



UNIVERSIDAD CENTRAL DE VENEZUELA  
FACULTAD DE CIENCIAS  
ESCUELA DE MATEMÁTICA

## Estimación del porcentaje de variación obtenido en ACP usando Bootstrap.

Trabajo Especial de Grado presentado ante  
la ilustre Universidad Central de Venezuela  
por el **Br. Hairol Pacheco Aranaga** para  
optar al título de Licenciado en Matemática.  
**Tutora: Dra. Mairene Colina**

Caracas - Venezuela  
25 de febrero de 2011

Nosotros, los abajo firmantes, designados por la Universidad Central de Venezuela como integrantes del Jurado Examinador del Trabajo Especial de Grado titulado “**Estimación del porcentaje de variación obtenido en ACP usando Bootstrap**”, presentado por el **Br. Hairol Pacheco Aranaga**, titular de la Cédula de Identidad **V-14.203.843**, certificamos que este trabajo cumple con los requisitos exigidos por nuestra Magna Casa de Estudios para optar al título de **Licenciado en Matemática**.

---

**Dra. Mairene Colina**  
**Tutora**

---

**M.Sc. José Hernández**  
**Jurado**

---

**M.Sc. Luis Paredes**  
**Jurado**

# Dedicatoria

Dedico el presente trabajo a mi madre, que además de darme la vida ha sido mi pilar de apoyo, mi consejera, mi amiga, mi padre, mi confidente y compañera en todos los buenos y no tan buenos momentos. A la memoria de mi abuela Isabel, quien estuvo siempre pendiente de mi crecimiento personal y académico y me llenó de cariño incondicional. A mi hermana, quien me ha acompañado durante toda su vida y me ha dado la oportunidad de aprender juntos y de compartir momentos especiales. A mis sobrinas, quienes llenan de luz mis ojos, mis pensamientos y mi corazón. A toda mi familia, que siempre creyó en mí y me dio la fortaleza necesaria para seguir adelante.

# Agradecimientos

A Dios, por darme la oportunidad de transitar y aprender en esta vida maravillosa.

A mi tutora, Dra. Mairene Colina por su persistente guía y su infinita paciencia.

A mis fieles, leales e inseparables amigos, Carlos, Germán, Marcelo, Marco, Olga, José Gregorio, Julio Cesar, Nobraskja y Patricia, por el simple hecho de ser y de estar, por dejarme caer y ayudarme a ponerme de pie para seguir adelante, por mostrarme el valor inconmensurable de la amistad sincera y verdadera, por todos los inolvidables momentos que hemos pasado juntos y por todos los consejos, enseñanzas y ayuda que me han dado en todo este tiempo.

A los Profesores Juan Guevara, Margarita Olivares, Inés Núñez, Francisco Tovar, José Gregorio Mijares, Mercedes Arriojas, Glaysar Castro, José Rafael León, Laura Galindo, Ángel Padilla y a todos los demás profesores que participaron en mi formación académica por su ayuda y por impartir sus conocimientos.

A Joana Marquez, quien fue mi pareja y compañera incondicional durante dos años, por haberme apoyado e impulsado a alcanzar la meta que hoy logro, por haberse sacrificado junto a mí, por haber sido mi soporte para no darme por vencido, por haberme entregado su amor y por enseñarme tantas cosas bellas y maravillosas en el tiempo que estuvimos juntos.

Al Lic. Hugo Villarroel, por sus valiosos aportes en la realización de este trabajo.

A la Lic. Jackeline Godoy, por haberme apagado el video beam durante la defensa de este trabajo.

A Ana, Johan, Teresa y a toda la familia Moreno, por brindarme su inmenso cariño y apoyo.

Y en especial a todos aquellos que no fueron nombrados, pero que dejaron una huella imborrable en mi alma durante su transitar en nuestras vidas.

¡¡¡Lo Logramos!!!

# Índice general

<b>Introducción</b>	<b>1</b>
<b>1. Análisis de Componentes Principales</b>	<b>3</b>
1.1. Introducción . . . . .	3
1.2. Definición del Análisis de Componentes Principales . . . . .	3
1.3. Obtención de las Componentes Principales y Problemas de Optimización . . . . .	4
1.3.1. Determinación del Número de Componentes Principales . . . . .	8
1.4. Propiedades Algebraicas de las Componentes Principales y sus Implicaciones Estadísticas . . . . .	9
1.5. Propiedades Geométricas de las Componentes Principales y sus Implicaciones Estadísticas . . . . .	17
1.6. Componentes Principales Usando una Matriz de Correlaciones . . . . .	20
1.7. Componentes Principales con Varianzas Nulas ó con Varianzas Iguales . . . . .	23
<b>2. Bootstrap</b>	<b>24</b>
2.1. Introducción . . . . .	24
2.2. Definición de Bootstrap . . . . .	24
2.3. Ejemplo de Estimación Bootstrap del Error Típico Para la Media . . . . .	27
2.4. Función de Distribución Empírica y el Principio Plug-in . . . . .	30
2.4.1. El Principio Plug-in . . . . .	31
2.5. Error Típico de la Media y Estimación del Error Típico de la Media . . . . .	32
2.5.1. El Error Típico de la Media . . . . .	32
2.5.2. Estimación del Error Típico de la Media . . . . .	34
2.6. La Estimación Bootstrap del Error Típico . . . . .	34
2.7. El Número $B$ de Réplicas Bootstrap . . . . .	35
2.8. Estimación del Porcentaje de Variación obtenido en ACP usando Bootstrap . . . . .	36
<b>3. Implementación de las Herramientas R y MATLAB Para realizar ACP y Bootstrap</b>	<b>39</b>
3.1. Introducción . . . . .	39
3.1.1. Implementación de la Herramienta R para la realización del ACP . . . . .	39

3.1.2. Implementación de la Herramienta MATLAB para la realización del re- muestreo Bootstrap . . . . .	45
3.2. Conclusiones . . . . .	47

# Introducción

La Estadística es la ciencia del aprendizaje a través de la experiencia. Es la ciencia con base matemática referente a la recolección, análisis e interpretación de datos, que busca explicar condiciones regulares en fenómenos de tipo aleatorio. Originándose alrededor del año 1650, esta rama de la matemática ha evolucionado de manera tal que muchas de sus técnicas forman las bases de los métodos analíticos de diferentes áreas de estudio como la biomedicina, psicología, educación, economía, teoría de las comunicaciones, sociología, genética, epidemiología, geología, física, astronomía, etc., ya que estas son áreas que demandan eficiencia informativa.

La estadística inferencial está apoyada en fuertes bases teóricas matemáticas, las cuales permiten tener un sin número de herramientas que hacen posible el trabajo del analista para realizar trabajos de investigación. Sin embargo, algunas veces el uso de estas herramientas teóricas hace difícil y tedioso el trabajo de investigación, lo cual en ocasiones genera pérdidas de tiempo considerables.

La aparición de las computadoras ha marcado de por vida el quehacer científico, puesto que con el poder y velocidad de cómputo, ha resultado mucho más fácil y económico realizar investigaciones donde el ser humano por sí solo, aun teniendo disponible toda la base teórica y las técnicas necesarias para poner en práctica esas herramientas, le resultaría a veces hasta interminable realizar algunas conclusiones acerca del estudio a realizarse, cosa que con una computadora podría tardar solo minutos.

Existen dos técnicas estadísticas muy útiles y poderosas desde el punto de vista computacional a la hora de realizar investigaciones en un área determinada. Una de ellas es el Análisis de Componentes Principales (ACP), una de las más viejas y conocidas técnicas del Análisis Multivariante. Propuesta por el matemático Karl Pearson (1867-1936), consiste en reducir la dimensión de un conjunto de variables aleatorias que se miden sobre un banco de datos, perdiendo así la menor cantidad de información. Para esto, el ACP construye un nuevo sistema de coordenadas para el conjunto original de datos. Al representar la nube de puntos en este nuevo sistema, la varianza de mayor tamaño es capturada en el primer eje coordenado (llamado Primer Componente Principal), la segunda varianza más grande en el segundo eje y así sucesivamente. La construcción del sistema de coordenadas viene dada por una transformación lineal

que se obtiene previamente de la construcción de la matriz de covarianzas  $\Sigma$ . Como esta matriz es simétrica y definida positiva, existe una base de autovectores  $\{\alpha_1, \alpha_2, \dots, \alpha_d\}$  asociados a un conjunto de autovalores  $\lambda_1, \lambda_2, \dots, \lambda_d$ , por lo que se puede decir que las componentes principales son una combinación lineal de las variables originales y además son independientes entre sí. Los valores  $\lambda_1, \lambda_2, \dots, \lambda_d$  son utilizados para calcular el porcentaje de varianza explicado por las primeras  $q$  componentes principales, el cual ayuda a determinar la dimensión del subespacio donde “caen” los datos originales.

La otra técnica estadística es el Bootstrap, un método de remuestreo propuesto por Bradley Efron en 1979, el cual se basa en el uso de un gran número de cálculos repetitivos para obtener medidas de precisión de estimaciones estadísticas (errores estándar, sesgos, intervalos de confianza, etc.). Se utiliza para estimar la distribución de un estadístico. Este remuestreo se realiza a partir de una muestra aleatoria con una función de probabilidad desconocida  $F$ . Para esto, se construye una distribución de probabilidad empírica  $\hat{F}(x)$  a partir de la muestra  $\mathbf{x}$ , asignando una probabilidad  $1/n$  a cada punto  $x_1, x_2, \dots, x_n$  de la misma, esta función será el estimador de la función de distribución de la población,  $F(x)$ . A partir de esta distribución de probabilidad empírica se extrae una muestra aleatoria  $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$  de tamaño  $n$  con reposición. Esta muestra se denomina muestra bootstrap. Luego se calcula el estadístico de interés a partir de la muestra bootstrap, obteniendo lo que se denomina la réplica bootstrap  $\hat{\theta}^*$ . Se repite este proceso una cantidad “B” de veces, donde “B” es un valor grande (entre 50 y 200 para estimar errores estándar). Finalmente se construye una distribución de probabilidad a partir de las B réplicas bootstrap asignando una probabilidad de  $1/B$  a cada punto,  $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$ , para poder realizar inferencias sobre el estadístico de interés.

El objetivo del presente trabajo es explicar cómo la combinación de la técnica del ACP y la técnica del Bootstrap resulta útil para determinar la precisión del ACP cuando se trabaja con un conjunto pequeño de datos, con lo cual es posible determinar cuán fiables son los resultados obtenidos en el mismo. En particular, se estimará el porcentaje de variación explicado por la primera componente principal, puesto que ésta variable posee la propiedad de minimizar la suma de los cuadrados de las distancias euclídeas entre ella y los datos representados en el nuevo espacio de variables componentes principales. Se aplicará la técnica del ACP a un conjunto de datos y se calculará el porcentaje de variación explicado por la primera componente principal. Se aplicará la técnica del bootstrap para estimar este porcentaje mediante la utilización de doscientas réplicas bootstrap y se calculará el promedio de estas doscientas réplicas para ser comparado con el porcentaje de variación observado.

# Capítulo 1

## Análisis de Componentes Principales

### 1.1. Introducción

En este capítulo se definirá el método del Análisis de Componentes Principales, se explicará el proceso de obtención de las componentes principales resolviendo problemas de optimización y se mencionarán los dos métodos utilizados comúnmente para determinar el número de componentes principales adecuado (cálculo del porcentaje de variación y gráfica autovalor versus número del autovalor). También se enunciarán y demostrarán propiedades algebraicas y geométricas de las componentes principales, con sus respectivas implicaciones estadísticas. Por último, se mencionarán brevemente dos casos particulares del Análisis de Componentes Principales (Análisis de componentes Principales usando una matriz de correlaciones y Análisis de Componentes Principales con varianzas nulas ó con varianzas iguales).

### 1.2. Definición del Análisis de Componentes Principales

El Análisis de Componentes Principales (ACP) es una técnica estadística que se emplea en análisis exploratorio de datos y en construcción de modelos predictivos. Consiste en reducir la dimensión de un conjunto de variables correlacionadas, perdiendo la menor cantidad de información posible, y obteniendo un nuevo conjunto de variables no correlacionadas llamadas componentes principales.

La obtención de las componentes principales (en lo que sigue, CP) viene dada por una función lineal que se obtiene previamente de la construcción de la matriz de covarianzas.

Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad, donde  
 $\Omega$  es un espacio muestral arbitrario  
 $\mathcal{F}$  es la  $\sigma$ -álgebra de subconjuntos de  $\Omega$   
 $P$  es la medida de probabilidad.

Entonces  $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{pmatrix} \in \mathbb{R}^d$ , con  $X_i : \Omega \rightarrow \mathbb{R}$  se denomina vector aleatorio con segundo momento finito si  $\mathbb{E}(X_i^2) < \infty, \forall i = 1, 2, \dots, d$ .

Sea  $\Sigma$  la matriz de covarianzas del vector aleatorio  $\mathbf{X}$ :

$$\Sigma = \begin{pmatrix} Cov(X_1, X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_d) \\ Cov(X_2, X_1) & Cov(X_2, X_2) & \cdots & Cov(X_2, X_d) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_d, X_1) & Cov(X_d, X_2) & \cdots & Cov(X_d, X_d) \end{pmatrix}$$

Se busca una función lineal  $\alpha'_1 \mathbf{X}$  de la forma:

$$\alpha'_1 \mathbf{X} = \alpha_{11}X_1 + \alpha_{12}X_2 + \cdots + \alpha_{1d}X_d.$$

Donde  $\alpha_1 = \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \vdots \\ \alpha_{1d} \end{pmatrix}$  es un vector de constantes en  $\mathbb{R}^d$  y  $\alpha'_1$  denota el vector traspuesto de  $\alpha_1$ .

Esta función lineal debe ser tal que:

- $Var(\alpha'_1 \mathbf{X})$  sea máxima.
- $\|\alpha_1\|^2 = 1$

A continuación se busca otra función lineal  $\alpha'_2 \mathbf{X}$  que esté no correlacionada con  $\alpha'_1 \mathbf{X}$  y que maximice la varianza. Se sigue este proceso hasta conseguir una función lineal  $\alpha'_k \mathbf{X}$  que no esté correlacionada con  $\alpha'_1 \mathbf{X}, \alpha'_2 \mathbf{X}, \dots, \alpha'_{k-1} \mathbf{X}; k \leq d$ , y que maximice la varianza. La primera variable  $\alpha'_1 \mathbf{X}$  se denomina la primera componente principal, la segunda variable  $\alpha'_2 \mathbf{X}$  se denomina la segunda componente principal y así sucesivamente.

### 1.3. Obtención de las Componentes Principales y Problemas de Optimización

Como se mencionó anteriormente, la obtención de las componentes principales se logra a partir de la construcción de la matriz de covarianzas, la cual posee dos propiedades importantes,

a saber:

- $\Sigma = \Sigma^t$ .
- $\alpha' \Sigma \alpha \geq 0$ .

La primera propiedad es una consecuencia de la definición de covarianza. Por lo tanto solo se probará la segunda propiedad.

Sea  $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{pmatrix} \in \mathbb{R}^d$ , se quiere probar que  $\mathbf{v}' \Sigma \mathbf{v} \geq 0$ .

$$\begin{aligned}
 \mathbf{v}' \Sigma \mathbf{v} &= (v_1, v_2, \dots, v_d) \begin{pmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_d) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \cdots & \text{Cov}(X_2, X_d) \\ \vdots & \vdots & & \vdots \\ \text{Cov}(X_d, X_1) & \text{Cov}(X_d, X_2) & \cdots & \text{Cov}(X_d, X_d) \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{pmatrix} \\
 &= \left( \sum_{k=1}^d v_k \text{Cov}(X_k, X_1), \sum_{k=1}^d v_k \text{Cov}(X_k, X_2), \dots, \sum_{k=1}^d v_k \text{Cov}(X_k, X_d) \right) \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{pmatrix} \\
 &= \left( \sum_{k=1}^d v_k \text{Cov}(X_k, X_1) \right) v_1 + \left( \sum_{k=1}^d v_k \text{Cov}(X_k, X_2) \right) v_2 + \dots + \left( \sum_{k=1}^d v_k \text{Cov}(X_k, X_d) \right) v_d \\
 &= \sum_{l=1}^d \sum_{k=1}^d v_k \text{Cov}(X_k, X_l) v_l \\
 &= \sum_{l=1}^d \sum_{k=1}^d \text{Cov}(X_k, X_l) v_k v_l \\
 &= \sum_{l=1}^d \sum_{k=1}^d \mathbb{E}[(x_k - \mathbb{E}[x_k]) (x_l - \mathbb{E}[x_l])] v_k v_l \\
 &= \mathbb{E} \left[ \sum_{l=1}^d \sum_{k=1}^d (x_k - \mathbb{E}[x_k]) (x_l - \mathbb{E}[x_l]) v_k v_l \right] \\
 &\geq \mathbb{E} \left[ \sum_{k=1}^d (x_k - \mathbb{E}[x_k]) v_k \right]^2 \geq 0.
 \end{aligned}$$

Como consecuencia de lo demostrado anteriormente, se tiene que si existen vectores no nulos  $w_k \in \mathbb{R}^d$  y escalares  $\lambda_k \in \mathbb{R}$ ;  $k = 1, 2, \dots, d$  tales que:

$$\Sigma w_k = \lambda_k w_k; \quad k = 1, 2, \dots, d,$$

Entonces se tiene que  $\lambda_k \geq 0 \quad \forall k = 1, 2, \dots, d$ .

La primera CP es aquella que maximiza la varianza:

$$Var(\alpha_1' \mathbf{X}) = Cov(\alpha_1' \mathbf{X}, \alpha_1' \mathbf{X}) = \alpha_1' \Sigma \alpha_1.$$

Supongamos que  $\alpha_1$  es un autovector de  $\Sigma$  asociado al autovalor  $\lambda$ . Entonces se cumple que:

$$\alpha_1' \Sigma \alpha_1 = \alpha_1' \lambda \alpha_1.$$

Sujeto a las restricciones:

- $\|\alpha_1\|^2 = 1$
- $\max_k(\alpha_{1k}) = 1$

Para maximizar  $\alpha_1' \Sigma \alpha_1$  sujeto a estas restricciones, se utiliza el método de multiplicadores de Lagrange.

Sea  $\lambda$  un multiplicador de Lagrange. Se quiere maximizar:

$$L(\alpha_1) = \alpha_1' \Sigma \alpha_1 - \lambda(\alpha_1' \alpha_1 - 1).$$

Derivando con respecto a  $\alpha_1'$  se obtiene:

$$\begin{aligned} \Sigma \alpha_1 - \lambda \alpha_1 &= 0 \\ \Rightarrow (\Sigma - \lambda I_d) \alpha_1 &= 0. \end{aligned}$$

Donde  $I_d$  es la matriz identidad de orden  $d$ .

De este modo,  $\lambda$  es un autovalor y  $\alpha_1$  el autovector correspondiente.

Se debe tener en cuenta que la cantidad que se maximiza es:

$$\boldsymbol{\alpha}'_1 \boldsymbol{\Sigma} \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}'_1 \lambda \boldsymbol{\alpha}_1 = \lambda \boldsymbol{\alpha}'_1 \boldsymbol{\alpha}_1 = \lambda_1.$$

Así,  $Var(\boldsymbol{\alpha}'_1 \mathbf{X}) = \lambda_1$  es el autovalor de  $\boldsymbol{\Sigma}$  más grande, y  $\boldsymbol{\alpha}_1$  el autovalor correspondiente a  $\lambda_1$ .

La segunda CP es la que maximiza la varianza:

$$Var(\boldsymbol{\alpha}'_2 \mathbf{X}) = Cov(\boldsymbol{\alpha}'_2 \mathbf{X}, \boldsymbol{\alpha}'_2 \mathbf{X}) = \boldsymbol{\alpha}'_2 \boldsymbol{\Sigma} \boldsymbol{\alpha}_2 = \boldsymbol{\alpha}'_2 \lambda \boldsymbol{\alpha}_2.$$

Sujeto a las restricciones:

- $\|\boldsymbol{\alpha}_2\|^2 = 1$
- $Cov(\boldsymbol{\alpha}'_1 \mathbf{X}, \boldsymbol{\alpha}'_2 \mathbf{X}) = 0$

En la segunda restricción se observa que:

$$Cov(\boldsymbol{\alpha}'_1 \mathbf{X}, \boldsymbol{\alpha}'_2 \mathbf{X}) = \boldsymbol{\alpha}'_1 \boldsymbol{\Sigma} \boldsymbol{\alpha}_2 = \boldsymbol{\alpha}'_1 \lambda \boldsymbol{\alpha}_2 = \lambda \boldsymbol{\alpha}'_1 \boldsymbol{\alpha}_2 = 0.$$

Como  $\lambda$  es no negativo, en particular si  $\lambda > 0$  entonces se tiene que  $\boldsymbol{\alpha}'_1 \boldsymbol{\alpha}_2 = 0$ .

Aplicando el operador traspuesto a  $\boldsymbol{\alpha}'_1 \boldsymbol{\Sigma} \boldsymbol{\alpha}_2$  se tiene que:

$$Cov(\boldsymbol{\alpha}'_1 \mathbf{X}, \boldsymbol{\alpha}'_2 \mathbf{X}) = \boldsymbol{\alpha}'_1 \boldsymbol{\Sigma} \boldsymbol{\alpha}_2 = \boldsymbol{\alpha}'_2 \boldsymbol{\Sigma} \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}'_2 \lambda \boldsymbol{\alpha}_1 = \lambda \boldsymbol{\alpha}'_2 \boldsymbol{\alpha}_1 = 0.$$

Aplicando el razonamiento dado anteriormente se obtiene que  $\boldsymbol{\alpha}'_2 \boldsymbol{\alpha}_1 = 0$ .

Cualquiera de estos dos resultados puede usarse para especificar la correlación entre  $\boldsymbol{\alpha}'_1 \mathbf{X}$  y  $\boldsymbol{\alpha}'_2 \mathbf{X}$ . Considerando multiplicadores de Lagrange  $\lambda$  y  $\phi$  y el segundo de los resultados anteriores, entonces la cantidad a maximizar es:

$$\boldsymbol{\alpha}'_2 \boldsymbol{\Sigma} \boldsymbol{\alpha}_2 - \lambda(\boldsymbol{\alpha}'_2 \boldsymbol{\alpha}_2 - 1) - \phi \boldsymbol{\alpha}'_2 \boldsymbol{\alpha}_1.$$

Derivando con respecto a  $\boldsymbol{\alpha}'_2$  se obtiene:

$$\boldsymbol{\Sigma} \boldsymbol{\alpha}_2 - \lambda \boldsymbol{\alpha}_2 - \phi \boldsymbol{\alpha}_1 = 0.$$

Multiplicando por  $\boldsymbol{\alpha}'_1$  se obtiene:

$$\alpha_1' \Sigma \alpha_2 - \alpha_1' \lambda \alpha_2 - \alpha_1' \phi \alpha_1 = 0$$

Donde  $\alpha_1' \Sigma \alpha_2 = 0$  dado que  $Cov(\alpha_1' \mathbf{X}, \alpha_2' \mathbf{X}) = 0$ . Además:

$$\lambda \alpha_1' \alpha_2 = 0$$

$$\alpha_1' \alpha_1 = 1$$

$$\Rightarrow \phi = 0.$$

Por lo tanto

$$\Sigma \alpha_2 - \lambda \alpha_2 = 0$$

$$\Rightarrow (\Sigma - \lambda I_d) \alpha_2 = 0.$$

Por lo que  $\lambda$  es nuevamente un autovalor de  $\Sigma$  y  $\alpha_2$  el autovector correspondiente.

Nuevamente al considerar que la cantidad que se maximiza es:

$$\alpha_2' \Sigma \alpha_2 = \alpha_2' \lambda \alpha_2 = \lambda \alpha_2' \alpha_2 = \lambda_2.$$

Se observa que  $\lambda_2$  es tan grande como sea posible. Asumiendo que  $\Sigma$  no tiene autovalores repetidos, se tiene que  $\lambda_2 \neq \lambda_1$ , y como  $\lambda_1$  es el autovalor más grande,  $\lambda_2$  es el segundo mayor autovalor de  $\Sigma$  y  $\alpha_2$  el autovector correspondiente.

Siguiendo este proceso, se tiene que para la tercera, cuarta,...,  $d$ -ésima CP, los vectores  $\alpha_3, \alpha_4, \dots, \alpha_d$  son los autovectores de  $\Sigma$  correspondientes a los autovalores  $\lambda_3, \lambda_4, \dots, \lambda_d$ , los cuales satisfacen  $\lambda_3 > \lambda_4 > \dots > \lambda_d$  y además:

$$Var(\alpha_k' \mathbf{X}) = \lambda_k; \quad k = 1, 2, \dots, d.$$

### 1.3.1. Determinación del Número de Componentes Principales

Cuando se lleva a cabo un ACP, es necesario determinar el número de CP que tienen varianzas positivas. Este número corresponde a la dimensión del espacio en el que “caen” los datos.

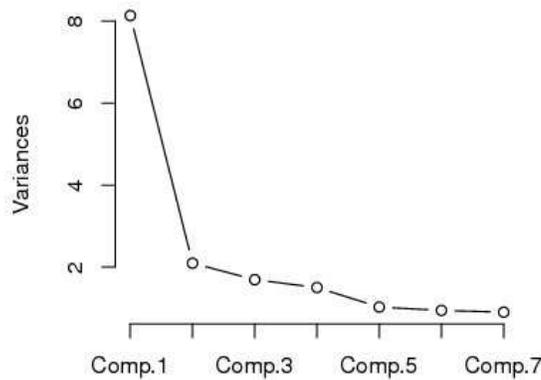
Para esto, se dispone de dos métodos, los cuales se basan en los autovalores de la matriz  $\Sigma$ .

**Método 1:** Se considera la cantidad:

$$\mathfrak{F}_q = \frac{\sum_{k=1}^q \lambda_k}{\sum_{k=1}^d \lambda_k} \quad ; \quad q < d; \quad 0 \leq \mathfrak{F}_q \leq 1$$

Entonces, el número de CP a elegir viene dado por el valor de  $q$  para el cual  $\mathfrak{F}_q > \alpha$ , donde  $\alpha \in [0, 1]$  es un valor elegido convenientemente.

**Método 2:** Se construye una gráfica situando los puntos  $(1, \lambda_1), (2, \lambda_2), \dots, (q, \lambda_q)$ . Cuando los puntos de la gráfica tienden a nivelarse, estos autovalores tienden a cero, por lo que pueden ser descartados. Por este método se supone que el número de CP a elegir viene dado por el autovalor grande más pequeño. Un ejemplo de este método se muestra en la siguiente figura:



Como podemos observar en la gráfica, a partir del autovalor 4 en adelante todos los demás autovalores son casi cero, lo que sugiere que el número de CP a elegir es 4.

Cabe destacar que en la práctica es conveniente considerar en forma simultánea ambos métodos.

## 1.4. Propiedades Algebraicas de las Componentes Principales y sus Implicaciones Estadísticas

Sea  $\mathbf{z}$  el vector cuyo  $k$ -ésimo elemento,  $z_k$ , es la  $k$ -ésima componente principal con la  $k$ -ésima varianza más grande donde  $k = 1, 2, \dots, d$ . Luego:

$$\mathbf{z} = \mathbf{A}'\mathbf{X}. \quad (1.1)$$

Donde  $\mathbf{A}$  es la matriz cuya  $k$ -ésima columna es el  $k$ -ésimo autovector de  $\mathbf{\Sigma}$ ,  $\boldsymbol{\alpha}_k$ . Por lo tanto las componentes principales se pueden definir por una transformación ortonormal lineal de  $\mathbf{X}$ .

Por la derivación de las componentes principales dada en la sección 1.2 se sabe que:

$$\mathbf{\Sigma}\mathbf{A} = \mathbf{A}\mathbf{\Lambda}. \quad (1.2)$$

Donde  $\mathbf{\Lambda}$  es una matriz diagonal cuyo  $k$ -ésimo elemento es el  $k$ -ésimo autovalor de  $\mathbf{\Sigma}$ ,  $\lambda_k$ , y además:

$$\lambda_k = \text{Var}(\boldsymbol{\alpha}'_k \mathbf{X}) = \text{Var}(z_k).$$

Como  $\mathbf{A}$  es ortogonal, entonces su matriz traspuesta es igual a su matriz inversa, por lo que podemos expresar  $\mathbf{\Sigma}$  y  $\mathbf{\Lambda}$  de las maneras siguientes:

$$\mathbf{A}'\mathbf{\Sigma}\mathbf{A} = \mathbf{\Lambda}. \quad (1.3)$$

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{\Lambda}\mathbf{A}'. \quad (1.4)$$

La transformación lineal ortonormal de  $\mathbf{X}$  que define a  $\mathbf{z}$  en (1.1) tiene una serie de propiedades óptimas, a saber:

**Propiedad A1:** para cualquier entero  $q$ ;  $1 \leq q \leq d$ ; considere la transformación ortonormal lineal

$$\mathbf{y} = \mathbf{B}'\mathbf{X}. \quad (1.5)$$

Donde  $\mathbf{y}$  es un vector de  $q$  elementos y  $\mathbf{B}'$  es una matriz de orden  $(q \times d)$ . Sea  $\mathbf{\Sigma}_y = \mathbf{B}'\mathbf{\Sigma}\mathbf{B}$  la matriz de covarianzas para  $\mathbf{y}$ . Entonces  $\text{Tr}(\mathbf{\Sigma}_y)$  es maximizada cuando  $\mathbf{B} = \mathbf{A}_q$ , donde  $\mathbf{A}_q$  es una matriz conformada por las primeras  $q$  columnas de  $\mathbf{A}$ .

**Demostración:** Sea  $\boldsymbol{\beta}_k$  la  $k$ -ésima columna de  $\mathbf{B}$ . Como las columnas de  $\mathbf{A}$  forman una base para el espacio de dimensión  $d$  se tiene que:

$$\boldsymbol{\beta}_k = \sum_{j=1}^d c_{jk} \boldsymbol{\alpha}_j; \quad k = 1, 2, \dots, q.$$

Donde  $c_{jk}$ ;  $k = 1, 2, \dots, q$ ;  $j = 1, 2, \dots, d$  son constantes escogidas convenientemente. Luego:

$$\mathbf{B} = \mathbf{A}\mathbf{C}.$$

Donde  $\mathbf{C}$  es la matriz cuyo  $(j, k)$ -ésimo elemento es  $c_{jk}$ . Entonces:

$$\begin{aligned} \mathbf{B}'\Sigma\mathbf{B} &= \mathbf{C}'\mathbf{A}'\Sigma\mathbf{A}\mathbf{C} \\ &= \mathbf{C}'\Lambda\mathbf{C} \\ &= \sum_{j=1}^d \mathbf{c}'_j \lambda_j \mathbf{c}_j \\ &= \sum_{j=1}^d \lambda_j \mathbf{c}'_j \mathbf{c}_j. \end{aligned}$$

Donde  $\mathbf{c}'_j$  es la  $j$ -ésima fila de  $\mathbf{C}$ . Por lo tanto:

$$\begin{aligned} \text{Tr}(\mathbf{B}'\Sigma\mathbf{B}) &= \text{Tr}\left(\sum_{j=1}^d \lambda_j \mathbf{c}'_j \mathbf{c}_j\right) \\ &= \sum_{j=1}^d \lambda_j \text{Tr}(\mathbf{c}'_j \mathbf{c}_j) \\ &= \sum_{j=1}^d \lambda_j \mathbf{c}_j \mathbf{c}'_j \\ &= \sum_{j=1}^d \sum_{k=1}^q \lambda_j c_{jk}^2. \end{aligned} \tag{1.6}$$

Ahora bien:

$$\begin{aligned} \mathbf{A}'\mathbf{B} &= \mathbf{C} \\ \Rightarrow \mathbf{C}'\mathbf{C} &= \mathbf{B}'\mathbf{A}\mathbf{A}'\mathbf{B} = \mathbf{B}'\mathbf{B} = \mathbf{I}_q. \end{aligned}$$

Luego, las columnas de  $\mathbf{C}$  son ortonormales (puesto que  $\mathbf{A}$  es ortogonal, y las columnas de  $\mathbf{B}$  son ortonormales) y además:

$$\sum_{j=1}^d \sum_{k=1}^q c_{jk}^2 = q. \quad (1.7)$$

Como  $\mathbf{C}$  es una matriz de orden  $(d \times q)$ , se puede decir que esta formada por las primeras  $q$  columnas de una cierta matriz  $\mathbf{D}$  de orden  $(d \times d)$ . Las filas de  $\mathbf{D}$  son ortonormales y satisfacen que  $\mathbf{d}'_j \mathbf{d}_j = 1$ ;  $j = 1, 2, \dots, d$ . Como las filas de  $\mathbf{C}$  están compuestas por los primeros  $q$  elementos de las filas de  $\mathbf{D}$  se tiene que  $\mathbf{c}'_j \mathbf{c}_j \leq 1$ ,  $j = 1, 2, \dots, d$ . Esto es:

$$\sum_{k=1}^q c_{jk}^2 \leq 1. \quad (1.8)$$

Ahora bien,  $\sum_{k=1}^q c_{jk}^2$  es el coeficiente de  $\lambda_j$  en (1.6). La suma de estos coeficientes es  $q$  en (1.7) y

$$\max_j \left( \sum_{k=1}^q c_{jk}^2 \right) = 1.$$

Como  $\lambda_1 > \lambda_2 > \dots > \lambda_d$ , entonces es posible encontrar valores  $c_{jk}$  tales que:

$$\sum_{k=1}^q c_{jk}^2 = \begin{cases} 1, & j = 1, 2, \dots, q \\ 0, & j = q + 1, q + 2, \dots, d \end{cases} \quad (1.9)$$

Entonces la cantidad  $\sum_{j=1}^d \left( \sum_{k=1}^q c_{jk}^2 \right) \lambda_j$  alcanza su valor máximo. Luego, tomando  $\mathbf{B}' = \mathbf{A}'_q$  se tiene que:

$$c_{jk} = \begin{cases} 1, & 1 \leq j = k \leq q \\ 0, & \text{otro caso} \end{cases}$$

Lo cual satisface (1.9). Por lo que  $Tr(\Sigma_y)$  alcanza su valor máximo cuando  $\mathbf{B} = \mathbf{A}_q$ .

**Propiedad A2:** Consideremos la expresión ortonormal descrita en la propiedad anterior.

$$\mathbf{y} = \mathbf{B}'\mathbf{X}.$$

Entonces  $Tr(\Sigma_y)$  es minimizada cuando  $\mathbf{B} = \mathbf{A}_q^*$ , donde  $\mathbf{A}_q^*$  es una matriz conformada por las últimas  $q$  columnas de  $\mathbf{A}$ .

La demostración de esta propiedad es análoga a la utilizada para demostrar la propiedad A1, considerando el hecho de que la derivación de las CP descrita en la sección (1.2) se puede convertir con la finalidad de buscar funciones lineales cuyas varianzas sean mínimas, y que esten sujetas a que se correlacionen con las funciones lineales anteriores. En este caso, también se buscan los autovectores de la matriz  $\Sigma$ , pero empezando por los más pequeños.

Esta propiedad es importante desde el punto de vista estadístico, puesto que los últimos CP tienen varianza mínima, y esta característica los hace bastante útiles para determinar relaciones casi-constantas entre los elementos de  $\mathbf{X}$ , para realizar análisis de regresión, determinar subconjuntos de variables de  $\mathbf{X}$  y determinación de valores atípicos.

**Propiedad A3:** *La descomposición espectral.*

$$\Sigma = \lambda_1 \alpha_1 \alpha_1' + \lambda_2 \alpha_2 \alpha_2' + \cdots + \lambda_d \alpha_d \alpha_d' \quad (1.10)$$

**Demostración:** Se sabe de (1.4) que:

$$\Sigma = \mathbf{A} \mathbf{\Lambda} \mathbf{A}'.$$

Desarrollando el lado derecho de esta igualdad se obtiene:

$$\begin{aligned} \Sigma &= \begin{pmatrix} \alpha_{11} & \alpha_{21} & \cdots & \alpha_{d1} \\ \alpha_{12} & \alpha_{22} & \cdots & \alpha_{d2} \\ \vdots & \vdots & & \vdots \\ \alpha_{1d} & \alpha_{2d} & \cdots & \alpha_{dd} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_d \end{pmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1d} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2d} \\ \vdots & \vdots & & \vdots \\ \alpha_{d1} & \alpha_{d2} & \cdots & \alpha_{dd} \end{pmatrix} \\ &= \begin{pmatrix} \alpha_{11} \lambda_1 & \alpha_{21} \lambda_2 & \cdots & \alpha_{d1} \lambda_d \\ \alpha_{12} \lambda_1 & \alpha_{22} \lambda_2 & \cdots & \alpha_{d2} \lambda_d \\ \vdots & \vdots & & \vdots \\ \alpha_{1d} \lambda_1 & \alpha_{2d} \lambda_2 & \cdots & \alpha_{dd} \lambda_d \end{pmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1d} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2d} \\ \vdots & \vdots & & \vdots \\ \alpha_{d1} & \alpha_{d2} & \cdots & \alpha_{dd} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{k=1}^d \alpha_{k1} \lambda_k \alpha_{k1} & \sum_{k=1}^d \alpha_{k1} \lambda_k \alpha_{k2} & \cdots & \sum_{k=1}^d \alpha_{k1} \lambda_k \alpha_{kd} \\ \sum_{k=1}^d \alpha_{k2} \lambda_k \alpha_{k1} & \sum_{k=1}^d \alpha_{k2} \lambda_k \alpha_{k2} & \cdots & \sum_{k=1}^d \alpha_{k2} \lambda_k \alpha_{kd} \\ \vdots & \vdots & & \vdots \\ \sum_{k=1}^d \alpha_{kd} \lambda_k \alpha_{k1} & \sum_{k=1}^d \alpha_{kd} \lambda_k \alpha_{k2} & \cdots & \sum_{k=1}^d \alpha_{kd} \lambda_k \alpha_{kd} \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
&= \begin{pmatrix} \lambda_1 \alpha_{11} \alpha_{11} & \lambda_1 \alpha_{11} \alpha_{12} & \cdots & \lambda_1 \alpha_{11} \alpha_{1d} \\ \lambda_1 \alpha_{12} \alpha_{11} & \lambda_1 \alpha_{12} \alpha_{12} & \cdots & \lambda_1 \alpha_{12} \alpha_{1d} \\ \vdots & \vdots & & \vdots \\ \lambda_1 \alpha_{1d} \alpha_{11} & \lambda_1 \alpha_{1d} \alpha_{12} & \cdots & \lambda_1 \alpha_{1d} \alpha_{1d} \end{pmatrix} + \cdots + \begin{pmatrix} \lambda_d \alpha_{d1} \alpha_{d1} & \lambda_d \alpha_{d1} \alpha_{d2} & \cdots & \lambda_d \alpha_{d1} \alpha_{dd} \\ \lambda_d \alpha_{d2} \alpha_{d1} & \lambda_d \alpha_{d2} \alpha_{d2} & \cdots & \lambda_d \alpha_{d2} \alpha_{dd} \\ \vdots & \vdots & & \vdots \\ \lambda_d \alpha_{dd} \alpha_{d1} & \lambda_d \alpha_{dd} \alpha_{d2} & \cdots & \lambda_d \alpha_{dd} \alpha_{dd} \end{pmatrix} \\
&= \lambda_1 \boldsymbol{\alpha}_1 \boldsymbol{\alpha}'_1 + \lambda_2 \boldsymbol{\alpha}_2 \boldsymbol{\alpha}'_2 + \cdots + \lambda_d \boldsymbol{\alpha}_d \boldsymbol{\alpha}'_d.
\end{aligned}$$

Luego,  $\boldsymbol{\Sigma} = \sum_{k=1}^d \lambda_k \boldsymbol{\alpha}_k \boldsymbol{\alpha}'_k$ .

Este resultado tiene las siguientes implicaciones:

- Las varianzas de los elementos de  $\mathbf{X}$  se pueden descomponer en las contribuciones decrecientes debidas a cada componente principal.

$$Var(x_j) = \sum_{k=1}^d \lambda_k \alpha_{kj}^2$$

- Toda la matriz de covarianzas se puede descomponer en las contribuciones debidas a cada componente principal (aunque no de manera estrictamente decreciente)

$$Cov(X_m, X_n) = \sum_{k=1}^d \lambda_k \alpha_{km} \alpha_{kn}; \quad m = 1, 2, \dots, d; \quad n = 1, 2, \dots, d.$$

- Las componentes principales son eficientes para explicar tanto los elementos que se encuentran en la diagonal de  $\boldsymbol{\Sigma}$ , como los que se encuentran fuera de ella. Esto es particularmente cierto cuando en la derivación de las CP se utiliza una matriz de correlaciones, pero cuando se utiliza una matriz de covarianzas y las varianzas de  $\mathbf{X}$  son muy diferentes el resultado es menos válido.
- La matriz de covarianzas (ó correlaciones) se puede construir con exactitud dando los coeficientes y las varianzas de las primeras  $r$  CP, donde  $r$  es el rango de la matriz de covarianzas.

**Propiedad A4:** Consideremos de nuevo la expresión ortonormal descrita en las propiedades A1 y A2.

$$\mathbf{y} = \mathbf{B}'\mathbf{X}.$$

Si  $\text{Det}(\boldsymbol{\Sigma}_y)$  denota el determinante de la matriz de covarianzas para  $\mathbf{y}$ , entonces  $\text{Det}(\boldsymbol{\Sigma}_y)$  se maximiza cuando  $\mathbf{B} = \mathbf{A}_q$ .

**Demostración:** Sea  $k$  un valor entero tal que  $1 \leq k \leq q$ . Sea  $S_k$  el subespacio generado por los vectores ortogonales  $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_{k-1}$ . Luego:

$$\dim(S_k) = k - 1.$$

Donde  $\dim(S_k)$  denota la dimensión de  $S_k$ . El  $k$ -ésimo autovalor  $\lambda_k$  de  $\boldsymbol{\Sigma}$  satisface:

$$\lambda_k = \sup_{\substack{\boldsymbol{\alpha} \in S_k \\ \boldsymbol{\alpha} \neq 0}} \left\{ \frac{\boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}}{\boldsymbol{\alpha}'\boldsymbol{\alpha}} \right\}$$

Supongamos que  $\mu_1 > \mu_2 > \dots > \mu_q$  son los autovalores de  $\boldsymbol{\Sigma}_y$  y que  $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \dots, \boldsymbol{\gamma}_q$  son los autovectores correspondientes. Sea  $T_k$  el subespacio generado por los vectores ortogonales  $\boldsymbol{\gamma}_{k+1}, \boldsymbol{\gamma}_{k+2}, \dots, \boldsymbol{\gamma}_q$ , con  $\dim(T_k) = d - [(k-1) + 1] = d - k$ . Luego, para cualquier vector  $\boldsymbol{\gamma} \in T_k$  no nulo se tiene que:

$$\frac{\boldsymbol{\gamma}'\boldsymbol{\Sigma}_y\boldsymbol{\gamma}}{\boldsymbol{\gamma}'\boldsymbol{\gamma}} \geq \mu_k.$$

Sea  $\tilde{S}_k$  el subespacio generado por los vectores de la forma  $\mathbf{B}\boldsymbol{\gamma}$  con  $\boldsymbol{\gamma} \in T_k$ . Entonces:

$$\dim(\tilde{S}_k) = \dim(T_k) = d - k.$$

Por un resultado del álgebra lineal (el Teorema de la Dimensión) se tiene que:

$$\dim(S_k \cap \tilde{S}_k) + \dim(S_k + \tilde{S}_k) = \dim(S_k) + \dim(\tilde{S}_k).$$

Pero  $\dim(S_k \cap \tilde{S}_k) \leq d$ ;  $\dim(S_k) = k - 1$  y  $\dim(\tilde{S}_k) = d - k$ . Por lo tanto:

$$\dim(S_k \cap \tilde{S}_k) \geq 1.$$

Por consiguiente, el vector  $\boldsymbol{\alpha} \in S_k$  no nulo es de la forma  $\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\gamma}$  para  $\boldsymbol{\gamma} \in T_k$ , y se deduce que:

$$\mu_k \leq \frac{\boldsymbol{\gamma}'\boldsymbol{\Sigma}_y\boldsymbol{\gamma}}{\boldsymbol{\gamma}'\boldsymbol{\gamma}} = \frac{\boldsymbol{\gamma}'\mathbf{B}'\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\gamma}}{\boldsymbol{\gamma}'\mathbf{B}'\mathbf{B}\boldsymbol{\gamma}} = \frac{\boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}}{\boldsymbol{\alpha}'\boldsymbol{\alpha}} \leq \lambda_k.$$

Así, el  $k$ -ésimo autovalor de  $\boldsymbol{\Sigma}_y$  es menor o igual al  $k$ -ésimo autovalor de  $\boldsymbol{\Sigma}$  para  $k = 1, 2, \dots, q$ . Luego:

$$\text{Det}(\boldsymbol{\Sigma}_y) = \prod_{k=1}^q \mu_k \leq \prod_{k=1}^q \lambda_k.$$

Pero si  $\mathbf{B} = \mathbf{A}_q$  entonces los autovalores de  $\boldsymbol{\Sigma}_y$  son  $\lambda_1, \lambda_2, \dots, \lambda_q$ , de modo que:

$$\text{Det}(\boldsymbol{\Sigma}_y) = \prod_{k=1}^q \lambda_k.$$

Por lo tanto,  $\text{Det}(\boldsymbol{\Sigma}_y)$  se maximiza cuando  $\mathbf{B} = \mathbf{A}_q$ .

Este resultado es estadísticamente importante puesto que la varianza generalizada (determinante de una matriz de covarianzas) puede ser usada como medida única de la propagación de un vector aleatorio multivariante. Por ejemplo, si  $\mathbf{X}$  tiene distribución normal multivariada, entonces la raíz cuadrada de la varianza generalizada es proporcional al volumen que encierra una proporción fija en la distribución de probabilidad de  $\mathbf{X}$ , y las primeras componentes principales son por lo tanto  $q$  funciones lineales de  $\mathbf{X}$  cuya distribución de probabilidad conjunta tiene contornos de probabilidad fija que se adjuntan al volumen máximo.

**Propiedad A5:** *Supongamos que queremos predecir cada variable aleatoria  $X_j$  de  $\mathbf{X}$  por una función lineal de  $\mathbf{y}$ , donde  $\mathbf{y} = \mathbf{B}'\mathbf{X}$ . Si  $\sigma_j^2$  es la varianza residual de  $\mathbf{y}$  en la predicción de  $X_j$ ,*

*entonces  $\sum_{j=1}^d \sigma_j^2$  se minimiza cuando  $\mathbf{B} = \mathbf{A}_q$ .*

La implicación estadística de este resultado es que si se desea obtener el mejor predictor lineal de  $\mathbf{X}$  en un subespacio de dimensión  $q$  con el objeto de minimizar la suma de las varianzas residuales de los elementos de  $\mathbf{X}$ , entonces el subespacio óptimo a considerar es el generado por las primeras CP.

## 1.5. Propiedades Geométricas de las Componentes Principales y sus Implicaciones Estadísticas

**Propiedad G1:** Sea

$$\mathbf{X}'\Sigma^{-1}\mathbf{X} = ctte. \quad (1.11)$$

Una familia de elipsoides de dimensión  $d$ . Entonces las CP definen los ejes principales de estos elipsoides.

**Demostración:** Se sabe de (1.1) que:

$$\mathbf{z} = \mathbf{A}'\mathbf{X}.$$

Como  $\mathbf{A}$  es ortogonal, entonces la transformación inversa viene dada por:

$$\mathbf{Az} = \mathbf{X}. \quad (1.12)$$

Sustituyendo en (1.11) se obtiene:

$$\begin{aligned} (\mathbf{Az})'\Sigma^{-1}(\mathbf{Az}) &= ctte \\ \Rightarrow \mathbf{z}'\mathbf{A}'\Sigma^{-1}\mathbf{Az} &= ctte. \end{aligned}$$

Como  $\Sigma^{-1}$  y  $\Sigma$  tienen los mismos autovectores y los autovalores de  $\Sigma^{-1}$  son los recíprocos de los de  $\Sigma$  (asumiendo que todos ellos son estrictamente positivos), entonces se tiene de (1.3) que:

$$\mathbf{A}'\Sigma^{-1}\mathbf{A} = \mathbf{\Lambda}^{-1}.$$

Y por lo tanto:

$$\mathbf{z}'\mathbf{\Lambda}^{-1}\mathbf{z} = ctte$$

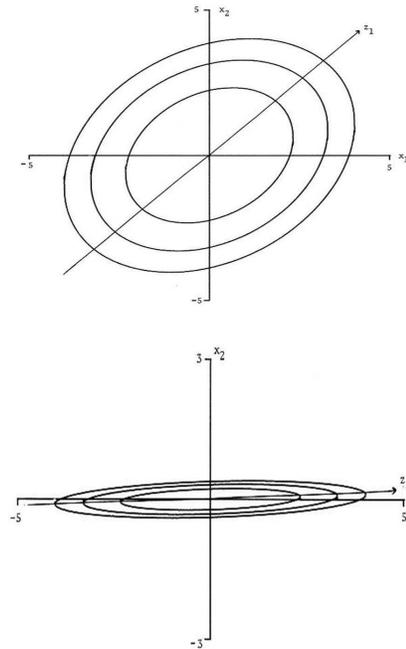
$$\Rightarrow (z_1, z_2, \dots, z_d) \begin{pmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_d} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_d \end{pmatrix} = ctte$$

$$\Rightarrow \left( \frac{z_1}{\lambda_1}, \frac{z_2}{\lambda_2}, \dots, \frac{z_d}{\lambda_d} \right) \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_d \end{pmatrix} = ctte$$

$$\Rightarrow \sum_{k=1}^d \frac{z_k^2}{\lambda_k} = ctte. \quad (1.13)$$

Esta es la ecuación de un elipsoide en términos de sus ejes principales. Aquí se observa que los valores de los semi-ejes del elipsoide son  $\lambda_1^{\frac{1}{2}}, \lambda_2^{\frac{1}{2}}, \dots, \lambda_d^{\frac{1}{2}}$ .

La implicación estadística de este resultado es que si  $\mathbf{X}$  tiene una distribución normal multivariante, entonces los elipsoides dados en (1.11) definen contornos de probabilidad constante para la distribución de  $\mathbf{X}$ . El primer eje principal de estos elipsoides (el más grande) define la dirección en la cual la varianza estadística es mayor, lo cual es otra manera de expresar la definición algebraica de la primera CP dada en la sección 1.2. El segundo eje principal maximiza la varianza, sujeto a la restricción de ser ortogonal con el primero, lo que corresponde nuevamente a la definición algebraica.



En las figuras mostradas anteriormente se ilustra la dirección de la primera CP que define el primer eje principal de los elipsoides de probabilidad constante.

**Propiedad G2** Supongamos que  $\mathbf{X}_1, \mathbf{X}_2$  son dos vectores aleatorios independientes con la misma distribución de probabilidad. Supongamos que ambos vectores están sujetos a la misma transformación lineal

$$\mathbf{y}_i = \mathbf{B}'\mathbf{X}_i; \quad i = 1, 2.$$

Si  $\mathbf{B}$  es una matriz de orden  $(d \times q)$  de columnas ortonormales escogidas para maximizar  $\mathbb{E}[(\mathbf{y}_1 - \mathbf{y}_2)'(\mathbf{y}_1 - \mathbf{y}_2)]$ , entonces  $\mathbf{B} = \mathbf{A}_q$ .

**Demostración:** Como  $\mathbf{X}_1$  y  $\mathbf{X}_2$  tienen la misma distribución de probabilidad, entonces ambos tienen vector de medias  $\boldsymbol{\mu}$  y matriz de covarianzas  $\boldsymbol{\Sigma}$ . Por lo tanto los vectores  $\mathbf{y}_1$  y  $\mathbf{y}_2$  tienen la misma media  $\mathbf{B}'\boldsymbol{\mu}$  y matriz de covarianzas  $\boldsymbol{\Sigma}_y$  respectivamente. Luego:

$$\begin{aligned} \mathbb{E}[(\mathbf{y}_1 - \mathbf{y}_2)'(\mathbf{y}_1 - \mathbf{y}_2)] &= \mathbb{E}\{[(\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu}) - (\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})]' [(\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu}) - (\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})]\} \\ &= \mathbb{E}\{[(\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu})' - (\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})'] [(\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu}) - (\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})]\} \\ &= \mathbb{E}\{[(\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu})' - (\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})'](\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu}) \\ &\quad - [(\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu})' - (\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})'](\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})\} \\ &= \mathbb{E}\{(\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu})'(\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu}) - (\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})'(\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu}) \\ &\quad - (\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu})'(\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu}) + (\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})'(\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})\} \end{aligned}$$

Como  $\mathbf{X}_1, \mathbf{X}_2$  son vectores independientes, entonces  $\mathbf{y}_1, \mathbf{y}_2, (\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu}), (\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})$  también son vectores independientes y por lo tanto:

$$\begin{aligned} (\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})'(\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu}) &= 0; \\ (\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu})'(\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu}) &= 0. \end{aligned}$$

Luego:

$$\begin{aligned} \mathbb{E}[(\mathbf{y}_1 - \mathbf{y}_2)'(\mathbf{y}_1 - \mathbf{y}_2)] &= \mathbb{E}\{(\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu})'(\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu}) + (\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})'(\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})\} \\ &= \mathbb{E}\{(\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu})'(\mathbf{y}_1 - \mathbf{B}'\boldsymbol{\mu})\} + \mathbb{E}\{(\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})'(\mathbf{y}_2 - \mathbf{B}'\boldsymbol{\mu})\} \end{aligned}$$

Sea  $\mathbf{z}_i \in \mathbb{R}^q$  tal que  $\mathbf{z}_i = \mathbf{y}_i - \mathbf{B}'\boldsymbol{\mu}; i = 1, 2$ . Entonces:

$$\begin{aligned}
\mathbb{E}\{\mathbf{z}'_i \mathbf{z}_i\} &= \mathbb{E}\{z_{i1}^2 + z_{i2}^2 + \cdots + z_{iq}^2\} \\
&= \mathbb{E}\{Tr[\mathbf{z}_i \mathbf{z}'_i]\} \\
&= Tr\{\mathbb{E}[\mathbf{z}_i \mathbf{z}'_i]\} \\
&= Tr(\boldsymbol{\Sigma}_y).
\end{aligned}$$

Pero  $Tr(\boldsymbol{\Sigma}_y)$  se maximiza cuando  $\mathbf{B} = \mathbf{A}_q$  por la propiedad A1.

Existe una propiedad equivalente la cual enuncia que bajo las mismas hipótesis de la propiedad G2, la cantidad:

$$Det\{\mathbb{E}[(\mathbf{y}_1 - \mathbf{y}_2)(\mathbf{y}_1 - \mathbf{y}_2)']\},$$

se maximiza cuando  $\mathbf{B} = \mathbf{A}_q$ . Esta propiedad muestra que  $\mathbf{B} = \mathbf{A}_q$  hace a la varianza generalizada de  $\mathbf{y}_1 - \mathbf{y}_2$  tan grande como sea posible. La varianza generalizada puede ser vista como una medida alternativa de la distancia entre  $\mathbf{y}_1$  y  $\mathbf{y}_2$  en un espacio de dimensión  $q$ , aunque es una medida un poco menos intuitiva que la distancia euclídea al cuadrado.

La propiedad G2 se puede “invertir” en el sentido de que si  $\mathbb{E}[(\mathbf{y}_1 - \mathbf{y}_2)(\mathbf{y}_1 - \mathbf{y}_2)']$  ó bien  $Det\{\mathbb{E}[(\mathbf{y}_1 - \mathbf{y}_2)(\mathbf{y}_1 - \mathbf{y}_2)']\}$  se deben minimizar, esto se lleva a cabo tomando  $\mathbf{B} = \mathbf{A}_q^*$ .

La implicación estadística de este resultado es que la distancia euclídea al cuadrado entre dos vectores de  $d$  variables aleatorias con la misma distribución en un subespacio de dimensión  $q$  se puede hacer tan grande como sea posible si el subespacio es generado por las primeras  $q$  CP.

## 1.6. Componentes Principales Usando una Matriz de Correlaciones

Recordemos que  $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{pmatrix} \in \mathbb{R}^d$  es un vector aleatorio, es decir:

$$X_i : \Omega \rightarrow \mathbb{R}; \forall i = 1, 2, \dots, d.$$

Sea  $\mathbf{R} \in M_{d \times d}$  la matriz de correlaciones del vector aleatorio  $\mathbf{X}$ . Esto es:

$$r_{ij} = \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i)Var(X_j)}}; i, j = 1, 2, \dots, d.$$

La derivación y las propiedades de las CP descritas anteriormente están basadas en los autovalores y autovectores de una matriz de covarianzas. En la práctica es más común definir las CP de la siguiente manera:

$$\mathbf{z} = \mathbf{A}'\widehat{\mathbf{X}}. \quad (1.14)$$

Donde  $\mathbf{A}$  es la matriz cuya  $k$ -ésima columna es el  $k$ -ésimo autovector de la matriz de correlaciones, y  $\widehat{\mathbf{X}}$  esta formado por variables estandarizadas de la forma:

$$\widehat{X}_j = \frac{X_j - \mu_j}{\sigma_{jj}^{\frac{1}{2}}}; \quad j = 1, 2, \dots, d.$$

$X_j$  es el  $j$ -ésimo elemento de  $\mathbf{X}$ , y  $\sigma_{jj}$  es la varianza de  $X_j$ . Luego la matriz de correlaciones de  $\mathbf{X}$  es la matriz de covarianzas de  $\widehat{\mathbf{X}}$ . Las propiedades que se discutieron en la secciones anteriores también son válidas para las matrices de correlaciones.

Se puede llegar a suponer que las CP de una matriz de correlaciones se pueden obtener de manera directa a partir de las CP de la matriz de covarianzas, pero esto no es cierto. Si bien  $\mathbf{X}$  y  $\widehat{\mathbf{X}}$  se relacionan a través de una simple transformación, los autovalores y autovectores de la matriz de correlaciones no se relacionan de manera simple con los autovalores y autovectores correspondientes de la matriz de covarianzas. En particular, si las CP encontradas a partir de la matriz de correlaciones se expresaran en términos de  $\mathbf{X}$  por la transformación inversa de  $\mathbf{X}$  a  $\widehat{\mathbf{X}}$ , entonces estas CP no serian las mismas que las encontradas a partir de  $\mathbf{\Sigma}$ , excepto en circunstancias muy especiales. Si bien las CP son invariantes bajo transformaciones ortogonales de  $\mathbf{X}$ , generalmente no son invariantes bajo otras transformaciones, y la transformación de  $\mathbf{X}$  a  $\widehat{\mathbf{X}}$  no es ortogonal. Por lo tanto, las CP de la matriz de covarianzas y las CP de la matriz de correlaciones no pueden derivarse unas de las otras.

A continuación se discuten de manera breve tres propiedades interesantes de las CP derivadas de una matriz de correlaciones.

La primera propiedad es que las CP dependen no de los valores absolutos de las correlaciones, sino de sus cocientes. Esto es así porque la multiplicación de todos los elementos fuera de la diagonal de la matriz de correlaciones por la misma constante deja a los autovectores de la matriz invariantes.

La segunda propiedad es que si en vez de la restricción de normalización  $\boldsymbol{\alpha}'_k \boldsymbol{\alpha}_k = 1$  se utiliza:

$$\widetilde{\boldsymbol{\alpha}}'_k \widetilde{\boldsymbol{\alpha}}_k = \lambda_k; \quad k = 1, 2, \dots, d. \quad (1.15)$$

Entonces  $\tilde{\alpha}_{kj}$ , el  $j$ -ésimo elemento de  $\tilde{\alpha}_k$ , es la correlación entre la  $j$ -ésima variable estandarizada  $\hat{X}_j$  y la  $k$ -ésima CP. Para ver esto, se puede ver que para  $k = 1, 2, \dots, d$ :

$$\tilde{\alpha}_k = \lambda_k^{\frac{1}{2}} \alpha_k; \quad Var(z_k) = \lambda_k.$$

Y el vector  $\Sigma \alpha_k$  de  $d$  elementos tiene como  $j$ -ésimo elemento la covarianza entre  $\hat{X}_j$  y  $z_k$ . Como  $\Sigma \alpha_k = \lambda_k \alpha_k$ , entonces la covarianza entre  $\hat{X}_j$  y  $z_k$  es  $\lambda_k \alpha_{kj}$ .

También se tiene que  $Var(\hat{X}_j) = 1$  y la correlación entre  $\hat{X}_j$  y  $z_k$  es:

$$\frac{\lambda_k \alpha_{kj}}{[Var(\hat{X}_j)Var(z_k)]^{\frac{1}{2}}} = \lambda_k^{\frac{1}{2}} \alpha_{kj} = \tilde{\alpha}_{kj}.$$

La tercera propiedad se enuncia de la siguiente manera:

**Propiedad A6:** Para cualquier entero  $q$ ,  $1 \leq q \leq d$ , considere la transformación lineal ortonormal

$$\mathbf{y} = \mathbf{B}'\mathbf{X}. \tag{1.16}$$

Sea  $R_{jq}^2$  el coeficiente de correlación múltiple al cuadrado entre  $X_j$  y las  $q$  variables  $y_1, y_2, \dots, y_q$  definidas por los elementos de  $\mathbf{y}$ . El valor

$$SC = \sum_{j=1}^d R_{jq}^2,$$

se maximiza cuando  $y_1, y_2, \dots, y_q$  son las primeras  $q$  CP de la matriz de correlaciones. El valor  $SC$  maximizado es igual a la suma de los  $q$  autovalores más grandes de la matriz de correlaciones.

Como las componentes principales son no correlacionadas, el valor de  $\sum_{j=1}^d R_{jq}^2$  se puede escribir como:

$$SC = \sum_{j=1}^d \sum_{k=1}^q cr_{jk}^2.$$

Donde  $cr_{jk}^2$  es la correlación al cuadrado entre la  $j$ -ésima variable y la  $k$ -ésima CP. Este valor se maximiza sucesivamente para  $q = 1, 2, \dots, d$  cuando las CP son derivadas de una matriz de correlaciones.

## 1.7. Componentes Principales con Varianzas Nulas ó con Varianzas Iguales

A continuación se discutirán de manera breve dos problemas que pueden presentarse en la teoría, pero que son poco comunes en la práctica:

**Problema 1:** *Cuando se tienen varianzas iguales.*

La igualdad de algunos autovalores, y por lo tanto la igualdad de algunas varianzas de CP, ocurre para ciertas matrices. La consecuencia de esto es que para un grupo de  $q$  autovalores iguales, los  $q$  autovectores correspondientes generan un subespacio único de dimensión  $q$ . Aun siendo ortogonales entre sí, estos vectores son arbitrarios. Por ejemplo, desde el punto de vista geométrico, si  $q = 2$  ó  $q = 3$  entonces los ejes principales de un círculo (ó una esfera) no pueden ser definidos de manera única; lo mismo ocurre si  $q > 3$ . Así las CP que corresponden a autovalores iguales no se definen de manera única. Desde el punto de vista estadístico, se vuelve realmente difícil hacer cualquier tipo de inferencias si se presenta este problema.

**Problema 2:** *Cuando se tienen varianzas iguales a cero.*

Esta complicación surge con más frecuencia que el problema anterior, pero aún así sigue siendo inusual. Si  $q$  autovalores son iguales a cero, entonces el rango de  $\Sigma$  es  $(d - q)$ , y a causa de esto se necesita hacer modificaciones a algunas propiedades dadas en la sección 1.4. Cualquier CP con varianza cero define una relación lineal constante entre los elementos de  $\mathbf{X}$ . Si tal relación existe, entonces esto implica que una variable es redundante, ya que su valor puede determinarse con exactitud a partir de los valores de las otras variables que aparecen en la relación. Sin embargo, si se buscan las posibles relaciones lineales antes de hacer el ACP, entonces es posible reducir el número de variables de  $d$  a  $(d - q)$  sin pérdida de la información. Alternativamente, se puede usar cualquier relación lineal o cuasi-lineal dadas por las últimas CP para seleccionar un subconjunto de variables que contengan la mayor parte de la información disponible en la totalidad de las variables originales.

Una observación general es que el número de autovalores iguales a cero de una matriz de correlaciones es igual al número de autovalores iguales a cero de la matriz de covarianzas correspondiente, de modo que las relaciones lineales entre los elementos de  $\mathbf{X}$  claramente implican a las relaciones lineales entre las variables estandarizadas y viceversa. Sin embargo, la equivalencia no es la misma cuando se consideran CP de varianzas iguales. La igualdad de algunos autovalores de la matriz de covarianzas no necesariamente implica que cualquiera de los autovalores de la matriz de correlaciones sean iguales, y viceversa.

# Capítulo 2

## Bootstrap

### 2.1. Introducción

En este capítulo se definirá el método Bootstrap, como parte de los métodos de remuestreo “Monte Carlo”. Se darán conceptos básicos involucrados con el método Bootstrap (Función de Distribución Empírica, muestra bootstrap, réplica bootstrap y estimación ideal bootstrap). Se dará un ejemplo de la estimación Bootstrap del error típico de la media aritmética. Se definirá el Principio Plug-in, se explicará cuál es su relación con la Función de Distribución Empírica y cómo resulta útil el uso del Principio Plug-in para estimar el error típico de la media aritmética. Se hará mención de manera breve a la idea de la estimación bootstrap del error típico para cualquier estadístico  $\hat{\theta}$  y a la determinación del valor  $B$  de réplicas bootstrap que se usan en la estimación de tal error. Finalmente, se explicará en detalle cómo el método Bootstrap resulta útil para determinar cuán exacto es el porcentaje de variación o varianza obtenido en ACP que es capturado por la primera componente principal.

### 2.2. Definición de Bootstrap

El Bootstrap es una técnica de remuestreo que se basa en un gran número de cálculos repetitivos para estimar la forma de la distribución de algún parámetro de interés  $\theta$ . Con esta técnica se obtienen medidas de precisión de estimaciones estadísticas (errores típicos, intervalos de confianza, etc.), sin necesidad de realizar fuertes hipótesis distribucionales ni fórmulas analíticas. Su fuerza radica en la potencia y rapidez de las computadoras. El Bootstrap realiza lo que se denomina el “muestreo Monte Carlo”, que es un remuestreo con reposición de los datos obtenidos en una muestra. Se ejecuta el remuestreo un gran número de veces para generar una estimación empírica de la distribución muestral del estadístico  $\hat{\theta}$ .

Sea  $\mathbf{x} = x_1, x_2, \dots, x_n$  una muestra aleatoria con función de probabilidad  $F$  desconocida. A partir de esta muestra se construye una “distribución de probabilidad empírica”,  $\hat{F}$ , que es la

función que asigna probabilidad  $1/n$  a cada punto de  $\mathbf{x}$ . Es decir:

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{x_i \leq x\}}(x),$$

donde  $I_A$  denota la función indicatriz del evento  $A$ .

**Proposición 2.1.** *Teorema de Kolmogórov-Smirnov:*

$$\widehat{F}_n(x) \xrightarrow[n \rightarrow \infty]{} F(x).$$

**Demostración:** Sea  $x$  fijo. Como:

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{x_i \leq x\}}$$

entonces, para cada  $n \geq 1$ , se tiene que  $\widehat{F}_n(x)$  es una variable aleatoria, puesto que las funciones indicatrices son variables aleatorias y  $\widehat{F}_n(x)$  es una suma finita de funciones indicatrices, todas con distribución  $F$ . Además, se tiene que  $\widehat{F}_n(x)$  es una variable aleatoria independiente para cada  $n \geq 1$ .

Ahora bien:

$$\begin{aligned} \mathbb{E}(\widehat{F}_n(x)) &= \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n I_{\{x_i \leq x\}}\right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(I_{\{x_i \leq x\}}) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{P}(x_i \leq x) \\ &= \frac{1}{n} \sum_{i=1}^n F_{x_i}(x) \\ &= \frac{1}{n} \sum_{i=1}^n F(x) \\ &= \frac{1}{n} n F(x) = F(x) < \infty. \end{aligned}$$

En particular,  $\mathbb{E}(\widehat{F}_1(x)) = F(x) < \infty$ . En virtud de la Ley Fuerte de los Grandes Números, existe  $c \in \mathbb{R}$  tal que:

$$\widehat{F}_n(x) \xrightarrow[n \rightarrow \infty]{} c.$$

Si  $c = F(x)$  entonces:

$$\widehat{F}_n(x) \xrightarrow[n \rightarrow \infty]{} F(x)$$

Sea  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  un conjunto de observaciones independientes de tamaño  $n$  con función de probabilidad  $F$  desconocida.

**Definición 2.2.** *Se define una muestra bootstrap como una muestra aleatoria de tamaño  $n$  tomada de  $\mathbf{x}$ .*

$$\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*),$$

donde cada  $x_i^*$ ;  $i = 1, 2, \dots, n$  representa una observación tomada aleatoriamente y con reposición de  $\mathbf{x}$

Esto significa que algunos elementos de la muestra original podrían aparecer una, dos, tres veces, etc., o bien podrían no aparecer.

**Definición 2.3.** *Se define la réplica bootstrap como el estadístico de interés que se calcula a partir de una muestra bootstrap.*

$$\hat{\theta}^* = \mathbf{s}(\mathbf{x}^*),$$

donde  $\mathbf{s}(\mathbf{x}^*)$  es la cantidad que resulta de aplicar la misma función  $\mathbf{s}(\cdot)$  a  $\mathbf{x}^*$  que se le aplica a  $\mathbf{x}$ .

Por ejemplo, si  $\mathbf{s}(\mathbf{x})$  es la media aritmética de  $\mathbf{x}$  ( $\bar{\mathbf{x}}$ ), entonces  $\mathbf{s}(\mathbf{x}^*)$  es la media aritmética de la muestra bootstrap.

$$\bar{\mathbf{x}}^* = \frac{1}{n} \sum_{i=1}^n x_i^*.$$

**Definición 2.4.** *Se define la estimación ideal bootstrap del error típico del estadístico  $\hat{\theta}$  como el error típico de  $\hat{\theta}$  para el conjunto de datos de tamaño  $n$  tomados aleatoriamente de la distribución empírica  $\widehat{F}$ .*

Si se denota por  $se_F(\hat{\theta})$  el error típico de  $\hat{\theta}$ , entonces la estimación ideal bootstrap de  $se_F(\hat{\theta})$  viene dada por:

$$se_{\hat{F}}(\hat{\theta}^*).$$

**Notación:** La estimación ideal bootstrap de  $se_F(\hat{\theta})$  se denota  $\hat{se}_\infty$ .

Básicamente, la técnica bootstrap trabaja de la siguiente manera:

- Se toma un gran número  $B$  de muestras bootstrap independientes  $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*B}$ , donde  $25 \leq B \leq 200$  (Para estimación del error típico).
- Se evalúan las réplicas bootstrap correspondientes a cada muestra bootstrap.

$$\hat{\theta}^*(b) = \mathbf{s}(\mathbf{x}^{*b}), \quad b = 1, 2, \dots, B$$

- Se estima el error típico  $se_F(\hat{\theta})$  a partir de la desviación estándar muestral de las  $B$  réplicas.

$$\hat{se}_B = \sqrt{\sum_{b=1}^B \frac{(\hat{\theta}^*(b) - \hat{\theta}^*(\cdot))^2}{B-1}},$$

donde  $\hat{\theta}^*(\cdot) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^*(b)$ .

## 2.3. Ejemplo de Estimación Bootstrap del Error Típico Para la Media

Sea  $\bar{x}$  la media aritmética de la muestra aleatoria  $\mathbf{x}$ .

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Entonces el error típico estimado de  $\bar{x}$  viene dado por:

$$\sqrt{\frac{s_1^2}{n}}, \quad (2.1)$$

donde  $s_1^2$  es la varianza centrada de la muestra  $\mathbf{x}$ .

$$s_1^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{\mathbf{x}})^2.$$

Observación: si el valor numérico del error típico estimado es pequeño (menor que 1), entonces el valor de  $\bar{\mathbf{x}}$  es cercano a su valor esperado.

Si  $\hat{\theta} = \mathbf{s}(\mathbf{x}) = \bar{\mathbf{x}}$  es el estadístico de interés, entonces las réplicas bootstrap correspondientes a cada muestra bootstrap vienen dadas por:

$$\hat{\theta}^*(b) = \mathbf{s}(\mathbf{x}^{*b}) = \bar{\mathbf{x}}^{*b}, \quad b = 1, 2, \dots, B.$$

Luego:

$$\hat{s}e_B = \sqrt{\sum_{b=1}^B \frac{(\bar{\mathbf{x}}^{*b} - \hat{\theta}^*(\cdot))^2}{B-1}}, \quad (2.2)$$

donde  $\hat{\theta}^*(\cdot) = \frac{1}{B} \sum_{b=1}^B \bar{\mathbf{x}}^{*b}$ .

**Proposición 2.5.**

$$\lim_{B \rightarrow +\infty} \hat{s}e_B = \sqrt{\frac{1}{n^2} \sum_{i=1}^n (x_i - \bar{\mathbf{x}})^2}$$

Demostración: Veamos que  $\mathbb{E}(\hat{s}e_B^2) < \infty; \forall B \geq 1$ .

$$\begin{aligned}
\mathbb{E}(\widehat{se}_B^2) &= \mathbb{E} \left( \left( \sqrt{\frac{\sum_{b=1}^B (\bar{\mathbf{x}}^{*b} - \hat{\theta}^*(\cdot))^2}{B-1}} \right)^2 \right) \\
&= \mathbb{E} \left( \sum_{b=1}^B \frac{(\bar{\mathbf{x}}^{*b} - \hat{\theta}^*(\cdot))^2}{B-1} \right) \\
&= \frac{1}{B-1} \sum_{b=1}^B \mathbb{E}(\bar{\mathbf{x}}^{*b} - \hat{\theta}^*(\cdot))^2 \\
&= \frac{1}{B-1} \sum_{b=1}^B \mathbb{E} \left[ (\bar{\mathbf{x}}^{*b})^2 + 2\bar{\mathbf{x}}^{*b}\hat{\theta}^*(\cdot) + (\hat{\theta}^*(\cdot))^2 \right] \\
&= \frac{1}{B-1} \sum_{b=1}^B \mathbb{E}(\bar{\mathbf{x}}^{*b})^2 + 2\mathbb{E}(\bar{\mathbf{x}}^{*b}\hat{\theta}^*(\cdot)) + \mathbb{E}(\hat{\theta}^*(\cdot))^2.
\end{aligned}$$

Ahora bien:

- $\mathbb{E}(\bar{\mathbf{x}}^{*b}) < \infty$  puesto que  $\bar{\mathbf{x}}^{*b}$  es la media aritmética de la  $b$ -ésima muestra bootstrap, la cual es una cantidad finita.
- $\mathbb{E}(\hat{\theta}^*(\cdot)) < \infty$  ya que  $\hat{\theta}^*(\cdot)$  es la media aritmética de las  $B$  réplicas bootstrap, la cual también es una cantidad finita.
- $\mathbb{E}(\bar{\mathbf{x}}^{*b}\hat{\theta}^*(\cdot)) < \infty$  ya que el producto de dos cantidades finitas es finito.

Luego:

$$\frac{1}{B-1} \sum_{b=1}^B \mathbb{E}(\bar{\mathbf{x}}^{*b})^2 + 2\mathbb{E}(\bar{\mathbf{x}}^{*b}\hat{\theta}^*(\cdot)) + \mathbb{E}(\hat{\theta}^*(\cdot))^2 < \infty.$$

Es decir:

$$\mathbb{E}(\widehat{se}_B^2) < \infty; \quad \forall B \geq 1.$$

Un corolario de la Ley Fuerte de los Grandes Números enuncia que si existe una sucesión  $\{X_n\}_{n \geq 1}$  de variables aleatorias independientes e idénticamente distribuidas tal que:

$$\mathbb{E}(X_1^2) < \infty,$$

entonces la varianza muestral converge en forma casi segura a la varianza teórica. Como  $\mathbb{E}(\widehat{se}_1^2) < \infty$ , entonces en virtud de este corolario se tiene que:

$$\lim_{B \rightarrow +\infty} \widehat{se}_B = \sqrt{\frac{1}{n^2} \sum_{i=1}^n (x_i - \bar{\mathbf{x}})^2}.$$

Este ejemplo resulta interesante por dos motivos: el primero es que la media observada es el estimador de la media poblacional o valor esperado, lo que lo hace un estadístico realmente importante. El segundo es que a diferencia de otros estimadores, la media observada es el único que posee una fórmula que permite obtener el error típico estimado. Esto significa que no es estrictamente necesario utilizar métodos computacionales, como Bootstrap, para medir la precisión o exactitud de este estimador.

El hecho de que ningún otro estimador aparte de la media posea una fórmula que permita calcular su error típico estimado, no significa que este no pueda ser calculado. Bootstrap puede ofrecer un buen estimador del error típico para cualquier estadístico que sea de interés.

## 2.4. Función de Distribución Empírica y el Principio Plug-in

La función de distribución empírica  $\widehat{F}$  se define como una distribución discreta que asigna probabilidad  $1/n$  a cada valor  $x_1, x_2, \dots, x_n$  de una muestra aleatoria. En otras palabras,  $\widehat{F}$  asigna a un subconjunto  $A$  del espacio muestral de  $\mathbf{x}$  su probabilidad empírica:

$$\mathbb{P}_{\widehat{F}}(A) = \frac{\#(x_i \in A)}{n}.$$

Recordemos que un parámetro es una función de la distribución de probabilidad  $F$ .

$$\theta = t(F).$$

Por ejemplo, si  $F$  es una distribución de probabilidad en la recta real, entonces se puede pensar en la esperanza como un parámetro:

$$\theta = t(F) = \mathbb{E}_F(x).$$

Un estadístico es una función de una muestra aleatoria  $\mathbf{x}$  con distribución de probabilidad  $F$ , que a su vez resulta ser una variable aleatoria.

$$\hat{\theta} = t(\mathbf{x}).$$

Por ejemplo, si  $\mathbf{x} = x_1, x_2, \dots, x_n$  es una muestra aleatoria, entonces la media observada es un estadístico de  $\mathbf{x}$

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n x_i.$$

### 2.4.1. El Principio Plug-in

El principio Plug-in es un método simple que sirve para estimar parámetros a partir de una muestra aleatoria en lugar de su distribución de probabilidad. El estimador Plug-in de un parámetro  $\theta = t(F)$  se define por:

$$\hat{\theta} = t(\hat{F}).$$

En otras palabras, se puede pensar en el estimador Plug-in como en el mismo parámetro definido para la distribución de probabilidad  $F$  pero tomando en lugar de esta la distribución de probabilidad empírica  $\hat{F}$ .

Cuando la única información disponible acerca de la distribución de probabilidad  $F$  proviene de la muestra  $\mathbf{x}$ , el estimador Plug-in es un excelente estimador del parámetro  $\theta$ . Por ejemplo, el estimador Plug-in de la esperanza  $\theta = \mathbb{E}_F(x)$  viene dado por:

$$\hat{\theta} = \mathbb{E}_{\hat{F}}(x) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{\mathbf{x}}.$$

El método Bootstrap tiene la virtud de producir sesgos y errores típicos de los estimadores Plug-in de manera automática, que luego podrán ser estudiados para determinar la exactitud de tales estimadores. De hecho, Bootstrap por sí mismo es una aplicación del principio Plug-in.

## 2.5. Error Típico de la Media y Estimación del Error Típico de la Media

### 2.5.1. El Error Típico de la Media

Sea  $x$  una variable aleatoria con distribución de probabilidad  $F$ . Sean:

$$\mu_F = \mathbb{E}_F(x), \quad \sigma_F^2 = \text{Var}_F(x) = \mathbb{E}_F[(x - \mu_F)^2]$$

La esperanza y la varianza sobre  $F$  de la variable aleatoria  $x$  respectivamente.

**Proposición 2.6.** *Sea  $\mathbf{x} = x_1, x_2, \dots, x_n$  una muestra aleatoria de tamaño  $n$  con distribución  $F$ . Entonces la media de la muestra*

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n x_i,$$

*tiene esperanza  $\mu_F$  y varianza  $\sigma_F^2/n$ .*

**Demostración:** Veamos que  $\mathbb{E}(\bar{\mathbf{x}}) = \mu_F$

$$\begin{aligned} \mathbb{E}(\bar{\mathbf{x}}) &= \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(x_i) \\ &= \frac{1}{n} n \mu_F \\ &= \mu_F. \end{aligned}$$

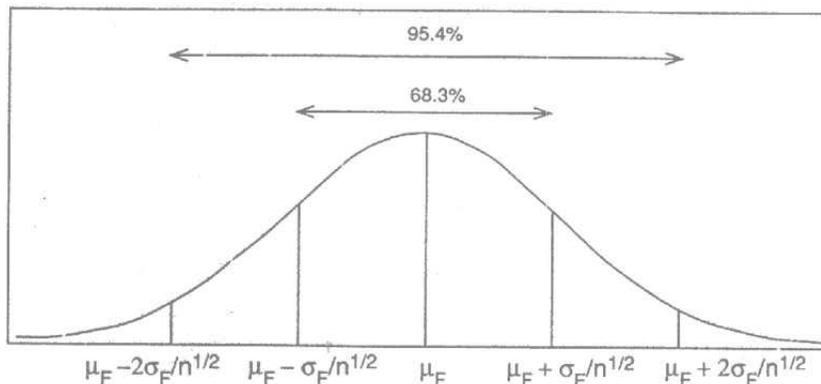
Veamos que  $\text{Var}(\bar{\mathbf{x}}) = \sigma_F^2/n$ ,

$$\begin{aligned}
\text{Var}(\bar{x}) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) \\
&= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(x_i) \\
&= \frac{1}{n^2} n \sigma_F^2 \\
&= \sigma_F^2/n.
\end{aligned}$$

El error típico de la media viene dado por:

$$se_F(\bar{x}) = \sqrt{\text{Var}_F(x)} = \frac{\sigma_F}{\sqrt{n}}. \quad (2.3)$$

El error típico es el parámetro más usado para indicar la precisión estadística de algún estimador. Por lo general, se espera que para la media  $\bar{x}$  el error típico sea menor que 1 en el 68 por ciento de las veces y menor que dos en el 95 por ciento de las veces, donde los porcentajes 68 y 95 están basados en el teorema central del límite, el cual estipula que bajo ciertas condiciones generales de  $F$ , la distribución de  $\bar{x}$  se aproxima a una distribución normal con parámetros  $\mu_F$  y  $\sigma_F^2/n$  siempre que  $n$  tienda a infinito. Lo anterior se ilustra en la siguiente figura:



Una de las ventajas de Bootstrap es que no se tiene que depender totalmente del teorema central del límite. Si el valor de  $n$  es más bien pequeño, (el tamaño de la muestra disponible), entonces los valores descritos anteriormente pierden precisión, por lo que una estimación del error típico nos da una buena idea de la precisión de nuestro estadístico.

## 2.5.2. Estimación del Error Típico de la Media

Supongamos que disponemos de una muestra aleatoria  $x_1, x_2, \dots, x_n$  con distribución  $F$  desconocida. Luego de calcular el estadístico  $\bar{\mathbf{x}}$  de la esperanza  $\mu_F$  resulta interesante conocer el valor del error típico de  $\bar{\mathbf{x}}$ . Sin embargo, como  $F$  es desconocida, no resulta útil utilizar (2.3), por lo cual se utiliza el principio Plug-in de la siguiente manera: Se sustituye  $\hat{F}$  por  $F$  en (2.3) y se considera el estimador Plug-in del error típico.

$$\hat{\sigma} = \sigma_{\hat{F}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{\mathbf{x}})^2}, \quad (2.4)$$

obteniendo así la estimación del error típico  $\widehat{se}(\bar{\mathbf{x}}) = se_{\hat{F}}(\bar{\mathbf{x}})$ .

$$\widehat{se}(\bar{\mathbf{x}}) = \frac{\sigma_{\hat{F}}}{\sqrt{n}} = \sqrt{\frac{1}{n^2} \sum_{i=1}^n (x_i - \bar{\mathbf{x}})^2}.$$

Lo cual es muy similar a (2.1).

Para realizar esta estimación, el principio Plug-in es utilizado dos veces, una para estimar la esperanza  $\mu_F$  por  $\mu_{\hat{F}} = \bar{\mathbf{x}}$  y luego para estimar el error típico  $se_F(x)$  por  $se_{\hat{F}}(\bar{\mathbf{x}})$ . La estimación bootstrap del error típico se refiere a utilizar el principio Plug-in para estimar el error típico de un estadístico arbitrario  $\hat{\theta}$ . Aquí se muestra que si  $\hat{\theta} = \bar{\mathbf{x}}$ , entonces esta aproximación tiende a la estimación usual del error típico para la media.

## 2.6. La Estimación Bootstrap del Error Típico

La estimación bootstrap del error típico de un estadístico  $\hat{\theta}$  es una estimación Plug-in que utiliza la función de distribución empírica en lugar de la distribución  $F$  desconocida.

**Definición 2.7.** *La estimación ideal bootstrap de  $se_F(\hat{\theta})$  se define por:*

$$se_{\hat{F}}(\hat{\theta}^*). \quad (2.5)$$

En otras palabras, la estimación ideal bootstrap de  $se_F(\hat{\theta})$  es el error típico de  $\hat{\theta}$  calculado sobre el conjunto de datos de tamaño  $n$  tomados aleatoriamente de  $\hat{F}$ .

Como ya vimos, el último paso en el algoritmo bootstrap se basa en estimar el error típico  $se_F(\hat{\theta})$  del parámetro  $\hat{\theta}$  a través de la desviación estándar de la muestra conformada por las  $B$  réplicas bootstrap.

$$\widehat{se}_B = \sqrt{\frac{1}{B-1} \sum_{b=1}^B [\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)]^2}.$$

Donde  $\hat{\theta}^*(\cdot) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^*(b)$ .

Un hecho interesante es que el límite de  $\widehat{se}_B$  cuando  $B$  tiende a infinito es la estimación ideal bootstrap de  $se_F(\hat{\theta})$ .

$$\lim_{B \rightarrow \infty} \widehat{se}_B = se_{\widehat{F}}(\hat{\theta}^*). \quad (2.6)$$

El hecho de que  $\widehat{se}_B$  se aproxime a  $se_{\widehat{F}}$  cuando  $B$  tiende a infinito equivale a decir que la desviación estándar empírica se aproxima a la desviación estándar de la población cuando el número de réplicas bootstrap aumenta. La población en este caso es la población de valores  $\hat{\theta}^* = \mathbf{s}(\mathbf{x}^*)$  donde  $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$  es una muestra aleatoria tomada de  $\widehat{F}$ .

## 2.7. El Número $B$ de Réplicas Bootstrap

Uno de los puntos más importantes del método Bootstrap es determinar el valor  $B$  de réplicas bootstrap que serán usadas para evaluar  $\widehat{se}_B$ . La estimación ideal bootstrap considera el caso  $B \rightarrow \infty$  para el cual  $\widehat{se}_B$  es igual al estimador plug-in  $se_{\widehat{F}}(\hat{\theta}^*)$ . Es importante la determinación de este valor puesto que el tiempo de cómputo del algoritmo depende del tiempo que se demore evaluar las réplicas bootstrap, y además porque se busca el mismo buen comportamiento que se obtuvo en la estimación del error típico para la estimación de otras cantidades de interés como el sesgo y la desviación estándar.

**Proposición 2.8.**  *$\widehat{se}_B$  siempre tiene desviación estándar mayor que  $se_{\widehat{F}}(\hat{\theta}^*)$*

Para determinar el valor  $B$  de réplicas bootstrap, se considera el coeficiente de variación de  $\widehat{se}_B$ .

$$cv(\widehat{se}_B) = \sqrt{cv(\widehat{se}_\infty)^2 + \frac{\mathbb{E}(\Delta) + 2}{4B}}, \quad (2.7)$$

donde  $\Delta$  es un parámetro que mide la longitud de cola de la distribución de  $\hat{\theta}^*$ . El aumento del

coeficiente de variación refleja el aumento de la variabilidad debido a la detención del algoritmo después de  $B$  réplicas bootstrap, en lugar de ir hasta el infinito.

$\Delta$  es 0 para la distribución normal, -2 para las distribuciones de colas muy cortas y valores arbitrariamente grandes cuando  $F$  es de cola larga (en la práctica no pasa de 10).

La estimación ideal  $\widehat{se}_\infty = se_{\widehat{F}}(\widehat{\theta}^*)$  no es perfecta, puesto que puede tener una considerable variabilidad como estimación de  $se_F(\widehat{\theta})$  debido a la variabilidad de  $\widehat{F}$  como estimación de  $F$ .

La fórmula (2.7) tiene una importante consecuencia práctica: para los valores de  $cv(\widehat{se}_\infty)$  y  $\Delta$  que pudieran surgir en la práctica,  $cv(\widehat{se}_B)$  no es mucho mayor que  $cv(\widehat{se}_\infty)$  para  $B \geq 200$ , aunque rara vez se necesitan más de 200 réplicas para la estimación del error típico. Incluso un número pequeño de réplicas bootstrap, digamos 25, es usualmente informativa mientras que  $B = 50$  es a menudo suficiente para dar una buena estimación de  $se_F(\widehat{\theta})$ .

Cabe mencionar que en la práctica resulta muy útil revisar tanto los datos reales como los datos bootstrap, y realizar un histograma para ambos.

## 2.8. Estimación del Porcentaje de Variación obtenido en ACP usando Bootstrap

Sea:

$$M(X) = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{pmatrix},$$

una matriz de datos donde cada una de las filas  $x_{1k}, x_{2k}, \dots, x_{nk}$  representa a cada uno de los  $n$  individuos de la población;  $k = 1, 2, \dots, d$ . Sea:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{pmatrix} \in \mathbb{R}^d,$$

el vector de variables aleatorias que se medirán sobre esta población. Consideremos la matriz de covarianzas para el vector aleatorio  $\mathbf{X}$ :

$$\Sigma = \begin{pmatrix} Cov(X_1, X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_d) \\ Cov(X_2, X_1) & Cov(X_2, X_2) & \cdots & Cov(X_2, X_d) \\ \vdots & \vdots & & \vdots \\ Cov(X_d, X_1) & Cov(X_d, X_2) & \cdots & Cov(X_d, X_d) \end{pmatrix}.$$

Sean  $\alpha_1, \alpha_2, \dots, \alpha_d$  los autovectores de la matriz de covarianzas  $\Sigma$  y  $\lambda_1, \lambda_2, \dots, \lambda_d$  los autovalores correspondientes. Entonces el porcentaje de variación capturado por las primeras  $q$  componentes principales viene dado por:

$$\mathfrak{F}_q = \frac{\sum_{k=1}^q \lambda_k}{\sum_{k=1}^d \lambda_k} \quad ; \quad q < d.$$

Como se dijo en el Capítulo 1, el número de componentes principales a considerar viene dado por el valor de  $q$  para el cual  $\mathfrak{F}_q > \alpha$ , donde  $\alpha$  se escoge a conveniencia.

Supongamos que nos interesa saber el porcentaje de variación capturado por la primera componente principal (cercanía de los puntos proyectados a la primera componente principal).

$$\mathfrak{F}_1 = \frac{\lambda_1}{\sum_{k=1}^d \lambda_k}.$$

Para saber que tan exacto es el valor  $\mathfrak{F}_1$ , se utiliza el método Bootstrap para estimar este valor. Para esto, consideramos la matriz:

$$X^* = \begin{pmatrix} x_{1k}^* \\ x_{2k}^* \\ \vdots \\ x_{nk}^* \end{pmatrix}, \quad k = 1, 2, \dots, d,$$

donde las filas  $x_{1k}^*, x_{2k}^*, \dots, x_{nk}^*$  son tomadas aleatoriamente con reposición, de las  $n$  filas de la matriz de datos  $X$ .

A partir de esta matriz  $X^*$ , y considerando nuevamente el vector aleatorio  $\mathbf{X}$ , calculamos la matriz de covarianzas  $\Sigma^*$

$$\Sigma^* = \begin{pmatrix} Cov^*(X_1, X_1) & Cov^*(X_1, X_2) & \cdots & Cov^*(X_1, X_d) \\ Cov^*(X_2, X_1) & Cov^*(X_2, X_2) & \cdots & Cov^*(X_2, X_d) \\ \vdots & \vdots & \ddots & \vdots \\ Cov^*(X_d, X_1) & Cov^*(X_d, X_2) & \cdots & Cov^*(X_d, X_d) \end{pmatrix}.$$

La notación  $Cov^*(X_i, X_j)$  indica que las entradas de  $\Sigma^*$  están calculadas en base a  $X^*$ .

Sean  $\alpha_1^*, \alpha_2^*, \dots, \alpha_d^*$  los autovectores de la matriz  $\Sigma^*$ , y  $\lambda_1^*, \lambda_2^*, \dots, \lambda_d^*$  los autovalores correspondientes. Entonces la réplica bootstrap de  $\mathfrak{F}_1$  es:

$$\mathfrak{F}_1^* = \frac{\lambda_1^*}{\sum_{k=1}^d \lambda_k^*}.$$

Como ya se sabe, para determinar el sesgo de  $\mathfrak{F}_1^*$ , se calcula el error típico  $\widehat{se}_B$ . Para determinar la exactitud de  $\mathfrak{F}_1^*$ , se calcula la media de las  $B$  réplicas bootstrap. Si  $\widehat{se}_B$  es pequeño (digamos menor que 1) y la media de las  $B$  réplicas bootstrap se aproxima considerablemente al valor de  $\mathfrak{F}_1$ , entonces podemos decir que  $\mathfrak{F}_1$  es un estimador consistente cuando  $B$  tiende a infinito.

# Capítulo 3

## Implementación de las Herramientas R y MATLAB Para realizar ACP y Bootstrap

### 3.1. Introducción

En este capítulo se explicará cómo la implementación de las herramientas de software matemático R y MATLAB pueden combinarse para realizar un Análisis de Componentes Principales (ACP) a un conjunto real de datos y un remuestreo usando el algoritmo Bootstrap para comprobar la exactitud de los resultados obtenidos en el ACP. En la sección (3.1.1) se realizará un ACP a un conjunto real de datos utilizando la herramienta R, se mostrarán gráficos y tablas que ayudarán al análisis e interpretación de los resultados obtenidos y se darán las conclusiones pertinentes. En la sección (3.1.2) se aplicará el algoritmo Bootstrap al conjunto de datos utilizando la herramienta MATLAB. Se mostrarán los comandos utilizados y los scripts creados para tal fin y se analizarán los resultados obtenidos.

#### 3.1.1. Implementación de la Herramienta R para la realización del ACP

Para esta investigación se dispone de un conjunto de datos, referentes al número de préstamos por estado del país que se atrasaron en mas de cuatro cuotas y que estan con ejecutivos especiales de cobro, donde:

- “cant client” es el número de clientes en ese estado con créditos atrasados.
- “cant prest” es el número de préstamos provenientes de ese estado con mas de cuatro cuotas atrasadas.

- “salcap” es el monto total de la deuda.
- “vcdo” es el monto que debe cancelar para poner sus deudas al día.
- “pag” es el monto pagado.
- “salcapven” es el monto vencido de la deuda.

	Edo	cant_client	cant_prest	salcap	vcdo	pag	salcapven
1	Amazonas	52	72	701421.53	282133.90	62238.58	260713.16
2	Anzoátegui	1080	1645	24215358.37	3829730.17	1130686.29	4964935.01
3	Apure	184	317	1673335.20	686226.20	61310.22	749730.87
4	Aragua	1532	2244	34146328.58	7548987.65	1372755.72	11536956.17
5	Barinas	266	383	5086754.08	1836262.86	136575.33	2066411.74
6	Bolívar	789	1154	5496945.36	1680281.73	316685.26	1962354.52
7	Carabobo	1971	2728	33777553.67	6037610.77	867531.76	10040840.56
8	Cojedes	208	241	2097221.52	815229.12	114073.20	900332.34
9	Delta Amacuro	21	31	152982.24	62913.80	24960.46	40613.27
10	Dependencias Federales	12	13	21525.91	16139.63	8495620.27	13621.51
11	Distrito Capital	3927	6505	44174738.98	11128609.16	2624280.23	13447293.15
12	Falcón	426	596	8218959.77	1992965.27	466997.55	2502688.21
13	Guárico	365	574	17597337.81	5792458.56	827863.60	6717448.05
14	Lara	842	1336	21084992.99	4434026.77	818533.12	6290837.98
15	Mérida	392	540	15847793.80	2170765.33	860080.25	1964377.70
16	Miranda	5708	8855	79920943.53	19355210.92	4673818.08	23988985.49
17	Monagas	587	849	11382308.65	1837069.71	331976.88	3200195.94
18	Nueva Esparta	376	606	3327320.10	965966.25	107936.35	1270278.29
19	Portuguesa	542	657	7359104.35	2183962.97	572228.45	2287984.23
20	Sucre	267	371	6548764.12	1540596.66	398587.63	1115453.46
21	Táchira	367	615	11819030.21	3264374.73	781938.30	3981343.30
22	Trujillo	353	447	7441360.11	1211845.98	334198.57	1897145.09
23	Vargas	696	887	5209914.11	2206020.67	191813.31	2270514.11
24	Yaracuy	130	191	1877862.95	545401.48	217751.00	244284.33
25	Zulia	3826	5445	77681031.69	15033761.08	2976908.88	25890697.12

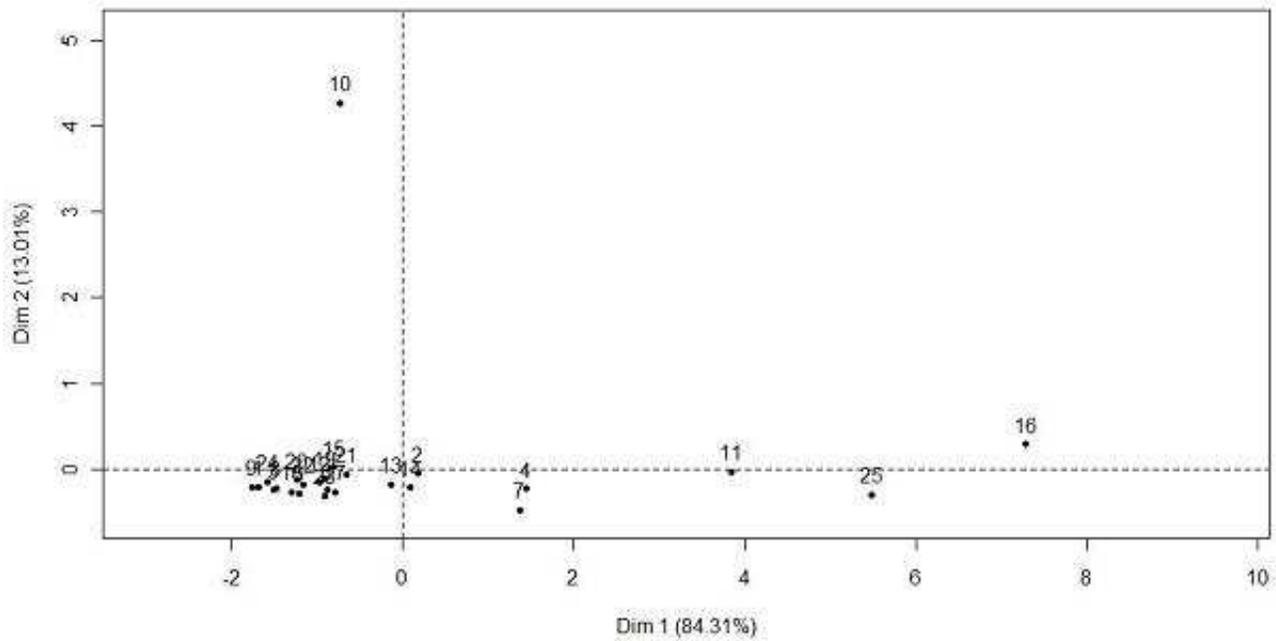
Para realizar un ACP a este conjunto de datos, se utiliza el plugin “FactoMineR”, el cual incluye la función “Principal Component Analysis (PCA)”. Esta función proporciona de manera directa los autovalores de la matriz de covarianzas, las varianzas capturadas por las componentes principales, las contribuciones que aportan las variables originales a cada una de las componentes principales y gráficos que ayudan al análisis e interpretación de los datos.

En la siguiente tabla se muestran los autovalores de la matriz de covarianzas, el porcentaje de varianza capturado por cada componente principal y el porcentaje de varianza acumulado:

```
> res$eig
```

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	5.0587889433	84.31314905	84.31315
comp 2	0.7803662113	13.00610352	97.31925
comp 3	0.1339920897	2.23320149	99.55245
comp 4	0.0172352881	0.28725480	99.83971
comp 5	0.0086271594	0.14378599	99.98349
comp 6	0.0009903082	0.01650514	100.00000

En esta tabla se puede ver que cerca de 97% de la varianza total es capturada por las dos primeras componentes principales, lo que indica que es posible proyectar los datos en un sistema de coordenadas de dos ejes, donde el eje horizontal corresponde a la primera componente principal y el eje vertical a la segunda componente principal. Esto se puede ver en el siguiente gráfico:



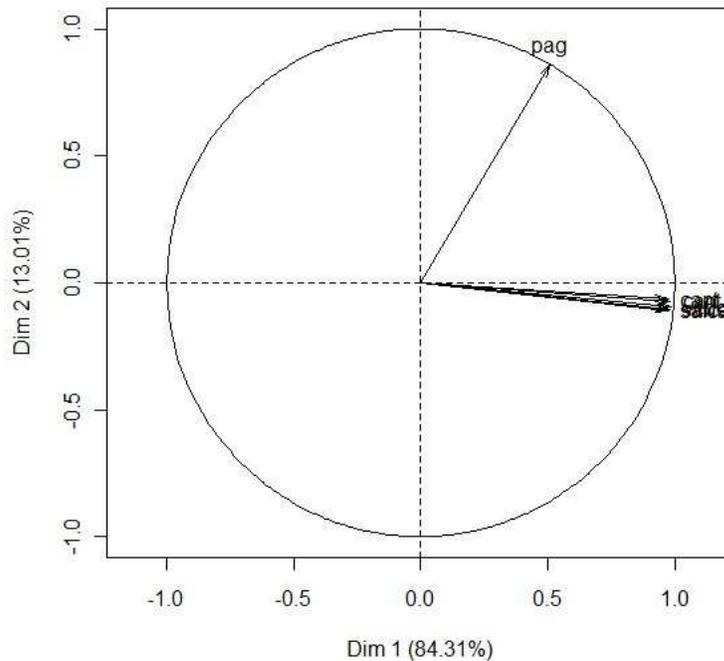
En ACP es importante estudiar la manera en que cada una de las variables aleatorias originales contribuyen con cada componente principal, puesto que estas contribuciones determinan que tan explicativas pueden ser las componentes principales. En la siguiente tabla se observan los aportes de cada una de las seis variables originales a las primeras cinco componentes principales:

\$contrib	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
cant_client	19.04164	0.7095636	22.0708207	5.163862e+00	2.595262e+00
cant_prest	18.75036	0.5399076	34.9140358	5.425234e-02	1.718457e-02
salcap	19.00411	1.3738751	15.0744084	2.495000e+01	3.941835e+01
vcdo	19.28286	1.1954104	2.0159126	6.905929e+01	6.613144e+00
pag	5.15996	94.6653065	0.1734542	4.280541e-05	1.187247e-04
salcapven	18.76108	1.5159367	25.7513683	7.725510e-01	5.135594e+01

Esta tabla indica que todas las variables originales ofrecen un gran aporte a las cinco primeras componentes principales. Considerando que cerca del 97% de la varianza total es capturada por las primeras dos componentes principales, se observa que salvo la variable “pag” (monto pagado) todas las variables contribuyen de manera significativa con la primera componente principal, lo que indica que esta componente explica de manera directa a las variables involucradas con la cantidad de clientes, cantidad de préstamos con más de cuatro cuotas atrasadas en ese estado, y las deudas adquiridas por tales atrasos.

La variable “pag” contribuye de manera significativa con la segunda componente principal, lo cual indica que esta componente explica mejor a la variable relacionada con el monto pagado por los clientes en ese estado.

El siguiente gráfico muestra el aporte de cada una de las variables originales a las dos primeras componentes principales:



La siguiente tabla contiene las contribuciones individuales de los 25 estados a las primeras cinco componentes principales:

Aquí se puede observar que existen tres estados (Miranda, Zulia y Distrito Capital) que aportan de manera significativa a la primera componente principal.

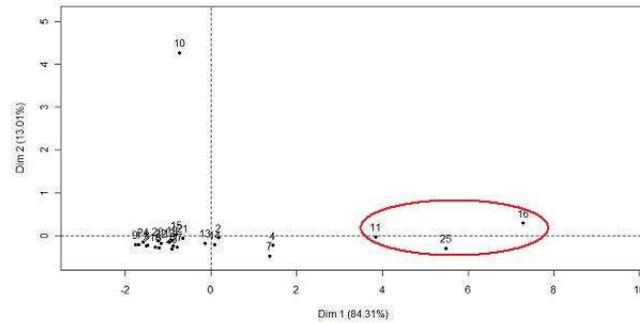
\$contrib					
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
1	2.259468004	0.226451013	2.369863e-01	6.432448e-02	0.607799214
2	0.026234647	0.015543330	5.137148e-02	7.784222e+00	22.303240451
3	1.807702035	0.321456854	5.379751e-01	4.698127e-04	1.352259982
4	1.633919036	0.279181838	7.823288e+00	1.327080e+00	0.590854593
5	1.146617281	0.421484931	7.711005e-03	2.024183e+00	0.750474631
6	0.608503512	0.309803225	4.110485e+00	2.930373e-01	3.831626650
7	1.475323747	1.179598792	9.811814e-02	1.326476e+01	0.280664684
8	1.739079440	0.269694199	2.846894e-01	1.987531e-02	1.292525561
9	2.444781902	0.236115787	3.057723e-01	2.706955e-01	0.639530032
10	0.417460433	93.588032131	2.575724e-01	2.731024e-01	0.920430999
11	11.715439840	0.005168940	2.965241e+01	3.655087e-02	1.256751652
12	0.736432157	0.126795747	8.574647e-04	4.818705e-02	0.017546117
13	0.013990603	0.160502527	1.212427e+01	4.107548e+01	0.664373566
14	0.005642225	0.239963304	2.005027e+00	9.846675e-02	1.202869594
15	0.507909375	0.001815827	6.178241e-01	3.241410e+00	38.119280140
16	41.988887854	0.432916875	7.857034e+00	7.436062e+00	5.490884184
17	0.489683984	0.382786060	8.145156e-03	3.590723e+00	0.114683141
18	1.327917980	0.379257867	1.189009e+00	1.928591e-01	2.220401763
19	0.664824930	0.058130662	1.620498e-01	6.090160e-01	0.067612423
20	1.200575126	0.088832022	2.404959e-02	9.096567e-02	3.441340518
21	0.339995658	0.020591709	1.725656e+00	4.867986e+00	0.168109108
22	1.065601975	0.175197195	6.763722e-03	1.623516e+00	0.001420894
23	0.658360253	0.503622191	1.666006e+00	1.395704e+00	3.811628760
24	1.963448382	0.113498468	4.041607e-01	2.999414e-02	0.008806741
25	23.762199622	0.463558507	2.884276e+01	1.034133e+01	10.844884600

Los contribuciones individuales que cada estado le otorga a la primera componente principal están directamente relacionados con todas las variables explicadas por dicha componente, de modo que como se muestra en la primera gráfica, mientras más grande es la contribución de cada uno de los estados a la primera componente principal, más alejada se encuentra su proyección del origen de coordenadas.

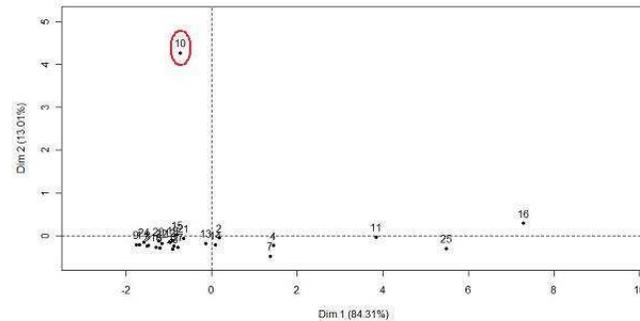
De todo lo explicado anteriormente se dan las siguientes conclusiones:

La primera componente principal separa a la nube de puntos en dos grupos: los estados que poseen una cantidad de clientes mayor a 2950 (estados Miranda, Zulia y Distrito Capital) y los que poseen una cantidad de clientes menor a 2950 (estados restantes).

Esto podría ser un indicativo de que en los estados que poseen gran cantidad de clientes se debe poner especial atención si uno de los objetivos de las instituciones que otorgan los préstamos es, por ejemplo, reducir los montos de las deudas.



La segunda componente principal separa a los estados en los que los montos pagados por los clientes superan a los montos de las deudas generadas por pagos atrasados (Dependencias Federales) de aquellos estados donde las deudas son superiores a los montos pagados de los préstamos (estados restantes).



Esto parece ser un indicativo de que en esta región los clientes cancelan la deuda despues de ser atendidos por los ejecutivos especiales de cobro.

A excepción de las Dependencias Federales, se observa que independientemente del número de clientes por estado, existe en todo el territorio nacional una tendencia por parte de los clientes a no pagar ó a retrasarse en los pagos de las cuotas de los préstamos solicitados.

Ahora bien, considerando que se dispone de un conjunto pequeño de datos (25 estados) y que estos fueron tomados en un momento determinado, el siguiente paso es verificar si estos resultados son confiables y si se mantiene esta tendencia a través del tiempo. Para esto hacemos uso del algoritmo Bootstrap, donde estimaremos el porcentaje de variación capturada por la primera componente principal utilizando un total de 200 réplicas bootstrap, apoyándonos en

la herramienta MATLAB.

### 3.1.2. Implementación de la Herramienta MATLAB para la realización del remuestreo Bootstrap

A partir de la tabla de datos de la sección (3.1.1), obtenemos la matriz de covarianzas  $\Sigma$ , haciendo uso de la función “*Cov(·)*”.

$$\Sigma = 1,0e + 014 * \begin{pmatrix} 0,0000 & 0,0000 & 0,0003 & 0,0001 & 0,0000 & 0,0001 \\ 0,0000 & 0,0000 & 0,0005 & 0,0001 & 0,0000 & 0,0001 \\ 0,0003 & 0,0005 & 4,8242 & 1,0384 & 0,1706 & 1,5118 \\ 0,0001 & 0,0001 & 1,0384 & 0,2327 & 0,0382 & 0,3286 \\ 0,0000 & 0,0000 & 0,1706 & 0,0382 & 0,0352 & 0,0531 \\ 0,0001 & 0,0001 & 1,5118 & 0,3286 & 0,0531 & 0,4836 \end{pmatrix}$$

Haciendo uso de la función “*roots(poly(·))*”, se calculan los autovalores de la matriz  $\Sigma$ , obteniéndose los siguientes autovalores:

$$\begin{aligned} \lambda_1 &= (1,0e + 014)5,5291 \\ \lambda_2 &= (1,0e + 014)0,0292 \\ \lambda_3 &= (1,0e + 014)0,0111 \\ \lambda_4 &= (1,0e + 014)0,0062 \\ \lambda_5 &= (1,0e + 014)0,0000 \\ \lambda_6 &= (1,0e + 014)0,0000 \end{aligned}$$

Y ahora para calcular el porcentaje de variación capturado por la primera componente principal ( $\mathfrak{F}_1$ ), utilizamos la siguiente secuencia de instrucciones:

$$\begin{aligned} f1 &= \text{lambda}(1)/(\text{sum}(\text{lambda})) \\ f1 &= 0,9916 \end{aligned}$$

Para crear la matriz  $X^*$  (muestra bootstrap), la cual es la matriz obtenida tomando aleatoriamente y con reposición las filas de la matriz de datos, se crea el script “*mind(·)*” con la siguiente secuencia de instrucciones:

$$\begin{aligned} \text{function } [C] &= \text{mind}(X) \\ [m, n] &= \text{size}(X); \end{aligned}$$

```

B = 1 : 1 : m;
for i = 1 : m
B(i) = floor((m) * rand(1,1) + 1);
end
C = X(B, :);

```

Ahora bien, este script será usado para generar 200 muestras bootstrap. A cada una de estas 200 muestras bootstrap se le calcularán tanto la matriz de covarianzas ( $\Sigma^*$ ) como los autovalores asociados a cada una de esas matrices, con lo cual se generarán las 200 réplicas bootstrap  $\mathfrak{F}_1^*$ . Para realizar este procedimiento se crea el script “*estim1*( $\cdot, \cdot$ )” con la siguiente secuencia de instrucciones:

```

function theta = estim1(A, l)
M = zeros(1, l);
for i = 1 : l
B = mind(A);
S1 = cov(B);
lambda = roots(poly(S1));
M(i) = lambda(1)/(sum(lambda));
end
S2 = sum(M)/l;
s = 0;
for i = 1 : l
s = (M(i) - S2).^2 + s;
end
Promedio = S2
theta = sqrt((1/(l - 1)) * s);

```

En este script también se incluye el cálculo del promedio de las 200 réplicas bootstrap y el cálculo de la estimación bootstrap del error típico  $\widehat{se}_B$ , que son valores que nos ayudarán a determinar cuán exacto es el porcentaje de variación obtenido en ACP.

Al utilizar el script “*estim1*( $\cdot, \cdot$ )” se obtienen los siguientes resultados:

$$estim1(M, 200)$$

$$Promedio = 0,9882$$

$$ans = 0,0168.$$

## 3.2. Conclusiones

Se tenía una muestra pequeña y se obtuvo una clasificación por ACP, la cual otorga un porcentaje de variación elevado cuando trabaja con matrices provenientes de datos reales. Esto queda sentado cuando se observa que para estos datos el porcentaje de variación capturado por las primeras dos componentes principales es cercano al 97 %.

Se sabe que las réplicas bootstrap, desde el punto de vista teórico, son un buen estimador para el coeficiente de variación, y en este caso se obtuvo un valor que difiere del valor verdadero en menos de 0,004, lo cual nos permite afirmar que los resultados obtenidos en ACP son confiables y que existe una tendencia por parte de los clientes de las instituciones financieras a atrasarse en más de cuatro cuotas en los pagos de los préstamos solicitados.

Destacamos la importancia del Bootstrap por dos motivos: primero, en el análisis estadístico es común realizar procedimientos en los que no se realizan supuestos sobre las poblaciones o sus parámetros (lo que comúnmente conocemos como "pruebas no paramétricas"), las cuales tienen la desventaja de que se pierde cierta cantidad de información, puesto que se trabaja con variables a nivel ordinal y los valores observados quedan reducidos a rangos de valores. El método Bootstrap soslaya este problema manteniendo los niveles de medida.

Segundo, la simplicidad del algoritmo Bootstrap, apoyado sobre bases computacionales, hace que este método constituya un enfoque atractivo en la práctica de la estadística, puesto que deja de lado las fuertes definiciones matemáticas y las suposiciones y logra aún así obtener medidas que permitan hacer análisis y conclusiones importantes del estudio que se esté realizando.

# Bibliografía

- [1] Efron, Bradley and Tibshirani, Robert J. An Introduction to the Bootstrap. Boca Raton. Chapman & Hall/CRC. 1993.
- [2] Jolliffe, Ian T. Principal Component Analysis. Second Edition. N.Y. Springer. 2002.
- [3] Resnick, Sidney I. A Probability Path. Boston. Birkhäuser. 1999.

Referencias electrónicas:

- [4] Terrádez Gurrea, Manuel. Análisis de componentes principales.  
*[http : //www.uoc.edu/in3/emath/docs/Componentes\\_principales.pdf](http://www.uoc.edu/in3/emath/docs/Componentes_principales.pdf).*
- [5] Cuadras, Carles M. Nuevos Métodos de Análisis Multivariante. 6 de Noviembre 2008.  
*[http : //www.usc.es/labca.f/es/system/files/CuadrasCM\\_MetodosAMultivariante.pdf](http://www.usc.es/labca.f/es/system/files/CuadrasCM_MetodosAMultivariante.pdf).*
- [6] Cuesta, Marcelino y Herrero, Francisco. Introducción al Bootstrap.  
*[http : //www.psico.uniovi.es/DptoPsicologia/metodos/tutor,9/boot1.html](http://www.psico.uniovi.es/DptoPsicologia/metodos/tutor,9/boot1.html).*  
*[http : //www.psico.uniovi.es/DptoPsicologia/metodos/tutor,9/boot2.html](http://www.psico.uniovi.es/DptoPsicologia/metodos/tutor,9/boot2.html).*  
*[http : //www.psico.uniovi.es/DptoPsicologia/metodos/tutor,9/boot3.html](http://www.psico.uniovi.es/DptoPsicologia/metodos/tutor,9/boot3.html).*