STATISTICAL MODEL USING DISCRIMINANT ANALYSIS TO CLASSIFY EYE REFRACTIVE PROBLEMS IN THE CITY OF BARRANQUILLA

MODELO ESTADÍSTICO UTILIZANDO ANÁLISIS DISCRIMINANTE PARA CLASIFICAR PROBLEMAS REFRACTIVOS OCULARES EN LA CIUDAD DE BARRANQUILLA

**Martín Díaz Rodríguez[[1]](#footnote-1).**

**José Vicente Barraza Angarita[[2]](#footnote-2)**

**Hernando Hernández Leal.[[3]](#footnote-3)**

ABSTRACT

The present work intends to build a multivariate statistical model to classify patients with ocular refractive problems (hyperopia, myopia, astigmatism, hyperopic astigmatism and myopic astigmatism) in terms of a set of variables that specify a type of eye refractive defect.

This study was developed in the North of Barranquilla, analyzing the medical histories provided by a clinic in Optometry.

In total 499 patients were selected at random to analyze their diagnostic. Their ages were in the range of 8 to 82, with different professions. The database has 15 quantitative variables and 5 qualitative variables.

The Multivariate analyzing techniques than were applied to the study, were the principal components factorial analysis with the aim of creating new variables than summarize all information which it may have in the original variables and the discriminant analysis with the purpose of classifying patients in some of the groups defined by the type of defects refractive manifested in his diagnostic.

The multivariate analysis is a tool that allows you to discover the interdependence

between variables that measure a visual refractive defect (Diopters in the spherical lens,

diopter in the cylindrical lens, axis of the cylinder, visual acuity) through the AF, the

dependence on the type of ocular refractive defect diagnosed by the physician, and this

set of variables or subset generated by AD in order to carry out the classification of

groups of patients according to the type of visual refractive defect manifesting in your

diagnosis and forecasting or allocation of a new patient in one of several groups defined.

**Key words**: Ocular refractive problems, medical histories, diagnostic (Diopters in the spherical lens, diopter in the cylindrical lens, visual acuity), factorial analysis, discriminant analysis.

**RESUMEN**

Este trabajo tiene como propósito construir un modelo estadístico multivariado para clasificar pacientes con problemas refractivos oculares, en términos de un conjunto de variables que especifican un tipo de defecto refractivo ocular.

El estudio se desarrolló en Barranquilla, Colombia, con las historias clínicas de 499 pacientes, de diferentes profesiones, con edades de 8 a 82 años. Los pacientes fueron seleccionados al azar de una base de datos facilitada para este fin por un consultorio optométrico de la ciudad.

La base de datos consta de 15 variables cuantitativas y 5 variables cualitativas.

Las técnicas de análisis multivariados aplicadas al estudio fueron: el Análisis Factorial por Componentes Principales1, con el objetivo de crear nuevas variables que resuman óptimamente toda la información contenida en las variables estudiadas, y luego el análisis discriminante,[[4]](#footnote-4),2,3,4con el propósito de clasificar a los pacientes en algunos de los grupos definidos por el tipo de defecto refractivo ocular5 manifestado en su diagnóstico.

El análisis multivariado es una herramienta que permite descubrir la interdependencia2 que pueda existir entre variables que miden un defecto refractivo visual5(Dioptrías en lentes esféricos, Dioptrías en lentes cilíndricos, Eje del cilindro, Agudeza visual) a través del análisis factorial, la dependencia del tipo de defecto refractivo ocular diagnosticado por el médico y el conjunto de variables o subconjunto generado de ellas mediante análisis discriminante1,2, clasificar los pacientes según el tipo de defecto refractivo visual que manifieste en su diagnóstico, y la predicción6 o asignación2 de un nuevo paciente en alguno de los grupos definidos.

**Palabras clave:** Problemas refractivos oculares, historias clínicas, diagnóstico (Dioptrías en lentes esféricos, Dioptrías en lentes cilíndricos, Eje del cilindro, Agudeza visual), análisis factorial, análisis discriminante. .

1. **INTRODUCCIÓN**

En la Costa Caribe, los oftalmólogos y optómetras recomiendan utilizar anteojos con filtro de protección contra los rayos ultravioletas del sol5. La alta incidencia de personas con predominio del sexo masculino que están padeciendo de Pterigion primario llama mucho la atención5.

Algunas enfermedades visuales son de carácter multifactorial7; por ejemplo, el glaucoma y la hipermetropía elevada. El estrabismo se constituye en un problema psicológico y antiestético y degenera en ambliopía8. Según su forma de aparición puede ser congénito o adquirido8. La catarata relacionada con la edad es una causa común de deterioro visual9; la prolongada experiencia de visión anormal en un niño menor de siete años de edad puede conducir a ambliopía.8,9,10,11,12,13 18,20,17

La frecuencia de aparición de miopía, es de destacar la influencia racial: en los países asiáticos hay un alto porcentaje de esta ametropía10,11. La miopía es muy común en las personas con una excesiva longitud axial10,11

En un estudio comparativo con muestras de 1000 pacientes de dos ciudades colombianas, Bogotá y Barranquilla, realizado por la Academia de Medicina12, se encontró que el 56% del grupo estudiado eran miopes en Bogotá, mientras que en Barranquilla era del 49%.

En este estudio realizado recientemente en el Norte de Barranquilla, se encontró que el 35.07% padecían de hipermetropía y astigmatismo, el 25.85% solo padecía de hipermetropía, el 20.24% de astigmatismo miópico, el 12.02% de miopía y el 6.81% de astigmatismo puro en uno o en ambos ojos.

En el caso de la hipermetropía sola o combinada con astigmatismo, se encontró que el 60.92% de los pacientes la padecen. Este defecto se presenta en la práctica con una frecuencia muy elevada (55%)7.

En las historias clínicas analizadas se encontraron pacientes con antecedentes familiares y antecedentes personales de todo tipo (médico, traumático, quirúrgico y alérgico).

Se encontraron, también, pacientes con baja visión, con dificultad para leer, ver televisión o desplazarse independientemente. Entre las causas de la baja visión están las enfermedades transmitidas por los padres, como la diabetes o la hipertensión arterial.5, 7, 8, 10,11

La presbicia,8,9,10,11,14,15,16,17por ejemplo, es una condición de la visión que consiste en una disminución gradual de la capacidad de enfocar objetos cercanos (30 a 40 cm), debida a una pérdida de la elasticidad del cristalino, que comienza a manifestarse después de los 40 años 8,10,11,12

**2. TÉCNICAS ESTADÍSTICAS MULTIVARIADAS APLICADAS**

Al aplicar pruebas de normalidad multivariada17,18,19 y de igualdad de matrices de varianzas-covarianzas19 sobre los 379 datos restantes, se encontró que se rechaza el supuesto de normalidad en cada uno de los grupos, al igual que el supuesto de igualdad de matrices de varianzas-covarianzas.

Al aplicar análisis factorial por componentes principales (AFCP) sobre todas las variables en el estudio, se observó un índice de esfericidad KMO1,20 mayor que 0.5, indicio de existencia de correlación significativa entre variables.

Identificadas las variables latentes21,22 (con AFCP) 1,17 que especifican el tipo de defecto refractivo ocular, se utilizó análisis discriminante (AD) lineal y cuadrático 19, 23 por el método de inclusión por pasos, para construir los modelos mediante los cuales se puedan clasificar los pacientes en una de las siguientes cinco categorías de la variable diagnóstico: astigmatismo (grupo 1), astigmatismo hipermétrope (grupo 2), astigmatismo miópico (grupo 3), hipermetropía (grupo 4) y miopía (grupo 5). Después de excluidos los datos atípicos 17,18 en la muestra y en cada uno de los grupos, quedaron 379 pacientes, de los cuales 27 pertenecen al grupo 1, 127 al grupo 2, 69 al grupo 3, 108 al grupo 4 y 48 al grupo 5.

Para especificar un defecto refractivo ocular, se utilizan diferentes tipos de variables, en su mayoría cuantitativas1, como lo son: dioptrías en el lente esférico ojo derecho e izquierdo (LESLD y LESLI); adición en las lentes ojo derecho e izquierdo (ADD OD y ADD OI); lente cilíndrico ojo derecho e izquierdo (LCILD, LCILI); eje del cilindro ojo derecho e izquierdo (EJED, EJEI); agudeza visual para visión de lejos ojo derecho e izquierdo (AVL OD, AVL OI), agudeza visual para visión próxima ojo derecho e izquierdo (AVP OD, AVP OI), distancias pupilares para visión de cerca y de lejos (DPC Y DPL), y la variable EDAD.

En la búsqueda de determinar unos pocos factores que retengan la mayor variabilidad contenida en los datos, utilizando el SPSS24 y aplicando el AFCP sobre la matriz de correlaciones, se tiene que con cinco componentes se explica un 88.9% de la variabilidad total1.

Al agrupar las variables en torno a cada componente aplicando el método de rotación Varimax1,21,22, se observa, en la tabla 1, lo siguiente:



Tabla 1. Matriz de componentes rotados

* En la primera componente están bien representadas las variables ADDd, ADDi y la variable Edad. Cuando se precisa hacer una adición dióptrica en una lente, es porque hay presencia de presbicia8,9, esto es bastante común en personas cuando pasan de los 40 años de edad8;la tercera componente será denominada de “corrección présbita”(CP).La presbicia se corrige con el empleo de lentes más para compensar la pérdida de potencia automática del cristalino9. Debe considerarse la corrección del defecto refractivo de base y calcular el valor de la ADD de acuerdo con los requerimientos visuales.8,13,14,15
* En la segunda componente están bien representadas las variables LCILD,LCILI, EJED y EJEI. Este es el tipo de lentes utilizados para corregir astigmatismo, aunque por lo general se usan lentes esferocilíndricas8, a diferencia de los defectos refractivos esféricos, el astigmatismo afecta la AV en proporción directa a su magnitud, en todas las distancias de trabajo8 .

En el astigmatismo regular hay dos meridianos principales9, cuando estos meridianos se encuentran dentro de 20º horizontal y vertical, el astigmatismo de clasifica en astigmatismo de regla o astigmatismo contra regla según el mayor poder de refracción, se encuentre ubicado en el meridiano vertical o horizontal(ejes) respectivamente.8,9,25

* En la tercera componente están bien representadas las variables AVLD y AVLI.

Como la agudeza visual para ver de lejos 8,9,10,11,14,15,16,17 es una medida de la capacidad del sistema visual para detectar, reconocer o detallar situaciones espaciales, la primera componente será denominada de “agudeza visual de lejos” (AVL).

* En la cuarta componente están bien representadas las variables LESLD y LESLI.

La formulación de lentes esféricas se da para corregir problemas de hipermetropías o de miopías.8,9,10,11 Por lo anterior, la segunda componente principal es denominada de “ubicación del objeto respecto a la retina” (UORR) o también de “Corrección esférica”.

● En la quinta componente están bien representadas las variables AVPD y AVPI

Para evaluar la agudeza visual de cerca se utilizan optotipos o cartillas, el más

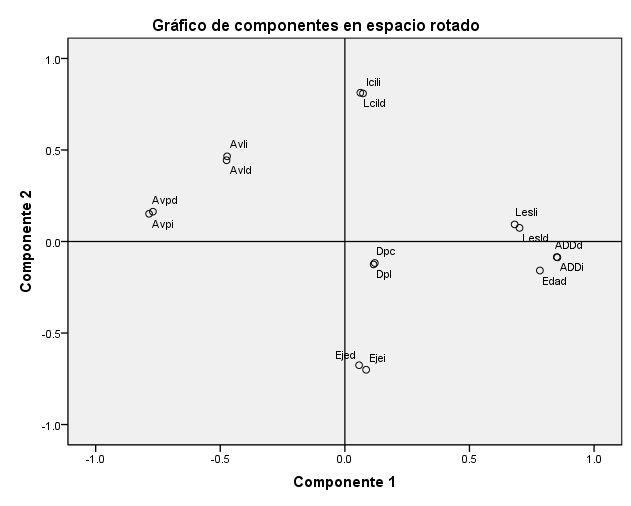
utilizado es el optotipo de Snellen8.El mejor nivel de AV se corresponde con

0.5m y el más bajo con 2.0 m8. El alejamiento del punto próximo de

acomodación obliga al paciente a retirar el objeto observado en VP, que es el

principal síntoma del présbita8,14, 15.

La gráfica 1 presenta con claridad qué grupos de variables están asociados a los componentes rotados; en ella se observa que:



Gráfica 1. Componentes en espacio rotado

* La primera componente está correlacionada positivamente con las variables LESD, LESLI, ADDD, ADDI, y la variable EDAD y negativamente, con las variables AVPD, AVPI, AVLI y AVLD.17,18,20,21
* Las variables LCILD, LCILI, EJED y EJEI están correlacionadas con la segunda componente, las dos primeras negativamente y las otras positivamente.
* Las variables DPC y DPL se encuentran muy cerca al origen, indicando que tienen una baja contribución en la formación de las dos primeras componentes.
  1. **APLICACIÓN DEL ANÁLISIS DISCRIMINANTE EN LOS PROBLEMAS REFRACTIVOS OCULARES**

El AD es utilizado para clasificar a los pacientes en uno de los cinco grupos:

1. Astigmáticos (A).

2. Astigmáticos hipermétropes (AH).

3. Astigmáticos miópicos (AM).

4. Hipermétropes (H).

5. Miopes (M).

Esta técnica proporciona los métodos que permiten establecer reglas que se puedan emplear para clasificar a otros pacientes en uno de los grupos anteriormente mencionados.

En el AD con una variable dependiente con cinco categorías y 14 variables explicativas, el número máximo de ejes discriminante es cuatro.1,18,6,26La clasificación de los pacientes entre los grupos se hace considerando las variables que más y mejor caracterizan el tipo de defecto refractivo ocular, y que en consecuencia diferencian a los grupos. Estas variables denominadas discriminantes, se presentan como combinaciones lineales(o cuadráticas) 1,27de las variables originales, y se expresan por una o más funciones discriminantes21.

Al aplicar AD por pasos28, sobre la variable diagnóstico con las variables latentes28 generadas por el AFCP, los resultados son los que se muestran en la Tabla 2.



Tabla 2. Variables en el modelo

Los resultados muestran que las variables de la tabla 2 son las que más discriminan en la clasificación de los pacientes en las diferentes categorías. El estadístico M de Box,1,6,26 que contrasta la hipótesis nula igualdad de medias entre los grupos, es rechazado muy significativamente, ya que el p-valor es igual a 0,000. Por lo tanto, se concluye que la información aportada por las variables latentes es estadísticamente significativa en la discriminación de las categorías.

El estadístico lambda de Wilks muestra que su menor valor se tiene cuando se incluyen las cinco variables.

En la matriz intra-grupo combinadas que se observa en la tabla 3, obtenida con la suma de matrices de sumas de cuadrados y productos cruzados de cada grupo por separado, sumando a continuación las matrices de todos los grupos y dividiendo finalmente por los grados de libertad. Esta matriz es utilizada para clasificar1,6,26a los pacientes en alguno de los problemas refractivos oculares o grupos definidos anteriormente.



Tabla 3. Matrices intra-grupo combinadas

* 1. **CONSTRUCCIÓN DEL MODELO CON LAS VARIABLES DISCRIMINANTES**

Identificadas las variables discriminantes, el paso siguiente es la construcción del modelo. En el caso lineal, la comparación de las funciones discriminantes genera desigualdades de la forma:

Donde el vector de medias de las variable discriminantes, es vector que define un paciente en el modelo, es la probabilidad a priori para la categoría es la matriz de varianzas covarianzas combinadas,29 en este caso, esa matriz tiene como elementos los mostrados en la tabla 4.



DIAG: diagnóstico

Tabla 4. Matrices varianza-covarianza por grupo

En el caso cuadrático, estas funciones generan desigualdades de la forma:

- 2

En ambos modelos, se clasifica al individuo , en la población , si se satisface la desigualdad para cada . 21

Las son las matrices de varianzas-covarianzas muestrales de cada categoría. En la tabla 4 se presentan cada una de estas matrices.

Construidas estas funciones, el interés se centra en determinar su importancia en la correcta clasificación de los pacientes. Para ello, bastaría comparar el porcentaje de personas correctamente clasificadas en cada grupo, en relación con el porcentaje de pacientes que se esperaría fueran correctamente clasificados aleatoriamente, teniendo en cuenta los tamaños de cada grupo.1,30

Las probabilidades previas de clasificación de pacientes en astigmáticos, astigmáticos hipermétropes, astigmáticos miópicos, hipermétropes y miopes fueron: 0.071, 0.335, 0.182, 0.285 y 0.127, respectivamente. No tiene sentido utilizar las funciones discriminantes para clasificar pacientes si el porcentaje de pacientes bien clasificados son valores cercanos a los anteriores.1,26

Las diferencias en los resultados de clasificación son notorias. Mientras que en el modelo lineal el porcentaje clasificado correctamente respecto a los casos agrupados originales fue de 81%, para el modelo cuadrático fue del 88.1%, como se observa en la tabla 5.

DIAG: diagnóstico

Tabla 5 Validación cruzada

La discriminación cuadrática muchas veces resulta bastante inestable, salvo que se tenga muestras grandes,29 además de que con frecuencia se presentan mejores resultados con el modelo lineal,29en este caso, sucedió todo lo contrario. Se observa, asimismo, la estabilidad de los dos modelos.6

Comparando los resultados de clasificación al aplicar estos modelos en comparación con una clasificación al azar,3 mediante el índice de significancia práctica (ISP),1se observa en el caso lineal un 74.74% de pacientes bien clasificados, en comparación con el porcentaje de pacientes que se esperaría fueran bien clasificados, al azar, en el caso cuadrático este porcentaje es de más del 84%.

1. **CONCLUSIONES**

El AFCP, aplicado a este estudio en el Norte de Barranquilla, facilitó descubrir la alta intercorrelación2entre las variables que tipifican este tipo de deficiencias visuales. Con cinco factores seleccionados, se logra explicar más del 88% de la variabilidad total contenida en las variables originales.

El AD es muy sensible a las desviaciones de la normalidad de los datos29 y, a pesar de que la discriminación lineal es más robusta, en este estudio, con el AD cuadrático se obtuvo un mejor ajuste en la clasificación de los pacientes. Con el AD lineal, se obtiene una reducción de más del 74% en el error de clasificación26, en relación con el porcentaje obtenido por una clasificación al azar. Con el AD cuadrático, se obtiene una reducción, de más del 84.1%, en el error de clasificación cuando este análisis se realiza sobre las variables latentes generadas por el AF. Y con un valor agregado en relación con el modelo lineal como es el hecho de la correcta clasificación de más pacientes en cada grupo en relación con el modelo lineal, más aún, en la tabla 6 se registra que solo 11 de los 307 pacientes bien clasificados por el modelo lineal fueron mal clasificados por el modelo cuadrático. Es decir, de los 307 pacientes bien clasificados por el mejor modelo lineal, fueron bien clasificados por el mejor modelo cuadrático 296 (más del 96%). Los resultados observados en la tabla 5, de pacientes bien clasificados al aplicar validación cruzada,4 hablan muy bien de la estabilidad de los dos modelos.



DIAG: diagnóstico

Tabla 6. Pacientes pronosticado por cada modelo

Para concluir, el mejor modelo lineal de clasificación de pacientes viene dado por:

, donde es la matriz de varianzas covarianzas combinadas. En el caso cuadrático, viene dado por:

Se clasifica al paciente , en la población , si para cada fijo.

REFERENCIAS BIBLIOGRÁFICAS

1. DÍAZ, R.M; GONZÁLEZ, A. A; HENAO, P.A; DÍAZ, M.M.E (2012). Una Introducción al Análisis Estadístico Multivariado Aplicado*.* Ediciones Uninorte, Barranquilla.
2. DILLON, W; GOLSTEIN M. Multivariate Analysis Methods and Applications.

John Wiley & Sons, Inc. 1984, USA.

1. KACHIGAN S. Multivariate Statistical Analysis. Ed 2° Radius Press, New York, USA. (1991)
2. TABACHNICK B. FIDELL L. Using Multivariate Statistics. Pearson Education, 5a ed. USA. (2007).

GÓMEZ, C; HERRERA, M; PADILLA, C(2004). Revista Cubana de Oftalmología.

1. GIL, J; GARCIA E. RODRIGUEZ G. Análisis Discriminante. Editorial La Muralla, S. A. (2001), Madrid, España.

TAMAYO; F. G; TAMAYO, F. M (2001). Errores refractivos y sus implicaciones genéticas.

<http://puj-portal.javeriana.edu.co/portal/page/portal/Facultad%20de%20Medicina/1documentos/>

Colecc\_der/Refractivos.pdf

GUERRERO, V. J.J (2012). Optometría clínica, 2ª edición. Edit. Kimpres Ltda, Bogotá.

RIORDAN, E.P; WITCHER, J. P (2004).Oftalmología general, 13ª edición, Editorial

El Manual Moderno S. A. de C. V. México.

KEITH, C. G; EMERY, A. E. H (1978). Oftalmología y Genética. Churchill-Livingstone,

Edinburgh; p-p 99-104.

EMERY, A. E.H; RIMOIN, D.L (1990). Principios y práctica de medicina genética.

Churchill-Livingstone, Edinburgh.

MIOPÍA, INTELIGENCIA Y ARTE. Revista Academia de Medicina.

[www.encolombia.com/medicina/academicina/academ24529-](http://www.encolombia.com/medicina/academicina/academ24529-)

elhombredelxxi-00.htm

ARENAS, E; MAGLIONE, A(1997). Laser in situ keratomileusis for astigmatism and

myopia after penetrating keratoplasty. J Retract Surg.

KANSKY, J.J (1992). Oftalmología clínica, 3ª edición, Doyma, Barcelona.

ALEMAÑY, M. J; VILLAR, V. R (2003). Oftalmología. Editorial Ciencias Médicas, Cuba.

BELMONTE, M. J. (1985). Oftalmología clínica básica, 1ª Ed. Universidad de Alicante.

1. JOHNSON R; WICHERN D. Applied Multivariate Statistical Analysis, 5ª Ed, Prentice- Hall, New Jersey. (2002).
2. URIEL E; ALDAS J. Análisis Multivariante Aplicado, Editorial Thomson, España, 2005.

KRZANOWSKI, W. J (1988). Principles of Multivariate Analysis. A user’s Perspective. Oxford University Press, N. Y, USA.

1. DALLAS J. Métodos multivariados aplicados al análisis de datos.

Editorial Thomson, 2000.

1. DÍAZ L. Estadística Multivariada: Inferencia y Métodos. 2ª Ed., Universidad Nacional de Colombia, Bogotá. (2007).

CUADRAS, C (2008). Nuevos Métodos de análisis Multivariante. CMC Editions, Barcelona, España.

1. GIRI N (2004). Multivariate Statistical Analysis. 2ª Ed, Marcel Dekker, Inc, New York.04

24. PÉREZ, C. (2001). Técnicas Estadísticas con SPSS. Pearson Educación, S. A., Madrid,

España.

JIMENO, P. et al (1998). La refracción en el niño. Edit. Mc Graw-Hill, Interamericana, Madrid.

1. LUQUE, T (2000). Técnicas de análisis de datos en investigación de mercados. Ediciones Pirámide, Madrid.
2. FINN J. A General Model Multivariate Analysis. Holt, Rinehart and Winston, Inc. (1974)

HAIR, J. F; ANDERSON, R. E; TATHAM, R. L; BLACK, W. C (1999). Análisis Multivariable.5ª Ed. Prentice-Hall, Madrid.

1. PEÑA, D. (2002). Análisis de datos multivariantes*.* Ed. McGraw-Hill, Interamericana de España.
2. TABACHNICK, B; FIDELL, L (2007). Using Multivariate Statistics, 5ª edición.

Pearson Education, Inc. U.S.A.

1. MSc. Matemáticas con énfasis en Estadística. Prof. Universidad del Norte, Barranquilla- Colombia. Tel

   575+ 3156934347. Email: mdiaz@uninorte.edu.co [↑](#footnote-ref-1)
2. Esp. Estadística. Prof. Universidad del Atlántico, Barranquilla-Colombia. Tel 575+3106126784. Email:

   josebarraza@mail.uniatlantico.edu.co [↑](#footnote-ref-2)
3. Optómetra. Prof. Universidad Metropolitana, Barranquilla-Colombia. Tel 575+3135508040. Email:

   optoacademico@unimetro.edu.co [↑](#footnote-ref-3)
4. [↑](#footnote-ref-4)