

TEOREMA DE BAYES

FABIOLA VÁZQUEZ

27 de octubre de 2020

1. Introducción

El objetivo de este estudio es trabajar con el *teorema de Bayes* y su aplicación en la interpretación de las pruebas diagnósticas de COVID-19.

2. Análisis

Cuando se realiza un tipo de prueba, se espera saber con certeza si el paciente está enfermo o no, pero dado que ninguna prueba es perfecta, puede haber errores. En particular, hay cuatro tipos de resultados posibles, que están descritos en el cuadro 1.

Los valores que representan los aciertos de las prueba son la sensibilidad, $P(+ \mid \text{Enfermo})$, y la especificidad, $P(- \mid \text{No enfermo})$. A partir de estos, otro tipo de valores se pueden obtener con el *teorema de Bayes* (teorema 1), como lo son $P(\text{Enfermo} \mid +)$ y $P(\text{No enfermo} \mid -)$.

Teorema 1. Sea B_1, B_2, \dots una partición del espacio muestral tal que $P(B_i) > 0$, para toda i , y A un evento cualquiera con probabilidad positiva, entonces

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{\infty} P(A \mid B_j)P(B_j)}. \quad (1)$$

Por el teorema anterior, se tiene que

$$P(\text{Enfermo} \mid +) = \frac{P(+ \mid \text{Enfermo}) \times P(\text{Enfermo})}{P(+ \mid \text{Enfermo}) \times P(\text{Enfermo}) + P(+ \mid \text{No enfermo}) \times P(\text{No enfermo})}, \quad (2)$$

Cuadro 1: Diferentes tipos de resultados en una prueba.

Resultado	No enfermo	Enfermo
+	Falso positivo	Verdadero positivo
-	Verdadero negativo	Falso negativo

Cuadro 2: Datos obtenidos asumiendo sensibilidad del 80 % y especificidad del 97.5 %.

Resultado	No enfermo	Enfermo	Total
+	74,691	3,501	78,192
−	18,672	136,546	155,218
Total	93,363	140,047	233,410

donde $P(+ | Enfermo)$ es la sensibilidad y $P(+ | No enfermo)$ es igual a 1−especificidad. Reescribiendo la ecuación 2, se tiene

$$P(Enfermo | +) = \frac{\text{sensibilidad} \times P(Enfermo)}{\text{sensibilidad} \times P(Enfermo) + (1 - \text{especificidad}) \times (1 - P(Enfermo))}. \quad (3)$$

2.1. COVID-19

En el estado de Nuevo León, se han realizado un total de 233,412 pruebas, resultando 78,389 de estas positivas y 155,023 negativas [3]. Según los indicadores del gobierno [4], en promedio un 40 % de las pruebas salen positivas.

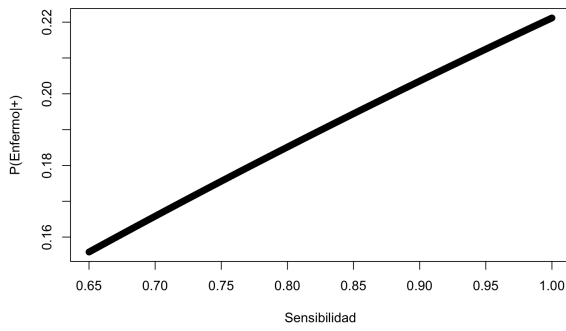
En el cuadro 2 se considera la realización de 233,410 pruebas, con una sensibilidad del 80 %, una especificidad del 97.5 % y una probabilidad marginal de tener COVID-19 del 40 %. Esto lleva a los resultados mostrados, cuyos valores se aproximan a los reales. Además, con dichos datos se calcula la probabilidad $P(Enfermo | +)$ y se obtiene que la probabilidad de estar enfermo, dado que la prueba de COVID-19 salió positiva, es de 95.5 %.

Si por el contrario consideramos la probabilidad de tener COVID-19 como la cantidad de casos confirmados entre la población total de 5.12 millones, tendríamos una probabilidad marginal del 1 %. Si además se asume una sensibilidad y una especificidad del 90 %, la $P(Enfermo | +)$ cambia y ahora se tiene una probabilidad del 8.3 %.

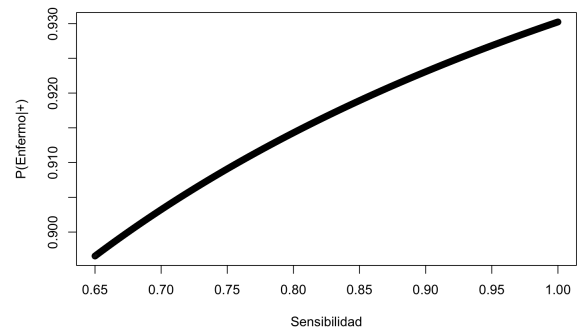
Para analizar las variaciones de $P(Enfermo | +)$, se consideran distintos valores de $P(Enfermo)$ y se varía el valor de la sensibilidad o de la especificidad de la prueba. Dicho experimento se realiza en el software R [6] en un cuaderno de Jupyter [5], y los resultados se observan en las figuras 1 y 2.

2.2. Influenza

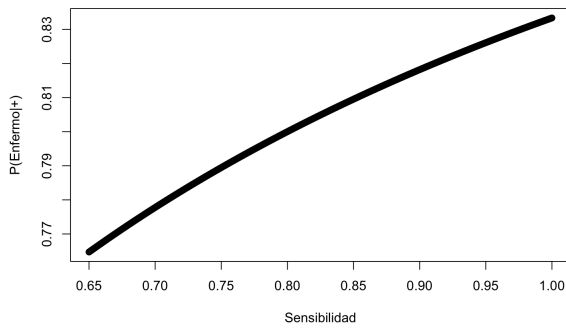
Con la temporada de influenza acercándose, y dada la similitud de sus síntomas con los del COVID-19, será de particular importancia distinguir entre estas enfermedades. De acuerdo con los datos del CDC, las pruebas rápidas de influenza tienen una sensibilidad de 50 % a 70 %, y una especificidad de 90 % a 95 % [2]. En el invierno 2018-2019, se estimaron 36 millones de casos de influenza en EE. UU. [1], con una población de 329 millones [7]. Con los datos mencionados, se tendría $P(Enfermo) = 36/329 \approx 0.10$, y una probabilidad condicional que va de $P(Enfermo|+) = .35$ a $P(Enfermo|+) = .60$.



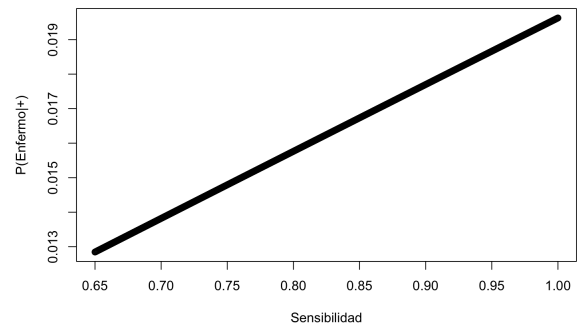
(a) $P(\text{Enfermo}) = 0.01$.



(b) $P(\text{Enfermo}) = 0.4$.

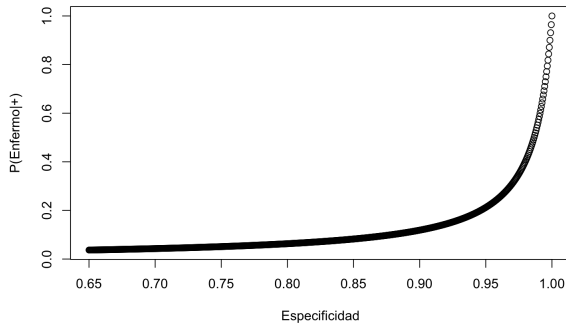


(c) $P(\text{Enfermo}) = 0.2$.

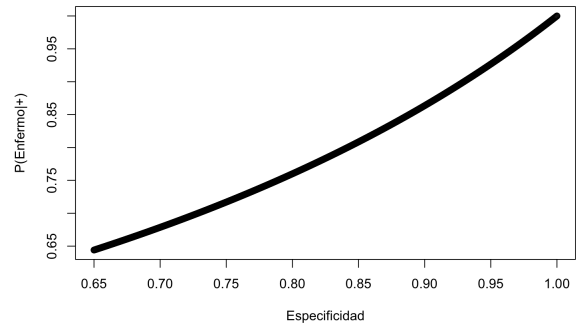


(d) $P(\text{Enfermo}) = 0.001$.

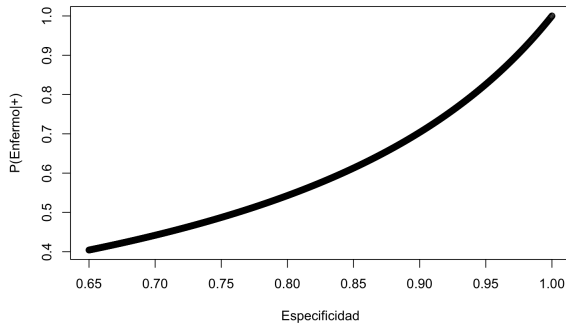
Figura 1: Considerando la especificidad como el 95 %.



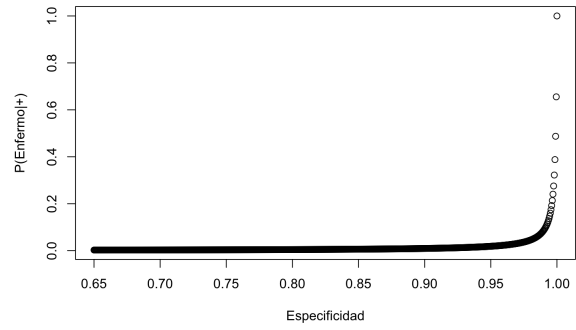
(a) $P(\text{Enfermo}) = 0.01$.



(b) $P(\text{Enfermo}) = 0.4$.



(c) $P(\text{Enfermo}) = 0.2$.



(d) $P(\text{Enfermo}) = 0.001$.

Figura 2: Considerando la sensibilidad del 95 %.

Referencias

- [1] CDC. Disease burden of influenza, 2020. <https://www.cdc.gov/flu/about/burden/index.html>.
- [2] CDC. Information for clinicians on Rapid Diagnostic Testing for influenza, 2020. <https://www.cdc.gov/flu/professionals/diagnosis/rapidclin.htm#table2>.
- [3] Gobierno del estado de Nuevo León. Casos de COVID-19 en Nuevo León. <https://www.nl.gob.mx/publicaciones/casos-de-covid-19-en-nuevo-leon>.
- [4] Gobierno del Estado de Nuevo León. Monitoreo de indicadores estatales de salud para la reapertura económica semana 42 (11 de octubre – 17 de octubre), 2020. <https://www.nl.gob.mx/presentacion-indicadoresymedidasdemitigacion-covid19-22-10-2020>.
- [5] Thomas Kluyver, Benjamin Ragan-Kelley, Pérez, et al. Jupyter notebooks—a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proceedings of the 20th International Conference on Electronic Publishing*, page 87. IOS Press, 2016.
- [6] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
- [7] United States Census Bureau. U.S. and world population clock, 2020. <https://www.census.gov/popclock/>.