

Análisis Exploratorio de Datos. Un enfoque hacia la Miopía.

Jackson Vera Pineda

Grupo C312

JACKSONVERAPINEDA@GMAIL.COM

Kevin Manzano Rodríguez

Grupo C312

KEVINMANZANO020610@GMAIL.COM

Roger Fuentes Rodríguez

Grupo C312

FUENTESROGER65@GMAIL.COM

Tutor(es):

Dr. José E. Valdez Castro, *Universidad de la Habana.*

MsC. Yanetsy E. Rodríguez, *Universidad de la Habana.*

Palabras Clave: Estadística, Análisis Exploratorio de Datos, Oftalmología, Miopía.

Tema: Estadística, Análisis Exploratorio de Datos.

1. Introducción

La miopía es una enfermedad de la refracción del ojo en el cual los rayos de luz paralelos convergen en un punto focal situado delante de la retina, en lugar de converger en la misma; es lo contrario de la hipermetropía, en la que los rayos de luz llegan a la retina antes de converger [1]. Afectando (la miopía) mayormente la visión lejana.

Entre los síntomas de esta condición pueden presentarse: visión borrosa en objetos que están lejos; necesidad de estremecer los ojos para ver con claridad; dolores de cabeza, y fatiga ocular. En niños, estos indicadores pueden afectar el comportamiento provocando, quizás, estremecimiento de los ojos constantemente, dificultad para notar objetos alejados, parpadeo y frotamiento frecuente de los ojos, necesidad de posicionarse cerca de la televisión u otros dispositivos.

Existen factores que frecuentemente se asocian con un alto riesgo de aparición de la enfermedad como son la herencia genética, realizar actividades prolongadas que requieran visión de cerca, más aún si estas actividades son frente a pantallas, y poca exposición a exteriores (luz natural y recreación al aire libre).

El presente artículo propone realizar un Análisis Exploratorio a un conjunto de datos del Estudio Longitudinal de Miopía de Orinda (OLSM por sus siglas en inglés) en California, Estados Unidos, comenzado en 1989 hasta 1999 [3], para desarrollar un perfil predictivo de la miopía a edades tempranas. Se evaluaron 618 niños de 5 a 9 años, inscritos en el estudio.

2. Desarrollo

2.1 Sobre los datos

Los datos fueron obtenidos de Kaggle [2]. Las mediciones se realizaron al ojo derecho de cada individuo

por tanto, cada observación está asociada a un sujeto independiente. Como denota del nombre OLSM, el estudio fue de forma longitudinal, es decir, se realizó un seguimiento a cada sujeto del estudio por varios años ¹, y en el conjunto actual se presenta la última revisión, como se especifica en el artículo por Zadnik K. [3]

2.2 Variables

A continuación se presentan las variables del conjunto de datos más relevantes para el presente análisis, y, en la Tabla 1, se expone un resumen descriptivo de cada una en la muestra :

MYOPIC : Variable booleana 0 si el sujeto desarrollara miopía en los siguientes 5 años o no

AGE : La Edad del sujeto (de 5 a 9 años)

GENDER : Género del sujeto 1 Femenino 0 Masculino,

SPHEQ : Equivalente Esférico de la Refracción (Spherical Equivalent Refraction) es una estimación del error refractivo de los ojos (Medición en D, dioptría).

AL : La Longitud Axial se refiere a la distancia entre la parte posterior y la parte delantera del ojo (Medición en milímetros mm).

VCD : La Profundidad de la Cámara Vítrea (Vitreous Chamber Depth, en inglés) en oftalmología se refiere a la distancia entre la parte posterior del cristalino y la retina.

MOMMY : Se refiere a si la madre es miope 0 negativo, 1 positivo.

1. La alternativa más común a estudio transversal, lo que significa que captura una instantánea de un grupo en un momento específico. Un estudio longitudinal observa a un grupo repetidamente a lo largo de un período de tiempo. [6]

Variables	Media \pm SD	Mínimo	Máximo
Edad	6.30 \pm 0.71	5	9
Equivalente Esférico de Refracción (D)	0.80 \pm 0.63	-0.70	4.37
Longitud Axial (mm)	22.50 \pm 0.68	19.90	24.56
Profundidad de la Cámara Anterior (mm)	3.58 \pm 0.23	2.77	4.25
Profundidad de la Cámara Vítrea (mm)	15.38 \pm 0.66	13.38	17.30

Figura 1: Resumen descriptivo de las variables numéricas.

DADMY : Se refiere a si el padre es miope 0 negativo, 1 positivo.

2.3 Visualización de Datos

Como puede observarse en la Figura 2a, la cantidad de pacientes de género masculino y femenino está equilibrada (variable GENDER). Sin embargo no puede afirmarse lo mismo del número de pacientes miopes Figura 2b, que es más de 6 veces menor que el número de pacientes sin la afección, este sesgo puede estar dado porque el comienzo de dicho problema con frecuencia se ubica entre los 6 y los 14 años, afecta aproximadamente al 5 % de los niños en edad preescolar, al 9 % de los niños en edad escolar [4].

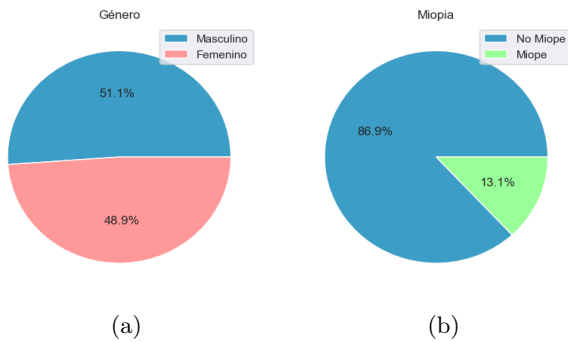


Figura 2: Gráfico de Pastel, Género (a), y Miopía (b).

En la Figura 3 se expone histogramas de las variables Equivalente Esférico de Refracción (SPHEQ) 3a, Profundidad de la Cámara Vítrea (VCD) 3b y la Longitud Axial (AL) 3c, según el género del paciente. Nótese que no se muestran diferencias importantes en las distribuciones y es que los infantes suelen tener características similares cuando son muy pequeños, pero después de los 9 años comienzan a diferenciarse mas por ejemplo: los niños tienden a mostrar longitudes axiales más largas que las niñas; los ojos asiáticos tienden a ser más largos que los ojos caucásicos.

El uso de datos de la AL es valioso tanto para determinar el riesgo para la salud ocular, como para juzgar la eficacia de un tratamiento para el control de la miopía. Varios estudios han demostrado la correlación entre la biometría ocular, especialmente la AL, con los errores refractivos: se encontró que el SPHEQ y la ACD afectaban los parámetros de la AL. Estas relaciones pueden distinguirse en la Figura 4, que representa un gráfico por pares de las variables AL, SPHEQ y VCD,

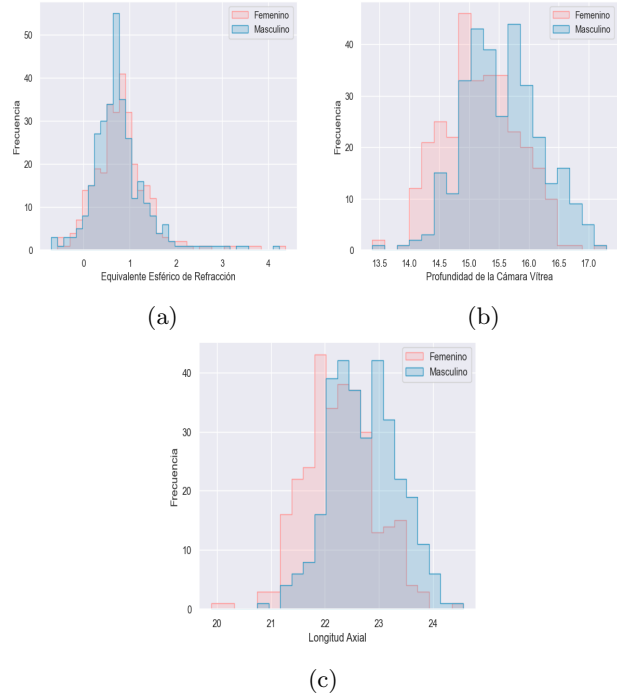


Figura 3: Histograma, SPHEQ (a), VCD (b) y AL (c) según el género del paciente.

con histograma de estas variables en la diagonal.

Cabe destacar en la Figura 4 cómo el SPHEQ a priori parece ser un elemento clave a tomar en cuenta en la predicción de la miopía, dado el posicionamiento de los pacientes convalecientes en los valores donde el SPHEQ es bajo. Además, es notoria la fuerte relación que existe entre la AL y VCD, lo cual es intuitivo dado que la Cámara Vitrea constituye gran parte del globo ocular.

2.4 Pruebas de Hipótesis

En la Figura 5 puede notarse que la Longitud Axial en los datos parece estar normalmente distribuida, de ahí que resulta interesante el análisis de esta variable para comprobar su normalidad, y conocer qué posibles propiedades cumple.

Se procede a realizar un gráfico Cuantil-Cuantil como se muestra en la Figura 6. Efectivamente, los cuantiles teóricos de una población normalmente distribuida coinciden con los cuantiles muestrales. Luego, se estiman la media y desviación estándar de la población ($\mu \approx 22.50$ y $\sigma \approx 0.69$), y se realiza una prueba de bondad de ajuste, Kolmogorov- Smirnov,

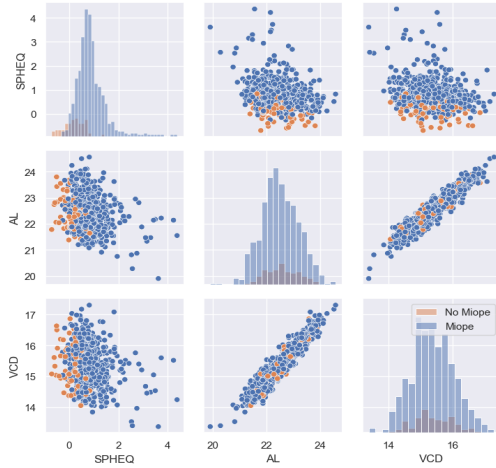


Figura 4: Gráfico por pares de las variables AL, VCD, SPHEQ, y la presencia de miopía (puntos en naranja).

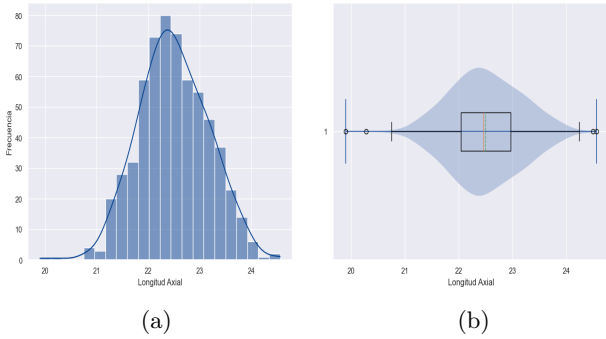


Figura 5: La distribución de la Longitud Axial, representada en un histograma (a), y gráfico de cajas, bigotes, y violín (b).

entre la muestra y una población normal teórica construida con los parámetros estimados previamente; arrojando $p\text{-value} = 0.77$, donde α es el nivel de significancia. Por tanto, no se tiene evidencia suficiente para rechazar la normalidad de la Longitud Axial.

En 2021 He X., Sankaridurg P. y un colectivo de científicos analizaron la correlación entre la AL en diferentes grupos etarios y la aparición de miopía en un estudio transversal en China. Dicho estudio incluyó un total de 14127 participantes (en su mayoría chinos), y de ellos 5742 niños entre 7 y 10 años [5]. Entre sus resultados determinaron la media de la AL, con un valor aproximado de $22.9 (\pm 0.8)$, desviación estándar para un rango de edades similar al de la muestra bajo el presente análisis.

De la estimación anteriormente calculada de la media muestral ($\bar{X} = 22.5$). Luego, se comprueba mediante una prueba de hipótesis t-student si H_0 : la media de la población de la que fue extraída la muestra coincide con la analizada por He X., H_1 : la media de la población es menor. El resultado arrojó un $p\text{-value} = 1.22e-42$, por tanto se rechaza la hipótesis nula; se concluye que la población de nuestra muestra tiene una media significativamente menor. Del resulta-

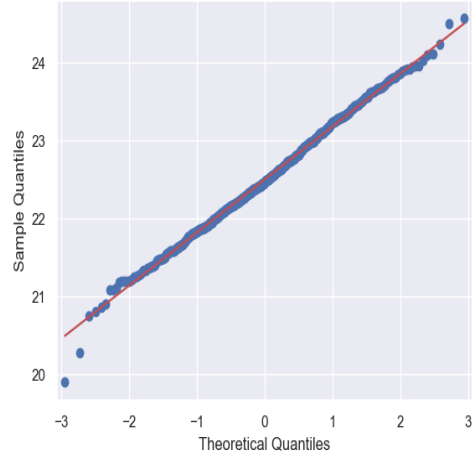


Figura 6: Gráfico Cuantil-Cuantil de la Longitud Axial, donde los cuantiles teóricos se refieren a una población normalmente distribuida.

do anterior puede comprobarse que efectivamente los ojos de niños caucásicos tienen una AL significativamente menor a los de asiáticos (al menos provenientes de China), aunque se debe tener en cuenta la brecha temporal entre ambos estudios.

Se comprueba, además, que la nuestros datos fueron extraídos de una población (normal, comprobado previamente) con media $\mu \in [22.44, 22.55]$, asumiendo que se desconoce la varianza población (por tanto calculando el intervalo de confianza con el estadístico t); y varianza $\sigma^2 \in [0.52, 0.41]$ (calculado con χ^2), con un 95 % de confianza.

2.5 Análisis de Varianza

La herencia genética así como factores anatómicos como la Longitud Axial, como ya se ha visto, juega un papel importante en la determinación de la Miopía en los pacientes, por tanto se propone analizar como afecta la herencia en esta característica particular en la descendencia. Se proponen dos enfoques al análisis de varianza ANOVA: primero para determinar si la herencia es significativamente determinante en la proporción de la Logitud Axial; y, más específicamente, si la herencia por parte de la madre o el padre (incluso ambos o ninguno), es determinante, en la característica en cuestión.

2.5.1 PRIMER ACERCAMIENTO

Se asume la herencia no sexual, es decir, se analizarán si las **características** de la AL están influenciadas por la miopía en la madre y el padre, como factores uniformemente determinantes. Esto particiona la muestra en tres grupos de interés, que son obtenidos al filtrar por la información en las variables MOMMY y DADMY:

1. Pacientes sin progenitores miopes.
2. Pacientes con un solo progenitor miope.

3. Pacientes con dos progenitores miopes.

2.5.2 SEGUNDO ACERCAMIENTO

Sin la **asunción** de la herencia no sexual se analiza como **varia** la *AL* teniendo en cuenta si es padre es miope, si la madre lo es, si ambos lo son, o ninguno, como grupos mutuamente excluyentes. De esta forma los grupos de interés son:

1. Pacientes sin progenitores miopes.
2. Pacientes con padre miope solamente.
3. Pacientes con madre miope solamente.
4. Pacientes con ambos progenitores miopes.

Para los dos enfoques, los resultados fueron similares *p-values* de 0.80 y 0.70 respectivamente, por tanto no existe evidencia suficiente en los datos para afirmar que la herencia es significativamente influyente en la Longitud Axial de la descendencia. Nótese en la Figura 7 que la distribución de estos grupos es bastante similar. Puede observarse en la Figura 8 una comparación entre la distribución de los grupos previamente analizados: entre las curvas de cada color no existe diferencia notoria en su forma y posición.

Se comprueba además, el cumplimiento los supuestos de ANOVA: normalidad, homocedasticidad (usando el test de Levene, H_0 : las varianzas de los grupos son iguales, que arrojó *p-values* de 0.37 y 0.56 para el primer y segundo acercamiento, respectivamente) e independencia de las muestras.

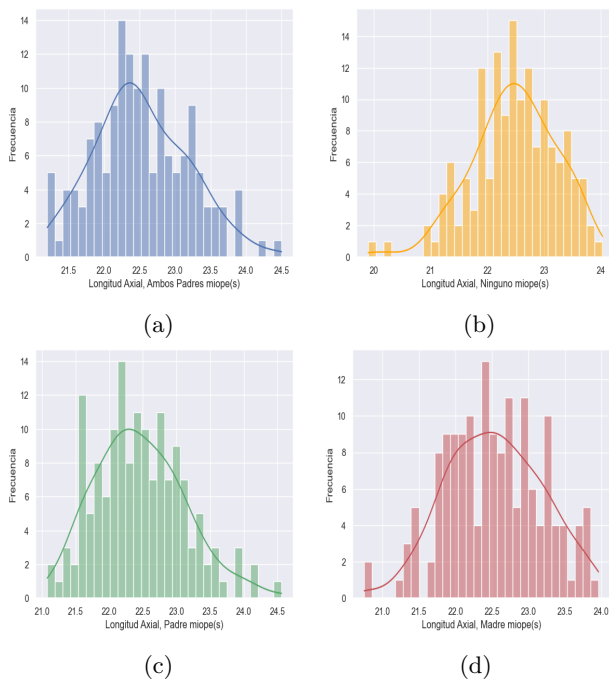


Figura 7: La distribución de la Longitud Axial en cada uno de los grupos de interés (usando histogramas y estimación de densidad).

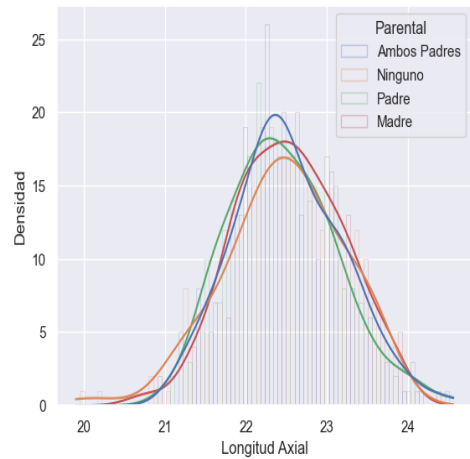


Figura 8: Comparación de la distribución de cada grupo (según la información parental).

3. Conclusiones

Se probó que, aunque el factor genético pueda influenciar la miopía, no existe evidencia suficiente para afirmar que la presencia de la enfermedad en los padres influencia notoriamente en la anatomía de los ojos de la descendencia a tempranas edades. Se comprobó además que la *AL* sigue una distribución normal, y que efectivamente factores étnicos son definitorios en el tamaño del globo ocular, y por tanto puede ser determinante en el riesgo de aparición de la enfermedad. Sentando con todos estos elementos, las bases para un posterior análisis predictivo de la enfermedad a edades tempranas.

4. Recomendaciones

Probadas las capacidades de las herramientas convencionales para el análisis de datos en este campo tan importante de la medicina, se recomienda realizar un estudio de este tipo sobre la Miopía en Cuba, teniendo en cuenta los factores étnicos que componen su sociedad. Debe tenerse en cuenta que la cantidad de pacientes diagnosticados con miopía, a pesar de ser lo suficientemente grande para plantear pruebas de hipótesis, es considerablemente baja teniendo en cuenta los pacientes emétopes; lo cual puede afectar considerablemente la precisión de cualquier modelo predictivo que se formule con los datos, siendo estos propensos a errores de tipo 2, falsos negativos, diagnóstico erróneo en pacientes enfermos.

Referencias

- [1] Sociedad Española de Medicina Interna (SEMI). URL: <https://www.fesemi.org>. Consultado en 28 de junio de 2024.
- [2] Kaggle: Myopic risk comparison. URL: <http://www.kaggle.com>

- [3] Zadnik, K., Mutti, D. O., Friedman, N. E., Qualley, P. A., Jones, L.A., Qui, P., Kim, H. S., Hsu, J. C., & Moeschberger, M. L. (1999). *Ocular predictors of the onset of juvenile myopia. Investigative ophthalmology & visual science*, 40(9), 1936–1943.
- [4] OftalmoLima. (n.d.). Miopía (visión de cerca) en niños y adolescentes. <https://oftalmolima.pe> Consultado en 28 de junio de 2024.
- [5] He X, Sankaridurg P, Naduvilath T, Wang J, Xiong S, Weng R, Du L, Chen J, Zou H, Xu X. *Normative data and percentile curves for axial length and axial length/corneal curvature in Chinese children and adolescents aged 4-18 years.* Br J Ophthalmol. 2023 Feb;107(2):167-175. doi: 10.1136/bjophthalmol-2021-319431. Epub 2021 Sep 16. PMID: 34531198; PMCID: PMC9887397.
- [6] Downey, A. B. (2014). *Think Stats: Exploratory Data Analysis in Python*. O'Reilly Media, Inc.