	79,2	1029,6
3		8949,6	20,8	8970,4
4		9900	100	10000

	B	C	D	E
6		0,923	0,0769	1
7		0,9976	2,3x10 ⁻³	1
8		0,99	0,01	1

Es decir que para este tipo de enfermedad y este tipo de prueba, la probabilidad de falsos positivos es del 92,3% mientras que la de falsos negativos es sólo del 0,23%. Este ejemplo muestra la importancia de conocer la probabilidad de falsos positivos, puesto que las pruebas de confirmación en caso de positivo de mamografía son un poco más agresiva ya que consisten en biopsias de extracción de tejido sospechoso.