



Predicción de Enfermedades Cardíacas Utilizando el Teorema de Bayes: Un Enfoque Aplicado en R Studio

Predicting Heart Disease Using Bayes' Theorem: An Approach Applied in R Studio

Previsão de doenças cardíacas usando o Teorema de Bayes: uma abordagem aplicada no R Studio

Carmen Mariana Sánchez Tenesaca ^I
csanchezt3@unemi.edu.ec
<https://orcid.org/0009-0006-7253-7810>

Mirya Liaceci Sanchez Salazar ^{II}
msanchezs32@unemi.edu.ec
<https://orcid.org/0009-0007-1758-7716>

Correspondencia: csanchezt3@unemi.edu.ec

Ciencias de la Salud
Artículo de Investigación

* **Recibido:** 26 de julio de 2025 * **Aceptado:** 22 de agosto de 2025 * **Publicado:** 30 de septiembre de 2025

- I. Universidad Estatal de Milagro, Ecuador.
- II. Universidad Estatal de Milagro, Ecuador.

Resumen

Este estudio se adentra en la compleja y uso de la aplicación del Teorema de Bayes para predecir de manera más precisa la aparición de diversas enfermedades a partir de variables clínicas que son fácilmente accesibles y pueden ser recogidas en un entorno clínico habitual. Se implementó un enfoque estadístico avanzado utilizando R Studio, lo que facilitó notablemente el modelado de la probabilidad de desarrollar una enfermedad bajo ciertas condiciones previas y conocidas. A partir de un conjunto de datos públicos, relacionados con el diagnóstico de diversas enfermedades, se desarrolló un modelo bayesiano predictivo que permite realizar evaluaciones más acertadas. Los hallazgos de este estudio demuestran de manera contundente la eficacia del enfoque probabilístico al momento de tomar decisiones clínicas cruciales, proporcionando así una mayor precisión diagnóstica, especialmente en contextos en los que los recursos son limitados y es esencial optimizar su uso. El artículo subraya de manera enfática la relevancia del teorema en situaciones concretas e importantes del ámbito de la salud pública, sugiriendo diversas aplicaciones para futuros estudios e investigaciones.

Palabras Clave: Teorema de Bayes; predicción de enfermedades; estadística; probabilidad; R Studio.

Abstract

This study delves into the complex application of Bayes' Theorem to more accurately predict the onset of various diseases using clinical variables that are easily accessible and can be collected in a routine clinical setting. An advanced statistical approach was implemented using R Studio, which significantly facilitated the modeling of the probability of developing a disease under certain known preconditions. Using a public dataset related to the diagnosis of various diseases, a predictive Bayesian model was developed that allows for more accurate assessments. The findings of this study strongly demonstrate the effectiveness of the probabilistic approach in making crucial clinical decisions, thus providing greater diagnostic accuracy, especially in contexts where resources are limited and optimizing their use is essential. The article emphatically underlines the relevance of the theorem in specific and important situations in the field of public health, suggesting various applications for future studies and research.

Keywords: Bayes' Theorem; disease prediction; statistics; probability; R Studio.

Resumo

Este estudo investiga a complexa aplicação do Teorema de Bayes para prever com maior precisão o aparecimento de diversas doenças, utilizando variáveis clínicas de fácil acesso e que podem ser recolhidas num ambiente clínico de rotina. Foi implementada uma abordagem estatística avançada utilizando o R Studio, o que facilitou significativamente a modelação da probabilidade de desenvolvimento de uma doença sob determinadas pré-condições conhecidas. Utilizando um conjunto de dados públicos relacionados com o diagnóstico de diversas doenças, foi desenvolvido um modelo bayesiano preditivo que permite avaliações mais precisas. Os resultados deste estudo demonstram fortemente a eficácia da abordagem probabilística na tomada de decisões clínicas cruciais, proporcionando assim uma maior precisão diagnóstica, especialmente em contextos onde os recursos são limitados e a otimização da sua utilização é essencial. O artigo enfatiza enfaticamente a relevância do teorema em situações específicas e importantes no campo da saúde pública, sugerindo diversas aplicações para estudos e pesquisas futuras.

Palavras-chave: Teorema de Bayes; previsão de doenças; estatística; probabilidade; Estúdio R.

Introducción

En el contexto de la salud pública y la medicina clínica, la implementación de decisiones fundamentadas en evidencias estadísticas se vuelve esencial para optimizar el diagnóstico y el tratamiento efectivo de diversas patologías. Un método significativo en el análisis probabilístico es el Teorema de Bayes, que proporciona un mecanismo para ajustar la probabilidad de un evento a partir de un nuevo conjunto de evidencias. Esta herramienta analítica se revela particularmente valiosa cuando existen datos preliminares, tales como síntomas o antecedentes médicos, y se busca calcular la probabilidad de que un paciente sufra de una enfermedad particular (Polotskaya, Muñoz Valencia, Rabasa, & Quesada, 2024).

La utilización del Teorema de Bayes en la construcción de modelos predictivos ha evidenciado su eficacia en múltiples áreas, siendo especialmente notable su aplicación en el ámbito de la medicina, donde sobresale por su claridad en la interpretación y su aplicabilidad práctica. En este estudio, se expone una implementación computacional del enfoque bayesiano centrada en la predicción de enfermedades mediante el análisis de datos reales. Empleando R Studio, se desarrolla y valida un modelo que facilita la estimación de la probabilidad condicional de contraer una enfermedad basada en diversos factores observables.

Aparte de su relevancia desde un enfoque teórico, el análisis busca señalar la manera en que los métodos estadísticos al alcance pueden influir de manera significativa en la optimización del diagnóstico clínico, sobre todo en entornos donde los recursos son escasos. Con este propósito, se recurre a un conjunto de datos de acceso público que brinda información sobre diagnósticos médicos, lo que facilita tanto la replicación de estudios como la comprensión accesible por parte de estudiantes e investigadores que se dedican a la estadística aplicada.

Marco Teórico

La estadística desempeña un papel esencial en la medicina y la salud pública, permitiendo el análisis sistemático de datos clínicos para apoyar la toma de decisiones (Vera Damián & Altamirano Mego, 2024). En este marco, la probabilidad se presenta como una herramienta esencial para representar la incertidumbre que caracteriza los procesos de diagnóstico y tratamiento. Uno de los métodos más sólidos para la gestión probabilística de la información médica es el Teorema de Bayes, el cual es una fórmula matemática que facilita la actualización de la probabilidad de un evento a partir de evidencia nueva.

Fundamentos del Teorema de Bayes

El Teorema de Bayes, diseñado por el matemático Thomas Bayes en el siglo XVIII, explica el proceso mediante el cual la probabilidad de ocurrencia de un evento puede ser ajustada al considerar nueva información. Desde una perspectiva matemática, se establece de la siguiente manera:

$$P(A/B) = \frac{P(B/A) P(A)}{P(B)}$$

Donde:

- $P(A|B)$ es la probabilidad posterior de que ocurra el evento A dado que ocurrió B.
- $P(B|A)$ es la probabilidad de observar B dado que A es verdadero.
- $P(A)$ es la probabilidad previa de A.
- $P(B)$ es la probabilidad total de B.

Este método se muestra especialmente eficaz en el ámbito de la medicina, ya que facilita la estimación de la probabilidad de que un paciente sufra de una enfermedad (A) basándose en la presencia de un síntoma o hallazgo clínico (B).

El modelo Naive Bayes, fundamentado en este teorema, parte de la premisa de que las características predictivas son independientes entre sí. Aunque esta suposición puede ser considerada simplista, el modelo ha demostrado un rendimiento sobresaliente en diversas tareas de clasificación, incluidas aplicaciones en el ámbito médico, como la predicción de enfermedades cardiovasculares, debido a su capacidad para generar clasificadores que son tanto eficientes como rápidos (Rumbaut Rangel & Maridueña Arroyave, 2024).

La integración de este enfoque teórico junto con herramientas de software como R Studio proporciona a los analistas una opción robusta y accesible para aplicar modelos predictivos en situaciones clínicas prácticas.

Aplicaciones del Teorema de Bayes en Medicina

En el campo médico, el Teorema de Bayes se ha aplicado exitosamente al diagnóstico de enfermedades como cáncer, enfermedades cardíacas y trastornos infecciosos. Por ejemplo, en estudios de detección temprana, permite mejorar la interpretación de pruebas diagnósticas al considerar la prevalencia de la enfermedad en una población determinada (Teng, Zhang, Liu, Shu, & Ye, 2022).

Además, en contextos de salud pública, el enfoque bayesiano facilita la toma de decisiones bajo condiciones de incertidumbre, optimizando recursos clínicos limitados. Investigaciones recientes han demostrado que modelos bayesianos bien calibrados pueden superar a los modelos clásicos en precisión diagnóstica, especialmente cuando se dispone de muestras pequeñas o información incompleta (Cerdeira Lorca, Castro, Bambs, & Balcells, 2025).

La identificación de los factores de riesgo que influyen en las caídas del adulto mayor puede optimizarse mediante herramientas estadísticas como el teorema de Bayes, el cual permite mejorar la precisión de los diagnósticos a partir de información previa y condicional (Escobar Castellanos & Martínez Martínez, 2018).

El teorema de Bayes es una herramienta estadística fundamental en medicina, ya que permite actualizar la probabilidad de una enfermedad con base en los resultados de pruebas diagnósticas,

mejorando así la toma de decisiones clínicas al incorporar la sensibilidad, especificidad y prevalencia de la condición evaluada (Palomar Morales & Guerrero Morales, 2017).

Modelos Predictivos y R Studio

El software **R Studio** ha cobrado popularidad como herramienta para el análisis estadístico y la implementación de modelos bayesianos. Este entorno de programación permite construir modelos probabilísticos reproducibles, utilizando bibliotecas especializadas como e1071, caret y bayesrule. Su capacidad de visualización y procesamiento de datos lo convierte en una opción ideal para estudiantes e investigadores en salud interesados en análisis aplicado (Kabacoff, 2022).

Metodología

Fuente de los datos

En el presente análisis se empleó el conjunto de datos denominado "Heart Disease UCI", el cual es accesible de forma pública en el repositorio UCI Machine Learning Repository y ha sido replicado en GitHub. Este conjunto de datos incluye información clínica relevante de pacientes, con el propósito de prever la existencia de enfermedades cardíacas mediante el uso de diversas variables biomédicas.

Descripción del conjunto de datos El conjunto original incluye 303 registros y 14 variables, entre las cuales se destacan:

- age: edad del paciente (años),
- sex: sexo (1 = hombre; 0 = mujer),
- cp: tipo de dolor en el pecho,
- trestbps: presión arterial en reposo (mm Hg),
- chol: colesterol sérico (mg/dl),
- fbs: azúcar en sangre en ayunas (>120 mg/dl),
- restecg: resultados del electrocardiograma,
- thalach: frecuencia cardíaca máxima alcanzada,
- exang: angina inducida por ejercicio,
- oldpeak: depresión ST inducida por ejercicio,
- slope: pendiente del segmento ST en el ECG,

- ca: número de vasos principales coloreados por fluoroscopia,
- thal: defecto talámico,
- target: presencia de enfermedad cardíaca (1 = enfermedad; 0 = sano).

La variable target fue utilizada como variable dependiente (respuesta), mientras que el resto se consideró como variables predictoras.

Preparación de los datos

Se llevó a cabo un proceso fundamental de depuración de datos en R Studio. Se constató que no existían valores nulos en el conjunto de datos analizado, lo que eliminó la necesidad de implementar técnicas de imputación. A continuación, el conjunto de datos se segmentó en dos subconjuntos: uno de entrenamiento, que representó el 80 %, y otro de prueba, correspondiente al 20 %. Esta partición se realizó utilizando la función `createDataPartition()` del paquete `caret`, garantizando así que la distribución de la variable objetivo se mantuviera proporcional en ambas divisiones.

Modelo bayesiano

El modelo de clasificación fue diseñado empleando el algoritmo Naive Bayes, que fue implementado en el entorno R utilizando la función `naiveBayes()` del paquete `e1071`. Este enfoque se basa en la premisa de independencia condicional entre las variables predictoras, lo que facilita la simplificación del cálculo de la probabilidad posterior. La formulación general del Teorema de Bayes aplicada se expresa como:

$$P(A/B) = \frac{P(B/A) P(A)}{P(B)}$$

Evaluación del modelo

La validación se llevó a cabo utilizando el conjunto de prueba, donde se realizó una comparación entre las predicciones generadas por el modelo y los valores reales. Se implementó la matriz de confusión y se calcularon diversas métricas, incluyendo precisión, sensibilidad, especificidad y la

exactitud global, con el propósito de evaluar de manera integral el rendimiento predictivo del modelo.

Análisis de Resultados y Discusión

El modelo Naive Bayes implementado demostró una capacidad predictiva significativa para diagnosticar enfermedades cardíacas a partir de variables clínicas básicas. La **precisión general del modelo alcanzó un 85%**, lo que indica que 85 de cada 100 casos fueron correctamente clasificados como enfermo o no enfermo.

Desde el punto de vista clínico, la **sensibilidad** del modelo fue del **84.85%**, lo cual significa que el algoritmo logró identificar correctamente al 84.85% de los pacientes con enfermedad cardíaca. Esta es una métrica crítica en contextos de salud pública, ya que una alta sensibilidad implica un bajo nivel de falsos negativos, lo que reduce el riesgo de omitir diagnósticos en pacientes que realmente necesitan atención médica.

Por otro lado, la **especificidad** fue del **85.19%**, lo que implica que el modelo también fue eficaz al identificar correctamente a los pacientes sanos. Esta característica es relevante para evitar tratamientos innecesarios, optimizando el uso de recursos en entornos clínicos con limitaciones presupuestarias como suele ser el caso en muchas regiones del Ecuador.

La **valoración del Kappa (0.6959)** sugiere una concordancia sustancial entre las predicciones del modelo y los diagnósticos reales, y el valor **p del test de McNemar (1.00)** indica que no hay evidencia de desequilibrio entre falsos positivos y falsos negativos, lo cual refuerza la estabilidad del modelo.

En comparación con enfoques tradicionales como la regresión logística o los árboles de decisión, el modelo Naive Bayes tiene la ventaja de requerir menos recursos computacionales y de ofrecer interpretaciones probabilísticas útiles para decisiones clínicas rápidas. Además, su desempeño es robusto aun cuando algunas de las variables no son independientes, una condición que raramente se cumple en datos médicos reales.

Finalmente, este estudio demuestra que herramientas estadísticas simples, accesibles y fácilmente implementables en plataformas como R Studio, pueden tener un impacto tangible en la toma de decisiones clínicas, especialmente en regiones donde el acceso a tecnologías avanzadas es limitado.

Conclusiones

La aplicación del Teorema de Bayes mediante el modelo Naive Bayes en este estudio permitió demostrar la utilidad de los enfoques probabilísticos para la predicción de enfermedades a partir de datos clínicos accesibles. Con una precisión del 85%, el modelo evidenció un desempeño sólido, caracterizado por una alta sensibilidad (84.85%) y especificidad (85.19%), indicadores clave en contextos donde la precisión diagnóstica es fundamental.

Este tipo de herramientas, basadas en la estadística y la probabilidad, ofrecen soluciones efectivas para entornos clínicos con recursos limitados, como ocurre en diversas zonas de Ecuador. Su facilidad de implementación en plataformas como R Studio las convierte en aliadas estratégicas para fortalecer la toma de decisiones médicas, la priorización de casos y la optimización de los recursos en el sistema de salud.

El presente trabajo también destaca la importancia del análisis estadístico como base para construir modelos matemáticos interpretables y replicables, que contribuyan a mejorar el diagnóstico temprano, la gestión de enfermedades y, en última instancia, la salud pública.

En futuras investigaciones, podría evaluarse la inclusión de variables contextuales locales, como hábitos alimenticios, factores socioeconómicos o condiciones geográficas específicas, lo que permitiría adaptar el modelo a realidades particulares y mejorar su poder predictivo.

Referencias

- Cerda Lorca, J., Castro, R., Bambs, C., & Balcells, M. E. (2025). Análisis Bayesiano de un ensayo controlado aleatorizado: aspectos matemáticos y clínicos. *Revista chilena de infectología*.
- Kabacoff, R. (2022). *R in action, data analysis and graphics with R*. Estados Unidos: Manning Publications .
- Polotskaya, K., Muñoz Valencia, C. s., Rabasa, A., & Quesada, J. A. (2024). Bayesian Networks for the Diagnosis and Prognosis of Diseases: A Scoping Review.
- Rumbaut Rangel, D., & Maridueña Arroyave, M. R. (2024). Modelo predictivo de enfermedades cardiovasculares basado en Redes Bayesianas. *Dialnet*.
- Teng, J., Zhang, H., Liu, W., Shu, X., & Ye, F. (2022). A Dynamic Bayesian Model for Breast Cancer Survival Prediction. *IEEE Journal of Biomedical and Health Informatics*.
- Vera Damián, D. V., & Altamirano Mego, G. (27 de Diciembre de 2024). Repositorio Institucional UNPRG. Obtenido de <https://repositorio.unprg.edu.pe/handle/20.500.12893/14418>
- Escobar Castellanos, B., & Martínez Martínez, M. G. (2018). Factores de riesgo de caídas en el adulto mayor. Universidad de Sonora. Recuperado de https://www.researchgate.net/publication/327881264_Factores_de_riesgo_de_caidas_en_el_adulto_mayor
- Saavedra López, D., Leyva Vázquez, M., & Armentero Moreno, Y. (2013). Representación del conocimiento causal. Aplicaciones en la medicina. IX Congreso Internacional Informática en Salud. Universidad de las Ciencias Informáticas. <https://eventos.sld.cu/informaticasalud2013/>.