

UNIVERSIDAD DE PANAMÁ
VICERRECTORÍA DE INVESTIGACIÓN Y POSTGRADO
PROGRAMA CENTROAMERICANO DE MAESTRÍA EN MATEMÁTICA

ANÁLISIS DE LAS COMPONENTES PRINCIPALES Y
SUS APLICACIONES

GONZALO ISAAC CARRASCO OBERTO

TESIS PRESENTADA COMO UNO DE LOS REQUISITOS PARA OPTAR
AL GRADO DE MAESTRO EN CIENCIAS CON ESPECIALIZACIÓN EN
ESTADÍSTICA MATEMÁTICA

PANAMÁ, REPÚBLICA DE PANAMÁ

1996

T.M



UNIVERSIDAD DE PANAMA

FACULTAD DE CIENCIAS NATURALES Y EXACTAS

Programa Centroamericano de Maestría en Matemática

27 NOV 1996

Aprobado por:

Aurora Mejia

PROF. AURORA MEJIA
Directora de Tesis

ms

PROF. GLADYS SEGURA
Miembro del Jurado

Manuel A. Tejada S.

PROF. MANUEL A. TEJADA S.
Miembro del Jurado

Fecha:

6 de Nov. de 1996

Segura

945682

DEDICATORIA

DEDICATORIA

Quiero dedicar este trabajo:

- A mis padres **Julia** y **Gonzalo** y a mis hermanos **Julia Edmeé** y **Armando** quienes con su comprensión y voz alentadora permitieron que alcanzara mi meta.
- A los profesores **Aurora Mejía** y **Julio Vallarino** por sus enseñanzas e infinita paciencia.

AGRADECIMIENTO

AGRADECIMIENTO

- **A nuestro Padre Dios y nuestra Madre Santísima por permitirnos la culminación exitosa de nuestro trabajo.**
- **A mis amigos: Carmen Cecilia, Elisa, Mitzi y Daniel por su gran apoyo brindado para la culminación de este trabajo.**
- **A la Dirección del Departamento de Estadística por su gran apoyo brindado para la culminación de este trabajo.**
- **Al Dr. César Villarreal por su cooperación recibida a través del Laboratorio de Investigación Erich Graetz.**
- **A todas aquellas personas que de una manera u otra hicieron posible la culminación de este trabajo.**

ÍNDICE GENERAL

INDICE

	Página
RESUMEN	1
INTRODUCCIÓN	2
 CAPTULO I: COMPONENTES PRINCIPALES	
Matrices de las componentes principales	5
Distribución de la matriz de las componentes principales	6
Propiedades de las componentes principales	8
Descomposición espectral de la matriz de covarianza Σ	11
Matriz de observaciones estandarizadas	11
Distribución de la matriz de las observaciones estandarizadas	12
Matriz de las componentes principales de observaciones estandarizadas	15
Distribución de la matriz de las componentes principales estandarizadas	16
Propiedades de las Componentes Principales estandarizadas	18
Matriz de covarianza muestral	21
Estimadores de máxima verosimilitud de los valores y vectores propios	21
Componentes principales de una matriz de datos con $X_{p \times n}$ matriz de covarianza de estructura especial	24
Componentes principales de la matriz de observaciones $X_{p \times n}$ con matriz de covarianza $\Sigma = \sigma^2 \rho$	26
 CAPTULO II: DISTRIBUCIÓN ASINTOTICA DE LOS VALORES PROPIOS	
Transformación de una matriz aleatoria $M \sim W_p(\Sigma, n)$	28
Relación entre la distribución Wishart Centrada y la distribución Ji-Cuadrada	30
Suma de matrices que siguen una distribución Wishart Centrada	31
Distribución de la matriz de covarianza muestral	34
Distribución de los valores propios	35

CAPITULO III: INTERPRETACIÓN DE LAS COMPONENTES PRINCIPALES

Proporción de la variabilidad total explicada por la k-ésima componente principal	40
Eficiencia del ajuste	41
Criterio gráfico	43
Criterio del promedio	43
Correlación entre las variables originales y las componentes principales	44
Coefficiente de determinación	46

CAPITULO IV: ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

Valores promedio y desviación estándar de cada una de las variables estudiadas en el estudio del pez <i>Dormitator latifrons</i> (Richardson, 1837) en la localidad de Mata de Corozo, Corregimiento de Pacora en los años 1992 y 1993	48
Matriz de covarianza	48
Matriz de correlación	48
Valores propios de la matriz de correlación	48
Proporción de la varianza total explicada por cada una de las componentes principales al usar la matriz de correlación	48
Criterio de selección	49
Vectores propios de la matriz de correlación	49
Proporción de la varianza original explicada por cada componente principal	49
Correlación de las variables originales con las componentes principales	50
Varianza total de las variables en estudio explicada por las componentes principales	51
Valores promedio y desviación estándar de cada una de las variables estudiadas en el estudio del pez <i>Dormitator latifrons</i> (Richardson, 1837) en la localidad de Quebrada Carrasquilla del Corregimiento de Panamá Viejo en los años 1992 y 1993	51
Matriz de covarianza	51
Matriz de correlación	52

Valores propios de la matriz de correlación	52
Proporción de la varianza total explicada por cada una de las componentes principales al usar la matriz de correlación	52
Criterio de selección	52
Vectores propios de la matriz de correlación	52
Proporción de la varianza original explicada por cada componente principal	52
Correlación de las variables originales con las componentes principales	53
Variación total de las variables en estudio explicada por las componentes principales	53
CONCLUSIONES Y RECOMENDACIONES	70
BIBLIOGRAFIA	74

ÍNDICE DE CUADROS Y FIGURAS

INDICE DE CUADROS

	Página
Cuadro I: Valores promedio y desviación estándar de las variables estudiadas en el pez <i>Dormitator latifrons</i> (Richardson, 1837) en la localidad de Mata de Corozo del Corregimiento de Pacora. Años: 1992-1993	54
Cuadro II: Matriz de Covarianza de la variables estudiadas en el pez <i>Dormitator latifrons</i> (Richardson, 1837) en la localidad de Mata de Corozo del Corregimiento de Pacora. Años: 1992-1993	55
Cuadro III: Matriz de correlación de las variables estudiadas en el pez <i>Dormitator latifrons</i> (Richardson, 1837) en la localidad de Mata de Corozo del Corregimiento de Pacora. Años: 1992-1993	56
Cuadro IV: Valores propios y proporción de la variación total explicada por las componentes principales en el estudio del pez <i>Dormitator latifrons</i> (Richardson, 1837) en la localidad de Mata de Corozo del Corregimiento de Pacora. Años: 1992-1993	57
Cuadro V: Componentes de los vectores propios de la matriz de correlación de las variables estudiadas en el pez <i>Dormitator latifrons</i> (Richardson, 1837) en la localidad de Mata de CorozodeCorregimiento de Pacora.Años: 1992-1993	58
Cuadro VI: Proporción de la variación original explicada por cada componente principal de la matriz de correlación de las variables en estudio del pez <i>Dormitator latifrons</i> (Richardson, 1837) en la localidad de Mata de Corozo del Corregimiento de Pacora. Años: 1992-1993	59

Cuadro VII:	Correlación de las variables originales con las tres primeras componentes principales en el estudio del pez <i>Dormitator latifrons</i> (Richardson, 1837) en la localidad de Mata de Corozo del Corregimiento de Pacora. Años: 1992-1993	60
Cuadro VIII:	Valores promedio y desviación estándar de las variables estudiadas en el estudio del pez <i>Dormitator latifrons</i> (Richardson, 1837) en la localidad Quebrada Carrasquilla del Corregimiento de Panamá Viejo. Años: 1992-1993	61
Cuadro IX:	Matriz de Covarianza de las variables estudiadas en el pez <i>Dormitator latifrons</i> (Richardson, 1837) en la localidad Quebrada Carrasquilla del Corregimiento de Panamá Viejo. Años: 1992-1993	62
Cuadro X:	Matriz de correlación de las variables estudiadas en el pez <i>Dormitator latifrons</i> (Richardson, 1837) en la localidad Quebrada Carrasquilla del Corregimiento de Panamá Viejo. Años: 1992-1993	63
Cuadro XI:	Valores propios y proporción de la variación explicada por las componentes principales en el estudio del pez <i>Dormitator latifrons</i> (Richardson, 1837) en la localidad Quebrada Carrasquilla del Corregimiento de Panamá Viejo. Años: 1992-1993	64
Cuadro XII:	Componentes de los vectores propios de la matriz de correlación de las variables estudiadas en el pez <i>Dormitator latifrons</i> (Richardson, 1837) en la localidad Quebrada Carrasquilla del Corregimiento de Panamá Viejo. Años: 1992-1993	65

Cuadro XIII: Proporción de la variación original explicada por cada componente principal de la matriz de correlación de las variables en estudio del pez *Dormitator latifrons* (Richardson, 1837) en la localidad Quebrada Carrasquilla del Corregimiento de Panamá Viejo. Años: 1992-1993 66

Cuadro XIV: Correlación de las variables originales con las tres primeras componentes principales en el estudio del pez *Dormitator latifrons* (Richardson, 1837) en la localidad Quebrada Carrasquilla del Corregimiento de Panamá Viejo. Años: 1992-1993 67

INDICE DE FIGURAS

	Página
Figura 1: Proporción de la variación total explicada por las componentes principales.	68
Figura 2: Proporción de la variación total explicada por las componentes principales	69

En este trabajo definimos las componentes principales de una matriz de datos multivariados. Se obtuvo la distribución de las mismas, demostrando algunas de sus propiedades. De igual manera se hace para las componentes principales estandarizadas. Se obtuvieron los estimadores de máxima verosimilitud de los valores y vectores propios y las componentes principales de la matriz de observaciones $X_{p \times n}$ con matriz de covarianza de estructura especial. Definimos la distribución Wishart Centrada. Se demostraron algunos teoremas relacionados a dicha distribución y se logró obtener la distribución asintótica de los valores propios. En este trabajo, además se presenta la variación explicada por las componentes principales; y los criterios para seleccionar el número de componentes a considerar en algunos estudios, la correlación entre las variables originales y las componentes principales, y el coeficiente de determinación. Basado en las suposiciones teórica arriba expuestas efectuamos un análisis de las componentes principales en el estudio biológico del *Dormitator latifrons* (pez), años de 1992 y 1993. Datos suministrados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas.

ABSTRACT

In the present work define the main components of a multivariate matrix data and its distribution; and we demonstrate as well some of its properties. The same has been done for the main standardized components. We obtained the maximum likelihood estimation of the proper vectors and values, and the main $X_{p \times n}$ observation matrix components with a covariance matrix of special structure. We defined the Centered Wishart distribution; also we demonstrated some related theorems of this distribution, obtaining the asymptotic distribution of the proper values. In this research, we introduced the variation explained by the main components and the criterion to select the numbers of components under consideration in some studies, also the correlation between the original variables and the main components; and the determinative coefficients. Based in the theoretical assumption above explained we made an analysis of the main components of a large biological research data of *Dormitator latifrons* (Pisces) provided by the Laboratory Erich Graetz of the Department of Animal Physiology and Behavior of the of Faculty Natural Sciences.

INTRODUCCIÓN

Es común encontrar muchas variables explicatorias o independientes. A veces es útil tener un gran número de ellas, pero también se corre el riesgo de duplicar la información. La duplicidad puede ser detectada a través de las correlaciones.

Una forma de confirmar la correlación, es estimar la matriz de correlación de todas las variables explicatorias y observar las interrelaciones entre cada dos variables. Existen interrelaciones entre más de dos variables, que no pueden ser detectadas por este método y se requiere del análisis de la estructura de los datos, utilizando técnicas como la descomposición vectorial. Para eliminar el efecto de la correlación de las variables, es necesario formar variables ortogonales (no correlacionadas) previamente a cualquier tipo de análisis (regresión, conglomerados, etc.). Una manera útil de lograr este tipo de combinaciones lineales ortogonales es la técnica de componentes principales.

Hotelling en 1933 fue el primero en formular el análisis de componentes principales, basándose en el trabajo publicado en 1901 por Karl Pearson sobre el ajuste de un multiespacio a una línea o a un

plano. El enfoque de Pearson se centra en el análisis de componentes principales que sintetiza la mayor variabilidad del sistema de puntos; ello explica el calificativo de "principal".

El análisis de componentes principales es un método que examina la dependencia estructural de datos multivariados obtenidos de una población, cuya distribución de probabilidades no es preciso conocer. Sin embargo, puede suponerse que la población mostrada tiene distribución multinormal, con lo que se podrían realizar las respectivas pruebas de hipótesis para extraer inferencias de la población en estudio.

Con el análisis de componentes principales se pretende generar nuevas variables que expresen la mayor parte de la información contenida en el conjunto original, reducir el número de variables para una mejor interpretación de los datos, y eliminar aquellos que aporten escasa información.

CAPÍTULO I.
COMPONENTES PRINCIPALES

COMPONENTES PRINCIPALES

Para n vectores X_α de orden $p \times 1$ distribuidos normalmente con media μ_α y matriz de covarianza Σ , diremos que la matriz $X_{p \times n} = (X_1, X_2, \dots, X_n)$ tiene una distribución normal con matriz de media $\Gamma_{p \times n} = (\mu_1, \mu_2, \dots, \mu_n)$ y matriz de covarianza $V \otimes \Sigma$ donde V es una matriz simétrica definida positiva de orden $n \times n$, tal que $\text{Cov}(X_\alpha, X_\beta) = v_{\alpha\beta} \Sigma$ y $V \otimes \Sigma$ denota el producto de Kronecker de la matriz V y la matriz Σ .

Lo anterior lo denotaremos por:

$$X \sim N(\Gamma, V \otimes \Sigma)$$

Si $V=I$ las columnas de X son mutuamente independientes.

DEFINICIÓN I.1. (Matriz de las Componentes Principales)

Sea $X_{p \times n}$ una matriz de datos multivariados donde para cada $1 \leq \alpha \leq n$, cada columna X_α de $X_{p \times n}$ corresponde al α -ésimo vector aleatorio de orden $p \times 1$ y L una matriz ortonormal donde cada $1 \leq i \leq p$, cada columna l_i corresponde a un vector propio de orden $p \times 1$ de Σ

y tal que $L'\Sigma L = \Lambda = \begin{pmatrix} \lambda_1 & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & & \lambda_p \\ & & & & 0 \end{pmatrix}$ con $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$

valores propios de Σ .

Definiremos la matriz de componentes principales como la matriz $Y_{p \times n} = L'X$, donde para cada $1 \leq \alpha \leq n$, las columnas $Y_\alpha = L'X_\alpha$ de la matriz Y , son vectores de orden $p \times 1$ que representan las componentes principales correspondientes al α -ésimo vector aleatorio X_α .

Además, definimos la i -ésima componente principal correspondiente al vector Y_α por:

$$Y_{i\alpha} = l_{i\alpha}' X_\alpha = \sum_{j=1}^p l_{ji} X_{j\alpha}$$

PROPOSICIÓN I.1:

Si $X \sim N(\Gamma, I \otimes \Sigma)$ y $L_{p \times p}$ la matriz ortonormal de la definición I.1.

entonces, $Y = L'X \sim N(L'\Gamma, I \otimes L'\Sigma L)$

Demostración:

Sean $X \sim N(\Gamma, I \otimes \Sigma)$ y $L_{p \times p}$ la matriz ortonormal de la definición I.1.

Como $X \sim N(\Gamma, I \otimes \Sigma)$ y $Y = L' X$, entonces para todo $1 \leq \alpha \leq n$

$Y_\alpha = L' X_\alpha$, representa el α -ésimo vector de orden $p \times 1$ de Y .

Además, para $1 \leq \alpha \leq n$ se tendrá;

$$\begin{aligned} E(Y_\alpha) &= E(L' X_\alpha) \\ &= L' E(X_\alpha) \\ &= L' \mu_\alpha \end{aligned}$$

Luego, $E(Y) = L' \Gamma$

Por otro lado, para todo $1 \leq \alpha, \beta \leq n$ tendremos que:

$$\begin{aligned} \text{Cov}(Y_\alpha, Y_\beta) &= E\{(Y_\alpha - L' \mu_\alpha)(Y_\beta - L' \mu_\beta)'\} \\ &= E\{(L' X_\alpha - L' \mu_\alpha)(L' X_\beta - L' \mu_\beta)'\} \\ &= E\{L'(X_\alpha - \mu_\alpha)(X_\beta - \mu_\beta)'L\} \\ &= L'E\{(X_\alpha - \mu_\alpha)(X_\beta - \mu_\beta)'\}L \\ &= \delta_{\alpha\beta} L'\Sigma L \end{aligned}$$

donde $\delta_{\alpha\beta}$ es el delta de Kronecker

Luego, si $1 \leq \alpha, \beta \leq n$ entonces,

$$\text{COV}(Y_\alpha, Y_\beta) = \begin{cases} L' \Sigma L & \text{si } \alpha = \beta \\ 0 & \text{si } \alpha \neq \beta \end{cases}$$

Por lo tanto la matriz de covarianza de Y es; $I \otimes L' \Sigma L$

Como X se distribuye normal, entonces $Y = L'X$ también se distribuye normal, con media $L'\Gamma$ y matriz de covarianza $I \otimes L' \Sigma L$, esto es $Y \sim N(L'\Gamma, I \otimes L' \Sigma L)$.

Por lo tanto para cada $1 \leq \alpha \leq n$, los vectores Y_α resultan independientes entre sí, con matriz de covarianza $L' \Sigma L$ y con distribución normal.

Propiedades de las componentes principales:

PROPOSICIÓN 1.2.

Si Y_{pm} es la matriz de componentes principales correspondientes a la matriz X_{pm} , con $Y \sim N(L'\Gamma, I \otimes L' \Sigma L)$, entonces

- (a) $\text{Var}(Y_{i\alpha}) = \lambda_i$ es el i -ésimo valor propio de Σ
- (b) $\text{Var}(Y_{1\alpha}) \geq \text{Var}(Y_{2\alpha}) \geq \dots \geq \text{Var}(Y_{p\alpha}) \geq 0$
- (c) para $1 \leq i, j \leq p, i \neq j$ $\text{Cov}(Y_{i\alpha}, Y_{j\alpha}) = 0$
- (d) $\sum_{i=1}^p \text{V}(Y_{i\alpha}) = \text{traza}(\Sigma)$
- (e) $\text{Cov}(X_\alpha, Y_\alpha) = \Sigma L$

Demostración:

Por la proposición I.1, para cada $1 \leq \alpha \leq n$ la matriz de covarianza de los vectores Y_α es $\Lambda = L'\Sigma L$.

Como L es una matriz ortonormal y los valores propios de Σ están definidos como las soluciones de la ecuación característica $|\Sigma - \lambda I| = 0$.

$$\begin{aligned} \text{Como } |\Sigma - \lambda I| &= |\Sigma - \lambda I| |L' L| \\ &= |L'| |\Sigma - \lambda I| |L| \\ &= |L'\Sigma L - \lambda L'L| \\ &= |L'\Sigma L - \lambda I| \end{aligned}$$

y por hipótesis, la matriz de covarianza de las componentes principales correspondiente al vector Y_α es $L'\Sigma L = \Lambda$, por consiguiente, la varianza de la i -ésima componente principal correspondiente al vector Y_α es $V(Y_{i\alpha}) = \lambda_i$ para cada $1 \leq i \leq p$, lo que demuestra (a).

Luego, $\text{Var}(Y_{1\alpha}) \geq \text{Var}(Y_{2\alpha}) \geq \dots \geq \text{Var}(Y_{p\alpha}) \geq 0$ lo que verifica (b).

Además, las componentes principales correspondientes al vector Y_α no están correlacionados, con lo cual para $1 \leq i, j \leq p$ $i \neq j$, $\text{Cov}(Y_{i\alpha}, Y_{j\alpha}) = 0$, lo que muestra (c).

Como $L'\Sigma L = \Lambda$, entonces

$$\begin{aligned} \text{traza} (L'\Sigma L) &= \text{traza} (\Lambda) \\ &= \sum_{j=1}^p \lambda_j \\ &= \sum_{j=1}^p V(Y_{j\alpha}) \end{aligned}$$

Por otro lado

$$\begin{aligned} \text{traza} (L'\Sigma L) &= \text{traza} (\Sigma LL') \\ &= \text{traza} (\Sigma I) \\ &= \text{traza} (\Sigma) \end{aligned}$$

De aquí que

$$\sum_{j=1}^p V(Y_{j\alpha}) = \text{traza} (\Sigma)$$

lo que verifica (d)

Por último, para cada $1 \leq \alpha \leq n$ sea $Y_\alpha = L'X_\alpha$ la α -ésima columna de Y y X_α la α -ésima columna de la matriz X . Luego, la matriz de covarianza entre X_α y Y_α es:

$$\begin{aligned} \text{Cov} (X_\alpha, Y_\alpha) &= E\{(X_\alpha - \mu_\alpha)(Y_\alpha - L'\mu_\alpha)'\} \\ &= E\{(X_\alpha - \mu_\alpha)(L'X_\alpha - L'\mu_\alpha)'\} \\ &= E\{(X_\alpha - \mu_\alpha)(X_\alpha - \mu_\alpha)'\}L \\ &= \Sigma L \end{aligned}$$

lo cual demuestra (e).

PROPOSICIÓN I.3.

Sea $\Sigma_{p \times p}$ una matriz simétrica con valores propios λ_i para cada $1 \leq i \leq p$ y L la matriz ortonormal cuyas columnas l_i son los vectores

propios de Σ entonces,
$$\sum_{i=1}^p \lambda_i l_i l_i' = \lambda_1 l_1 l_1' + \lambda_2 l_2 l_2' + \dots + \lambda_p l_p l_p'$$

representa la descomposición espectral de Σ .

Demostración:

De la definición I.1. se tiene

$$L' \Sigma L = \Lambda$$

entonces,
$$\Sigma L = L \Lambda$$

con lo cual,
$$\Sigma = L \Lambda L'$$

de donde,
$$\Sigma = \lambda_1 l_1 l_1' + \lambda_2 l_2 l_2' + \dots + \lambda_p l_p l_p'$$

y así
$$\Sigma = \sum_{j=1}^p \lambda_j l_j l_j'$$

DEFINICIÓN I.2. (Matriz de Observaciones Estandarizadas)

Sean $X_{p \times n} \sim N(\Gamma, I \otimes \Sigma)$, $(X - \Gamma)$ la matriz de las desviaciones y $V_{p \times p}^{1/2}$ la matriz diagonal donde los elementos de la diagonal principal

son las raíces de las varianzas pertenecientes a la matriz de covarianza Σ .

Definimos la matriz de observaciones estandarizadas como la matriz $Z_{p \times m} = (V^{1/2})^{-1}(X - \Gamma)$.

PROPOSICIÓN 1A:

Si $X_{p \times m} \sim N(\Gamma, I \otimes \Sigma)$ y

$$\rho = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{pmatrix}$$

donde para $1 \leq i, j \leq p, i \neq j$

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}} \sqrt{\sigma_{jj}}}$$

entonces $Z_{p \times m} \sim N(0, I \otimes \rho)$

Demostración:

Sean $X \sim N(\Gamma, I \otimes \Sigma)$ y para cada $1 \leq \alpha \leq n$ $Z_\alpha = (V^{1/2})^{-1}(X_\alpha - \mu_\alpha)$ el α -ésimo vector aleatorio estandarizado.

Entonces, para cada $1 \leq \alpha \leq n$ se tiene:

$$\begin{aligned}
 E(Z_\alpha) &= E((V^{1/2})^{-1} (X_\alpha - \mu_\alpha)) \\
 &= (V^{1/2})^{-1} E(X_\alpha - \mu_\alpha) \\
 &= (V^{1/2})^{-1} (E(X_\alpha) - \mu_\alpha) \\
 &= (V^{1/2})^{-1} (\mu_\alpha - \mu_\alpha) \\
 &= 0
 \end{aligned}$$

luego, $E(Z) = 0$

Por otro lado, para cada $1 \leq \alpha, \beta \leq n, \alpha \neq \beta$ tenemos;

$$\begin{aligned}
 \text{Cov}(Z_\alpha, Z_\beta) &= E((Z_\alpha - 0)(Z_\beta - 0)') \\
 &= E(Z_\alpha Z_\beta') \\
 &= E(((V^{1/2})^{-1} (X_\alpha - \mu_\alpha))((V^{1/2})^{-1} (X_\beta - \mu_\beta))') \\
 &= (V^{1/2})^{-1} E((X_\alpha - \mu_\alpha)(X_\beta - \mu_\beta)')(V^{1/2})^{-1} \\
 &= \delta_{\alpha\beta} (V^{1/2})^{-1} \Sigma (V^{1/2})^{-1}
 \end{aligned}$$

Como,

$$(V^{1/2})^{-1} \Sigma (V^{1/2})^{-1} = \begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \dots & 0 & \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \dots & 0 & \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\sqrt{\sigma_{pp}}} & \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\sqrt{\sigma_{pp}}} \end{pmatrix}$$

$$= \begin{vmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}} & \dots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}} & \left| \begin{array}{cccc} 1 & 0 & \dots & 0 \\ \sqrt{\sigma_{11}} & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \\ \sqrt{\sigma_{pp}} & \sqrt{\sigma_{pp}} & \dots & \sqrt{\sigma_{pp}} \end{array} \right. \\ \frac{\sigma_{21}}{\sqrt{\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}} & \dots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}} & \\ \dots & \dots & \dots & \dots & \\ \dots & \dots & \dots & \dots & \\ \frac{\sigma_{p1}}{\sqrt{\sigma_{pp}}} & \frac{\sigma_{p2}}{\sqrt{\sigma_{pp}}} & \dots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}} & \end{vmatrix}$$

$$= \begin{vmatrix} 1 & \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} & \dots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}\sigma_{pp}}} \\ \frac{\sigma_{21}}{\sqrt{\sigma_{22}\sigma_{11}}} & 1 & \dots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}\sigma_{pp}}} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \frac{\sigma_{p1}}{\sqrt{\sigma_{pp}\sigma_{11}}} & \frac{\sigma_{p2}}{\sqrt{\sigma_{pp}\sigma_{22}}} & \dots & 1 \end{vmatrix}$$

entonces, para cada $1 \leq \alpha, \beta \leq n$

$$\text{Cov}(Z_\alpha, Z_\beta) = \begin{cases} \rho & \text{si } \alpha = \beta \\ 0 & \text{si } \alpha \neq \beta \end{cases}$$

Luego la matriz de covarianza de Z es $I \otimes \rho$. Como X se distribuye normal, entonces $Z = (V^{1/2})^{-1} (X - \Gamma)$, también se distribuye normal, con matriz de media cero (0) y matriz de covarianza $I \otimes \rho$, esto

es, $Z \sim N(0, I \otimes \rho)$, con lo cual los vectores Z_α son independientes entre si, su matriz de covarianza es ρ y con distribución normal.

DEFINICIÓN 1.3. (Matriz de las componentes principales de observaciones estandarizadas)

Sea $Z_{pxn} \sim N(0, I \otimes \rho)$ y B_{pxp} la matriz ortonormal en donde para cada $1 \leq i \leq p$ la columna b_i corresponde a un vector propio de ρ de orden $p \times 1$ y tal que,

$$B' \rho B = \Omega = \begin{pmatrix} \theta_1 & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & & \theta_p \end{pmatrix} \text{ con } \theta_1 \geq \theta_2 \geq \dots \geq \theta_p \geq 0 \text{ valores propios de } \rho.$$

Definimos la matriz de componentes principales estandarizado como la matriz $U_{pxn} = B'Z$.

Para cada $1 \leq \alpha \leq n$ las columnas $U_\alpha = B'Z_\alpha$, de la matriz U son vectores de orden $p \times 1$ que representan las componentes principales estandarizados correspondientes al α -ésimo vector aleatorio estandarizado Z_α .

Definimos además, para cada $1 \leq i \leq p$, el i -ésimo componente principal estandarizado correspondiente al vector U_α , por:

$$U_{i\alpha} = b_i' Z_\alpha = \sum_{j=1}^p b_{ji} Z_{j\alpha}$$

PROPOSICIÓN I.5:

Si $Z_{pxn} \sim N(0, I \otimes \rho)$, B_{pxp} la matriz ortonormal de la definición I.3.

entonces,

$$U = B' Z \sim N(0, I \otimes B' \rho B)$$

Demostración:

Sean $Z \sim N(0, I \otimes \rho)$ y B_{pxp} la matriz ortonormal de la definición I.3.

Como $Z_{pxn} \sim N(0, I \otimes \rho)$ y $U = B'Z$, entonces, para cada $1 \leq \alpha \leq n$,

$U_\alpha = B'Z_\alpha$, representa α -ésimo vector de orden $p \times 1$, de U .

Entonces para cada $1 \leq \alpha \leq n$ se tendrá:

$$\begin{aligned} E(U_\alpha) &= E(B'Z_\alpha) \\ &= BE(Z_\alpha) \\ &= B'0 \\ &= 0 \end{aligned}$$

con lo cual

$$E(U) = 0$$

Por otro lado, para cada $1 \leq \alpha, \beta \leq n$, tenemos:

$$\begin{aligned}
 \text{Cov}(U_\alpha, U_\beta) &= E((U_\alpha - 0)(U_\beta - 0)') \\
 &= E(U_\alpha U_\beta') \\
 &= E(B'Z_\alpha(B'Z_\beta)') \\
 &= E(B'Z_\alpha Z_\beta' B) \\
 &= B'E(Z_\alpha Z_\beta')B \\
 &= \delta_{\alpha\beta} B'\rho B
 \end{aligned}$$

Así, para $1 \leq \alpha, \beta \leq n$

$$\text{Cov}(U_\alpha, U_\beta) = \begin{cases} B'\rho B & \text{si } \alpha = \beta \\ 0 & \text{si } \alpha \neq \beta \end{cases}$$

Por lo tanto, la matriz de covarianza de U es $I \otimes B'\rho B$.

Como Z se distribuye normal, entonces $U = B'Z$, también se distribuye normal con matriz de media cero (0) y matriz de covarianza $I \otimes B'\rho B$, esto es $U \sim N(0, I \otimes B'\rho B)$.

Por lo anterior, para cada $1 \leq \alpha \leq n$, los vectores U_α son independientes entre sí, su matriz de covarianza es $B'\rho B$ y con distribución normal.

**PROPIEDADES DE LAS COMPONENTES PRINCIPALES
ESTANDARIZADOS**

PROPOSICIÓN 1.6.

Si $U_{p \times m}$, es la matriz de componentes principales de las observaciones estandarizadas de la matriz $X_{p \times m}$, y $U \sim N(0, I \otimes B' \rho B)$, entonces,

(a) $V(U_{i\alpha}) = \theta_i$, donde θ_i es del i -ésimo valor propio de ρ .

(b) $V(U_{1\alpha}) \geq V(U_{2\alpha}) \geq \dots \geq V(U_{p\alpha}) \geq 0$

(c) $\text{Cov}(U_{i\alpha}, U_{j\alpha}) = 0 \quad i \neq j$

(d) $\sum_{i=1}^p V(U_{i\alpha}) = \text{traza}(\rho)$

(e) $\text{Cov}(Z_\alpha, U_\alpha) = \rho B$

Demostración:

Sabemos que para cada $1 \leq \alpha \leq n$, la matriz de covarianza de los vectores U_α es $B' \rho B$.

Como B es una matriz ortonormal y los valores propios de ρ están definidos como las soluciones de la ecuación característica $|\rho - \theta I| = 0$ y

$$\begin{aligned}
|\rho - \theta \mathbf{I}| &= |\rho - \theta \mathbf{I}| |\mathbf{B}'\mathbf{B}| \\
&= |\mathbf{B}'| |\rho - \theta \mathbf{I}| |\mathbf{B}| \\
&= |\mathbf{B}'\rho\mathbf{B} - \theta\mathbf{B}'\mathbf{B}| \\
&= |\mathbf{B}'\rho\mathbf{B} - \theta\mathbf{I}|
\end{aligned}$$

Como la matriz de covarianza de las componentes principales correspondiente al vector U_α es $\mathbf{B}'\rho\mathbf{B}$ donde $\mathbf{B}'\rho\mathbf{B} = \Omega$, tendremos que para todo $1 \leq i \leq p$ la varianza del i -ésimo componente principal estandarizado correspondiente al vector U_α es $V(U_{i\alpha}) = \theta_i$ lo que demuestra (a).

Luego, $V(U_{1\alpha}) \geq V(U_{2\alpha}) \geq \dots \geq V(U_{p\alpha}) \geq 0$ lo que verifica (b).

Además las componentes principales correspondientes al vector U_α no están correlacionados, con lo cual para $1 \leq i, j \leq p$ $i \neq j$, $\text{Cov}(U_{i\alpha}, U_{j\alpha}) = 0$ lo que demuestra (c).

Dado que $\mathbf{B}'\rho\mathbf{B} = \Omega$ entonces,

$$\begin{aligned}
\text{traza} (\mathbf{B}'\rho\mathbf{B}) &= \text{traza} (\Omega) \\
&= \sum_{j=1}^p \theta_j \\
&= \sum_{j=1}^p V(U_{j\alpha})
\end{aligned}$$

Por otro lado,

$$\begin{aligned} \text{traza } (\mathbf{B}'\rho\mathbf{B}) &= \text{traza } (\rho\mathbf{B}\mathbf{B}') \\ &= \text{traza } (\rho\mathbf{I}) \\ &= \text{traza } (\rho) \end{aligned}$$

Luego,

$$\sum_{i=1}^p \mathcal{V}(U_{i\alpha}) = \text{traza } (\rho)$$

lo que demuestra (d).

Finalmente, para cada $1 \leq \alpha \leq n$ sea $U_\alpha = \mathbf{B}'\mathbf{Z}_\alpha$ el vector correspondiente a la α -ésima columna de la matriz \mathbf{U} y \mathbf{Z}_α el vector aleatorio correspondiente a la α -ésima columna de la matriz \mathbf{Z} .

Entonces, la matriz de covarianza entre \mathbf{Z}_α y U_α es:

$$\begin{aligned} \text{Cov } (\mathbf{Z}_\alpha, U_\alpha) &= \mathbf{E}((\mathbf{Z}_\alpha)(U_\alpha)') \\ &= \mathbf{E}(\mathbf{Z}_\alpha (\mathbf{B}\mathbf{Z}_\alpha)') \\ &= \mathbf{E}(\mathbf{Z}_\alpha \mathbf{Z}_\alpha') \mathbf{B} \\ &= \rho \mathbf{B} \end{aligned}$$

lo que verifica (e)

**ESTIMADORES DE MÁXIMA VEROSIMILITUD
DE LOS VALORES Y VECTORES PROPIOS
DE LA MATRIZ DE COVARIANZA Σ**

DEFINICIÓN I.4. (Matriz de Covarianza Muestral)

Sean X_1, X_2, \dots, X_n n vectores de orden $p \times 1$ que representan una muestra aleatoria.

Definimos la matriz de covarianza muestral como la matriz:

$$S = \frac{1}{n-1} \sum_{\alpha=1}^n (X_{\alpha} - \bar{X})(X_{\alpha} - \bar{X})'$$

PROPOSICIÓN I.7.

Sean, X_1, X_2, \dots, X_n $n > p$ observaciones de $N(\mu, \Sigma)$ donde $\Sigma_{p \times p}$ es una matriz con valores propios $\lambda_1, \lambda_2, \dots, \lambda_p$ y vectores propios asociados l_1, l_2, \dots, l_p , $S_{p \times p}$ la matriz de covarianza muestral, W la matriz ortonormal donde, para cada $1 \leq i \leq p$ cada columna W_i corresponde a

un vector propio de S y $W'SW = D = \begin{pmatrix} \hat{\lambda}_1 & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & & \hat{\lambda}_p \end{pmatrix}$ con

$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p \geq 0$ valores propios de S y

$$\hat{\Sigma} = \frac{1}{n} \sum_{\alpha=1}^n (X_{\alpha} - \bar{X})(X_{\alpha} - \bar{X})' \quad (1)$$

$$\hat{\Sigma} = \frac{1}{n} (n-1) S$$

el estimador de máxima verosimilitud de Σ .

Entonces, las soluciones k_1, k_2, \dots, k_p de la ecuación característica $|\hat{\Sigma} - k_i I| = 0$ forman un conjunto de estimadores de máxima verosimilitud de $\lambda_1, \lambda_2, \dots, \lambda_p$ y los vectores propios asociados w_1, w_2, \dots, w_p que verifican

$$(\hat{\Sigma} - k_i I) w_i = 0$$

$$w_i' w_i = 1$$

forman un conjunto de estimadores de máxima verosimilitud de $\lambda_1, \lambda_2, \dots, \lambda_p$.

Demostración:

Sea $S_{p \times p}$ la matriz de covarianza muestral, con $\lambda_1, \lambda_2, \dots, \lambda_p$ valores propios y $W_{p \times p}$ la matriz ortonormal donde para cada $1 \leq i \leq p$ la columna w_i corresponde a un vector propio de S

$$W'SW = D = \begin{pmatrix} \hat{\lambda}_1 & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & & \hat{\lambda}_p \end{pmatrix} \quad \text{con } \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p \geq 0$$

(1) ANDERSON (1958)

y además, los valores propios de $\hat{\Sigma}$ están definidos como las soluciones de la ecuación característica $|\hat{\Sigma} - kI| = 0$ y como

$$\begin{aligned} |\hat{\Sigma} - kI| &= \left| \frac{n-1}{n} S - kI \right| |W'W| \\ &= \left| \frac{n-1}{n} W'SW - kW'W \right| \\ &= \left| \frac{n-1}{n} W'SW - kI \right| \\ &= \left| \frac{n-1}{n} D - kI \right| \end{aligned}$$

entonces, para cada $1 \leq i \leq p$ la i -ésima raíz característica de $\hat{\Sigma}$, será:

$$k_i = \frac{n-1}{n} \hat{\lambda}_i \quad \text{donde } k_1 \geq k_2 \geq \dots \geq k_p \geq 0,$$

luego, para cada $1 \leq i \leq p$ el estimador de máxima verosimilitud de

$$\lambda_i \quad \text{es} \quad k_i = \frac{n-1}{n} \hat{\lambda}_i$$

Como para cada $1 \leq i \leq p$ $k_i = \frac{n-1}{n} \hat{\lambda}_i$, es un valor propio de

$\hat{\Sigma}$ y $w_i \neq 0$ un vector propio ortonormal asociado, entonces para cada $1 \leq i \leq p$ satisface

$$(\hat{\Sigma} - k_i I) w_i = 0$$

$$\text{y} \quad w_i' w_i = 1,$$

Además como $\hat{\Sigma}$ es un estimador de máxima verosimilitud de Σ , tendremos que, para cada $1 \leq i \leq p$ w_i es un estimador de máxima verosimilitud de l_i .

**COMPONENTES PRINCIPALES DE UNA MATRIZ DE DATOS
CON MATRIZ DE COVARIANZA DE ESTRUCTURA ESPECIAL**

PROPOSICIÓN I.8.

Si $X_{pxm} \sim N(\Gamma, I \otimes \Sigma)$, donde Σ_{pxp} es una matriz simétrica definida positiva, entonces existe una transformación $P = CX$, tal que $P \sim N(C\Gamma, I \otimes D)$ donde D es una matriz diagonal, y además

$$Y = R'P \sim N(R'\Gamma, I \otimes R'DR), \text{ donde } T = C\Gamma.$$

Demostración:

Sea $X \sim N(\Gamma, I \otimes \Sigma)$ donde Σ_{pxp} es una matriz simétrica definida positiva luego, existe una matriz C_{pxp} ortogonal, tal que $C\Sigma C' = D$ es una matriz diagonal.

Consideremos la transformación $P = CX$

Como $X \sim N(\Gamma, I \otimes \Sigma)$ entonces $P \sim N(C\Gamma, I \otimes D)$.

Para cada $1 \leq \alpha \leq n$, $Y_\alpha = R'P_\alpha$ la α -ésima componente principal correspondiente a la α -ésima columna de la matriz aleatoria P , donde R_{pxp} representa la matriz ortonormal de los vectores propios de la matriz diagonal D .

Luego, para cada $1 \leq \alpha \leq n$, se tendrá;

$$\begin{aligned} E(Y_\alpha) &= E(R'P_\alpha) \\ &= R'E(P_\alpha) \\ &= R't_\alpha \quad \text{donde,} \quad t_\alpha = C\mu_\alpha \end{aligned}$$

$$\text{Luego } E(Y) = R'T \quad \text{donde} \quad T = CT$$

Ahora, para $1 \leq \alpha, \beta \leq n$ tendremos;

$$\begin{aligned} \text{Cov}(Y_\alpha, Y_\beta) &= E\{(Y_\alpha - R't_\alpha)(Y_\beta - R't_\beta)'\} \\ &= E\{(R'P_\alpha - R't_\alpha)(R'P_\beta - R't_\beta)'\} \\ &= E\{(R'P_\alpha - R't_\alpha)(P_\beta - t_\beta)'R\} \\ &= R'E\{(P_\alpha - t_\alpha)(P_\beta - t_\beta)\}R \\ &= \delta_{\alpha\beta} R'DR \end{aligned}$$

Luego, para $1 \leq \alpha, \beta \leq n$

$$\text{COV}(Y_\alpha, Y_\beta) = \begin{cases} R'DR & \text{si } \alpha = \beta \\ 0 & \text{si } \alpha \neq \beta \end{cases}$$

Por lo tanto, la matriz de covarianza de Y es $I \otimes R'DR$.

Como P se distribuye normal entonces $Y = R'P$, también se distribuye normal, con media $R'T$ y matriz de covarianza $I \otimes R'DR$ esto es:

$$Y \sim N(R'T, I \otimes R'DR)$$

PROPOSICIÓN 1.9:

Si $X_{pm} \sim N(\Gamma, I \otimes \Sigma)$ donde $\Sigma_{pp} = \sigma^2 \rho$, donde ρ es la matriz de correlación. Entonces $Y = L'X \sim N(L'\Gamma, I \otimes \sigma^2 L'\rho L)$.

Demostración:

Sabemos que $E(Y_\alpha) = L'\mu_\alpha$ y $E(Y) = L'\Gamma$

$$\begin{aligned} \text{Además Cov } (Y_\alpha, Y_\beta) &= E\{(Y_\alpha - L'\mu_\alpha)(Y_\beta - L'\mu_\beta)'\} \\ &= E\{(L'X_\alpha - L'\mu_\alpha)(L'X_\beta - L'\mu_\beta)'\} \\ &= L'E [(X_\alpha - \mu_\alpha)(X_\beta - \mu_\beta)']L \\ &= \delta_{\alpha\beta} L'\Sigma L \end{aligned}$$

$$\begin{aligned} \text{donde, } L'\Sigma L &= L'\sigma^2 \rho L \\ &= \sigma^2 L'\rho L. \end{aligned}$$

Luego, para cada $1 \leq \alpha, \beta \leq n$

$$\text{Cov } ((Y_\alpha, Y_\beta)) = \begin{cases} \sigma^2 L'\rho L & \text{si } \alpha = \beta \\ 0 & \text{si } \alpha \neq \beta \end{cases}$$

Por lo tanto la matriz de covarianza de Y es; $I \otimes \sigma^2 L'\rho L$.

Como X se distribuye normal, entonces $Y = L'X$, también se distribuye normal con media $L'\Gamma$ y matriz de covarianza $I \otimes \sigma^2 L'\rho L$, esto es $Y \sim N(L'\Gamma, I \otimes \sigma^2 L'\rho L)$.

CAPITULO II

DISTRIBUCIÓN ASINTÓTICA DE LOS VALORES PROPIOS

DISTRIBUCIÓN ASINTÓTICA DE LOS VALORES PROPIOS

DEFINICIÓN II.1. (Distribución Wishart Centrada)

Diremos que una matriz $M_{p \times p}$ sigue una distribución Wishart centrada, que denotaremos $W_p(\Sigma, n)$, si y sólo si, existe una matriz de datos multivariados $X_{p \times n}$ que sigue una distribución normal $X \sim N(0, I \otimes \Sigma)$ y tal que $M = XX'$.

PROPOSICIÓN II.1.

Si $M \sim W_p(\Sigma, n)$ y $B_{p \times q}$ es una matriz, entonces $B'MB \sim W_p(B'\Sigma B, n)$.

Demostración:

Sean $M \sim W_p(\Sigma, n)$ y $B_{p \times q}$ una matriz, entonces, por la definición II.1 existe $X \sim N(0, I \otimes \Sigma)$, tal que $M = XX'$.

Luego, $B'MB = B'XX'B = YY'$ donde $Y = B'X$.

Ahora, para cada $1 \leq \alpha \leq n$ se tendrá

$$\begin{aligned}
 E(Y_\alpha) &= E(B'X_\alpha) \\
 &= B' E(X_\alpha) \\
 &= 0
 \end{aligned}$$

Luego, $E(Y) = 0$

Además para $1 \leq \alpha, \beta \leq n$

$$\begin{aligned}
 \text{Cov}(Y_\alpha, Y_\beta) &= E[(Y_\alpha - 0)(Y_\beta - 0)'] \\
 &= E[(B'X_\alpha)(B'X_\beta)'] \\
 &= B' E(X_\alpha X_\beta') B \\
 &= \delta_{\alpha\beta} B' \Sigma B
 \end{aligned}$$

Luego, para $1 \leq \alpha, \beta \leq n$

$$\text{Cov}(Y_\alpha, Y_\beta) = \begin{cases} B' \Sigma B & \text{si } \alpha = \beta \\ 0 & \text{si } \alpha \neq \beta \end{cases}$$

Por lo tanto, la matriz de covarianza de Y es $I \otimes B' \Sigma B$.

Como X se distribuye normal entonces $Y = B'X$ también se distribuye normal con matriz de media cero (0) y matriz de covarianza $I \otimes B' \Sigma B$, esto es $Y \sim N(0, I \otimes B' \Sigma B)$.

Por otro lado $B'MB = B'XX'B = YY'$ tiene una distribución Wishart con matriz escalar $B' \Sigma B$ y n grados de libertad.

Así $B'MB \sim W_p(B' \Sigma B, n)$

PROPOSICIÓN II.2

Si $M \sim W_p(\Sigma, n)$ y $\mathbf{a} \neq 0$ es cualquier vector no aleatorio de orden $p \times 1$.

tal que $\mathbf{a}'\Sigma \mathbf{a} \neq 0$ entonces, $\frac{\mathbf{a}'M\mathbf{a}}{\mathbf{a}'\Sigma\mathbf{a}} \sim \chi^2$

Demostración:

Sea $M \sim W_p(\Sigma, n)$ y $\mathbf{a} \neq 0$ cualquier vector no aleatorio de orden $p \times 1$ tal que $\mathbf{a}'\Sigma \mathbf{a} \neq 0$ entonces, por definición, existe una matriz de datos multivariados $X_{p \times n}$ que sigue una distribución normal $X \sim N(0, I \otimes \Sigma)$, tal que $M = XX'$.

$$\text{Además } \mathbf{a}'M\mathbf{a} = \mathbf{a}'XX'\mathbf{a}$$

$$\text{donde } \mathbf{a}'X \sim N(0, \mathbf{a}'\Sigma \mathbf{a} \otimes I) \quad (2)$$

Como $\mathbf{a}'\Sigma \mathbf{a}$ es unidimensional, el producto de $\mathbf{a}'\Sigma \mathbf{a} \otimes I = (\mathbf{a}'\Sigma \mathbf{a})I$, por lo tanto $\mathbf{a}'X \sim N(0, (\mathbf{a}'\Sigma \mathbf{a})I)$.

Luego,

$$\frac{\mathbf{a}'M\mathbf{a}}{\mathbf{a}'\Sigma\mathbf{a}} \sim \chi^2 \text{ tiene una distribución Ji-cuadrada con } n \text{ grados de libertad} \quad (3)$$

(2) POLTRONIERI (1995)

(3) GRAYBILL (1961)

PROPOSICIÓN II.3

Sea $\{M_i\}_{i=1}^q$ una familia finita de matrices aleatorias mutuamente independientes, tales que, para cada $1 \leq i \leq q$, $M_i \sim W_p(\Sigma, n_i)$.

Entonces, $\sum_{i=1}^q M_i \sim W_p(\Sigma, n)$ donde $n = \sum_{i=1}^q n_i$

Demostración:

Sea $\{M_i\}_{i=1}^q$ una familia finita de matrices aleatorias mutuamente independientes, tales que, para cada $1 \leq i \leq q$, $M_i \sim W_p(\Sigma, n_i)$.

Entonces, para cada $1 \leq i \leq q$, existe una matriz aleatoria X_i , de orden $p \times n_i$, que sigue una distribución normal $X_i \sim N(0, I \otimes \Sigma)$ y tal que $M_i = X_i X_i'$.

Si para $1 \leq i \leq q$ y $1 \leq \alpha \leq n_i$, X^α corresponde a la α -ésima columna de X_i entonces, $X^\alpha \sim N(0, \Sigma)$.

Consideremos ahora $\sum_{i=1}^q M_i = \sum_{i=1}^q X_i X_i'$ y como para cada

$$1 \leq i \leq q \quad X_i X_i' = \sum_{\alpha=1}^{n_i} X^\alpha X^{\alpha'}$$

Renumerando las columnas de las matrices X_i con $1 \leq i \leq q$ por:

para cada $\sum_{j=1}^{i-1} n_j \leq \alpha \leq \sum_{k=1}^i n_k$ X^α corresponde a la $\alpha - \sum_{j=1}^{i-1} n_j$ -ésima

columna de X_i , tendremos,

$$\begin{aligned} \sum_{i=1}^q M_i &= \sum_{i=1}^q \sum_{\substack{j=1 \\ \alpha = \sum_{k=1}^j n_k}}^{n_i} X^\alpha X^{\alpha'} \quad \text{donde } n_0 = 1 \\ &= \sum_{\alpha=1}^n X^\alpha X^{\alpha'} \quad \text{donde } n = \sum_{i=1}^q n_i \end{aligned}$$

Como además M_1, M_2, \dots, M_q son mutuamente independientes entonces, para cada $1 \leq \alpha \leq n$ las matrices aleatorias X_i son mutuamente independientes lo que corresponde afirmar que, para cada $1 \leq \alpha \leq n$ los X^α son mutuamente independientes y $X^\alpha \sim N(0, \Sigma)$.

Consideremos la matriz aleatoria $X_{p \times n}$ tal que X^α corresponde a la α -ésima columna de X , entonces $X \sim N(0, I \otimes \Sigma)$.

Por lo tanto,

$$\sum_{i=1}^q M_i = XX' \sim W_p(\Sigma, n)$$

PROPOSICIÓN II.4

Sean $X_{p \times n}$ una matriz de datos multivariado tal que $X \sim N(0, I \otimes \Sigma)$, $B_{n \times n}$ una matriz ortonormal con su última fila igual a $1/\sqrt{n} (1, \dots, 1)$, $Z = XB'$ y, $A = XX' - n\bar{X}\bar{X}'$ entonces,

- (a) La última columna de Z es igual a $Z_n = \sqrt{n}\bar{X}$
- (b) $A = Z^* Z^{*'}$ donde Z^* es una matriz de orden $p \times (n-1)$

(c) $S = 1/(n-1) A$ es un estimador insesgado de Σ

Demostración:

Sean $X \sim N(0, I \otimes \Sigma)$, y B_{nm} la matriz ortonormal tal que

$$b_n = \frac{1}{\sqrt{n}} (1, \dots, 1)$$

Tenemos que: $XX' = ZZ'$ ⁽⁴⁾

Además como $Z = XB'$ tenemos que la n -ésima columna de Z estará dada por $Z_n = \sqrt{n}\bar{X}$, lo que demuestra (a).

Luego, $Z_n Z_n' = n\bar{X}\bar{X}'$

Como $A = XX' - n\bar{X}\bar{X}'$,

entonces $A = XX' - n\bar{X}\bar{X}'$

$$= ZZ' - Z_n Z_n'$$

$$= \sum_{\alpha=1}^n Z_\alpha Z_\alpha' - Z_n Z_n'$$

$$= \sum_{\alpha=1}^{n-1} Z_\alpha Z_\alpha'$$

$$= Z^* Z^{*'} \text{ lo que prueba (b)}$$

(4) ANDERSON (1958)

Por otro lado, como $Z = XB'$ y $X \sim N(0, I \otimes \Sigma)$, entonces $Z \sim N(0, I \otimes \Sigma)$ y en particular, $Z^* \sim N(0, I \otimes \Sigma)$.

Además, como que $1/n A$ es un estimador de máxima verosimilitud de Σ ⁽⁵⁾ veamos que $S = 1/(n-1) A$ es un estimador insesgado de Σ .

$$\text{Como } \hat{\Sigma} = \frac{1}{n} A = \frac{1}{n} Z^* Z^{*'}$$

entonces,

$$\begin{aligned} E(\hat{\Sigma}) &= E\left(\frac{1}{n} A\right) \\ &= \frac{1}{n} E(Z^* Z^{*'}) = \frac{1}{n} (\text{tr} I) \Sigma \quad (6) \\ &= \frac{n-1}{n} \Sigma \end{aligned}$$

luego, un estimador insesgado de Σ es:

$$\begin{aligned} \frac{n}{n-1} \hat{\Sigma} &= \frac{n}{n-1} \frac{1}{n} A \\ &= \frac{1}{n-1} A \\ &= S \end{aligned}$$

lo cual verifica (d).

(5) ANDERSON (1958)

(6) MEJIA (1986)

Consecuencia

Si $A = Z^* Z^{*'}$ con $Z^* \sim N(0, I \otimes \Sigma)$ entonces $(n-1) S \sim W_p(\Sigma, n-1)$ y por ende $A \sim W_p(\Sigma, n-1)$.

DISTRIBUCIÓN ASINTÓTICA DE LOS VALORES PROPIOS

Para muestras grandes es posible obtener, con la ayuda del teorema central del límite, la distribución asintótica de los valores propios de la matriz de covarianza muestral.

Este resultado nos será de mucha utilidad a la hora de hacer pruebas de hipótesis referentes a los valores propios de la matriz de covarianza de Σ .

PROPOSICIÓN II.5

Sean, $\Sigma_{p \times p}$ una matriz definida positiva con valores propios distintos, $A \sim W_p(\Sigma, n-1)$, $U = (n-1)^{-1} A$, $\Sigma = L \Lambda L'$ y $U = G D G'$ la descomposición espectral de Σ y U respectivamente y finalmente λ y d dos vectores de orden $p \times 1$ de elementos de la diagonal de Λ y D respectivamente.

Entonces $d \sim N_p(\lambda, (2/(n-1))\Lambda^2)$ esto es, los valores propios de U son asintoticamente normales, insesgados, mutuamente independientes y para cada $1 \leq i \leq q$, la varianza de d_i igual a $2\lambda_i^2 / (n-1)$.

Demostración:

Sea Σ una matriz definida positiva cuyos valores propios distintos son los elementos de la matriz diagonal Λ .

Consideremos la matriz A , definida en la proposición II.4, dada por, $A = Z^*Z^{*'}$ donde $Z^* \sim N_{p \times (n-1)}(0, I \otimes \Sigma)$. Luego, $A \sim W_p(\Sigma, n-1)$.

Definimos $M = L'AL$, entonces $M \sim W_p(\Lambda, n-1)$, pues $\Sigma = L\Lambda L'$ y Λ es la matriz de valores propios de Σ .

Tomando $U = (n-1)^{-1} M$ tendremos:

$$\begin{aligned} (1) \quad E(U) &= E[(n-1)^{-1} M] \\ &= E[(n-1)^{-1} L'AL] \\ &= L' E(S) L \\ &= L'\Sigma L \\ &= \Lambda \end{aligned}$$

Además, la matriz de covarianza de U está dada por:

$$\begin{aligned} (2) \quad \text{Var}(U) &= \text{tr } I_{n-1} (\Lambda \otimes \Lambda + \Lambda \dot{\otimes} \Lambda) \quad (7) \\ &= 1/(n-1) ((\Lambda \otimes \Lambda + \Lambda \dot{\otimes} \Lambda)) \end{aligned}$$

(7) MEJIA (1986)

De (1) y (2) tenemos que:

- $E(u_{ii}) = \lambda_i$ para cada $1 \leq i \leq p$
- $E(u_{ij}) = 0$ para $1 \leq i, j \leq p, i \neq j$
- $V(u_{ii}) = 2/(n-1) \lambda_i^2$ para cada $1 \leq i \leq p$
- $V(u_{ij}) = 1/(n-1) \lambda_i \lambda_j$ para $1 \leq i, j \leq p, i \neq j$
- $Cov(u_{ii}, u_{ij}) = 0$ para $1 \leq i, j \leq p, i \neq j$

Por proposición II.3, M puede ser representada como la suma de $n-1$ matrices independientes $W_p(\Lambda, 1)$. Luego, si consideramos el vector aleatorio.

$u' = (u_{11}, u_{22}, \dots, u_{pp})$ con vector de media $\lambda' = (\lambda_1, \lambda_2, \dots, \lambda_p)$ y matriz de covarianza $\frac{2}{n-1} \Lambda^2$, aplicando el teorema central del

límite tenemos que:

$$(n-1)^{1/2} (u - \lambda) \xrightarrow{D^{(*)}} Np(0, 2\Lambda^2) \quad (8)$$

y

(*) Convergencia en Distribución.

(8) MARDIA et. al. (1979)

$$(n-1)^{1/2} u_{ij} \xrightarrow{D^{(*)}} Np(0, \lambda_i, \lambda_j) \text{ para } 1 \leq i, j \leq p, i \neq j \quad (9)$$

luego $u = \lambda + O_p((n-1)^{-1/2})$ y $u_{ij} = O_p((n-1)^{-1/2})$.

Por lo que las raíces $l_1 > l_2 > \dots > l_p$ del polinomio $|U - dI|$ convergen en probabilidad a las raíces $\lambda_1 > \lambda_2 > \dots > \lambda_p$ del polinomio

$|\Lambda - dI|$; esto es:

$$u - d \xrightarrow{p^{(**)}} 0$$

de aquí que:

$$(n-1)^{1/2} (d - \lambda) = (n-1)^{1/2} (u - \lambda) + O_p((n-1)^{1/2}) \xrightarrow{D^{(*)}} Np(0, 2\Lambda^2)$$

de donde abusando del lenguaje tendremos que $l \sim N_p(\lambda, 2/(n-1)\Lambda^2)$.

(*) Convergencia en Distribución.

(9) MARDIA et. al. (1979)

(**) Convergencia en Probabilidad.

CAPTULO III
INTERPRETACIÓN DE LAS COMPONENTES
PRINCIPALES

INTERPRETACIÓN DE LAS COMPONENTES PRINCIPALES

En este capítulo presentaremos la variación explicada por las componentes principales, los criterios para seleccionar el número de componentes a considerar cuando se reduce la dimensión original, la correlación entre las variables originales y las componentes principales y el coeficiente de determinación.

Proporción de la variabilidad total explicada por la k-ésima componente principal.

Cada componente principal explica una proporción de la variabilidad total. Dicha proporción se define como el cociente entre el valor propio y la traza de Σ , esto es $\frac{\lambda_k}{\text{traza } \Sigma}$ y se denomina proporción de la variabilidad total explicada por la k-ésima componente principal.

Eficiencia del Ajuste:

Como los valores propios se pueden ordenar en forma creciente, es posible seleccionar los m valores propios $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ (siendo $m < p$) de manera que la eficiencia del ajuste de los datos originales por las nuevas m componentes principales se define como la proporción de la variación total explicada por la suma de los m primeros valores propios, así:

$$\text{Eficiencia del ajuste} = \frac{\sum_{j=1}^m \lambda_j}{\text{traza } \Sigma}$$

Aplicando la proposición I.2 (d) a las ecuaciones anteriores es posible expresar la variación explicada por cada componente principal o las m primeras componentes principales en función de la suma total de los valores propios. Así obtiene el porcentaje de la variación explicada por la k -ésima componente principal, dado por:

$$\frac{\lambda_k}{\sum_{i=1}^p \lambda_i} 100$$

y el porcentaje de la variación explicada por las m -primeras componentes principales, dado por:

$$\frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^p \lambda_j} 100$$

Cuando se consideran todos las componentes, es decir, si $m=p$, la proporción de la variación explicada es 1 y el porcentaje es 100%.

Si $m < p$ y la proporción de la variación explicada es 1, entonces, para cada $m+1 \leq j \leq p$, $\lambda_j = 0$.

Criterios de selección

Al decidir cuántas componentes principales se mantienen en una situación particular, deberán examinarse cuántas son necesarias incluir para que el porcentaje de variación explicada sea satisfactorio. Se recomiendan los siguientes criterios, si se desea reducir la dimensión original del problema dado.

Criterio Gráfico (Cattel, 1966)

Este criterio consiste en elaborar un gráfico donde se represente, en la ordenada, el porcentaje de variación explicada por cada componente principal y en la abscisa las correspondientes componentes principales. Por simple inspección visual al gráfico deberán considerarse las componentes anteriores al punto de la curva que refleja un bajo porcentaje de la variación total.

Criterio del Promedio (Kaiser)

Este criterio consiste en seleccionar aquellas componentes principales cuyos valores propios asociados sean superiores al promedio de los valores propios.

Observación:

Los criterios gráficos y del promedio poseen efectos contrapuestos en el número de componentes principales a considerar.

Mientras Cattel indica que el criterio gráfico tiende a incluir un número alto de componentes principales, kaiser establece que el criterio

del promedio tiende a incluir muy pocas componentes principales cuando el número de variables es inferior a 20.

Esta situación entre ambos criterios nos conduce a realizar una inspección detallada de las correlaciones entre las variables originales, las componentes principales y el coeficiente de determinación.

Correlaciones entre las variables originales y los componentes principales

Sea $X_{p \times n}$ una matriz aleatoria, $Y_{p \times n} = L'X$ la matriz de las componentes principales y, para cada $1 \leq \alpha \leq n$, X_α el vector aleatorio de orden $p \times 1$ correspondiente a la α -ésima columna de X y $Y_\alpha = L'X_\alpha$ el vector de las nuevas variables (componentes principales) de orden $p \times 1$ correspondiente a la α -ésima columna de Y .

Entonces, la correlación entre X_α y Y_α es:

$$\rho = \text{COV}(X_\alpha, Y_\alpha) (V^{1/2})^{-1} (\Lambda^{1/2})^{-1}$$

donde $V^{1/2}_{p \times p}$ es la matriz diagonal cuyos elementos de la diagonal principal son las raíces de las varianzas pertenecientes a Σ y $\Lambda^{1/2}_{p \times p}$ es una matriz diagonal donde los elementos de la diagonal principal son

las raíces de los valores propios pertenecientes a Σ .

Por la proposición I.2 (e) tenemos que:

$$\text{Cov}(X_\alpha, Y_\alpha) = \Sigma L$$

luego, $\text{Cov}(X_\alpha, Y_\alpha) = LL'\Sigma L = L\Lambda$

de donde $\rho = L\Lambda(V^{1/2})^{-1}(\Lambda^{1/2})^{-1}$

Así, la covarianza entre la j -ésima variable de X_α ($X_{j\alpha}$) y la k -ésima componente principal de Y_α , es el elemento en la posición (jk) de la matriz $L\Lambda$. Como la matriz L tiene la j -ésima componente del k -ésimo vector propio en la posición (jk) esto es, l_{jk} y la matriz diagonal Λ tiene en su diagonal principal los valores propios de la matriz Σ , entonces, para $1 \leq j, k \leq p$:

$$\text{Cov}(X_{j\alpha}, Y_{k\alpha}) = l_{jk} \lambda_k$$

Luego, la correlación entre $X_{j\alpha}$ y $Y_{k\alpha}$ será:

$$\rho_{jk} = \frac{l_{jk} (\lambda_k)^{1/2}}{\sqrt{\sigma_{jj}}}$$

donde σ_{jj} es la varianza de la j -ésima variable.

Coefficiente de determinación

El valor de ρ_{jk}^2 denominado, coeficiente de determinación, es una medida de la asociación entre las variables originales y los componentes principales y una manera de cuantificar la proporción de la variación total de una variable original explicada por la k-ésima componente principal.

En términos de la matriz de covarianza, el coeficiente de determinación será:

$$\rho_{jk}^2 = \frac{I_{jk}^2 \lambda_k}{\sigma_{jj}}$$

y para las m primeras componentes principales incluidas, luego de la selección, será:

$$\sum_{k=1}^m \rho_{jk}^2 = \frac{1}{\sigma_{jj}} \sum_{k=1}^m \lambda_k I_{jk}^2$$

CAPITULO IV

ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

El cuadro I muestra los valores promedio y desviación estándar de cada una de las variables estudiadas en el estudio del pez *Dormitator latifrons* (Richardson, 1837), en la localidad de Mata de Corozo corregimiento de Pacora en los años 1992 y 1993.

Se calculó la matriz de covarianza con las siete variables en estudio, la matriz resultante se presenta en el cuadro II.

El valor de la varianza para la variable peso total es mucho mayor que para las otras variables, lo cual se debe a que dicha variable tiene un rango de variación mayor. Esto implicaría que la variable peso total es la que más influye en el análisis. Se recomienda estandarizar las variables en estudio a fin de homogenizar las magnitudes.

La matriz de covarianza de los datos estandarizados es la matriz de correlación la cual se puede apreciar en el cuadro III.

En el cuadro IV se presentan los valores propios y la proporción de la variación total explicada por cada una de las componentes principales al usar la matriz de correlación.

El primer componente sintetiza el 38.934% de la variación total y junto con el segundo y tercer componente alcanzan el 87.37% de la variación total.

Aplicando el criterio de selección del promedio propuesto por Kaiser, se utilizarán los tres primeros componentes principales cuyos valores propios son mayores que el promedio.

El criterio de selección del gráfico propuesto por Cattell, se presenta en la figura 1, y conduce a la misma conclusión.

Ahora pasaremos a examinar los vectores propios y la proporción de la variación original explicada por cada componente. Los cuadros V y VI muestran los valores de las componentes de los vectores propios y la proporción de la variación original explicada por cada componente principal.

El cuadro V muestra que a través de los valores de las componentes de los vectores propios se pueden asociar las variables a los diferentes componentes principales. En nuestro estudio las variables largo total, peso total y peso del hígado están asociadas a la primera componente principal por tener un coeficiente más alto. De igual manera las variables estado gonadal, sexo y peso gonadal están asociadas a la

segunda componente principal y la variable mes se le asocia a la tercera componente principal.

Esta misma situación se refleja cuando examinamos la proporción de la variación original explicada por cada componente principal.

En el cuadro VI se observa que el 71.4%, 85.4% y 68.9% de la variabilidad total de las variables largo total, peso total y peso del hígado respectivamente es explicada por la primera componente principal.

El 82.9%, 81.5% y 50.0% de la variabilidad total de las variables estado gonadal, sexo y peso gonadal respectivamente es explicada por la segundo componente principal.

Por otro lado, el 95.2% de la variabilidad total de la variable mes es explicada por la tercera componente principal.

Podemos concluir que las siete variables en estudio quedan representadas por las tres primeras componentes principales.

Lo anterior confirma los resultados de Kaiser y Cattell.

En el cuadro VII se presenta la correlación de las variables originales con las tres primeras componentes principales y la variación total de las variables en estudio explicada por las tres primeras

componentes principales.

En el mismo cuadro se puede observar que las variables en estudio que tienen una variación total de más del 85.0% es explicada por las tres primeras componentes principales además de las variables largo total, peso total, peso gonadal cuya variación total es de 78.2%, 71.63% y 73.1% , respectivamente.

El cuadro VIII muestra los valores promedio y desviaciones estándar de cada una de las variables estudiadas en el estudio del pez *Dormitator latifrons* (Richardson, 1837) en la localidad Quebrada Carrasquilla del corregimiento de Panamá Viejo en los años 1992 y 1993.

Se calculó la matriz de covarianza con las ocho variables en estudio, cuya matriz resultante es presentada en el cuadro IX.

El valor de la varianza para la variable peso total es mucho mayor que para las otras variables, lo cual se debe fundamentalmente a que su rango de variación es mayor.

Para que el análisis no esté influido por este valor se recomienda realizar el estudio mediante los datos estandarizados, para que todas las variables tengan igual ponderación en el análisis en el cual se utilizará

la matriz de correlación que se presenta en el cuadro X.

En el cuadro XI se presentan los valores propios y la proporción de la variación total explicada por cada uno de las componentes principales al usar la matriz de correlación.

La primera componente principal sintetiza el 44.207% de la variación total y junto con la segunda componente principal alcanzan el 70.579% de la variación total. Aplicando el criterio de selección del promedio propuesto por Kaiser, se utilizarán los dos primeros componentes principales cuyos valores propios son superiores al promedio. El criterio de selección del gráfico propuesto por Cattell y representado en la figura 2, conduce a la misma conclusión.

Ahora pasaremos a examinar los vectores propios y la proporción de la variación original explicada por cada componente principal.

Los cuadros XII y XIII muestran los valores de las componentes de los vectores propios y la proporción de la variación original explicada por cada componente principal.

Al examinar los coeficientes de los vectores propios y las proporciones de la variación original podemos decir que las variables largo total, peso total, peso gonadal y largo gonadal están asociados a

la primera componente principal. De igual manera las variables estado gonadal, sexo y peso del hígado están asociados a la segunda componente principal y la variable mes asociada a la tercera componente principal. Esto agregaría un componente adicional a los resultados de kaiser y Cattell, donde el 82.0% de la variabilidad total es explicada por las tres primeras componentes principales. (Ver Cuadro XI).

El cuadro XIV presenta la correlación de las variables originales con las tres primeras componentes principales y la variación total de las variables en estudio explicado por las tres primeras componentes principales.

En el mismo cuadro se presenta que las variables en estudio que tienen una variación total de más del 80.0% es explicada por las tres primeras componentes principales además de las variables peso gonadal, largo gonadal, cuya variación total es de 77.9% y 57.1% respectivamente.

Cuadro I. VALORES PROMEDIO Y DESVIACIONES ESTÁNDAR DE LAS VARIABLES ESTUDIADAS EN EL *DORMITATOR LATIFRONS* (RICHARDSON, 1837) EN LA LOCALIDAD DE MATA DE COROZO DEL CORREGIMIENTO DE PACORA. AÑOS: 1992-1993.

VARIABLE	PROMEDIO	DESVIACIÓN ESTÁNDAR
Estado Gonadal	5.0227	0.9823
Mes	9.4318	1.8307
Sexo	1.4773	0.5023
Largo Total	25.4080	4.3978
Peso Total	269.4591	134.6944
Peso Gonadal	24.6397	13.6965
Peso del Hígado	7.2963	4.2462

Fuente: Datos proporcionados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas. Universidad de Panamá.

Cuadro II. MATRIZ DE COVARIANZA DE LAS VARIABLES ESTUDIADAS EN EL PEZ *DORMITATOR LATIFRONS* (RICHARDSON, 1837) EN LA LOCALIDAD DE MATA DE COROZO DEL CORREGIMIENTO DE PACORA. AÑOS: 1992-1993

VARIABLE	VARIABLE						
	Estado Gonadal	Mes	Sexo	Largo Total	Peso Total	Peso Gonadal	Peso del Hígado
Estado Gonadal	0.96499	-0.17	-0.471	-1.24	-21.0	5.3	-0.42
Mes	-0.1708	3.35	0.056	-1.47	-32.0	-0.5	2.12
Sexo	-0.4707	0.06	0.252	0.60	11.0	-2.5	0.14
Largo Total	-1.2380	-1.47	0.595	19.34	428.0	15.7	10.62
Peso Total	-20.7400	-31.56	10.522	427.98	18143.0	1006.2	424.23
Peso Gonadal	5.3418	-0.53	-2.534	15.69	1006.0	187.6	25.57
Peso del Hígado	-0.4156	2.12	0.144	10.62	424.0	25.6	18.03

Fuente: Datos proporcionados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas, Universidad de Panamá.

Cuadro III. MATRIZ DE CORRELACIÓN DE LAS VARIABLES ESTUDIADAS EN EL PEZ *DORMITATOR LATIFRONS* (RICHARDSON, 1837) EN LA LOCALIDAD DE MATA DE COROZO DEL CORREGIMIENTO DE PACORA. AÑOS: 1992-1993

VARIABLE	VARIABLE						
	Estado Gonadal	Mes	Sexo	Largo Total	Peso Total	Peso Gonadal	Peso del Hígado
Estado Gonadal	1.00	-0.09	-0.95	-0.29	-0.16	0.40	-0.10
Mes	-0.90	1.00	0.06	-0.18	-0.13	-0.02	0.27
Sexo	-0.95	0.06	1.00	0.27	0.16	-0.37	0.07
Largo Total	-0.29	-1.18	0.27	1.00	0.72	0.26	0.57
Peso Total	-0.16	-0.13	0.16	0.72	1.00	0.55	0.74
Peso Gonadal	0.40	-0.02	-0.37	0.26	0.55	1.00	0.44
Peso del Hígado	-0.10	0.27	0.07	0.57	0.74	0.44	1.00

Fuente: Datos proporcionados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas. Universidad de Panamá.

Cuadro IV. VALORES PROPIOS Y PROPORCIÓN DE LA VARIACIÓN TOTAL EXPLICADA POR LAS COMPONENTES PRINCIPALES EN EL ESTUDIO DEL PEZ *DORMITATOR LATIFRONS* (RICHARDSON, 1837) EN LA LOCALIDAD DE MATA DE COROZO DEL CORREGIMIENTO DE PACORA. AÑOS: 1992-1993

COMPONENTES	VALOR PROPIO	PROPORCIÓN DE LA VARIACIÓN TOTAL EXPLICADA	
		ABSOLUTA (%)	ACUMULADA (%)
1	2.725	38.934	38.934
2	2.227	31.819	70.753
3	1.163	16.617	87.370
4	0.443	6.329	93.699
5	0.267	3.815	97.514
6	0.131	1.872	99.386
7	0.043	0.614	100.000

Fuente: Datos proporcionados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas. Universidad de Panamá.

Cuadro V. COMPONENTES DE LOS VECTORES PROPIOS DE LA MATRIZ DE CORRELACIÓN DE LAS VARIABLES ESTUDIADAS EN EL PEZ *DORMITATOR LATIFRONS* (RICHARDSON, 1837) EN LA LOCALIDAD DE MATA DE COROZO DEL CORREGIMIENTO DE PACORA. AÑOS: 1992-1993

VARIABLES	VECTORES PROPIOS						
	1	2	3	4	5	6	7
Estado Gonadal	-0.211	0.610	0.018	0.231	0.015	-0.097	0.719
Mes	-0.001	-0.064	-0.905	0.046	0.327	-0.260	0.019
Sexo	0.205	-0.605	0.015	-0.329	-0.002	0.086	0.691
Largo Total	0.512	-0.034	0.238	0.465	0.672	0.105	0.024
Peso Total	0.560	0.121	0.104	-0.084	-0.279	-0.759	-0.009
Peso Gonadal	0.290	0.474	-0.044	-0.733	0.286	0.260	-0.053
Peso del Hígado	0.503	0.130	0.333	0.275	-0.530	0.511	0.038

Fuente: Datos proporcionados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas. Universidad de Panamá.

Cuadro VI. PROPORCIÓN DE LA VARIACIÓN ORIGINAL EXPLICADA POR CADA COMPONENTE PRINCIPAL DE LA MATRIZ DE CORRELACIÓN DE LAS VARIABLES EN ESTUDIO DEL PEZ *DORMITATOR LATIFRONS* (RICHARDSON, 1837) EN LA LOCALIDAD DE MATA DE COROZO DEL CORREGIMIENTO DE PACORA. AÑOS: 1992-1993

VARIABLES	COMPONENTES PRINCIPALES						
	1	2	3	4	5	6	7
Estado Gonadal	0.12100	0.82900	0.00040	0.02400	0.00006	0.00100	0.02200
Mes	0.00000	0.00900	0.95200	0.00090	0.02800	0.00900	0.00001
Sexo	0.11400	0.81500	0.00030	0.04800	0.00000	0.00090	0.02000
Largo Total	0.71400	0.00200	0.06600	0.09600	0.12000	0.00100	0.00002
Peso Total	0.85400	0.03300	0.01200	0.00300	0.02100	0.07500	0.00000
Peso Gonadal	0.22900	0.50000	0.00200	0.23800	0.02200	0.00900	0.00010
Peso del Hígado	0.68900	0.03800	0.12900	0.03400	0.07500	0.03400	0.00006

Fuente: Datos proporcionados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas. Universidad de Panamá.

Cuadro VII. CORRELACIÓN DE LAS VARIABLES ORIGINALES CON LAS TRES PRIMERAS COMPONENTES PRINCIPALES EN EL ESTUDIO DEL PEZ *DORMITATOR LATIFRONS* (RICHARDSON, 1837) EN LA LOCALIDAD DE MATA DE COROZO DEL CORREGIMIENTO DE PACORA. AÑOS: 1992-1993

VARIABLES	COMPONENTES PRINCIPALES			VARIACIÓN TOTAL EXPLICADA POR LAS TRES PRIMERAS COMPONENTES (%)
	1	2	3	
Estado Gonadal	-0.350	0.91	0.02	95.04
Mes	-0.002	-0.09	-0.98	96.10
Sexo	0.340	-0.90	0.02	92.93
Largo Total	0.840	-0.04	0.26	78.20
Peso Total	0.920	0.18	0.11	71.63
Peso Gonadal	0.450	0.71	-0.04	73.10
Peso del Hígado	0.83	0.19	-0.36	85.6

Fuente: Datos proporcionados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas. Universidad de Panamá.

Cuadro VIII. VALORES PROMEDIO Y DESVIACIONES ESTÁNDAR DE CADA UNA DE LAS VARIABLES ESTUDIADAS EN EL ESTUDIO DEL PEZ *DORMITATOR LATIFRONS* (RICHARDSON, 1837) EN LA LOCALIDAD QUEBRADA CARRASQUILLA DEL CORREGIMIENTO DE PANAMÁ VIEJO. AÑOS: 1992-1993

VARIABLE	PROMEDIO	DESVIACIÓN ESTÁNDAR
Estado Gonadal	4.5051	0.77425
Mes	7.9091	0.28894
Sexo	1.7071	0.45742
Largo Total	19.8152	2.96498
Peso Total	134.4263	53.19667
Peso Gonadal	5.8572	3.82756
Peso del Hígado	4.4720	2.29360
Largo Gonadal	5.9020	1.86252

Fuente: Datos proporcionados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas. Universidad de Panamá.

Cuadro IX. MATRIZ DE COVARIANZA DE LAS VARIABLES ESTUDIADAS EN EL PEZ *DORMITATOR LATIFRONS* (RICHARDSON, 1837) EN LA LOCALIDAD QUEBRADA CARRASQUILLA DEL CORREGIMIENTO DE PANAMÁ VIEJO. AÑOS: 1992-1993

VARIABLE	VARIABLE							
	Estado Gonadal	Mes	Sexo	Largo Total	Peso Total	Peso Gonadal	Peso del Hígado	Largo Gonadal
Estado Gonadal	0.5995	-0.076	-0.279	-0.600	-11.000	0.88	0.16	-0.56
Mes	-0.0760	0.083	0.055	0.200	3.000	0.22	0.02	0.10
Sexo	-0.2790	0.055	0.209	0.400	8.000	-0.28	0.03	0.35
Largo Total	-0.5980	0.150	0.446	8.800	144.000	4.68	4.08	2.88
Peso Total	-11.4000	3.209	8.374	143.800	2830.000	98.64	83.09	51.15
Peso Gonadal	0.8793	0.223	-0.278	4.700	99.000	14.65	5.50	2.44
Peso del Hígado	0.1628	0.021	0.034	4.100	83.000	5.50	5.26	1.88
Largo Gonadal	-0.5570	0.103	0.355	2.900	51.000	2.44	1.88	3.47

Fuente: Datos proporcionados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas, Universidad de Panamá.

Cuadro X. MATRIZ DE CORRELACIÓN DE LAS VARIABLES ESTUDIADAS EN EL PEZ *DORMITATOR LATIFRONS* (RICHARDSON, 1837) EN LA LOCALIDAD QUEBRADA CARRASQUILLA DEL CORREGIMIENTO DE PANAMÁ VIEJO. AÑOS: 1992-1993

VARIABLE	VARIABLE							
	Estado Gonadal	Mes	Sexo	Largo Total	Peso Total	Peso Gonadal	Peso del Hígado	Largo Gonadal
Estado Gonadal	1.00	-0.34	-0.79	-0.26	-0.28	0.30	0.09	-0.39
Mes	-0.34	1.00	0.41	0.18	0.21	0.20	0.03	0.19
Sexo	-0.79	0.41	1.00	0.33	0.34	-0.16	0.03	0.42
Largo Total	-0.26	0.18	0.33	1.00	0.91	0.41	0.60	0.52
Peso Total	-0.28	0.21	0.34	0.91	1.00	0.48	0.68	0.52
Peso Gonadal	0.30	0.20	-0.16	0.41	0.48	1.00	0.63	0.34
Peso del Hígado	0.09	0.03	0.03	0.60	0.68	0.63	1.00	0.44
Largo Gonadal	-0.39	0.19	0.42	0.52	0.52	0.34	0.44	1.00

Fuente: Datos proporcionados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas. Universidad de Panamá.

Cuadro XI. VALORES PROPIOS Y PROPORCIÓN DE LA VARIACIÓN TOTAL EXPLICADA POR LAS COMPONENTES PRINCIPALES EN EL ESTUDIO DEL PEZ *DORMITATOR LATIFRONS* (RICHARDSON, 1837) EN LA LOCALIDAD QUEBRADA CARRASQUILLA DEL CORREGIMIENTO DE PANAMÁ VIEJO. AÑOS: 1992-1993

COMPONENTES	VALOR PROPIO	PROPORCIÓN DE LA VARIACIÓN TOTAL EXPLICADA	
		ABSOLUTA (%)	ACUMULADA (%)
1	3.537	44.207	44.207
2	2.110	26.372	70.579
3	0.914	11.424	82.003
4	0.588	7.349	89.352
5	0.335	4.187	93.539
6	0.251	3.137	96.676
7	0.193	2.412	99.008
8	0.073	0.912	100.000

Fuente: Datos proporcionados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas. Universidad de Panamá.

Cuadro XII. COMPONENTES DE LOS VECTORES PROPIOS DE LA MATRIZ DE CORRELACIÓN DE LAS VARIABLES ESTUDIADAS EN EL PEZ *DORMITATOR LATIFRONS* (RICHARDSON, 1837) EN LA LOCALIDAD QUEBRADA CARRASQUILLA DEL CORREGIMIENTO DE PANAMÁ VIEJO. AÑOS: 1992-1993

VARIABLES	VECTORES PROPIOS							
	1	2	3	4	5	6	7	8
Estado Gonadal	-0.229	0.567	-0.142	0.021	0.098	0.140	0.746	0.141
Mes	0.206	-0.254	-0.854	0.145	0.092	0.368	0.007	0.004
Sexo	0.282	-0.516	0.028	-0.005	-0.404	-0.356	0.603	0.007
Largo Total	0.464	0.088	0.210	0.394	0.416	0.036	0.159	-0.615
Peso Total	0.482	0.105	0.153	0.374	0.148	-0.036	-0.068	0.752
Peso Gonadal	0.376	0.360	0.082	-0.029	-0.717	0.422	-0.100	-0.141
Peso del Hígado	0.286	0.441	-0.402	-0.214	-0.029	-0.692	-0.166	-0.100
Largo Gonadal	0.394	-0.066	0.121	-0.800	0.330	0.244	0.111	0.085

Fuente: Datos proporcionados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas. Universidad de Panamá.

Cuadro XIII. PROPORCIÓN DE LA VARIACIÓN ORIGINAL EXPLICADA POR CADA COMPONENTE PRINCIPAL DE LA MATRIZ DE CORRELACIÓN DE LAS VARIABLES EN ESTUDIO DEL PEZ *DORMITATOR LATIFRONS* (RICHARDSON, 1837) EN LA LOCALIDAD QUEBRADA CARRASQUILLA DEL CORREGIMIENTO DE PANAMÁ VIEJO. AÑOS: 1992-1993

VARIABLES	COMPONENTES PRINCIPALES							
	1	2	3	4	5	6	7	8
Estado Gonadal	0.185	0.678	0.0180	0.00260	0.0030	0.0050	0.10700	0.001000
Mes	0.150	0.136	0.6660	0.01200	0.0030	0.0340	0.00000	0.000001
Sexo	0.281	0.562	0.0007	0.00001	0.0550	0.0320	0.00500	0.000004
Largo Total	0.761	0.016	0.0400	0.09100	0.0580	0.0003	0.07000	0.028000
Peso Total	0.822	0.023	0.0210	0.08200	0.0070	0.0003	0.00500	0.041000
Peso Gonadal	0.500	0.273	0.0060	0.00050	0.1720	0.0450	0.00200	0.001000
Peso del Hígado	0.289	0.410	0.1480	0.02700	0.0003	0.1200	0.00500	0.007000
Largo Gonadal	0.547	0.009	0.0130	0.37600	0.0360	0.0150	0.00200	0.000500

Fuente: Datos proporcionados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas, Universidad de Panamá.

Cuadro XIV. CORRELACIÓN DE LAS VARIABLES ORIGINALES CON LAS TRES PRIMERAS COMPONENTES PRINCIPALES EN EL ESTUDIO DEL PEZ *DORMITATOR LATIFRONS* (RICHARDSON, 1837) EN LA LOCALIDAD QUEBRADA CARRASQUILLA DEL CORREGIMIENTO DE PANAMÁ VIEJO. AÑOS: 1992-1993

VARIABLES	COMPONENTES PRINCIPALES			VARIACIÓN TOTAL EXPLICADA POR LAS TRES PRIMERAS COMPONENTES (%)
	1	2	3	
Estado Gonadal	-0.430	0.82	-0.130	88.10
Mes	0.390	-0.37	-0.820	95.20
Sexo	0.530	-0.75	0.030	84.37
Largo Total	0.870	0.13	0.200	81.70
Peso Total	0.910	0.15	0.140	86.60
Peso Gonadal	0.710	0.52	0.080	77.90
Peso del Hígado	0.540	0.64	-0.380	84.70
Largo Gonadal	0.740	-0.09	0.110	57.10

Fuente: Datos proporcionados por el Laboratorio Dr. Erich Graetz del Departamento de Fisiología y Comportamiento Animal de la Facultad de Ciencias Naturales y Exactas. Universidad de Panamá.

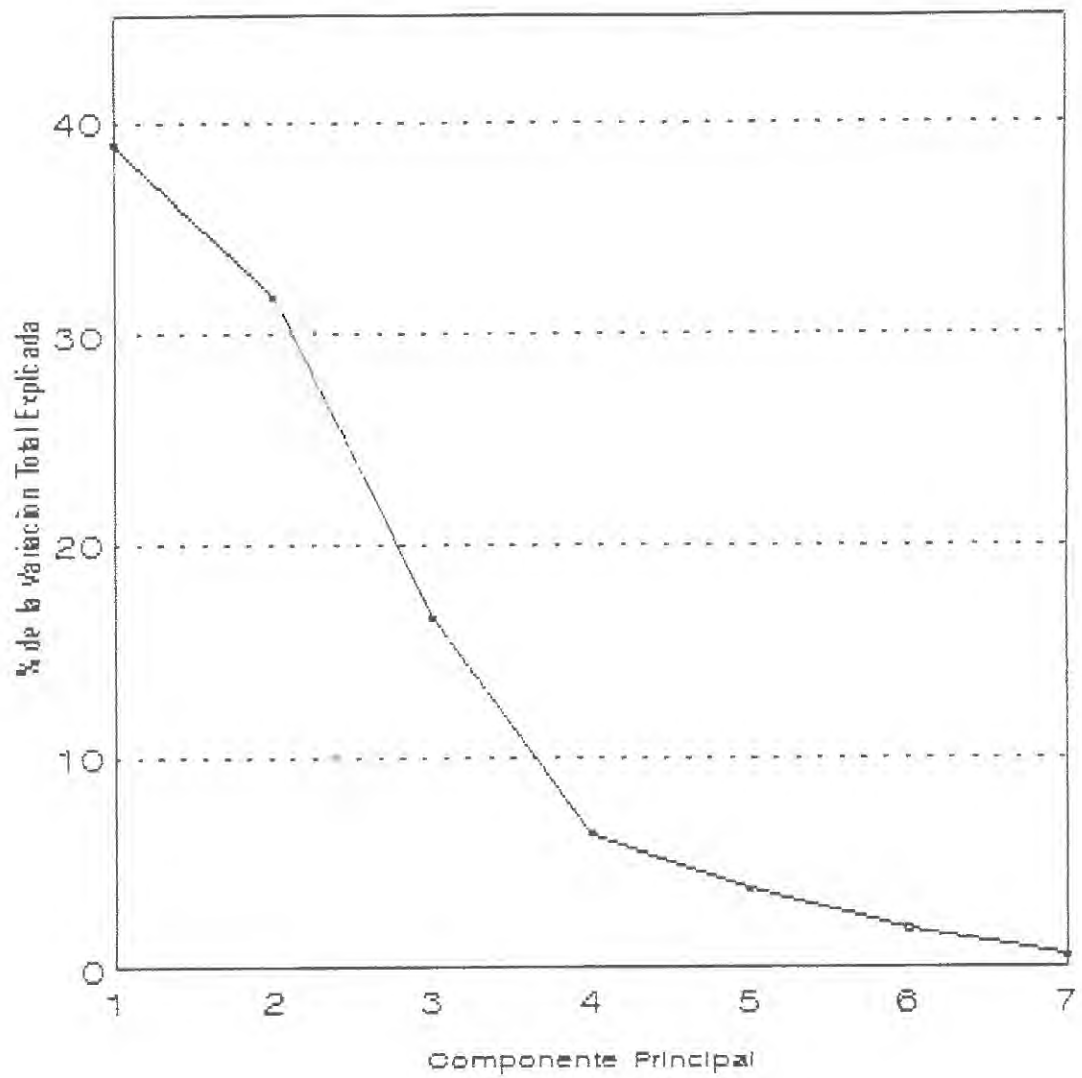


Fig. 1. Proporción de la variación total explicada por las componentes principales.

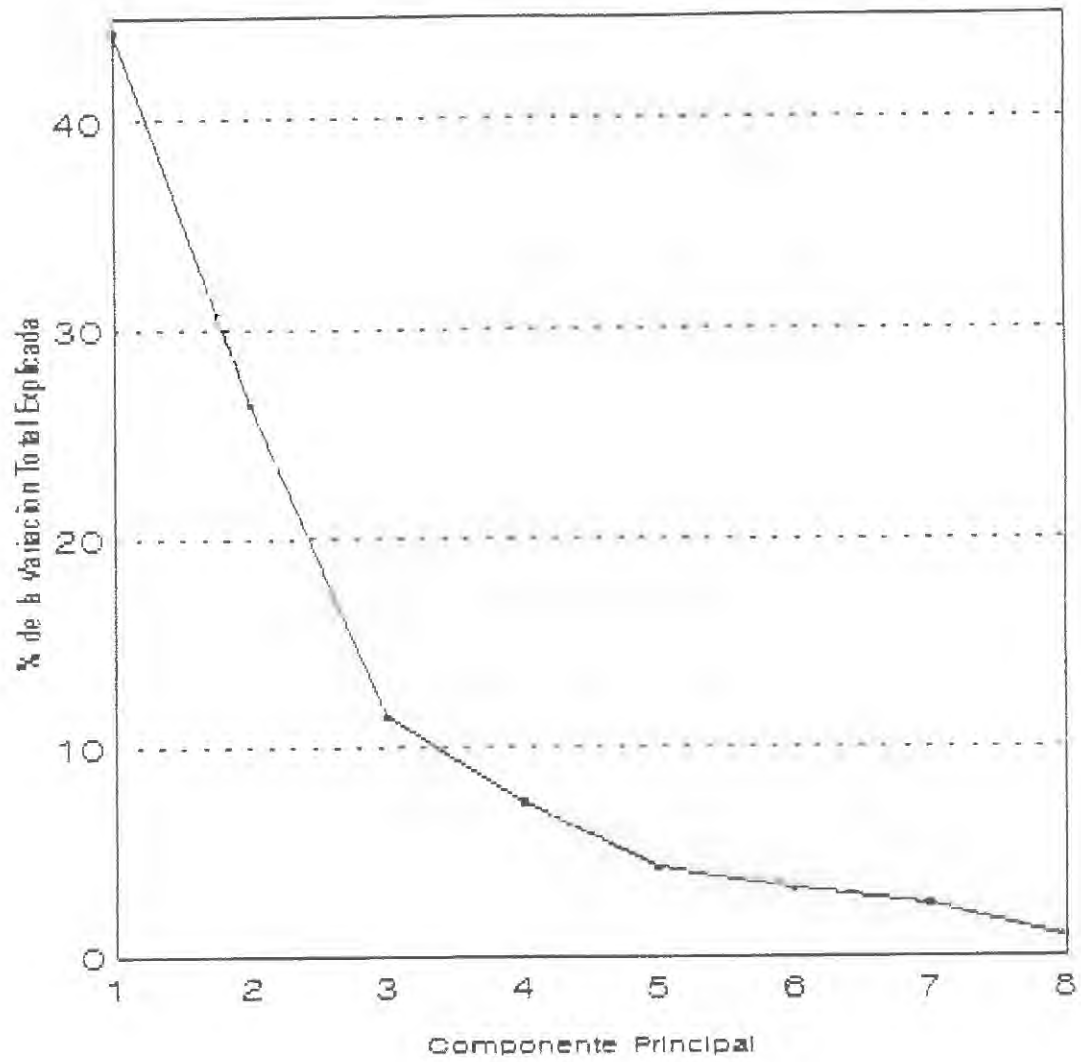


Fig. 2. Proporción de la variación total explicada por las componentes principales.

CONCLUSIONES Y RECOMENDACIONES

1. **En un análisis multivariado es común encontrar un gran número de variables independientes lo que puede ser perjudicial ya que se corre el riesgo de duplicidad de la información. El método de componentes principales nos permite eliminar este contratiempo.**
2. **Para adoptar una decisión definitiva acerca del número de componentes principales es necesario examinar los vectores propios y la proporción de la variación total de las variables originales explicadas por cada una de las componentes principales.**
3. **El método de componentes principales nos permite obtener nuevas variables aleatorias no correlacionadas lo que nos permite luego realizar cualquier tipo de análisis multivariado (regresión, conglomerados, etc.).**
4. **La distribución de la matriz de las componentes principales Y_{pm} es $N(L'T, I \otimes \Lambda)$ lo que nos indica que las componentes principales (y_{ij}) son mutuamente independientes con varianza λ_i que son los valores**

propios de Σ .

5. **La distribución de la matriz D, resulta ser asintóticamente normal, con matriz de media Λ , luego D es un estimador insesgado de Λ .**
6. **No se deben incluir en un modelo de regresión las últimas componentes principales ya que aumenta la varianza de los estimadores, sin embargo disminuye el sesgo si existe una elevada correlación de éstas con la variable dependiente.**
7. **En el estudio efectuado en la localidad de Mata de Corozo podemos recomendar que las siete variables en estudio pueden ser representadas a través de tres nuevas variables que corresponde a las tres primeras componentes principales.**
8. **Se puede observar que la variable mes tiene una alta correlación solamente con la variable estado gonadal (ver cuadro III). Lo cual exigirá al investigador una decisión respecto a la permanencia o exclusión de la variable en cuestión.**

9. **En el estudio realizado en la localidad de Quebrada Carrasquilla se recomienda agregar la tercera componente principal para que las variables analizadas estén representadas a través de las tres primeras componentes principales.**

BIBLIOGRAFIA

ANDERSON, T. W. 1958. An introduction multivariate analysis. 1st. ed. J. WILLEY and SONS INC. New York, 374 págs.

BRYAN, M. 1986. Multivariate statistical methods a primer. 1st. ed. CHAPMAN and HALL. New York, 460 págs.

DILLON, W. y GOLDSTEIN, M. 1984. Multivariate analysis methods and applications. 1st. ed. J. WILLEY and SONS INC. New York, 520 págs.

GRAYBILL, F. 1961. An introduction to linear statistical models. Vol I. 1st. ed. MCGRAW-HILL BOOK COMPANY INC. New York, 463 págs.

GHANADESIKAN, R. 1977. Mehods for statistics data analysis of multivariate observations. 1st. ed. J. WILLEY and SONS INC. New York, 311 págs.

HARRIS, R. 1975. A primer of multivariate statistics. 1st. ed. ACADEMIC PRESS. New York, 332 págs.

JOHNSON, R. y WICHERN, D. 1982. Applied multivariate statistical analysis. 1st. ed. PRENTICE HALL INC. New Jersey, 594 págs.

JOLLYFFE, I.T. 1986. Principal component analysis. 1st ed. SPRINGER-VERLAG. New York, 269 págs.

MARDIA, K.V. , Kent J.T. y BiBBY, J.M. 1979. Multivariate analysis. 1st. ed. ACADEMIC-PRESS. London, 521 págs.

MEJIA, A. 1986. Aspectos de la distribución Wishart en estadística multivariante. Tesis. Universidad de Panamá. Panamá, Panamá, 81 págs.

MORRISON, D.F. 1976. Multivariate statistical methods. 2nd. ed. McGRAW-HILL BOOK COMPANY. New York, 415 págs.

PLA, L. 1986. Análisis multivariado: Método de componentes principales. Organización de los Estados Americanos. Programa Regional de Desarrollo Científico Tecnológico. Washington, D.C, 94 págs.

POLTRONIERI, J. 1995. Contribución al estudio de formas cuadráticas en estadística multivariada. IX Simposio de Métodos Matemáticos aplicados a las Ciencias. Costa Rica, 133-142 págs.

QUIROZ, R. , ARCE, B. y HOLLE, M. 1991. Métodos de investigación con enfoque y análisis de datos de sistemas agropecuarios. Vol 41, N° 1. Costa Rica, 1-14 págs.

SISTACHS, V. , ACOSTA, L. y LINARES, G. 1986. Estadística Multivariada. 1ra. ed. UNIVERSIDAD DE LA HABANA, Cuba, 320 págs.

ZALDIVAR, A. y MENACHO, C. 1991. Metodología estadística para la caracterización de Fincas de cuyes. Vol. 41, N° 1. Costa Rica, 15-21 págs.