



**UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO**  
**ESCUELA DE POSGRADO**

**MAESTRÍA EN ESTADÍSTICA**

**PROYECTO DE TESIS**

**COMPARACIÓN DE LA INFERENCIA ESTADÍSTICA  
CLÁSICA CON LA INFERENCIA BAYESIANA  
MEDIANTE REDES BAYESIANAS PARA LA  
DECISIÓN EN LA DETECCIÓN DE QUERATOCONO**

**PARA OPTAR AL GRADO ACADÉMICO DE MAESTRO EN  
ESTADÍSTICA**

**AUTOR:**

Br. KEVIN HEBERTH HAQUEHUA APAZA

**ASESOR:**

Dr. NOMBRE ASESOR

CODIGO ORCID: 0002-0002-0002-7998

**CUSCO - PERÚ**

**2025**

## ÍNDICE

<b>ÍNDICE</b>	<b>II</b>
<b>ÍNDICE DE TABLAS</b>	<b>IV</b>
<b>ÍNDICE DE FIGURAS</b>	<b>V</b>
<b>1 PLANTEAMIENTO DEL PROBLEMA</b>	<b>1</b>
1.1 Situación problemática . . . . .	1
1.2 Formulación del problema . . . . .	4
1.2.1 Problema general . . . . .	4
1.2.2 Problemas específicos . . . . .	4
1.3 Justificación de la investigación . . . . .	4
1.4 Objetivos de la investigación . . . . .	5
1.4.1 Objetivo general . . . . .	5
1.4.2 Objetivos específicos . . . . .	5
<b>2 MARCO TEÓRICO CONCEPTUAL</b>	<b>7</b>
2.1 Bases Teóricas . . . . .	7
2.2 Marco conceptual . . . . .	13
2.3 Antecedentes . . . . .	13
<b>3 HIPÓTESIS Y VARIABLES</b>	<b>16</b>
3.1 Hipótesis . . . . .	16
3.2 Hipótesis general . . . . .	16
3.3 Hipótesis específicas . . . . .	16

3.4	Identificación de variables e indicadores . . . . .	16
3.4.1	Operacionalización de variables . . . . .	18
<b>4</b>	<b>METODOLOGÍA</b>	<b>21</b>
4.1	Ámbito de estudio: localización política y geográfica . . . . .	21
4.2	Tipo y nivel de investigación . . . . .	21
4.3	Unidad de análisis . . . . .	21
4.4	Población de estudio . . . . .	21
4.5	Tamaño de muestra . . . . .	22
4.6	Técnicas de selección de muestra . . . . .	22
4.7	Técnicas de recolección de información . . . . .	22
4.8	Técnicas de análisis e interpretación de la información . . . . .	22
4.9	Técnicas para demostrar la verdad o falsedad de las hipótesis planteadas . . . .	23
	<b>PRESUPUESTO</b>	<b>23</b>
	<b>CRONOGRAMA DE REALIZACIÓN DE TESIS</b>	<b>24</b>
	<b>BIBLIOGRAFÍA</b>	<b>25</b>
	<b>ANEXOS</b>	<b>28</b>
a.	Matriz de consistencia . . . . .	29

## ÍNDICE DE TABLAS

<b>Tabla 1</b>	Ejemplos de distribuciones conjugadas . . . . .	10
<b>Tabla 2</b>	Matriz de operacionalización de variables . . . . .	18
<b>Tabla 3</b>	Lista de productos y sus costos . . . . .	24
<b>Tabla 4</b>	Plan de actividades . . . . .	24

## ÍNDICE DE FIGURAS

<b>Figura 1</b>	Grafo de una red bayesiana entre enfermedades y hallazgos . . . . .	12
-----------------	---	----

# **CAPÍTULO I**

## **PLANTEAMIENTO DEL PROBLEMA**

### **1.1 Situación problemática**

La estadística como ciencia y base clave para resolver problemas reales ha sido de gran importancia a lo largo del tiempo (Spiegelhalter, 2023), en el cual la parte de la estadística aplicada apoyo a investigadores de otras áreas a poder realizar avances y descubrimientos en sus respectivas áreas. Sin embargo hubo un gran debate entre el uso de la estadística frecuentista y la estadística bayesiana para la resolución de problemas complejos hasta los tiempos actuales (Clayton, 2021).

Algo importante a tomar en cuenta es la filosofía de la estadística frecuentista acerca de la probabilidad tomado desde el ensayo de Bernoulli, seguido de la fuerte base matemática apoyada en Neyman, Karl Pearson y Ronald Fisher. Asimismo se tiene el enfoque bayesiano basado en la condición y actualización de la probabilidad desarrollado por el reverendo Thomas Bayes, apoyado de matemáticos como Pierre-Simon Laplace o Carl Friedrich Gauss, especialmente al querer solucionar problemas reales en la aplicación de la probabilidad (Clayton, 2021).

Por ejemplo Clayton (2021) toma el caso en el que indiques la probabilidad de adivinar el resultado de un lanzamiento de un dado. El enfoque de la estadística clásica y frecuentista te indica que es  $1/6$ , pero que pasa si se te adiciona información en el que en 5 lanzadas realizadas se tuvieron 1 número igual, dos número menores y dos números mayores al que se tiene que adivinar; tu probabilidad no puede seguir siendo  $1/6$  ya que es imposible que el número por adivinar sea 1 o 6 debido a que me salieron números mayores o menores, el siguiente gráfico muestra mejor esta distribución de las probabilidad con los dados sin la información actualizada y con información adicionada. Tomando en cuenta este ejemplo la estadística frecuentista te seguirá indicando que la probabilidad es  $1/6$  sin tomar en cuenta esta información adicional, extendiendo este caso a problemas donde actualmente se aplican las herramientas estadísticas para la resolución de otros problemas más complejos como lo es el del queratocono el cual no tiene hasta ahora una forma de diagnóstico exacto, ya que es basado mayormente en la

probabilidad.

El queratocono es una enfermedad que afecta a la córnea causando mala visión de forma progresiva y evolucia en pacientes jóvenes. Según la Organización Mundial de la Salud (OMS) esta enfermedad puede indicar cifras epidémicas. A pesar de que se considere una afección de impacto moderado y no tan grave que afecta a la función visual y calidad de vida, esta enfermedad puede resultar más perjudicial ya que puede afectar a individuos totalmente activos (THALASSELIS et al., 1988). El queratocono generalmente inicia en la pubertad hasta la tercera o casos de cuarta década de vida (Althomali et al., 2018), sus síntomas son variables y dependen de la etapa de progresión que se encuentre, en etapas iniciales no pueden presentarse síntomas hasta etapas más avanzadas en donde el paciente experimenta una distorsión notable en la visión acompañada de una pérdida visual severa. A pesar de eso no llegan a producir ceguera completa debido a esta enfermedad (Rabinowitz, 1998).

Los primeros estudios de prevalencia se llevó en la población de Indianapolis, USA encontró 600 casos por 100,000 (0.6 %). Un estudio poblacional más reciente informó una prevalencia de 375 por 100,000 (0.375 %) en Países Bajos (Godefrooij et al., 2017). Mientras que en Latino América los registros de prevalencia son muy pocos, teniendo en Paraná entre Ríos-Argentina una prevalencia aproximada de 229 por cada 100,000 habitantes con edad promedio de diagnóstico a los 24.5 años (Pussetto et al., 2011), en la ciudad de Quito en Ecuador se tuvo una prevalencia de (0.75 %) (Mansfield Rojas, 2017), por otra parte en el Hospital Regional Honorio Delgado, durante el periodo 2014 - 2017 en Arequipa-Perú se evidenció que la prevalencia de queratocono fue de 1.37 por cada mil pacientes atendidos (Ramos Almirón, 2018). Por último con respecto a la ciudad del Cusco en el trabajo de investigación de (Samaniego Mojica, 2023) reportó en el Centro de Salud Integral La Fuente una prevalencia de 3.32 % siendo esta mayor a las otras zonas estudiadas en el cual aún no se le esta dando la importancia por parte del estado peruano.

Asimismo en la clínica Oftalmo Salud del Perú indica que no se sabe con certeza porque algunas personas desarrollan Queratocono, aunque hay estudios con asociaciones a la genética o dicho hereditaria, asimismo con alergías y lesiones oculares provocado por la frotación excesiva de los ojos, el uso inapropiado de lentes de contacto y la presencia de algunas patologías como

la queratoconjuntivitis vernal, la retinitis pigmentaria o la retinopatía del prematura. Múltiples estudios han desarrollado técnicas para poder realizar la detección temprana de esta enfermedad ya sea por modelos de Machine Learning que indicaron que índices de PENTACAM fueron anormales para detección y exclusión de pacientes con queratocono (Zhao et al., 2024)

Teniendo esta situación problemática nos encontramos en un marco de incertidumbre sobre una forma de ver la probabilidad de que una persona desarrolle el queratocono. Como se mostró en los estudios anteriores generalmente la inferencia estadística clásica indica posibles asociaciones a diversos factores, especialmente las pruebas de hipótesis que se utilizan en la mayor parte de estudios de investigación. Pero ¿será necesario para indicar si una persona posee o no la enfermedad?. En cambio si combinas esta información para obtener una probabilidad a posteriori se acercaría mayormente a una inferencia causal. Los enfoques bayesianos para problemas bioestadísticos se volvieron comunes en aplicaciones epidemiológicas, el uso de la metodología bayesiana ha experimentado grandes avances, aumentando el enfoque del uso bayesiano (Lawson, 2018). Centrandonos en la construcción de una red bayesiana para relacionar estructura de datos complicados con preguntas científicas, verificando el ajuste de dicho modelo e investigando la sensibilidad de las conclusiones a supuestos de modelos razonables. Gelman et al. (2013) indican que una de las fortalezas del enfoque bayesiano reside en combinar información de diferentes fuentes y dar una explicación más amplia de la incertidumbre acerca de las incógnitas en un problema estadístico. Estudios de enfermedades raras ya aplican este enfoque como lo es de Fouarge et al. (2021) en el que se estudia las miopatías centronucleares con un modelo jerárquico bayesiano con distribución de probabilidad beta y una función logit de enlace, en el cual el autor menciona que este tipo de problemas pueden ser abordado bajo el enfoque bayesiano en el que se deban realizar más investigaciones con el diálogo continuo de autoridades reguladoras que permitan la aplicación de la estadística bayesiana. Asimismo en el trabajo de Férrez Rubio (2017) indicó que una herramienta para apoyar en la tomade decisiones de la medicina basada en la evidencia son las redes bayesianas ya que tiene una capacidad para representar las relaciones causales que tienen las variables involucradas en un problema, y que se visualice de una manera intuitiva.

En el cual el objetivo es comparar la inferencia de la estadística clásica con las pruebas de



hipótesis y la inferencia bayesiana de la probabilidad a posteriori mediante las redes bayesianas para cuantificar la incertidumbre de la detección de la enfermedad del queratocono en la ciudad del Cusco, tomando en cuenta los diferentes estudios de esta enfermedad ya sea en la asociación de variables clínicas, topográficas, tests, etc.

## **1.2 Formulación del problema**

### ***1.2.1 Problema general***

¿Cómo comparar la inferencia estadística clásica con la inferencia bayesiana?

### ***1.2.2 Problemas específicos***

El estudio toma en cuenta los siguientes problemas específicos:

- ¿Qué variables clínicas, demográficas y topográficas se asocian en la detección del queratocono?
- ¿Cómo se obtiene la inferencia de probabilidad en la presencia o ausencia de queratocono a través de las diversas variables?
- ¿Cómo afecta la decisión de usar la inferencia estadística clásica o bayesiana en la detección de queratocono?
- ¿Cómo implementar las inferencias de riesgo a nivel individual?

## **1.3 Justificación de la investigación**

El presente estudio aplica pruebas de hipótesis estadística evaluando su adecuado uso y procedimiento. Asimismo se usarán las redes bayesianas como herramientas metodológicas robustas para el análisis y detección del queratocono. La elección de enfoques bayesianos se fundamenta en su capacidad para incorporar conocimiento previo (experiencia clínica, estudios anteriores, o literatura científica) y actualizar continuamente las inferencias a medida que se dispone de nuevos datos. Por otro lado, las redes bayesianas facilitan la representación gráfica

y probabilística de relaciones causales o condicionales entre variables clínicas, permitiendo la detección temprana, el diagnóstico asistido por probabilidad, y la simulación de escenarios clínicos bajo distintos supuestos.

Asimismo, la implementación de estos modelos permite mejorar la precisión diagnóstica, optimizar los protocolos de atención clínica y reducir el riesgo de error médico en contextos de alta incertidumbre o recursos limitados. Al integrar evidencia previa con datos observacionales, se fortalece la toma de decisiones basada en datos, tanto a nivel individual (diagnóstico del paciente) como poblacional (definición de políticas de intervención). Además, estos enfoques pueden adaptarse a diversos contextos geográficos o epidemiológicos mediante su capacidad de actualización dinámica, permitiendo así protocolos clínicos personalizados y adaptativos. En conjunto, esta metodología no solo mejora la eficacia clínica, sino que también contribuye a una asignación más eficiente de los recursos sanitarios y al diseño de estrategias preventivas basadas en riesgo.

## **1.4 Objetivos de la investigación**

### ***1.4.1 Objetivo general***

Comparar la inferencia estadística clásica con la inferencia bayesiana para la detección de queratocono.

### ***1.4.2 Objetivos específicos***

- Determinar las variables clínicas, demográficas y topográficas asociadas que apoyen en la detección del queratocono.
- Construir una red bayesiana estructurada que infiera la probabilidad de presencia o ausencia del queratocono a partir de las variables observadas y como estas se actualizan al variar sus parámetros.
- Evaluar la detección de queratocono para futuros pacientes utilizando la inferencia estadística clásica y la inferencia bayesiana.

- Implementar un prototipo computacional que visualice las inferencias de riesgo de quetocono a nivel individual.

## CAPÍTULO II

### MARCO TEÓRICO CONCEPTUAL

#### 2.1 Bases Teóricas

##### Axiomas de probabilidad y notaciones básicas

**Definición 2.1.1. Notaciones** (Koski y Noble, 2011)

- El conjunto de todos los posibles resultados de un experimento aleatorio es denotado por  $\Omega$
- El resultado de un posible experimento aleatorio o un conjunto del resultado es representado por  $X$
- El espacio de parámetros es denotado por  $\tilde{\Theta}$ , tal que  $\Omega = X \times \tilde{\Theta}$
- Si  $A$  y  $B$  son dos conjuntos, luego  $A \cup B$  representa la unión, además que si  $A_1, A_2, \dots, A_n$  son una colección finita de conjuntos, entonces  $\bigcup_{i=1}^n A_i$  representa la unión de conjuntos.
- Si  $A$  y  $B$  son dos conjuntos, luego  $A \cap B$  representa la intersección, además que si  $A_1, A_2, \dots, A_n$  son una colección finita de conjuntos, entonces  $\bigcap_{i=1}^n A_i$  representa la intersección de conjuntos.
- $A \subset B$  denota que  $A$  es un subconjunto de  $B$ .  $A \subseteq B$  denota que  $A$  es un subconjunto de  $B$ , posiblemente igual a  $B$
- El conjunto vacío puede ser denotado por  $\emptyset$
- $A^c$  denota el complemento de  $A$

**Definición 2.1.2. Probabilidad** (Koski y Noble, 2011) La distribución de probabilidad  $P$  es una función que satisface los axiomas de Kolmogorov

1.  $P(\emptyset) = 0$  y  $P(\Omega) = 1$

2. Si  $A_1, A_2, \dots, A_n$  es una colección finita tal que para cada  $A_i$  como eventos que satisface  $A_j \cap A_k = \emptyset$  para todo  $j \neq k$ , entonces

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

3.  $0 \leq P(A) \leq 1$  para todo  $A$  que pertenece a  $\Omega$

**Definición 2.1.3. Distribución de probabilidad sobre  $X$**  (Koski y Noble, 2011) Si  $X$  contiene un número finito de elementos  $x$ , una distribución de probabilidad sobre  $X$  satisface

1.  $P(x) \geq 0$  para todo  $x \in X$

2.  $\sum_{x \in X} P(x) = 1$

**Definición 2.1.4. Regla de Bayes** (Koski y Noble, 2011) La regla de Bayes para dos eventos  $A$  y  $C$  es dada por

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

### Análisis de datos bayesiano

La característica esencial de los métodos bayesianos es uso de la probabilidad para cuantificar la incertidumbre en inferencias basadas en el análisis de datos estadísticos. En los que Gelman et al. (1995) los divide en estos tres pasos:

1. Configurar un modelo de probabilidad completo en el cual se debe de tener una distribución de probabilidad conjunta para todas cantidades observadas y no observadas en un problema, en el cual el modelo debe ser coherente con el conocimiento científico subyacente y el proceso de recopilación de datos.
2. El condicionamiento sobre datos observados en el que se debe calcular e interpretar la distribución a posteriori apropiada en el que la distribución de probabilidad no observada como último interés que va a estar dado por los datos observados.
3. Evaluación del ajuste del modelo y las implicaciones de las distribuciones a posteriori resultante, planteando las siguientes preguntas: ¿Qué tan bien se ajusta el modelo a los

datos? ¿Son razonables las conclusiones sustanciales? ¿Qué tan sensibles son los resultados a los supuestos del modelo del primer paso?. En respuesta se puede modificar o expandir el modelo y repetir los tres pasos.

## Inferencia Bayesiana

Las conclusiones estadísticas bayesianas acerca de un parámetro desconocido  $\theta$  de datos no observados  $\tilde{y}$  son explicados en términos de probabilidad que son condicionadas de los valores observados de  $y$ , en la que su notación puede ser escrito como  $P(\theta|x)$ , también se condiciona los valores conocidos de cualquier covariable  $x$  Gelman et al. (1995)

## Modelo de verosimilitud

(Lawson, 2018) La verosimilitud dado los datos  $y_i$  donde  $i = 1, \dots, m$  es definido como

$$L(\mathbf{y}|\theta) = \prod_{i=1}^m f(y_i|\theta)$$

Donde  $\theta$  es un vector de longitud  $p$ :  $\{\theta_1, \theta_2, \dots, \theta_p\}$  y  $f(\cdot|\cdot)$  es una función de densidad (o masa), es posible calcular el producto de las contribuciones individuales ya que los valores de la muestra  $\mathbf{y}$  y dado los parámetros son independientes. El logaritmo de verosimilitud es importante en el desarrollo de modelos y es definido como:

$$l(\mathbf{y}|\theta) = \sum_{i=1}^m \log f(y_i|\theta)$$

## Distribución a priori

Todos los parámetros de los modelos bayesianos son estocásticos y son asignados apropiadamente por las distribuciones de probabilidad, la distribución a priori se asigna al parámetro antes de ver los datos, además que las distribuciones a priori es que proporcionan datos adicionales a un problema y que pueden usarse para mejorar la estimación o la identificación de los parámetros (Lawson, 2018).

Para un vector de parámetros  $\theta$ , la distribución a priori puede ser denotada como  $\mathbf{g}(\theta)$

## Distribución a posteriori

Las distribuciones a priori y la verosimilitud proporcionan dos fuentes de información acerca de cualquier problema. La verosimilitud informa sobre el parámetro con los datos, mientras que la distribución a priori informa sobre las creencias o informaciones previas (Lawson, 2018). Cuando hay grandes cantidades de datos, es decir un tamaño de muestra grande, la probabilidad contribuirá más en la estimación del riesgo relativo, cuando se tienen pocos datos, la distribución a priori dominará el análisis.

El producto de la verosimilitud y la distribución a priori es llamado como la distribución a posteriori, en el cual esta distribución describe los parámetros dado los datos observados y poder realizar suposiciones previas, esta es definida como:

$$P(\theta|\mathbf{y}) = L(\mathbf{y}|\theta)\mathbf{g}(\theta)/C$$

Donde  $C = \int_p L(\mathbf{y}|\theta)\mathbf{g}(\theta)d\theta$  con  $\mathbf{g}(\theta)$  como la distribución conjunta de los vectores  $\theta$ . Alternativamente este vector puede ser especificada de manera proporcional:  $P(\theta|\mathbf{y}) \propto L(\mathbf{y}|\theta)\mathbf{g}(\theta)$

## Conjugación

Ciertas combinaciones de distribuciones a priori y verosimilitudes conducen a una distribución a posteriori que la distribución a priori. Esto genera ventajas en la inferencia, ya que la forma a posteriori se desprenderá de la especificación previa (Lawson, 2018). La Tabla 1 muestra unos ejemplos de conjugación.

<b>Verosimilitud</b>	<b>Priori</b>	<b>Posteriori</b>
$\mathbf{y} \sim \text{Poisson}(\theta)$	$\theta \sim G(\alpha, \beta)$	$\theta \mathbf{y} \sim G(\sum y_i + \alpha, m + \beta)$
$\mathbf{y} \sim \text{binomial}(\mathbf{p}, 1)$	$\mathbf{p} \sim \text{Beta}(\alpha_1, \alpha_2)$	$\mathbf{p} \mathbf{y} \sim \text{Beta}(\sum y_i + \alpha_1, m - \sum y_i + \alpha_2)$
$\mathbf{y} \sim N(\mu, \tau), \tau \text{ fijo}$	$\mu \sim N(\alpha_0, \tau_0)$	$\mu \mathbf{y} \sim N\left(\frac{\tau_0 \sum y_i + \alpha_0 \tau}{m\tau_0 + \tau}, \frac{\tau_0 \tau}{m\tau_0 + \tau}\right)$
$\mathbf{y} \sim \text{gamma}(1, \beta)$	$\beta \sim G(\alpha_0, \beta_0)$	$\beta \mathbf{y} \sim G(1 + \alpha_0, \beta_0 + \sum y_i)$

Tabla 1  
Ejemplos de distribuciones conjugadas

Por lo que la elección de la distribución a priori de los parámetros es muy importante ya que puede afectar a la distribución a posteriori significativamente, asimismo el equilibrio entre

la evidencia a priori y posteriori es relacionado a la verosimilitud y el tamaño de muestra.

## **Distribuciones predictivas**

Las distribuciones a posteriori resume la comprensión de los parámetros dado los datos observados y toma un rol fundamental en el modelado bayesiano. Sin embargo también se puede examinar otras distribuciones relacionadas que pueden ser útiles cuando se requieren la predicción de nuevos datos (Lawson, 2018), si definimos una nueva observación de  $y$  como  $y^*$ . Podemos determinar la distribución predictiva de  $y^*$  de dos formas. En general la distribución predictiva es definida como

$$P(y^*|\mathbf{y}) = \int L(y^*|\theta)P(\theta|\mathbf{y})d\theta$$

Aquí la predicción es basada en la marginal sobre los parámetros en la verosimilitud de los nuevos datos  $L(y^*|\theta)$  usando la distribución a posteriori  $P(\theta|\mathbf{y})$  para definir la contribución de los datos observados en la predicción. Definida como distribución predictiva posteriori. Una variante de esta definición usan la distribución a priori en lugar de la distribución a posteriori.

$$P(y^*|\mathbf{y}) = \int L(y^*|\theta)P(\theta)d\theta$$

Esto enfatiza la predicción basada solo en la distribución a priori (antes de ver algún dato), la cual simplemente es la distribución marginal de  $y^*$

## **Modelos Jerárquicos Bayesianos**

En el modelado bayesiano, los parámetros libres tienen distribuciones. Estas distribuciones controlan la forma de los parámetros y son especificadas por el investigador basado generalmente en creencias previas sobre su comportamiento (Lawson, 2018).

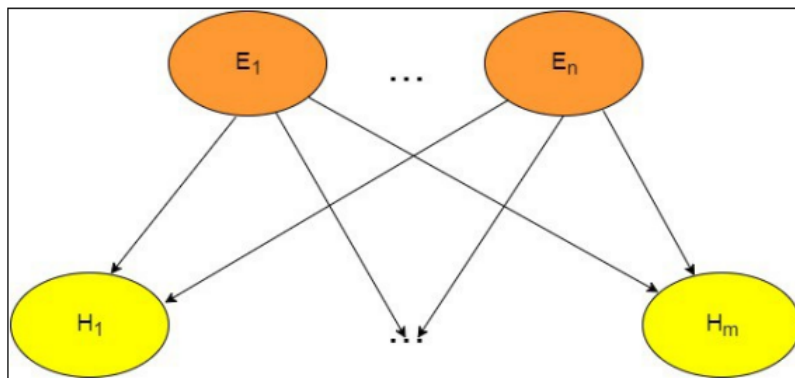
La idea de que los valores de los parámetros podrían surgir de distribuciones es una característica fundamental de la metodología bayesiana y conduce al uso de modelos donde los parámetros surgen dentro de jerarquías



## Redes bayesianas

Las redes bayesianas constituyen el refinamiento de un método de diagnóstico probabilístico conocido como el método probabilístico clásico (Koski y Noble, 2011), tomemos en cuenta para el diagnóstico de enfermedades los siguientes conceptos:

Si se desea diagnosticar  $n$  enfermedades  $E_1, E_2, E_3, \dots, E_n$  a partir de un conjunto de  $m$  hallazgos  $H_1, H_2, H_3, \dots, H_m$  mostrado en el grafo se tiene una representación en la Figura 1



*Figura 1*  
*Grafo de una red bayesiana entre enfermedades y hallazgos*

Entonces el objetivo sería calcular las probabilidades del tipo:

$$P(E_1, E_2, \dots, E_n | H_1, H_2, \dots, H_m)$$

Estas probabilidades refieren a las probabilidades de cada enfermedad si están presentes o ausentes cuando se conoce una serie de observaciones asociadas a cada uno de los distintos hallazgos, entonces es necesario conocer:

- Probabilidad a priori de las distintas enfermedades
- Probabilidad de que se obtengan determinadas observaciones asociadas a los distintos hallazgos conocida como la presencia o ausencia de las distintas enfermedades, es decir la verosimilitud.

En la cual integrando todas las probabilidades por la regla de Bayes se tiene:

$$P(E_1, E_2, \dots, E_n | H_1, H_2, \dots, H_m) = \frac{P(H_1, H_2, \dots, H_m | E_1, E_2, \dots, E_n) P(E_1, E_2, \dots, E_n)}{\sum_{E_1^c, \dots, E_n^c} P(H_1, H_2, \dots, H_m | E_1^c, E_2^c, \dots, E_n^c) P(E_1^c, E_2^c, \dots, E_n^c)}$$

## 2.2 Marco conceptual

- **Queratocono:** Transtorno corneal ectásico, caracterizado por adelgazamiento progresivo del estroma y debilitamiento estructural con protrusión apical en forma de cono, llevando a defectos refractivos como miopía progresiva, astigmatismo irregular, cicatrización y disminución de la visión (Caruso et al., 2024)
- **Modelos bayesianos:** Modelos basados en la regla de Bayes en las cuales se estima la probabilidad a posteriori basada en información previa y verosimilitud de los datos (Koski y Noble, 2011)
- **Redes bayesianas:** Representación gráfica de dependencias para razonamiento probabilístico, en la cual los nodos son las variables aleatorias y representa dependencia directa entre las variables (Sucar y Tonantzintla, 2006)

## 2.3 Antecedentes

### Antecedentes Internacionales

El estudio de Waddell (2023) aplicó las redes bayesianas para el apoyo en la toma de decisiones, tomando datos de diferentes fuentes para explorar fenotípicos específicos y resultados clínicos, en la cual se pueden tener deficiencias debido al conjunto de datos pequeños y cohortes desequilibrados llegando a realizar inferencias falsas probabilísticas. Algo importante de destacar es como el uso de redes bayesianas ayuda en la detección de enfermedades especialmente en incertidumbre compleja, asimismo el autor indicar de que a pesar que la medicina actualmente se encuentra en competencia con el desarrollo de la inteligencia artificial, esta a su vez debe ser poder interpretada por los médicos en el cual el verdadero potencial de las redes bayesianas

viene a ser la capacidad de hablar el lenguaje clínico, modelar preguntas relacionales causales y responder preguntas al explorar fenotipos individuales llegando a plantear preguntas cruciales como ¿por qué? y ¿que pasaría sí?.

El trabajo de Férrez Rubio (2017) presenta una mejor evidencia de la medicina basada en la evidencia en el cual se debe tener un uso consciente, explícito y juicioso de la evidencia científica disponible para tomar la decisión de los pacientes, en el cual comprender la realidad y expresarla de manera sistemática, inteligible y sintética. Integrar la experiencia de los profesionales con la evidencia científica disponible para mejorar la toma de decisiones clínicas. Para conseguir este propósito se tienen diferentes entes matemáticas en la que su estudio enfatizó a las redes bayesianas como una de las herramientas más valiosas en el proceso de toma de decisiones. El trabajo planteo conceptos probabilísticos enfocado al análisis bayesiano en los cuales se fundamentan las redes bayesianas, para posteriormente crear dos redes a ejemplos prácticos, en la cual la red bayesiana tiene una capacidad para representar la relación causa entre variables involucradas en un problema y visualizarlas de una manera muy intuitiva, el cual es el modelo gráfico probabilístico más utilizado en el marco de la toma de decisión diagnóstica.

### **Antecedentes Nacionales**

La investigación de Lopez Esquivel (2023) para optar el título de maestro en Estadística de la Pontífice Universidad Católica del Perú, se basó acerca de las infecciones respiratorias, en la cual buscó establecer una relación entre la incidencia de infecciones respiratorias agudas (IRA) y la incidencia de neumonía en el Perú en la cual además de ver si estas variables están correlacionadas quiso dar un enfoque espacial a nivel provincial en la cual se esperaría que la incidencia de (IRA) y neumonía sea mayor en provincias vecinas. Por lo cual se estudió la distribución espacial entre la incidencia de (IRA) y neumonía a nivel provincial en el Perú a través de un modelo espacial multivariado con efectos aleatorios condicionales autoregresivos utilizando inferencia bayesiana en el modelo jerárquico espacial multivariado a través del método de integración anidada de Laplace (INLA). El cual mostró una bondad de ajuste del modelo y su eficacia en la simulación realizada teniendo una optimización de tiempo menor, asimismo

se vio una correlación positiva directa entre las personas con (IRA) sin neumonía y neumonía. Además de que se tuvo evidencia estadística entre las precipitaciones y la neumonía, indicando que el aumento de precipitaciones en una provincia reduce significativamente la incidencia de IRA sin neumonía, mientras el aumento de precipitación en una provincia aumenta la incidencia de neumonías en dicha provincia, teniendo una autocorrelación espacial moderada con respecto a las tasas de enfermedad entre provincias vecinas, recomendado que se puede trabajar con más covariables y factores que expliquen mejor la causalidad del modelo, de la misma forma incluir un modelo espacio-temporal empleando INLA que tiene una mejor eficiencia computacional al trabajar con más datos.

## **CAPÍTULO III**

### **HIPÓTESIS Y VARIABLES**

#### **3.1 Hipótesis**

#### **3.2 Hipótesis general**

La inferencia bayesiana indica una causalidad más aproximada para la detección de queratocono.

#### **3.3 Hipótesis específicas**

1. Las variables más influyentes son las evaluadas en estudios previos de factores asociados al queratocono.
2. Las redes bayesianas representarán la relación causalidad del queratocono entre variables.
3. Se evaluará la importancia de usar la inferencia bayesiana frente a la inferencia estadística clásica para determinar si una persona tiene o no tiene queratocono.
4. El prototipo computacional apoyará en las inferencias de riesgo a nivel individual.

#### **3.4 Identificación de variables e indicadores**

Las variables independientes viene a ser las probabilidades de generar queratocono en el nivel que se encuentre, asimismo como la clasificación de presencia o ausencia de esta enfermedad, que va a estar dependiendo principalmente de variables clínicas, demográficas y topográficas.

##### **1. Variables dependientes**

- Probabilidad de generar queratocono.
- Clasificación de presencia o ausencia de queratocono.

## **2. Variables independientes**

- Variables clínicas
- Variables demográficas
- Variables topográficas

### 3.4.1 Operacionalización de variables

Tabla 2  
 Matriz de operacionalización de variables

Variable	Definición	Indicador	Tipo	Escala
<b>Dependiente</b>				
Probabilidad de generar queratocono	Posibilidad de generar queratocono	Probabilidad queratocono	Continua	Razón
Presencia o ausencia de queratocono	Clasificación de la presencia o ausencia del queratocono	Presencia o ausencia	Dicotómica	Nominal
<b>Independiente</b>				
Variables clínicas	Variables tomadas en la revisión de la historia clínica usando test o cuestionarios.	Ojo seco	Dicotómica	Nominal
		Alergias	Dicotómica	Nominal
		Frotamiento de ojos	Dicotómica	Nominal
		Antecedentes familiares con queratocono	Dicotómica	Nominal
Variables demográficas	Información general de los pacientes	Género	Dicotómica	Nominal
		Edad	Continua	Razón

Variable	Definición	Indicador	Tipo	Escala
<b>Independiente</b>				
Variables topográficas	Son indicadores claves que se obtienen en los equipos de oftalmología de forma general	Altitud vive	Continua	Razón
		Clasificación altitud	Politómica	Ordinal
		Esfera topográfica	Continua	Intervalar
		Cilindro topográfico	Continua	Razón
		Eje topográfico	Continua	Razón
		Esfera manifiesta	Continua	Intervalar
		Cilindro manifiesto	Continua	Razón
		Eje manifiesto	Continua	Razón
		Emetropía	Politómica	Nominal
		Queratometría media	Continua	Razón
		Paquimetría	Continua	Razón
		Queratometría media	Continua	Razón
		BAD-D	Continua	Razón
		RMS	Continua	Razón
		IPP media	Continua	Razón



Variable	Definición	Indicador	Tipo	Escala
<b>Independiente</b>				
		Elevación frontal	Continua	Razón
		Elevación posterior	Continua	Razón

## **CAPÍTULO IV**

### **METODOLOGÍA**

#### **4.1 Ámbito de estudio: localización política y geográfica**

El estudio se realizará en el centro de salud integral La Fuente en la ciudad del Cusco, ubicado en el distrito de San Jerónimo, provincia de Cusco, departamento de Cusco en el Perú.

#### **4.2 Tipo y nivel de investigación**

La investigación será de tipo aplicada no experimental retrospectivo y prospectivo ya que tomará en cuenta la información de los pacientes atendidos en el centro de salud en el año 2025 para evaluar los factores influyentes aplicando el análisis bayesiano de tal forma que se aumente el conocimiento científico acerca de la metodología estadística bayesiana y su aplicación, como la evaluación para futuros pacientes (Hernández-Sampieri y Mendoza, 2020). Asimismo el trabajo de investigación tiene un nivel relacional causal ya que se estudiará la relación y la causalidad que puedan tener las diferentes variables con respecto a la presencia del queratocono.

#### **4.3 Unidad de análisis**

La unidad de análisis serán los pacientes entre 10 a 40 años atendidos en el centro de salud integral La Fuente en el año 2025.

#### **4.4 Población de estudio**

La población de estudio serán todos los pacientes atendidos en el centro de salud integral la fuente que cumplan con los criterios de inclusión.

## **Criterios de inclusión**

1. Edad entre 10 a 40 años.
2. Tengan información clínica de interés completa.
3. Tengan información demográfica de interés completa.
4. Tengan los exámenes topográficos de interés completos.

## **4.5 Tamaño de muestra**

Para la investigación se obtuvo una cantidad total de 471 productos registrados en almacén del centro de salud integral. De la cual el único criterio de inclusión a tomar en cuenta para seleccionar los productos que se analizarán mediante los modelos de inventarios, serán aquellos que se encuentren en el grupo A de la clasificación de actividades basadas en costos (ABC), la justificación y uso de esta clasificación se detalla en el marco teórico del apartado

## **4.6 Técnicas de selección de muestra**

La técnica empleada es la revisión de documentos y/o análisis de datos secundarios del registro que se tienen de los productos de almacén del centro de salud integral.

## **4.7 Técnicas de recolección de información**

Las técnicas de recolección de información será a través de la revisión de documentos y/o análisis de datos secundarios, teniendo como instrumento la historia clínica y resultados del topógrafo de PENTACAM.

## **4.8 Técnicas de análisis e interpretación de la información**

Los datos serán recopilados en un archivo de extensión *.xlsx* (Archivo Excel), en el cual los análisis se realizarán en el lenguaje de programación R bajo el entorno de desarrollo RStudio

debido a su alta capacidad para manejar análisis estadístico desde los descriptivos hasta la realización de modelos robustos.

#### **4.9 Técnicas para demostrar la verdad o falsedad de las hipótesis planteadas**

Las hipótesis de investigación planteadas en el estudio serán comprobadas a través de los resultados a nivel inferencial que se obtenga mediante el modelo bayesiano establecido, asimismo por la inferencia que indiquen las redes bayesianas. Utilizando los criterios como es el criterio de AKAIKE (AIC), criterio bayesiano (BIC), entre otros.

## Presupuesto

Tabla 3

*Lista de productos y sus costos*

Descripción	Cantidad	Precio Unitario	Costo total
Laptop	1	S/.2000.00	S/.2000.00
Impresora	1	S/.750.00	S/.750.00
USB	1	S/.20.00	S/.20.00
Papel bond (Paquete)	1	S/.15.00	S/.15.00
Lapiceros	3	S/.1.00	S/.3.00
Documentos	120	S/.0.20	S/.24.00
<b>TOTAL</b>	-	-	<b>S/.2812.00</b>

## Cronograma

Tabla 4

*Plan de actividades*

Actividades	Octubre 2025	Noviembre 2025	Diciembre 2025	Enero 2026	Febrero 2026
Elaboración y aprobación del proyecto de tesis	✓	✓	✓		
Análisis y proceso de la información	✓	✓	✓	✓	
Redacción del informe final			✓	✓	✓
Presentación				✓	✓
Sustentación					✓

## BIBLIOGRAFÍA

- Althomali, T. A., Al-Qurashi, I. M., Al-Thagafi, S. M., Mohammed, A., y Almalki, M. (2018). Prevalence of keratoconus among patients seeking laser vision correction in Taif area of Saudi Arabia. *Saudi Journal of Ophthalmology*, 32(2):114–118.
- Caruso, C., D’Andrea, L., Troisi, M., Rinaldi, M., Piscopo, R., Troisi, S., y Costagliola, C. (2024). Corneal collagen cross-linking in patients with keratoconus from the Dresden protocol to customized solutions: theoretical basis. *International Journal of Ophthalmology*, 17(5):951.
- Clayton, A. (2021). *Bernoulli’s fallacy: Statistical illogic and the crisis of modern science*. Columbia University Press.
- Férez Rubio, J. A. (2017). Redes bayesianas aplicadas a la Medicina.
- Fouarge, E., Monseur, A., Boulanger, B., Annoussamy, M., Seferian, A. M., De Lucia, S., Lilien, C., Thielemans, L., Paradis, K., Cowling, B. S., et al. (2021). Hierarchical Bayesian modelling of disease progression to inform clinical trial design in centronuclear myopathy. *Orphanet Journal of Rare Diseases*, 16:1–11.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., y Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.
- Gelman, A., Carlin, J. B., Stern, H. S., y Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.
- Godefrooij, D. A., De Wit, G. A., Uiterwaal, C. S., Imhof, S. M., y Wisse, R. P. (2017). Age-specific incidence and prevalence of keratoconus: a nationwide registration study. *American journal of ophthalmology*, 175:169–172.
- Hernández-Sampieri, R. y Mendoza, C. (2020). Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta.

- Koski, T. y Noble, J. (2011). *Bayesian networks: an introduction*. John Wiley & Sons.
- Lawson, A. B. (2018). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. Chapman and Hall/CRC.
- Lopez Esquivel, M. A. (2023). Modelamiento bayesiano espacial multivariado para datos de áreas.
- Mansfield Rojas, N. (2017). El queratocono en pacientes de una institución privada de la ciudad de Quito, Ecuador en el período de Enero de 2015 a octubre de 2016. B.S. thesis, Quito: Universidad de las Américas, 2017.
- Pussetto, A. P., Rossi, M. L., Ciani, J. L., y Magurno, M. G. (2011). Alta prevalencia de pacientes con queratocono en la población de la ciudad de Paraná, Entre ríos. *Oftalmol Clin Exp*, 4(4).
- Rabinowitz, Y. S. (1998). Keratoconus. *Survey of ophthalmology*, 42(4):297–319.
- Ramos Almirón, V. A. (2018). Prevalencia y Factores Asociados al Queratocono en Pacientes Atendidos en el Hospital Regional Honorio Delgado, Arequipa, 2014-2017.
- Samaniego Mojica, A. C. (2023). Prevalencia y factores asociados al queratocono en pacientes entre los 10 a 30 años de edad atendidos en el Centro de Salud Integral la Fuente, Cusco-Perú durante el año 2019.
- Spiegelhalter, D. (2023). *El arte de la estadística*. Capitán Swing Libros.
- Sucar, L. E. y Tonantzintla, M. (2006). Redes bayesianas. *Aprendizaje Automático: conceptos básicos y avanzados*, 77:100.
- THALASSELIS, A., TAIE, H. F., ETCHEPAREBORDA, J., y SELIM, A. (1988). Keratoconus, magnesium deficiency, type A behavior, and allergy. *Optometry and Vision Science*, 65(6):499–505.
- Waddell, T. H. (2023). *Applying Bayesian-networks to model multi-organ pathologies*. Tesis de doctorado, University of Oxford.

Zhao, S. H., Berkowitz, C., Ralay Ranaivo, H., Laurenti, K., Bohnsack, B. L., Basti, S., y Rossen, J. L. (2024). Evaluation of parameters for early detection of pediatric keratoconus. *BMC ophthalmology*, 24(1):463.



# **ANEXOS**

## a. Matriz de consistencia

Problemas	Objetivos	Hipótesis	Variables	Metodología
<b>General</b>			<b>Dependientes</b>	<ul style="list-style-type: none"> <li>◦ <b>Ámbito de estudio:</b> La Fuente Centro de Salud Integral - Cusco - Perú</li> <li>◦ <b>Tipo y nivel:</b> aplicada no experimental retrospectiva, relacional-causal</li> <li>◦ <b>Unidad de análisis:</b> Pacientes atendidos en el centro de salud integral</li> <li>◦ <b>Población:</b> Pacientes atendidos que cumplan los criterios de inclusión</li> <li>◦ <b>Muestreo:</b> Aleatorio simple</li> <li>◦ <b>Instrumento recolección de datos:</b> Historia clínica y resultados topográfico de PENTACAM</li> <li>◦ <b>Análisis e interpretación de datos:</b> Archivo de extensión .xlsx y RStudio.</li> </ul>
¿Cómo comparar la inferencia estadística clásica con la inferencia bayesiana?	Comparar la inferencia estadística clásica con la inferencia bayesiana para la detección del queratocono	La inferencia bayesiana indica una causalidad más aproximada para la detección de queratocono.	<ul style="list-style-type: none"> <li>◦ Probabilidad de presencia de queratocono</li> <li>◦ Clasificación de presencia o ausencia de queratocono</li> </ul>	
<b>Específicos</b>			<b>Independientes</b>	
<ul style="list-style-type: none"> <li>◦ ¿Qué variables clínicas, demográficas y topográficas se asocian en la detección del queratocono?</li> <li>◦ ¿Cómo se obtiene la inferencia de probabilidad en la presencia o ausencia de queratocono a través de las diversas variables?</li> <li>◦ ¿Cómo afecta la decisión de usar la inferencia estadística clásica o bayesiana en la detección de queratocono?</li> <li>◦ ¿Cómo implementar las inferencias de riesgo a nivel individual?</li> </ul>	<ul style="list-style-type: none"> <li>◦ Determinar las variables clínicas, demográficas y topográficas asociadas que apoyen en la detección del queratocono.</li> <li>◦ Construir una red bayesiana estructurada que infiera la probabilidad de presencia o ausencia del queratocono a partir de las variables observadas y como estas se actualizan al variar sus parámetros.</li> <li>◦ Evaluar la detección de queratocono para futuros pacientes utilizando la inferencia estadística clásica y la inferencia bayesiana.</li> <li>◦ Implementar un prototipo computacional que visualice las inferencias de riesgo de queratocono a nivel individual.</li> </ul>	<ul style="list-style-type: none"> <li>◦ Las variables más influyentes son las evaluadas en estudios previos de factores asociados al queratocono.</li> <li>◦ Las redes bayesianas representarán la relación causalidad del queratocono entre variables.</li> <li>◦ Se evaluará la importancia de usar la inferencia bayesiana frente a la inferencia estadística clásica para determinar si una persona tiene o no tiene queratocono.</li> <li>◦ El prototipo computacional apoyará en las inferencias de riesgo a nivel individual.</li> </ul>	<ul style="list-style-type: none"> <li>◦ Variables demográficas.</li> <li>◦ Variables clínicas.</li> <li>◦ Variables topográficas.</li> </ul>	