



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

**ESCUELA DE POSGRADO**

**MAESTRÍA EN ESTADÍSTICA**

**EJERCICIOS DISTRIBUCIÓN NORMAL MULTIVARIADA**

**MODELOS LINEALES**

**AUTOR:**

Br. KEVIN HEBERTH HAQUEHUA APAZA

**DOCENTE:**

Mgt. CARLA PATRICIA ZUÑIGA VILCA

**CUSCO - PERÚ**

**NOVIEMBRE - 2024**

## 1. Problema

Sea el vector aleatorio  $\mathbf{v} \sim N_4(\boldsymbol{\mu}, \Sigma)$ ,

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 5 \\ -2 \\ 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 9 & 0 & 3 & 3 \\ 0 & 1 & -1 & 2 \\ 3 & -1 & 6 & -3 \\ 3 & 2 & -3 & 7 \end{pmatrix}.$$

Si  $\mathbf{v}$  es particionada como  $\mathbf{v} = (y_1, y_2, x_1, x_2)'$ ,

$$\boldsymbol{\mu}_y = \begin{pmatrix} 2 \\ 5 \end{pmatrix}, \quad \boldsymbol{\mu}_x = \begin{pmatrix} -2 \\ 1 \end{pmatrix},$$

$$\Sigma_{yy} = \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Sigma_{yx} = \begin{pmatrix} 3 & 3 \\ -1 & 2 \end{pmatrix}, \quad \Sigma_{xx} = \begin{pmatrix} 6 & -3 \\ -3 & 7 \end{pmatrix}.$$

Obtener

### 1.1. Esperanza

$\mathbb{E}(\mathbf{y}|\mathbf{x})$

#### 1.1.1. Solución

Utilizando el teorema 4:

**Teorema 4:** Si  $y$  y  $x$  son normales multivariadas conjuntas con  $\Sigma_{yx} \neq 0$ , entonces la distribución condicional de  $y$  dado  $x$ ,  $f(y|x)$ , es normal multivariada con vector de medias y matriz de covarianzas.

$$\mathbb{E}(\mathbf{y}|\mathbf{x}) = \boldsymbol{\mu}_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$$

1. **Invertir**  $\Sigma_{xx} = \Sigma_{xx}^{-1}$ , usando la fórmula de inversa de una matriz

$$\Sigma_{xx}^{-1} = \frac{1}{\det(\Sigma_{xx})} \begin{pmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{pmatrix}.$$

Calculamos el determinante de  $\Sigma_{xx}$

$$\det(\Sigma_{xx}) = (6)(7) - (-3)(-3) = 42 - 9 = 33.$$

De tal forma que tenemos

$$\Sigma_{xx}^{-1} = \frac{1}{33} \begin{pmatrix} 7 & 3 \\ 3 & 6 \end{pmatrix} = \begin{pmatrix} \frac{7}{33} & \frac{3}{33} \\ \frac{3}{33} & \frac{6}{33} \end{pmatrix}$$

2. **Multiplicar**  $\Sigma_{yx} \Sigma_{xx}^{-1}$

$$\Sigma_{yx} \Sigma_{xx}^{-1} = \begin{pmatrix} 3 & 3 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} \frac{7}{33} & \frac{3}{33} \\ \frac{3}{33} & \frac{6}{33} \end{pmatrix} = \begin{pmatrix} \frac{30}{33} & \frac{27}{33} \\ -\frac{1}{33} & \frac{9}{33} \end{pmatrix}$$

3. **Calcular**  $(\mathbf{x} - \mu_x)$

$$\mathbf{x} - \mu_x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} x_1 + 2 \\ x_2 - 1 \end{pmatrix}$$

4. **Calcular**  $\Sigma_{yx} \Sigma_{xx}^{-1} (\mathbf{x} - \mu_x)$

$$\Sigma_{yx} \Sigma_{xx}^{-1} (\mathbf{x} - \mu_x) = \begin{pmatrix} \frac{30}{33} & \frac{27}{33} \\ -\frac{1}{33} & \frac{9}{33} \end{pmatrix} \begin{pmatrix} x_1 + 2 \\ x_2 - 1 \end{pmatrix}$$

$$\Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \mu_x) = \begin{pmatrix} \frac{30}{33}(x_1 + 2) + \frac{27}{33}(x_2 - 1) \\ (-\frac{1}{33})(x_1 + 2) + \frac{9}{33}(x_2 - 1) \end{pmatrix} = \begin{pmatrix} \frac{30}{33}x_1 + \frac{27}{33}x_2 + 1 \\ -\frac{1}{33}x_1 + \frac{9}{33}x_2 - \frac{1}{3} \end{pmatrix}$$

## 5. Sumar $\mu_y$

$$\mathbb{E}(\mathbf{y}|\mathbf{x}) = \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \mu_x)$$

$$\mathbb{E}(\mathbf{y}|\mathbf{x}) = \begin{pmatrix} 2 \\ 5 \end{pmatrix} + \begin{pmatrix} \frac{30}{33}x_1 + \frac{27}{33}x_2 + 1 \\ -\frac{1}{33}x_1 + \frac{9}{33}x_2 - \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{30}{33}x_1 + \frac{27}{33}x_2 + 3 \\ -\frac{1}{33}x_1 + \frac{9}{33}x_2 + \frac{154}{33} \end{pmatrix}$$

## 1.2. Covarianza

$$\text{cov}(\mathbf{y}|\mathbf{x})$$

### 1.2.1. Solución

De la misma forma utilizamos la definición del teorema 4 de covarianza.

$$\text{Cov}(\mathbf{y}|\mathbf{x}) = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$$

1. **Hallar la transpuesta** del vector que tenemos  $\Sigma_{yx}^T$

$$\Sigma_{xy} = \Sigma_{yx}^T = \begin{pmatrix} 3 & -1 \\ 3 & 2 \end{pmatrix}$$

2. **Calcular**  $\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$

$$\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} = \begin{pmatrix} 3 & 3 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} \frac{7}{33} & \frac{3}{33} \\ \frac{3}{33} & \frac{6}{33} \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 3 & 2 \end{pmatrix}$$

$$\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} = \begin{pmatrix} \frac{21}{33} + \frac{9}{33} & \frac{9}{33} + \frac{18}{33} \\ -\frac{7}{33} + \frac{6}{33} & -\frac{3}{33} + \frac{12}{33} \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} \frac{30}{33} & \frac{27}{33} \\ -\frac{1}{33} & -\frac{9}{33} \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 3 & 2 \end{pmatrix}$$

### 1.2.2. Problemas específicos

1. ¿Qué variables están asociadas al riesgo de poder realizarse un Retratamiento LASIK en pacientes sometidos a cirugía refractiva LASIK?
2. ¿Podrán describir los componentes las características clínicas y demográficas en las pruebas preoperatorias de pacientes sometidos a cirugía refractiva LASIK?
3. ¿Cómo afectan estos componentes en la predicción de realizarse un retratamiento LASIK?

### 1.3. Justificación de la investigación

El estudio aplica técnicas bioestadística como el análisis de componentes principales de tal forma que pueda explicar las variables clínicas de pacientes sometidos a cirugía refractiva Laser In Situ Keratomileusis y ver su asociación en la incidencia de tener un retratamiento LASIK. De tal forma que nuevos estudios tomen en cuenta la importancia de las técnicas bioestadísticas que se puedan emplear en la cirugía refractiva ya sea para la predicción preoperatorio si un paciente tendrá la probabilidad de desarrollar un Retratamiento LASIK así como la influencia que tenga para el desarrollo de nomogramas posteriores.

### 1.4. Objetivos de la investigación

#### 1.4.1. Objetivo general

Establecer si existe asociación entre el perfil composicional de pacientes sometidos para cirugía refractiva Laser In Situ Keratomileusis con respecto a ser sometido a un Retratamiento LASIK.

### 1.4.2. Objetivos específicos

1. Identificar variables y factores asociados al riesgo de poder realizarse un Retratamiento LASIK en pacientes sometidos a cirugía refractiva LASIK.
2. Interpretar los componentes con respecto a sus variables originales de pacientes sometidos a cirugía refractiva LASIK.
3. Evaluar la asociación de componentes con respecto al Retratamiento LASIK.

## 2. Marco teórico conceptual

### 2.1. Bases teóricas

#### 2.1.1. Conceptos básicos

**Definición 2.1.** (?) Si se tiene  $m$  y  $n$  como números enteros positivos. Decimos a una **matriz de tamaño**  $m$  por  $n$ , o una **matriz**  $m \times n$ , como un arreglo con  $mn$  números agrupados en  $m$  filas y  $n$  columnas, expresado en la siguiente manera

$$\begin{bmatrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1n} \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2n} \\ X_{31} & X_{32} & X_{33} & \cdots & X_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & X_{m3} & \cdots & X_{mn} \end{bmatrix}$$

Donde el primer sufijo representa a la **fila** y el segundo sufijo representa a la **columna**, tal que  $x_{ij}$  aparece en la  $i$ -ésima fila y la  $j$ -ésima columna. De forma abreviada se tiene:

$$X = [x_{ij}]_{m \times n}$$

Donde  $X$  es la matriz de tamaño  $m \times n$  donde  $x_{ij}$  es el  $(i, j)$ -elemento o la  $(i, j)$ -entrada de la matriz.

**Definición 2.2.** (?) Sea  $A = [a_{ij}]_{m \times n}$  y  $B = [b_{ij}]_{n \times p}$ . Se define al **producto**  $AB$  a la matriz  $m \times p$  donde el  $(i, j)$ -ésimo elemento es:

$$[AB]_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{in}b_{nj}$$

Es decir que los  $(i, j)$  elementos del producto  $AB$  es obtenido por la suma de productos de los elementos de la  $i$ -ésima fila de  $A$  con los correspondientes elementos de la  $j$ -ésima columna de  $B$ , expresado también como la siguiente sumatoria:

$$[AB]_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$$

**Definición 2.3.** (?) Una matriz es llamada *cuadrada* si el tamaño es  $n \times n$  es decir que tiene el mismo número de filas y columnas

**Definición 2.4.** (?) Para cada matriz  $A_{n \times n}$  existe una única matriz  $I_n$  de tamaño  $n \times n$  tal que  $AI = A = IA$  llamada como **matriz identidad**

**Definición 2.5.** (?) Si  $A$  es una matriz  $m \times n$ , se llama **matriz transpuesta** a una matriz  $A^T$  donde el  $(i, j)$ -ésimo elemento es el  $(j, i)$ -ésimo elemento de  $A$ . Mas precisamente decir si  $A = [a_{ij}]_{m \times n}$  entonces  $A^T = [a_{ji}]_{n \times m}$

**Definición 2.6.** (?) Una matriz  $A_{n \times n}$  es llamado **ortogonal** si:

$$AA^T = I_n = A^T A$$

**Definición 2.7.** (?) Se entiende como **sistema de  $m$  ecuaciones lineales con las  $n$  incógnitas**

$x_1, \dots, x_n$  tal que la lista de ecuaciones se encuentra de la forma:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n = b_2$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots + a_{3n}x_n = b_3$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n = b_m$$

Donde los  $a_{ij}$  y los  $b_i$  son números, expresados también de la siguiente forma:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots + a_{3n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n \end{bmatrix}$$

Donde se observa un sistema expresado de la forma matricial como  $\mathbf{Ax} = \mathbf{b}$  donde  $A = [a_{ij}]_{m \times n}$

y  $\mathbf{x}, \mathbf{b}$  son las siguientes matrices columnas

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

La matriz  $A_{m \times n}$  es llamada **matriz de coeficientes** del sistema.

**Definición 2.8.** (?) Sea una matriz  $A_{m \times n}$  es llamado **matriz inversa izquierda de A** a una matriz  $X_{n \times m}$  si satisface la ecuación  $XA = I_n$ ; y una **matriz inversa derecha de A** si satisface la ecuación  $AX = I_m$



**Definición 2.9.** (?) Un **espacio vectorial** es un conjunto  $V$  de  $F$  en el cual se definen dos operaciones “adición” y “multiplicación de escalares”. Si los escalares son números reales sería un **espacio vectorial real**, y si los escalares son números complejos sería un **espacio vectorial complejo**. Por lo tanto  $F$  puede pertenecer a los números reales o los números complejos.

**Definición 2.10.** (?) Si  $V$  y  $W$  son espacios vectoriales sobre un cuerpo  $F$  entonces por **aplicación lineal** (o **transformación lineal**) de  $V$  a  $W$  nos referimos a una aplicación  $f : V \rightarrow W$  tal que

$$1. (\forall x, y \in V) \quad f(x + y) = f(x) + f(y)$$

$$2. (\forall x \in V)(\forall \lambda \in F) \quad f(\lambda x) = \lambda f(x)$$

**Definición 2.11.** (?) Un **autovalor** (o **valor latente**) de  $A$  es un escalar  $\lambda$  para el que existe una matriz no cero  $\mathbf{x}_{n \times 1}$  tal que  $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ . Tal que las columnas de la matriz  $\mathbf{x}$  son llamados **autovectores** (o **vectores latentes**) asociados con  $\lambda$

### 2.1.2. Análisis de componentes principales

(?) El análisis de componentes principales (PCA) toma las variables incorrelacionadas entre sí y que puedan ordenarse de acuerdo con la información que llevan incorporada para crear nuevas variables llamadas componentes.

1. **Componentes principales de dos variables:** Tomemos en cuenta que se tienen dos variables distribuidas normalmente y que están correlacionadas entre sí, de tal forma que su gráfico de dispersión formará una elipse como se muestra en la figura 1. En la cual se pueda trazar dos líneas perpendiculares que midan la longitud y la altura de la elipse. Estas líneas vienen a ser los autovectores de la matriz de correlaciones, además de que las líneas al ser perpendiculares vienen a ser los autovectores. Si se añadiera una tercera variable la elipse se convertiría en un balón de rugby y se tendrían tres líneas perpendiculares, tres autovectores y así sucesivamente.

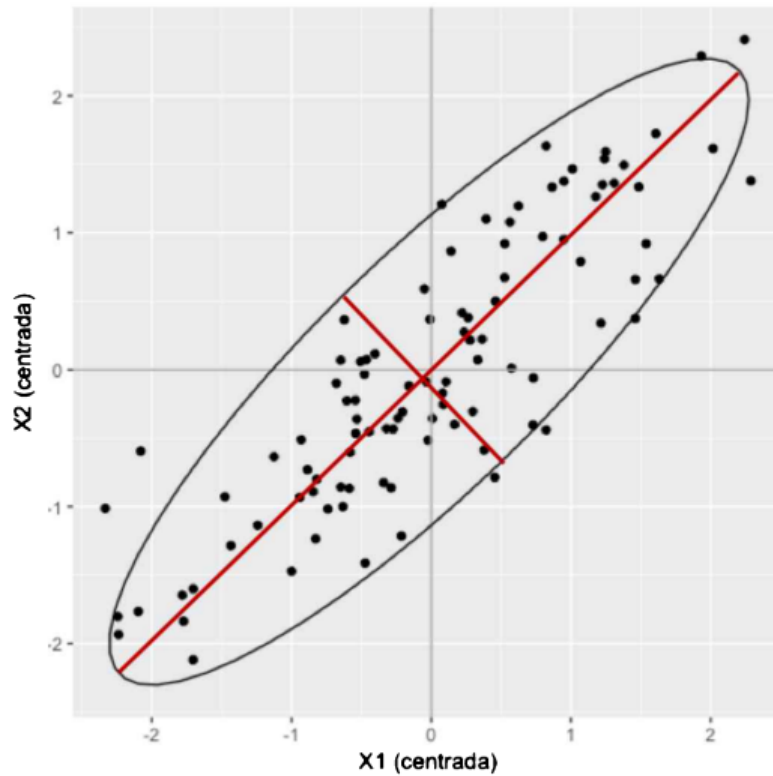


Figura 1: Interpretación gráfica de autovectores y autovalores (?)

Cada autovector tienen un asociado un autovalor. El autovalor da una medida de longitud del autovector y observando ese valor se puede ver si los datos están distribuidos de forma homogénea o heterogénea. Entonces se parte de la matriz de correlaciones en el cual se muestra que existe un nivel de correlación razonable entre las variables que justifique el uso de un PCA.

Si tomamos en cuenta los autovectores calculados se tendría la siguiente matriz:

$$\mathbf{u}_1 = \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix}; \mathbf{u}_2 = \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix}$$

La aplicación del procedimiento de componentes principales requiere calcular los autovalores y autovectores de la matriz de covarianzas. En el cual los autovectores representan el giro que se ha producido entre los ejes. Así genéricamente se tiene que aplicar a las

variables originales los coeficientes de vectores  $\mathbf{u}_1$  y  $\mathbf{u}_2$  expresados de la siguiente manera:

$$x_1^* = u_{11}X_1 + u_{12}X_2$$

$$x_2^* = u_{21}X_1 + u_{22}X_2$$

Así mismo deben de cumplir las condiciones arbitrarias pero necesarias de que los ejes sean ortogonales (1) y fijar las escalas de medida (2)

$$u_{11}^2 + u_{12}^2 = 1 \quad ; \quad u_{21}^2 + u_{22}^2 = 1 \quad (1)$$

$$u_{11}u_{21} + u_{12}u_{22} = 0 \quad (2)$$

En el análisis de componentes principales es importante conocer la correlación de cada variable con las componentes para tener una medida de lo importante que es cada variable en la interpretación del compoente principal. A este se le denomina **carga factorial**  $l_{hj}$  entre la componente  $h$ -ésima y la variable  $j$ -ésima viene dada por:

$$l_{hj} = \frac{u_{hj}}{\hat{s}_j} \sqrt{\lambda_h} \quad (3)$$

Donde la única notación no conocida es  $\hat{s}_j$  que viene a ser la desviación típica de la variable  $j$ -ésima

2. **Componentes principales para el caso general:** La obtención de componentes principales es un caso típico de cálculo de autovalores y autovectores de una matriz simétrica.

a) **Obtención de la primera componente** Considere que se dispone de una muestra de tamaño  $n$  acerca de las  $p$  siguientes variables  $X_1, X_2, \dots, X_p$  y que las observaciones están expresadas bien en desviaciones respecto a la media (centradas) o bien como

variables tipificadas. La primera componente, de la misma forma que el resto, se expresa como combinación lineal de las variables originales:

$$Z_{1i} = u_{11}X_{1i} + u_{12}X_{2i} + \dots + u_{1p}X_{pi} \quad (4)$$

Puede comprobarse fácilmente que tanto si las  $X_j$  variables están expresadas en desviaciones como si están tipificadas se obtiene que la media muestral de  $Z_1$  es igual a 0. Para el conjunto de las  $n$  observaciones muestrales esta ecuación se puede expresar matricialmente de la siguiente forma:

$$\begin{bmatrix} Z_{11} \\ Z_{12} \\ \vdots \\ Z_{1n} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{p1} \\ X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \cdots & X_{pn} \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1p} \end{bmatrix} \quad (5)$$

o en notación matricial compacta:

$$\mathbf{z}_1 = \mathbf{X}\mathbf{u}_1 \quad (6)$$

La primera componente se obtiene de forma que su varianza sea máxima, sujeta a la restricción de que la suma de los pesos  $u_{ij}$  al cuadrado sea igual a la unidad. La varianza del primer componente, teniendo en cuenta que su media es 0, vendrá dada por:

$$var(Z_1) = \frac{\sum_{i=1}^n Z_{1i}^2}{n} = \frac{1}{n} \mathbf{z}_1' \mathbf{z}_1 = \frac{1}{n} \mathbf{u}_1' \mathbf{X}' \mathbf{X} \mathbf{u}_1 = \mathbf{u}_1' \left[ \frac{1}{n} \mathbf{X}' \mathbf{X} \right] \mathbf{u}_1 \quad (7)$$

Si las variables están expresadas en desviaciones respecto a la media,  $(1/n)\mathbf{X}'\mathbf{X}$  es la matriz de covarianzas muestral a la que denominaremos  $\mathbf{V}$ ; para variables tipificadas  $(1/n)\mathbf{X}'\mathbf{X}$  es igual a la matriz de correlaciones  $\mathbf{R}$ . Las componentes pueden obtenerse

para ambos tipos de variables, aunque en los paquetes de ordenador para análisis multivariante se utiliza la matriz de correlación en la mayor parte de los casos. A efectos expositivos y para darle mayor generalidad vamos a suponer, sin embargo, que utilizamos la matriz de covarianzas. Por lo tanto, la varianza a maximizar es:

$$var(Z_1) = \mathbf{u}_1' \mathbf{V} \mathbf{u}_1 \quad (8)$$

La restricción señalada analíticamente de que la suma de los cuadrados de los pesos sea la unidad viene dada por:

$$\sum_{j=1}^p u_{1j}^2 = \mathbf{u}_1' \mathbf{u}_1 = 1 \quad (9)$$

En consecuencia, incorporando esta restricción se forma el siguiente lagrangiano:

$$\mathbf{L} = \mathbf{u}_1' \mathbf{V} \mathbf{u}_1 - \lambda (\mathbf{u}_1' \mathbf{u}_1 - 1) \quad (10)$$

Para maximizar el valor lagrangiano derivamos respecto a  $\mathbf{u}_1$  e igualmente a 0:

$$\frac{\partial \mathbf{L}}{\partial \mathbf{u}_1} = 2\mathbf{V} \mathbf{u}_1 - 2\lambda \mathbf{u}_1 = \mathbf{0} \quad (11)$$

es decir:

$$(\mathbf{V} - \lambda \mathbf{I}) \mathbf{u}_1 = \mathbf{0} \quad (12)$$

Al resolver la ecuación  $|\mathbf{V} - \lambda \mathbf{I}| = 0$ , se obtienen  $p$  autovalores. Si se toma la raíz característica mayor ( $\lambda_1$ ), se halla el autovector asociado a la misma  $\mathbf{u}_1$ , aplicando la regla de normalización dada en (9). Así el vector de ponderaciones que se aplica a las variables iniciales para obtener la primera componente principal es el autovector

asociado al mayor autovalor de la matriz  $\mathbf{V}$ .

- b) Obtención de las restantes componentes** Una componente genérica expresada en forma matricial para todas las observaciones viene dada, de forma análoga a (6), por:

$$\mathbf{z}_h = \mathbf{X}\mathbf{u}_h \quad (13)$$

Para su obtención, además de la restricción de:

$$\mathbf{u}_h' \mathbf{u}_h = 1 \quad (14)$$

se imponen las restricciones adicionales ya expuestas de que:

$$\mathbf{u}_h' \mathbf{u}_1 = \mathbf{u}_h' \mathbf{u}_2 = \dots = \mathbf{u}_h' \mathbf{u}_{h-1} = 0 \quad (15)$$

Por lo tanto, se imponen las restricciones adicionales de que el vector característico asociado a la componente  $h$ -ésima sea ortogonal a todos los vectores característicos calculados previamente. Su obtención, a parte de tener en cuenta las restricciones (15), plantea problemas conceptuales nuevos. En todo caso señalaremos que el vector  $\mathbf{u}_h$  está asociado a la raíz  $h$ -ésima, una vez ordenadas de mayor a menor. En definitiva, las  $p$  componentes principales que se pueden calcular son una combinación lineal de las variables originales, en la que los coeficientes de ponderación son los correspondientes vectores característicos asociados a la matriz  $\mathbf{V}$ .

- c) Varianzas de las componentes** Si se tienen en cuenta las propiedades de los autovalores y como ilustramos en el caso de dos componentes, la varianza de la componente  $h$ -ésima coincide su autovalor, es decir:

$$Var(Z_h) = \mathbf{u}_h' \mathbf{V} \mathbf{u}_h = \lambda_h \quad (16)$$

Adoptando como medida la variabilidad de las variables originales la suma de sus varianzas, dicha variabilidad será igual a la traza de la matriz  $\mathbf{V}$ . Ahora bien, la traza de la matriz  $\mathbf{V}$  se puede expresar así teniendo en cuenta:

$$\frac{\lambda_h}{traza\mathbf{V}} = \frac{\lambda_h}{\sum_{h=1}^p \lambda_h} \quad (17)$$

y en el caso particular en los datos estuviera tipificados y la matriz de covarianzas fuera la de correlación  $\mathbf{R}$ , entonces dado que  $traza(\mathbf{R}) = p$ , la proporción de variabilidad correspondiente a la componente  $h$ -ésima es:

$$\frac{\lambda_h}{p} \quad (18)$$

- d) Correlación entre las componentes principales y las variables originales:** Primeramente se debe calcular la covarianza entre la variable  $X_j$  y el componente  $Z_h$ . Definamos los vectores muestrales de la componente  $Z_h$  y variable  $X_j$  por:

$$\mathbf{x}_j = \begin{bmatrix} X_{j1} \\ X_{j2} \\ \vdots \\ X_{jn} \end{bmatrix} \quad \mathbf{z}_h = \begin{bmatrix} Z_{h1} \\ Z_{h2} \\ \vdots \\ Z_{hn} \end{bmatrix} \quad (19)$$

La covarianza muestral entre  $X_j$  y  $Z_h$  viene dada por:

$$cov(X_j, Z_h) = \frac{1}{n} \mathbf{x}'_j \mathbf{z}_h \quad (20)$$

El vector  $\mathbf{x}_j$  se puede expresar en función de la matriz  $\mathbf{X}$ , utilizando el vector de orden  $p$ , al que designaremos por  $\delta$ , que tiene un 1 en la posición  $j$ -ésima y 0 en las

posiciones restantes. Así,

$$\mathbf{x}'_j = \delta' \mathbf{X} = [0 \quad \dots \quad 1 \quad \dots \quad 0] \begin{bmatrix} X_{11} & \dots & X_{1i} & \dots & X_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{j1} & \dots & X_{ji} & \dots & X_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{p1} & \dots & X_{pi} & \dots & X_{pn} \end{bmatrix} \quad (21)$$

Teniendo en cuenta la expresión anterior y (13), la covarianza dada en (20) puede expresarse de la siguiente forma:

$$cov(X_j, Z_h) = \frac{1}{n} \delta' \mathbf{X}' \mathbf{X} \mathbf{u}_h = \delta' \mathbf{V} \mathbf{u}_h = \delta' \lambda_h \mathbf{u}_h = \lambda_h \delta' \mathbf{u}_h = \lambda_h u_{hj} \quad (22)$$

En consecuencia, la correlación existente entre la variable  $X_j$  y la componente  $Z_h$ , que recordemos denominábamos **carga factorial**, es la siguiente:

$$l_{jh} = \frac{cov(X_j, Z_h)}{\sqrt{var(X_j)} \sqrt{var(Z_h)}} = \frac{\lambda_h u_{hj}}{\sqrt{var(X_j)} \sqrt{\lambda_h}} = \frac{u_{hj}}{\sqrt{var(X_j)}} \sqrt{\lambda_h} \quad (23)$$

expresión que coincide con la que derivamos en su momento (3). En el caso en que las variables estén tipificadas (varianza 1), la expresión anterior adopta la forma:

$$l_{hj} = u_{hj} \sqrt{\lambda_h} \quad (24)$$

que es la que se corresponde con las llamadas *matrices factoriales* en la mayoría de programas estadísticos.

3. **Número de componentes principales a extraer:** La decisión a tomar es cuánta información (varianza) estamos dispuestos a sacrificar para ganar facilidad de interpretación.



Aunque depende del juicio del investigador, las reglas mostradas serán orientativas.

- a) **Criterio del autovalor superior a la unidad:** El criterio de retener aquellas componentes cuyo autovalor sea superior a la unidad es la opción por defecto de la mayoría de programas estadísticos. Si una componente no es capaz de explicar más información que una variable, no va a facilitar la reducción de datos, es decir, facilitar la interpretabilidad de la información. El criterio planteado de manera general es retener aquellas componentes cuyos autovalores sean superiores a la media de los autovalores

$$\lambda_h > \bar{\lambda} = \frac{\sum_{j=1}^p \lambda_j}{p} \quad (25)$$

pero, como cuando trabajamos con variables tipificadas, se verifica que la suma de los autovalores coincide con la matriz de correlaciones, que es  $p$ , entonces:

$$\sum_{j=1}^p \lambda_j = p \quad (26)$$

por lo que el criterio (25) se puede expresar como decíamos al principio:

$$\lambda_h > 1 \quad (27)$$

Se demostró que el criterio funciona bastante bien salvo que se tenga un gran número de variables ( $> 40$ ) siendo especialmente preciso con un número pequeño de variables (10-15). La varianza de una variable explicada por el componente (comunalidad) también influye en la precisión del criterio.

- b) **Criterio del gráfico de sedimentación (*scree plot*)** En esta parte se propone representar en un gráfico los autovalores correspondientes a las distintas componentes. Como los autovalores siempre son decrecientes, lo será este gráfico de líneas. El

criterio es retener aquellas componentes anteriores a donde la línea comienza a nivelarse (si fuera la ladera de una montaña y cayesen rocas, donde estas comenzarían a pararse o sedimentar, de ahí el nombre del gráfico), aunque si no se usa con precaución se puede descartar componentes que pueden ser relativas en la relevancia teórica de las mismas.

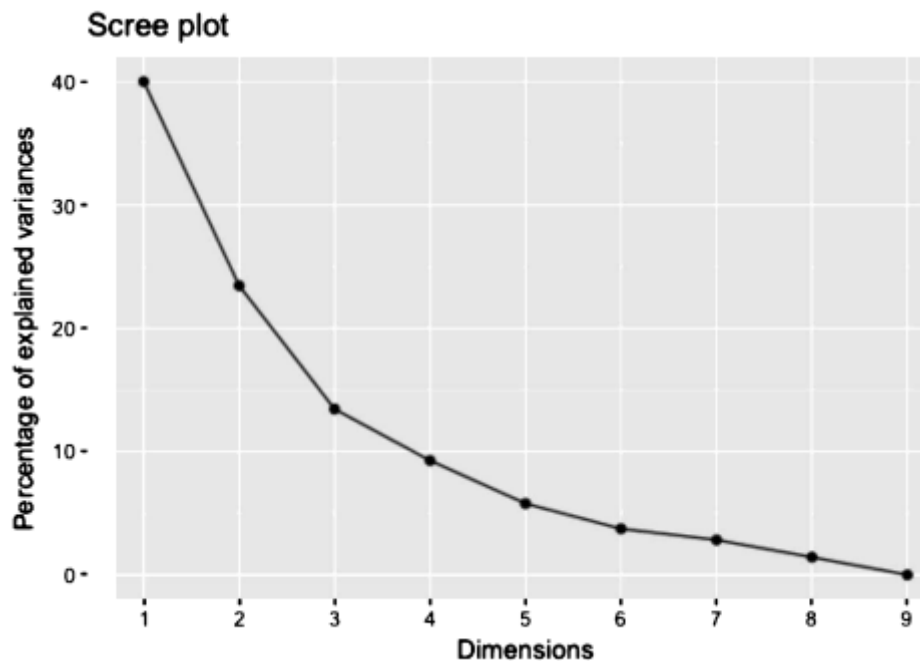


Figura 2: Gráfico de sedimentación (?)

- c) **Método paralelo:** Es obvio que hay mucha subjetividad en decidir dónde se produce el “codo” en la línea que marcaría el corte, sobre todo cuando la caída es progresiva. En la figura 2 no tendríamos muy claro si considerar tres o cuatro componentes al menos. Por esta razón Horn (1965)<sup>1</sup> sugirió un procedimiento llamado análisis paralelo para objetivar este criterio en el cual indica que cuando se tienen datos totalmente incorrelacionados, un PCA debería generar autovalores iguales a 1 (cada autovalor explica la varianza de una única variable). Sin embargo, los errores de muestreo generan correlaciones espúreas que hace que unos autovalores sean superiores a la unidad y otros inferiores. La estrategia de Horn es contrastar los

<sup>1</sup>A rationale and a test for the number of factors in factor analysis. *Psychometrika*, 30:179-185

autovalores obtenidos mediante un PCA paralelo aplicado a muestras aleatorias de variables no correlacionadas con el mismo número de variables y observaciones que la muestra original que generarán autovalores que se han ajustado para evitar el error muestral. Solo se retendría aquellas componentes con autovalores ajustados superiores a la unidad. El autovalor ajustado vendrá dado por la expresión:

$$\lambda_p - (\bar{\lambda}_p^r - 1) \quad (28)$$

donde  $\lambda_p$  es el autovalor  $p$ -ésimo de los datos originales y  $\bar{\lambda}_p^r$  es la media de las estimaciones de ese mismo autovalor en las submuestras aleatorias generadas.

- d) **Test estadístico:** El test utilizado es el de Bartlett (1951)<sup>2</sup> para analizar si la matriz de correlaciones es la identidad, adolece de las mismas limitaciones - básicamente ser muy sensible al tamaño muestral - y no recomiendan su uso pues suele aconsejar la retención de demasiadas componentes principales. Se considera que los  $p - m$  últimos autovalores poblacionales son iguales a 0. Si los autovalores muestrales que observamos correspondientes a estas componentes no son exactamente igual a 0, se debe a los problemas del azar. Por ello, bajo el supuesto de que las variables originales siguen una distribución normal multivariante, se pueden formular las siguientes hipótesis relativas a los autovalores poblacionales:

$$H_0 : \lambda_{m+1} = \lambda_{m+2} = \dots = \lambda_p = 0 \quad (29)$$

El estadístico para el contraste de esta hipótesis es el siguiente

$$Q^* = \left\{ n - \frac{2p + 11}{6} \right\} \left\{ (p - m) \ln \bar{\lambda}_{p-m} - \sum_{j=m+1}^p \ln \lambda_j \right\} \quad (30)$$

---

<sup>2</sup>The effect of standardization on a chi square. Approximation in factor analysis. *Biometrika*, 38(3/4):337-344

Bajo la hipótesis nula, el estadístico anterior se distribuye como una ji-cuadrado con  $(p - m + 2)(p - m + 1)/2$  grados de libertad. Existen algunas variantes de este test básicamente referidas al cálculo de los grados de libertad, como son las de Anderson (1963)<sup>3</sup> y Lawley (1956)<sup>4</sup>.

4. **Interpretación de las componentes principales:** Para interpretar las dos componentes extraídas es necesario fijarse en la contribución de cada variable. Al trabajar con datos tipificados, los valores estandarizados (cargas) y sin estandarizar coinciden. Cuanto mayor es la carga, mayor es la influencia que ha tenido esa variable en la formación de la componente. Por lo tanto podemos analizar cuáles son las cargas más altas y usar las variables a las que corresponden para dar una interpretación al eje.

## 2.2. Marco conceptual

Los siguientes conceptos se extraerán de (?)

- **Cirugía refractiva:** La cirugía refractiva es una subespecialidad de la oftalmología que engloba diferentes tipos de procedimientos quirúrgicos (incisionales, fotoablativos y lenticulares) que producen la corrección de los defectos de refracción del ojo humano.
- **Fisiología óptica:** El segmento anterior del ojo está compuesto de una estructura óptica esencial para la visión:
  - **La córnea:** Órgano avascular y transparente compuesto por cinco diferentes capas histológicas. La función primordial de la córnea consiste en enfocar las imágenes provenientes del exterior sobre la porción neurosensorial del ojo.
  - **La retina:** El desvío de los rayos luminosos que componen las imágenes hacia la retina se denomina refracción.

---

<sup>3</sup>Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics*, 34(1):122-148

<sup>4</sup>Tests of significance for the latent roots of covariance and correlation matrix. *Biometrika*, 43(1/2):128-136

Cuando las imágenes proyectadas desde la parte anterior del ojo se forman exactamente sobre la retina, el individuo es capaz de percibir nítidamente estas imágenes: a esa condición se denomina **emetropía**. Sin embargo, cuando los rayos de luz se enfocan fuera de la retina, se produce un defecto en la percepción visual denominado **ametropía** o defecto refractivo.

1. **Miopía:** Ocurre cuando el ojo es demasiado largo y/o la curvatura de la córnea es demasiado curva. Las imágenes que entran al ojo miope se enfocan **delante de la retina**, de tal manera que se produce una visión borrosa o desenfocada de las mismas para la visión de lejos. En general, el paciente con miopía tiene una excelente visión de cerca.
  2. **Astigmatismo:** Ocurre cuando la córnea se curva más en una dirección que la otra. Es como si la superficie anterior del ojo pareciese una pelota de rugby en lugar de a una pelota de fútbol. Si el astigmatismo es significativo, las imágenes que llegan a la retina a partir de un objeto son **múltiples** y en **diferentes focos**, produciéndose una visión distorsionada del objeto.
  3. **Hipermetropía:** Se produce cuando el ojo es corto y/o la curvatura corneal es muy plana. Las imágenes que entran al ojo se enfocan por detrás de la retina, de tal manera que se produce una visión borrosa o desenfocada de las mismas, sobre todo en la visión cercana, aunque en grados significativos de hipermetropía la visión lejana también puede estar afectada.
- **Excimer Láser:** LASER es una abreviación que significa amplificación de la luz por emisión estimulada de radiación. Básicamente consiste en una fuente de radiación (sólida, líquida o gaseosa) que luego de ser estimulada apropiadamente emite una forma de energía luminosa que es amplificada y dirigida hacia el tejido blanco. De acuerdo al tipo de fuente de radiación usada se obtienen diferentes tipos de energía LASER, cada una de las cuales

tienen un determinado efecto sobre el ojo; fotoablación, fotodisrupción o fotocoagulación. El término Excimer es una contracción de dímero excitado, una molécula energizada de dos componentes. El Excimer utiliza los gases del argón y del flúor para generar un haz de luz ultravioleta con una longitud de onda de 193 nm. La energía de esta longitud de onda de la luz es suficiente para romper los enlaces moleculares en la córnea y quitar una capa muy fina de tejido. Este proceso es denominado fotoablación<sup>5</sup>. La ablación de la córnea con el Excimer Láser produce que el retiro de entre 0.1 a 0.25 micras de tejido por cada aplicación, con mínimo o ningún daño al tejido circundante. Esta precisión submicrónica es la clave de la seguridad y eficacia obtenidas con este procedimiento. De esta manera, el Excimer Láser es capaz de moldear sobre la córnea, una nueva superficie refractiva, permitiendo de esta manera que las imágenes lleguen en forma precisa hasta los órganos sensoriales.

- **Fundamentos de la Cirugía Refractiva con Excimer Láser:** El objetivo de la cirugía refractiva es **modificar la curvatura de la córnea**, y así cambiar su poder de convergencia. Un ojo hipermetrope tiene un déficit de convergencia. Para incrementar la convergencia, es necesario aumentar la curvatura corneal en el centro, lo que se logra eliminando tejido en la periferia, en forma de anillo. Un ojo miope tiene un exceso de convergencia que para disminuirla, es necesario reducir la curvatura corneal, lo que se logra removiendo tejido en forma de círculo en el centro de la córnea. Un astigmatismo tendrá más poder de convergencia en un eje de la córnea. Para la corrección del astigmatismo se llevan a cabo diferentes ablaciones en la parte central como en la periferia de la córnea dando como resultado el aplanamiento del meridiano más curvo y el encurvamiento del meridiano más plano.

- **Evaluación preoperatoria:** Al ser un procedimiento electivo, la selección del paciente

---

<sup>5</sup>Azar Dimitri T. Refractive Surgery. Second edition. *Mosby Elseiver*, 2007

que será sometido a cirugía refractiva es de fundamental importancia. Existen condiciones que constituyen contraindicaciones absolutas para cirugía refractiva con Excimer Láser:

- Enfermedades de la córnea y superficie ocular
- Enfermedades oculares inflamatorias crónicas
- Embarazo y lactancia
- Enfermedades sistémicas autoinmunes
- Epilepsia
- Transtornos psiquiátricos
- Enfermedades sistémicas descompensadas (diabetes, infecciones)

1. **Historia clínica:** Los pacientes serán sometidos a un interrogatorio exhaustivo para determinar si cumplen con los requisitos necesarios para recibir el tratamiento. A continuación se detallan los factores que deben considerarse en la historia clínica con el paciente:

- Edad
- Tipo de ametropía, grado y la estabilidad de la misma
- Uso previo de lentes de contacto (blandas o rígidas, forma de uso)
- Medicación en uso: anticonceptivos orales, isotretinoin
- Ocupación o trabajo
- Antecedentes de infecciones corneales (queratitis herpética), cirugía corneal previa, trauma ocular previo
- Antecedentes patológicos personales: diabetes mellitus, colagenopatías, procesos inflamatorios o inmunoalérgicos, e infecciones sistémicas (HIV, hepatitis)
- Antecedentes oftalmológicos familiares: queratocono

2. **Examen oftalmológico:** El examen oftalmológico debe ser completo; los pacientes deben suspender el uso de lentes de contacto blandos por 15 días y los rígidos por un mes:

- *Refracción con y sin cicloplejia:* Esta exploración debe ser realizada de la manera más exacta posible ya que es de importancia superlativa para lograr la corrección deseada a través de los cálculos de programación del aparato de Excimer Láser. Es fundamental detectar la agudeza visual no corregida, la máxima agudeza visual corregida y el grado exacto de ametropía.
- *Pupila:* Medición del tamaño pupilar que debe ser realizada en condiciones fotópicas y escotópicas, importante a la hora de calcular la zona óptica de tratamiento.
- *Examen de la musculatura ocular:* Detecta cualquier alteración del equilibrio motor que puede ser descompensada por la cirugía, por lo que deberá ser tratada previamente.
- *Paquimetría corneal:* Medida del grosor corneal central y periférico, de vital importancia para la selección del paciente y la cantidad de ametropía que pueda ser corregida.
- *Topografía corneal:* Estudia la Queratometría y además realiza un estudio cualitativo de la cornea. Permite detectar la presencia de irregularidades, asimetrías y otras alteraciones de la superficie corneal. Es el elemento más importante para la detección de queratocono, aún en sus fases más incipientes.
- *Aberrometría:* Detecta las aberraciones de bajo y alto orden del sistema óptico del ojo, con lo que es posible personalizar el patrón de ablación para cada paciente.
- *Tensión ocular:* Descartará la existencia de hipertensión ocular o glaucoma.



- *Fondo de ojo*: Con oftalmoscopia indirecta es crucial en el examen de la retina central y periférica, permitiendo descartar desgarros, distrofias o agujeros que deben ser tratados antes de la cirugía.
- *Examen de biomicroscopia*: El estudio en la lámpara de hendidura proporciona información detallada fundamental para la cirugía.
  - Exploración de la lágrima mediante el tiempo de ruptura de la película lagrimal y la tinción con fluoresceína con el fin de descartar el ojo seco.
  - El examen de la conjuntiva permitirá determinar la existencia de procesos cicatrizales y reacciones tisulares que indiquen infecciones o inflamaciones.
  - Un examen meticuloso de la córea es de gran valor en el momento de la selección del paciente, puesto que es la estructura sobre la cual se aplicará la ablación. La exploración de la córnea debe descartar:
    - ◊ Epitelio: Erosión corneal recurrente, distrofias de la membrana basal
    - ◊ Estroma: Cicatrices, opacidades, adelgazamientos
    - ◊ Endotelio: Distrofia de Fuchs, depósitos queráticos
  - La evaluación del cristalino permite detectar catarata.
- **Asesoramiento al paciente**: Una parte fundamental para el éxito de la cirugía refractiva es la comunicación adecuada entre el médico y paciente antes de decidir el tratamiento. El oftalmólogo debe describir en detalle cuáles son las limitaciones y potenciales complicaciones del procedimiento. El paciente debe ser consciente que se trata de una cirugía electiva en un órgano sensorial que busca mejorar su calidad de vida y que debe ser parte responsable en la decisión asumiendo los riesgos que podría ocurrir.

- **Cirugía lamelar: LASIK** Son las siglas de *Laser-Assisted in Situ Keratomileusis*. Es un

procedimiento quirúrgico que se realiza en dos partes: la creación del colgajo corneal seguida de la fotoablación estromal<sup>67</sup> El colgajo corneal es una capa de tejido corneal que se corta en forma de casquete y se aparta para poder realizar la ablación estromal, dejando intacto el epitelio, la membrana de Bowman y el estroma superficial. El grosor del colgajo comprende entre 100 y 160 micras y su diámetro oscila de 8 a 9,5 mm. El corte no se realiza en toda la extensión del casquete sino que el colgajo queda unido a la córnea por un pedículo, que normalmente queda en posición nasal o superior, lo que permite, una vez acabada la cirugía, volver a recolocararlo en su sitio sin necesidad de suturas. Para realizar el colgajo se utiliza un microquerátomo. El primero en salir al mercado era mecánico, y fue desarrollado por Barraquer en 1958, y desde entonces hasta ahora han ido evolucionando obteniendo mejores resultados en su uso. Hace unos años aparecieron los aparatos Láser de Femtosegundo (haciendo alusión al tiempo que dura cada pulso) que emiten un haz infrarrojo de longitud de onda de 1053 nm que provoca fotodisrupción. Este es un proceso que transforma el tejido en plasma, la presión y temperatura elevada hace que se produzcan numerosas micro cavidades en el estroma, generando el colgajo. Una vez desplazado el colgajo creado con el microquerátomo o con el láser de Femtosegundo, se procede a la fotoablación con el Excimer Láser y al terminar esta, el colgajo es devuelto a su posición original. La cirugía lamelar presenta una recuperación visual más rápida con mínimos o nulos síntomas de discomfort post operatorio en comparación con la ablación de superficie. Sin embargo, existe una chance mayor de complicaciones relacionadas con la creación del colgajo como se muestra el figura 3.

- **Retratamiento LASIK: (?)** Aunque la predictabilidad de LASIK es alto e incluso mayor que otros procedimientos quirúrgicos refractivos, la naturaleza de la cirugía sobre el tejido

---

<sup>6</sup>Pallikaris IG, Papatzanaki ME, Stathi EZ, Frenschok O, Georgiadis A. Laser in situ keratomileusis *Laser Surg Med*, 1990;10: 463-68.

<sup>7</sup>Buratto L, Ferrari M, Genísi C. Myopic keratomileusis with the excimer laser: one-year follow up. *Refract Corneal Surg*, 1993; 9: 12-19

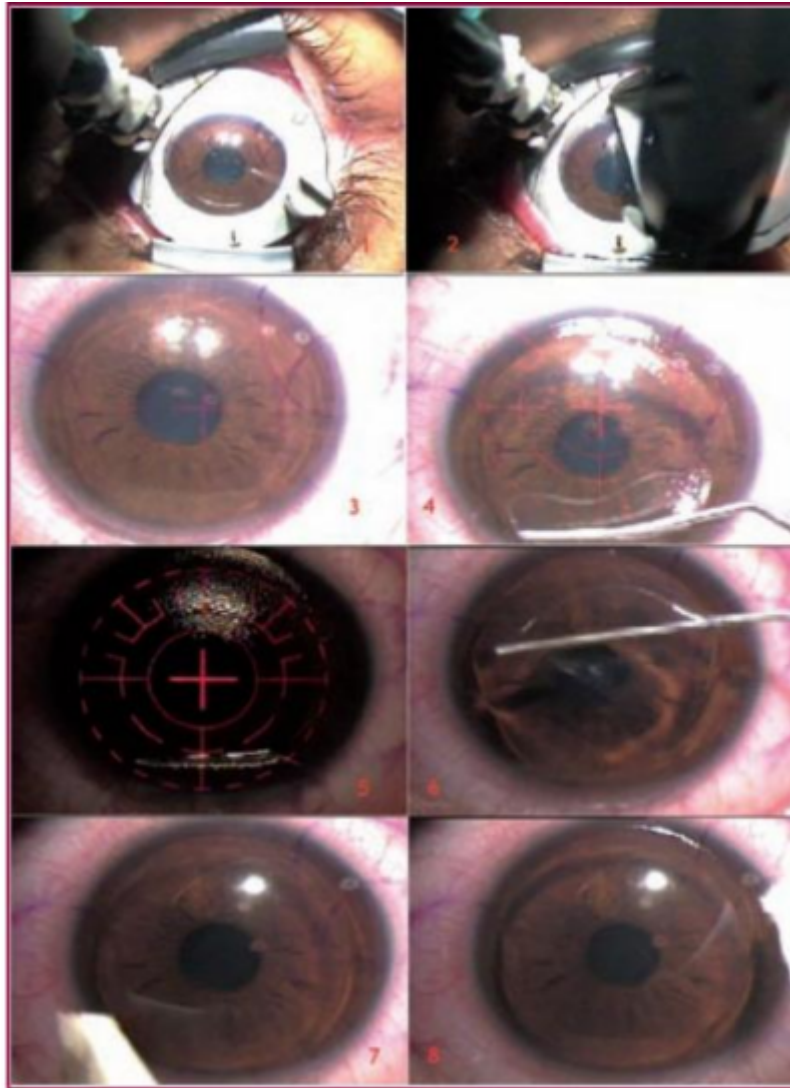


Figura 3: Técnica quirúrgica - LASIK ?

vivo es impredecible. Aunque la sobrecorrección ocurre raramente tras la aplicación de LASIK, aunque hipocorrecciones o regresiones en dichos pacientes resultan a veces inevitables. Tal como el LASIK, reLASIK actúa sobre el estroma corneal, separando previamente el epitelio y la membrana de Bowman, es decir, el flap original, o bien realizando un nuevo flap. Tras la separación se realiza una segunda ablación<sup>8</sup>

<sup>8</sup>Bragheeth, M., Fares, U., Dua, H.. Re-Treatment After Laser In Situ Keratomileusis for Correction of Myopia and Myopic Astigmatism *The British Journal of Ophthalmology*, 92; 11: 1506-1510

## **2.3. Antecedentes empíricos de la investigación**

### **2.3.1. Antecedentes internacionales**

A nivel mundial se ha encontrado los siguientes artículos que hablan sobre aplicaciones bioestadísticas como:

En 2023 (?) en su trabajo “Perfiles composicionales de la fórmula leucocitaria como predictores de enfermedad severa en pacientes con infección por SARS-CoV2” quiso determinar si existía asociación entre la composición de la fórmula leucocitaria al momento del diagnóstico confirmado por infección de SARS-CoV2 al desarrollo de la enfermedad severa, crítica o muerte de pacientes mayores de 18 años, en el cual realizó un agrupamiento no supervisado y ajustó un modelo de regresión logística tomando como variable predictora al grupo identificado por la composición describiendo las variables de acuerdo a su nivel de medición. En la cual se mostró que existe asociación entre la composición de la fórmula leucocitaria al diagnóstico y el estado vital en pacientes con infección por SARS-CoV2. El grupo con mayor riesgo presenta una mayor proporción de neutrófilos y menor proporción de linfocitos.

En 2016 el estudio de (?) “The Analysis of Human Serum Albumin Proteoforms Using Compositional Framework” explicó el marco de análisis de datos composicionales es ideal para explorar y analizar la abundancia relativa de proteoformas medidas mediante ensayos inmunoespectrométricos de masas (MSIA). Se hizo la aplicación de los conceptos explorando la asociación de las modificaciones postraduccionales de la albúmina sérica humana y la función renal en pacientes con diabetes mellitus tipo 2. Concluyendo que los cambios en las proporciones de las proteoformas de albúmina cisteiniladas pueden revelar información sobre el estado de la enfermedad renal crónica en un individuo o indicar problemas con el almacenamiento y el manejo de datos, en el cual el análisis implica que los ensayos MSIA se pueden utilizar para explorar el papel clínico de las modificaciones postraduccionales de una proteína.

### **2.3.2. Antecedentes nacionales**

A nivel nacional, se encontró el siguiente trabajo de investigación:

En 2023 el trabajo de investigación “Algoritmos de machine learning para predecir la anemia en niños de 6 a 35 meses de edad en Cusco” presentado por (?) identificó el algoritmo de Machine Learning más eficiente para predecir la anemia en niñas y niños de 6 a 35 meses de edad a partir de factores sociodemográficos, biológicos, etc. en Cusco utilizando la Encuesta Demográfica y Discovery in Databases (KDD). Evaluando los algoritmos, destacó a los siguientes: K-Nearest Neighbor (KNN), Support Vector Machine (SVM) y Redes Neuronales siendo estos los que poseen mayor precisión, sensibilidad y especificidad. Finalmente determinó como mejor algoritmo a la máquina de soporte vectorial o Support Vector Machine (SVM) que poseó una sensibilidad del 100 %.

## **3. Hipótesis y variables**

### **3.1. Hipótesis**

#### **3.2. Hipótesis general**

Los perfiles composicionales de los pacientes sometidos a cirugía refractiva Laser In Situ Keratomileusis se asocian con respecto a si el paciente será sometido a un Retratamiento LASIK.

#### **3.3. Hipótesis específicas**

- Las variables asociadas a un Retratamiento LASIK son la refracción a ocurrir y datos demográficos.
- Los componentes agruparán las características clínicas, demográficas entre otros de los pacientes sometidos a cirugía refractiva LASIK.

- Los componentes explicarán una mejor regresión con respecto a la predicción de realizarse un Retratamiento LASIK.

### 3.4. Identificación de variables e indicadores

### 3.5. Operacionalización de variables

## 4. Metodología

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

1. **Localización política:** El estudio se desarrollará en el departamento del Cusco ubicada en la división política de Perú. Los pacientes que serán tomados en cuenta para el estudio pueden ser de la localización política así como de otras localizaciones. Otros departamentos de Perú, y así como a nivel internacional.
2. **Localización geográfica:** Departamento de Cusco que posee sus características geográficas propias, como una superficie de 71987 km<sup>2</sup>, ubicada en la parte sur-oriental del territorio nacional y limitada con los departamentos de Junín y Ucayali por el norte, Madre de Dios y Puno por el este, Arequipa por el sur-oeste y, Apurímac y Ayacucho por el oeste. La ciudad capital de Cusco está ubicada a 3339 m.s.n.m.<sup>9</sup>

### 4.2. Tipo y nivel de investigación

#### 4.2.1. Tipo de investigación

El método a desarrollar es de investigación no experimental, en el cual solo observará los resultados de los pacientes sometidos a cirugía refractiva LASIK para poder ser analizados. Además que es una investigación transversal porque los datos fueron recogidos de cirugías refractiva LASIK anteriores realizadas.

<sup>9</sup>Caracterización del departamento de Cusco. *Banco Central de Reserva del Perú Sucursal Cusco*, recuperado de <https://www.bcrp.gob.pe/docs/Sucursales/Cusco/cusco-caracterizacion.pdf>

#### **4.2.2. Nivel de investigación**

El nivel de esta investigación es exploratorio - descriptivo y correlacional. Esto debido a que se aplicará técnicas de análisis multivariado con respecto a la cirugía refractiva como lo es el análisis de componentes principales, también descriptivo ya que se explicarán los resultados de estos componentes y por último correlacional ya que evaluará la asociación o relación que posean estos componentes hacia la posibilidad de tener que hacerse un retratamiento LASIK.

#### **4.3. Unidad de análisis**

Todos los pacientes sometidos a cirugía refractiva Laser In Situ Keratomileusis “LASIK” incluidos aquellos que se sometieron a un Retratamiento LASIK.

#### **4.4. Población de estudio**

Pacientes del Centro de Salud Integral “La Fuente” que fueron aptos y sometidos a cirugía refractiva Laser In Situ Keratomileusis “LASIK”

#### **4.5. Técnicas de recolección de información**

Las técnicas para la recolección de datos serán los instrumentos mecánicos o electrónicos que se obtuvieron de equipos como PENTACAM y MS-39 que miden las variables del examen oftalmológico mencionados anteriormente, además de que se usará la revisión de las historias clínicas de los pacientes sometidos a cirugía refractiva LASIK

#### **4.6. Técnicas de análisis e interpretación de la información**

Se utilizarán técnicas de análisis multivariado estadístico como lo es el análisis de componentes principales, seguidamente se realizará un análisis descriptivo de las variables estudiadas así como la de los componentes asociado y por último se realizará un análisis correlacional al

momento de evaluar la asociación entre los componentes hacia la posibilidad de realizarse un Retratamiento LASIK.

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

- Análisis de componentes principales
- Análisis de regresión o clasificación (generalmente la regresión logística)
- Pruebas de normalidad y test de comparación entre poblaciones paramétricas o no paramétricas



## 5. Presupuesto

RUBROS	COSTO(S/.)	TOTAL (S/.)
<b>A) Recursos humanos:</b>		
Consultoria	2000.00	
Papel bond A4 80 <sup>0</sup>	100.00	
USB	60.00	
CD	20.00	
Plumones cargador de plumones	50.00	
Papel Bulking	20.00	
Corrector	10.00	
Lapiceros	30.00	
Adquisición de bibliografía	600.00	2 790
<b>B) SERVICIOS:</b>		
Movilidad	700.00	
Tinta de impresora	240.00	
Impresión	150.00	
Empastados	300.00	
Copias	200.00	
Anillados	100.00	
Imprevistos	400.00	1 990
	<b>TOTAL</b>	<b>4 980</b>

Cuadro 1: Presupuesto.

## 6. Cronograma de la realización de la tesis

Actividades	Año-2024									
	1	2	3	4	5	6	7	8	9	10
Revisión Bibliográfica	X									
Proyecto de Tesis		X	X	X	X	X	X	X	X	X
Capítulo 1		X	X	X	X	X				
Capítulo 2				X	X	X				
Capítulo 3					X	X	X			
Conclusiones						X	X	X	X	
Presentación y Revisión										X

Cuadro 2: Actividades del año 2024.

## **7. Anexos**

### **7.1. Matriz de consistencia**