Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Análisis de la expansión de la frontera agrícola en la región transfronteriza entre Tabasco y Chiapas

Tesis presentada como requisito para obtener el título de:

Licenciada en Matemáticas Aplicadas

Presenta: Blanca Xochilt Muñoz Vargas

Directores: Dra. Lucía Cervantes Gómez y
Dr. Bulmaro Juárez Hernández

Puebla, Pue., Junio 2015

Agradecimientos

Expreso mi profundo agradecimiento a las siguientes Instituciones y especialistas por el apoyo otorgado:

Secretaría de Educación Pública (SEP) por la beca otorgada bajo el Programa Nacional de Becas 2014 en su modalidad Titulación.

Facultad de Ciencias Físico Matemáticas de la BUAP por las facilidades y el apoyo para realizar mis estudios de Licenciatura en Matemáticas Aplicadas.

Vicerrectoría de Investigación y Estudios de Posgrado de la BUAP por las becas otorgadas mediante sus programas de divulgación de la investigación para estudiantes en distintas modalidades.

CA de Ecuaciones Diferenciales y Modelación Matemática de la BUAP por una beca otorgada a través del Apoyo a las redes temáticas de la SEP que me permitió concluir los estudios.

Centro de Investigación en Geografía y Geomática “Ing. Jorge L. Tamayo” A.C. (CentroGeo), en especial al Ing. Yosu Rodríguez Aldabe y al M.C. Mauricio Galeana Pizaña, por la información facilitada del problema ambiental que se estudió en esta tesis.

Dra. Lucía Cervantes Gómez, por todas sus enseñanzas, consejos y apoyo a lo largo de mi carrera y el tiempo dedicado a la dirección de esta tesis.

Dr. Bulmaro Juárez Hernández por su apoyo, paciencia y dedicación a la dirección de esta tesis.

Lic. en Cs. Ambientales Myriam Poisot Cervantes por la asesoría en el planteamiento del problema en términos ambientales.

Dra. Gladys Linares Fleites, el Dr. Víctor Hugo Vázquez Guevara y M.C. Julio Erasto Poisot Macías por su dedicación en la lectura y revisión, sus correcciones y contribuciones.

Dra. María Elena Franco Carcedo, por la revisión, corrección de estilo y sus valiosas aportaciones a la estructura de este trabajo.

Dedico esta tesis a todos mis seres queridos.

Siglas

ACP	Análisis de componentes principales
CentroGeo	Centro de Investigación en Geografía y Geomática “Ing. Jorge L. Tamayo” A.C.
CONABIO	Comisión Nacional para el Conocimiento y Uso de la Biodiversidad
COPLADET	Comité de Planeación para el Desarrollo del Estado de Tabasco
EM	*Millennium Ecosystem Assessment* (Evaluación de los Ecosistemas del Milenio)
FAO	*Food and Agriculture Organization of the United Nations* (Organización de las Naciones Unidas para la Agricultura y la Alimentación)
INEGI	Instituto Nacional de Estadística y Geografía
PROCAMPO	Programa de Apoyos Directos al Campo
SAGARPA	Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación
SNIM	Sistema Nacional de Información Municipal

Resumen

En esta tesis se presentan los métodos de análisis de regresión lineal múltiple y análisis de componentes principales, su aplicación al problema de la identificación de las principales variables que determinan la expansión de la frontera agrícola en la región transfronteriza entre Tabasco y Chiapas.

En el Capítulo 1 se desarrolla el método de análisis de regresión lineal múltiple y en el Capítulo 2 el método de análisis de componentes principales. Ambos se presentan de manera exhaustiva, con sus principales resultados y demostraciones, reuniendo la información necesaria para facilitar una comprensión profunda y completa de los mismos que permita identificar los supuestos necesarios para su aplicación correcta. En el Capítulo 3 se presenta la aplicación al análisis de la expansión de la frontera agrícola en la región transfronteriza entre Tabasco y Chiapas.

Índice general

Introducción

Un problema grave a nivel mundial es la deforestación, en particular la deforestación en la región transfronteriza entre Tabasco y Chiapas es una de las causas de las graves inundaciones que sufre periódicamente la región. En México una de las principales causas de deforestación es la expansión de la frontera agropecuaria. En este trabajo presentamos un análisis para determinar la influencia de diferentes posibles causas de la expansión de la frontera agrícola en la región transfronteriza entre Tabasco y Chiapas; el análisis de este problema motiva el uso de los métodos estadísticos de Análisis de regresión lineal múltiple y de Análisis de componentes principales, por lo que en esta tesis se presentan de manera exhaustiva, incluyendo sus resultados y demostraciones.

Cabe mencionar que, aunque los métodos de Análisis de regresión lineal múltiple y de Análisis de componentes principales son muy útiles y utilizados, sus resultados completos y demostraciones no se encuentran usualmente en los libros de texto, por lo que una importante aportación de este trabajo es la inclusión de los resultados completos junto con las demostraciones que se realizaron para este propósito, en particular, las pruebas de razón de verosimilitud generalizadas para construir los estadísticos de prueba de diferentes pruebas de hipótesis que se presentan para ambos métodos.

Si se tiene un conjunto de variables explicativas y una variable respuesta, para determinar si la variable respuesta depende de una manera lineal de las variables explicativas, se puede usar el análisis de regresión lineal múltiple. Este análisis nos proporciona una estimación de esta relación y, además, bajo ciertas suposiciones se pueden hacer pruebas de hipótesis y predicciones.

Otro de los aspectos que interesan al analizar datos múltiples es describir las principales tendencias de la variación de los objetos con respecto a todas las variables medidas, no solo a algunas de ellas. Observar los diagramas de dispersión de los objetos con respecto a todos los posibles pares de variables es una aproximación tediosa, que por

lo general no aclara el problema. Sin embargo, existe un método que permite estudiar la variabilidad de los datos a pesar de que sean múltiples: el Análisis de Componentes Principales (ACP), que construye nuevas variables que son combinaciones lineales de las variables originales y conservan el mayor porcentaje de la variabilidad de los datos. Las variables construidas permiten disminuir el número de variables en el análisis sin gran pérdida de información, por lo cual son utilizables en otros análisis en lugar de las variables originales. Estas nuevas variables son no correlacionadas.

El ACP se puede utilizar para eliminar la correlación entre las variables explicativas y posteriormente realizar un análisis de regresión con una variable respuesta y las componentes principales.

El presente trabajo está organizado en 3 capítulos, la conclusión y 4 apéndices. El Capítulo 1 se dedica al Análisis de regresión lineal múltiple, en él se define el modelo de regresión lineal múltiple y se describe (Secciones 1.1 y 1.2); además se hallan estimaciones y conceptos básicos que se utilizan en el análisis (Secciones 1.3 y 1.4); más adelante se trabaja con un caso particular: el caso de la distribución normal, para el cual se demuestran propiedades de los estimadores, pruebas de hipótesis, se hace inferencia, predicción y se mencionan métodos de comprobación de hipótesis (Secciones 1.6, 1.8, 1.9, 1.10 y 1.11). Por último, se proporcionan algunos métodos de elección de modelos, que ayudan a elegir el mejor modelo (Sección 1.13).

El capítulo 2 contiene el desarrollo del Análisis de componentes principales, primero se describen algunos aspectos útiles, posteriormente los objetivos del análisis, luego se hace el desarrollo del análisis para el caso poblacional para el cual se construyen las componentes principales, se da la definición formal de las componentes y se demuestran algunas propiedades (Secciones 2.1, 2.2 y 2.3). Para continuar, se hace notar que los datos con unidades de medición no comparables o varianzas muy diferentes afectan el análisis, por lo cual se desarrollan aspectos del análisis aplicado a datos estandarizados (Sección 2.4). Como lo anterior realizó para el caso poblacional, se hace algo similar para el caso muestral (Secciones 2.5 y 2.6). Una de las virtudes de este análisis es que se puede reducir el número de variables con las que se esté trabajando, por lo que se enuncian diferentes criterios para elegir el número de componentes principales, se sigue

con algunas representaciones gráficas, propiedades muestrales de las componentes, pruebas de hipótesis y por último se dan algunas aplicaciones del ACP entre las cuales está la de regresión múltiple (Secciones 2.7, 2.8, 2.9, 2.10 y 2.11).

En el capítulo 3 se realiza una aplicación de los conceptos presentados en los capítulos 1 y 2 al problema de la expansión de la frontera agrícola, el cual es una de las principales causas de deforestación en México, mediante un análisis que pretende ver si existe una relación lineal entre la expansión con el crecimiento poblacional, la marginación y el subsidio de PROCAMPO, este último es uno de los principales subsidios al campo. Este análisis se realiza para tres periodos: 1993-2002; 2003-2007 y 2008-2011.

En el Apéndice A se explica el procedimiento que se llevó a cabo para la recopilación de información y la construcción de las bases de datos empleadas en el Capítulo 3. El Apéndice B muestra algunos conceptos matemáticos que se emplean a lo largo de los capítulos 1 y 2, y en el Apéndice C se enlistan diferentes definiciones y resultados estadísticos con los cuales se desarrollan los capítulos 1 y 2. Por último, el Apéndice D proporciona una lista de comandos de R con los cuales se puede efectuar el Análisis de regresión y el ACP.

Capítulo 1

Análisis de Regresión lineal múltiple

Los datos de variables múltiples se presentan en todas las ramas de la ciencia. Casi todos los datos reunidos actualmente por los investigadores se pueden clasificar como de variables múltiples. Por ejemplo, un investigador de mercados podría querer identificar las características de los individuos que le permitirían determinar si es probable que cierta persona compre un producto específico, o un agricultor que cultiva trigo podría interesarse en otras características del cultivo y no sólo en el rendimiento de algunas variedades nuevas de trigo, también podría interesarse en la resistencia de estas variedades al daño por insectos y sequías. También, un investigador de las ciencias sociales podría querer estudiar las relaciones entre los comportamientos en las citas de las adolescentes y las actitudes de sus padres. Por último, en ecología, usualmente se observan varios descriptores para cada objeto bajo estudio. Cada uno de estos esfuerzos está relacionado con datos de variables múltiples [8].

Los métodos de regresión lineal simple son aplicables cuando se desea ajustar un modelo lineal al relacionar el valor de una variable dependiente Y con el valor de una sola variable independiente x. Sin embargo, hay muchos casos en los que una sola variable independiente no es suficiente. En algunas situaciones hay varias variables independientes, $x_1, x_2, \ldots, x_p$, relacionadas con una variable dependiente Y. Si la relación entre las variables dependiente e independiente es lineal, se puede usar la técnica de **regresión lineal múltiple** [12].

En este capítulo se hace un estudio de procedimientos inferenciales que se pueden

usar cuando una variable aleatoria Y, llamada *variable dependiente*, tiene una media que es función de una o más variables no aleatorias $x_1, x_2, \ldots, x_p$, llamadas *variables independientes* (en este contexto, los términos *independiente* y *dependiente* se usan en su sentido matemático, no hay relación con el concepto probabilístico de variables aleatorias independientes).

El análisis de regresión es importante cuando los valores de una variable se modifican de acuerdo con los de otras que pueden ser manipuladas por nosotros; trata de investigar la naturaleza de la relación y construir modelos que la describan con el propósito de predecir el comportamiento de una de ellas a partir de valores de las otras, aunque en general se puede decir que el objetivo global es hacer inferencias óptimas sobre la variable que interesa primordialmente.

1.1. Modelos estadísticos lineales

Aun cuando se puede usar un número infinito de funciones diferentes para modelar el valor medio de la variable de respuesta Y como función de una o más variables independientes, la atención se centrará en un conjunto de modelos llamados *modelos estadísticos lineales* [16].

Definición 1.1. *Un **modelo estadístico lineal** o **modelo de regresión lineal múltiple** relaciona una respuesta aleatoria Y con un conjunto de variables independientes $x_1, x_2, \ldots, x_p$ de la forma:*

$$Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon, \tag{1.1}$$

donde β_0, β_1, ..., β_p son parámetros desconocidos; ε es una variable aleatoria y las variables $x_1, x_2, \ldots, x_p$ toman valores conocidos. Se supondrá que $E(\varepsilon) = 0$ y por tanto:

$$E(Y) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p.$$

Desde el punto de vista práctico, ε reconoce la incapacidad para dar un modelo exacto por naturaleza. En experimentación repetida, Y varía alrededor de $E(Y)$ de un modo aleatorio, ya que no se ha incluido en el modelo toda la gran cantidad de variables que

pueden afectar a Y. Por fortuna, muchas veces el efecto neto de estas variables no medidas y con mucha frecuencia desconocidas es hacer que Y varíe de manera que puedan ser aproximadas adecuadamente mediante una suposición de comportamiento aleatorio [16].

Hay algunos casos especiales del modelo de regresión lineal múltiple que con frecuencia se utilizan en la práctica.

- **Modelo de regresión lineal simple:**

  Este modelo relaciona la $E(Y)$ como una función lineal de β_0 y β_1 únicamente. El modelo de regresión lineal simple es:

$$Y = \beta_0 + \beta_1 x + \varepsilon.$$

- **Modelo de regresión polinomial:**

  En este modelo las variables independientes son potencias de una sola variable. El modelo de regresión polinomial de grado p es:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_p x^p + \varepsilon. \tag{1.2}$$

- **Modelo cuadrático:**

  Los modelos de regresión múltiple también se pueden hacer con potencias de diversas variables. Por ejemplo, un modelo de regresión polinomial de grado 2, también llamado **modelo cuadrático**, en dos variables x_1 y x_2 está dado por:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \varepsilon. \tag{1.3}$$

  Como a una variable producto de las otras dos se llama **interacción** en el modelo cuadrático (1.3) la variable $x_1 x_2$ es la **interacción**, de x_1 y x_2.

Los **modelos de regresión polinomial (1.2)** y **cuadrático (1.3)** se consideran lineales, aunque contengan términos no lineales de las variables independientes. La razón de que continúen siendo modelos lineales es que *son lineales en los coeficientes* β_i y *no* necesariamente en las variables independientes.

1.2. Descripción del modelo de regresión lineal múltiple

A continuación se describirá el modelo de regresión múltiple. Supongamos que se tiene una muestra de n elementos, y para cada uno se ha medido una variable dependiente Y y p variables independientes $x_1, x_2, \ldots, x_p$. El i-ésimo elemento de la muestra, por tanto, tiene el conjunto ordenado $(Y_i, x_{i1}, x_{i2}, \ldots, x_{ip})$ [12]. Ahora, supongamos que se tiene el modelo lineal dado por la Ecuación (1.1)

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon,$$

en consecuencia, se puede escribir la observación Y_i como:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i,$$

donde x_{ij} es el valor de la j-ésima variable independiente para la i-ésima observación, $i = 1, 2, \ldots, n$ [16].

Se definen las siguientes matrices con $x_{i0} = 1$, $i = 1, 2, \ldots, n$:

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{10} & x_{11} & x_{12} & \cdots & x_{1p} \\ x_{20} & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{n0} & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \quad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entonces, las n ecuaciones que representan Y_i como función de las x, las β y las ε se pueden escribir simultáneamente como:

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}. \tag{1.4}$$

1.3. Estimación de los coeficientes

En cualquier modelo de regresión múltiple, los estimadores para β_0, β_1, ..., β_p, que se denotarán con $\widehat{\beta}_0$, $\widehat{\beta}_1$,...,$\widehat{\beta}_p$ respectivamente, se calculan por medio de mínimos cuadrados o usando máxima verosimilitud si se asume alguna distribución para $\underline{\varepsilon}$. Considerando $\widehat{\underline{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_p)'$, se tiene que las ecuaciones de mínimos cuadrados y las soluciones presentadas en notación matricial son las siguientes:

Teorema 1.2. *Las ecuaciones de mínimos cuadrados conocidas como ecuaciones normales son* $(X'X)\underline{\widehat{\beta}} = X'\underline{Y}$*, y sus soluciones para un modelo de regresión lineal múltiple son* $\underline{\widehat{\beta}} = (X'X)^{-1}X'\underline{Y}$*, si* $X'X$ *es no singular.*

Demostración. Los estimadores de mínimos cuadrados son los valores de β_0, β_1, ..., β_p, los cuales minimizan la suma de cuadrados de los errores:

$$\begin{aligned} \sum_{i=1}^{n} \varepsilon_i^2 &= \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}))^2 \\ &= (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta}) \\ &= \underline{Y}'\underline{Y} - \underline{Y}'X\underline{\beta} - \underline{\beta}'X'\underline{Y} + \underline{\beta}'X'X\underline{\beta}. \end{aligned}$$

Además, $\underline{Y}'_{1\times n} X_{n\times(p+1)} \underline{\beta}_{(p+1)\times 1}$ y $\underline{\beta}'_{1\times(p+1)} X'_{(p+1)\times n} \underline{Y}_{n\times 1}$ son escalares, por lo tanto son iguales. Luego:

$$g(\underline{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \underline{Y}'\underline{Y} - 2\underline{\beta}'X'\underline{Y} + \underline{\beta}'X'X\underline{\beta}. \tag{1.5}$$

Derivando respecto a $\underline{\beta}$ se tiene que:

$$g'(\underline{\beta}) = -2X'\underline{Y} + 2X'X\underline{\beta} \tag{1.6}$$

e igualando a $\underline{0}$, se obtiene $-2X'\underline{Y} + 2X'X\underline{\beta} = \underline{0}$ o equivalentemente:

$$X'X\underline{\beta} = X'\underline{Y}.$$

Este conjunto de ecuaciones se conoce como ecuaciones normales, consisten en $p+1$ ecuaciones con $p+1$ incógnitas, y se pueden presentar dos casos:

- Algunas ecuaciones dependen de otras, así que hay menos de $p+1$ ecuaciones independientes con $p+1$ incógnitas, luego $X'X$ es singular, por lo tanto $(X'X)^{-1}$ no existe. Entonces el modelo debe ser expresado en términos de menos parámetros o bien, adicionalmente, se deben dar o asumir otras restricciones en los parámetros.

- Si las $p+1$ ecuaciones normales son independientes, $X'X$ es no singular y su inversa existe. Entonces $\underline{\beta} = (X'X)^{-1}X'\underline{Y}$.

Derivando nuevamente la función dada en (1.5) con respecto a $\underline{\beta}$, se tiene $g''(\underline{\beta}) = 2X'X$, que se define positiva para toda $\underline{\beta}$, pues $(2X'X)' = 2X'X$ y dado $\underline{w} \in \mathbb{R}^{p+1} - \{\underline{0}\}$, se cumple que $\underline{w}'(2X'X)\underline{w} = 2(X\underline{w})'(X\underline{w}) > 0$.

Por lo tanto $\underline{\widehat{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_p)' = (X'X)^{-1}X'\underline{Y}$ minimiza la suma de los cuadrados residuales. □

Los estimadores por mínimos cuadrados permiten dar las siguientes definiciones y resultados:

Definición 1.3. *La ecuación $\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \ldots + \widehat{\beta}_p x_p$ se llama **ecuación de mínimos cuadrados** o **ecuación de regresión ajustada**.*

Definición 1.4. *Se define $\widehat{Y}_i$ como la coordenada $\widehat{Y}$ de la ecuación de mínimos cuadrados correspondiente a los valores de $x_{i1}, \ldots, x_{ip}$ ($\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \ldots + \widehat{\beta}_p x_{ip}$). Además, se tiene que $\underline{\widehat{Y}} = (\widehat{Y}_1, \widehat{Y}_2, \ldots, \widehat{Y}_n)' = X\underline{\widehat{\beta}}$.*

Definición 1.5. *Los **residuos** representan las cantidades $\epsilon_i = Y_i - \widehat{Y}_i$, que constituyen las diferencias entre los valores observados $\underline{Y}$ y los valores estimados $\underline{\widehat{Y}}$ que proporciona la ecuación de mínimos cuadrados. El vector de los residuos de mínimos cuadrados es $\underline{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)' = \underline{Y} - \underline{\widehat{Y}} = \underline{Y} - X\underline{\widehat{\beta}}$.*

Teorema 1.6. *Bajo las condiciones del modelo de regresión lineal múltiple, el vector de los residuos de mínimos cuadrados cumple que $X'\underline{\epsilon} = \underline{0}$.*

Demostración. $X'\underline{\epsilon} = X'(\underline{Y} - X\underline{\widehat{\beta}}) = X'\underline{Y} - X'X(X'X)^{-1}X'\underline{Y} = X'\underline{Y} - X'\underline{Y} = \underline{0}$. □

Corolario 1.7. *La suma de los residuos es cero.*

Demostración. Como la primera columna de X está formada por 1's y $X'\underline{\epsilon} = \underline{0}$, se sigue que:

$$\underline{0} = X'\underline{\epsilon} = \begin{bmatrix} \underline{1}' \\ (X^*)' \end{bmatrix} \underline{\epsilon} = \begin{bmatrix} \underline{1}'\underline{\epsilon} \\ (X^*)'\underline{\epsilon} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n \epsilon_i \\ (X^*)'\underline{\epsilon} \end{bmatrix}$$

De forma que $\sum_{i=1}^n \epsilon_i = 0$. □

1.4. Suma de cuadrados

Gran parte del análisis en la regresión múltiple se basa en tres cantidades fundamentales, que se definen a continuación.

Definición 1.8 (Suma de los cuadrados)**.** *En el modelo de regresión lineal múltiple se definen las siguientes sumas de cuadrados:*

- *Suma de los cuadrados de la regresión:* $SCR = \sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2$
- *Suma de los cuadrados del error:* $SCE = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$
- *Suma total de cuadrados de las desviaciones:* $STC = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$*.*

Teorema 1.9. *Las sumas de cuadrados que se definen para el modelo de regresión lineal múltiple cumplen las siguientes igualdades:*

i) $SCR = \underline{Y}' M \underline{Y} - n\overline{Y}^2 = \underline{Y}'(M - J)\underline{Y}$

ii) $SCE = \underline{Y}'(I - M)\underline{Y}$

iii) $STC = \underline{Y}'\underline{Y} - n\overline{Y}^2 = \underline{Y}'(I - J)\underline{Y}$

donde $M = X(X'X)^{-1}X'$ *y* $J = n^{-1}\underline{1}\underline{1}'$*.*

Demostración. Como $\overline{Y} = n^{-1}\underline{Y}'\underline{1}$, donde $\underline{1}$ es un vector columna de n unos, se sigue que $\overline{Y}^2 = \overline{YY}' = n^{-1}\underline{Y}'n^{-1}\underline{1}\underline{1}'\underline{Y} = n^{-1}\underline{Y}'J\underline{Y}$. Dado que $\sum_{i=1}^{n} \epsilon_i = 0$, se tiene que $\sum_{i=1}^{n} \widehat{Y}_i = \sum_{i=1}^{n} (Y_i - \epsilon_i) = \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \epsilon_i = \sum_{i=1}^{n} Y_i$. Ahora se calculará SCR, SCE y por último STC.

$$\begin{aligned}
SCR &= \textstyle\sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2 = \underline{\widehat{Y}}'\underline{\widehat{Y}} - 2\overline{Y}\sum_{i=1}^{n} \widehat{Y}_i + n\overline{Y}^2 \\
&= \underline{\widehat{\beta}}' X'X \underline{\widehat{\beta}} - 2\overline{Y}\textstyle\sum_{i=1}^{n} Y_i + n\overline{Y}^2 \\
&= \underline{Y}'X(X'X)^{-1}X'X(X'X)^{-1}X'\underline{Y} - 2n\overline{Y}^2 + n\overline{Y}^2 \\
&= \underline{Y}'X(X'X)^{-1}X'\underline{Y} - n\overline{Y}^2 = \underline{Y}'M\underline{Y} - \underline{Y}'J\underline{Y} = \underline{Y}'(M - J)\underline{Y}
\end{aligned}$$

$$\begin{aligned}
SCE &= \textstyle\sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2 = (\underline{Y} - \underline{\widehat{Y}})'(\underline{Y} - \underline{\widehat{Y}}) \\
&= (\underline{Y}' - \underline{\widehat{\beta}}' X')(\underline{Y} - X\underline{\widehat{\beta}}) = \underline{Y}'\underline{Y} - \underline{Y}'X\underline{\widehat{\beta}} - \underline{\widehat{\beta}}' X'\underline{Y} + \underline{\widehat{\beta}}' X'X\underline{\widehat{\beta}}) \\
&= \underline{Y}'\underline{Y} - \underline{Y}'X\underline{\widehat{\beta}} - \underline{Y}'X(X'X)^{-1}X'\underline{Y} + \underline{Y}'X(X'X)^{-1}X'X\underline{\widehat{\beta}} \\
&= \underline{Y}'\underline{Y} - \underline{Y}'M\underline{Y} = \underline{Y}'(I - M)\underline{Y}
\end{aligned}$$

$$\begin{aligned} STC &= \sum_{i=1}^{n}(Y_i-\overline{Y})^2 = \underline{Y}'\underline{Y} - 2\overline{Y}\sum_{i=1}^{n} Y_i + n\overline{Y}^2 \\ &= \underline{Y}'\underline{Y} - 2n\overline{Y}^2 + n\overline{Y}^2 = \underline{Y}'\underline{Y} - n\overline{Y}^2 = \underline{Y}'(I-J)\underline{Y}. \end{aligned}$$

□

Corolario 1.10. *Las sumas de cuadrados que se definen para el modelo de regresión lineal múltiple cumplen $STC = SCR + SCE$, esta ecuación se llama **identidad del análisis de la varianza**.*

1.5. Supuestos para los errores en modelos lineales

El análisis se restringirá al caso más simple, en el cual se satisface que los errores ε_1, ε_2, ..., ε_n son variables aleatorias independientes e idénticamente distribuidas en forma normal con media cero y varianza σ^2. Simbólicamente esto puede escribirse como $\varepsilon_i \sim iidN(0,\sigma^2)$, $i = 1, 2, \ldots, n$. Este supuesto implica los siguientes resultados:

Teorema 1.11. *Las Y_i son variables aleatorias independientes, y cada Y_i tiene una distribución normal con media $\mu_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$ y varianza $\sigma_i^2 = \sigma^2$, es decir, $Y_i \sim N(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}, \sigma^2)$.*

Demostración. Como $\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$ es una constante y $\varepsilon_i \sim N(0,\sigma^2)$, entonces $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i \sim N(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}, \sigma^2)$. Como ε_1, ε_2, ..., ε_n son independientes, se sigue que Y_1, Y_2, ..., Y_n son independientes. □

Cada coeficiente β_i representa el cambio en la media de Y_i relacionado con un aumento de una unidad en el valor de x_i, cuando las otras variables x se mantienen constantes.

El siguiente lema proporciona la función de verosimilitud (que se denotará por $L(\textit{parámetros}; \textit{muestra})$) de $Y_1, \ldots, Y_n$ bajo los supuestos dados.

Lema 1.12. *Cuando $\varepsilon_i \sim iidN(0,\sigma^2)$, $i = 1, 2, \ldots, n$, la función de verosimilitud de $\underline{Y} = (Y_1, \ldots, Y_n)'$ es $L(\underline{\theta}; \underline{y}) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n}(y_i-\mu_i)^2/2\sigma^2}$ y $l(\underline{\theta}; \underline{y}) = ln(L(\underline{\theta}; \underline{y})) = -nln(\sqrt{2\pi}) - nln(\sigma) - \sum_{i=1}^{n}(y_i-\mu_i)^2/2\sigma^2$.*

Demostración. La función de verosimilitud de $\underline{Y} = (Y_1, \ldots, Y_n)'$ es:

$$\begin{aligned} L(\underline{\theta}; \underline{y}) &= \prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}} e^{-(y_i-\mu_i)^2/2\sigma_i^2} = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(y_i-\mu_i)^2/2\sigma^2} \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n}(y_i-\mu_i)^2/2\sigma^2}. \end{aligned}$$

Por lo tanto

$$\begin{aligned} l(\underline{\theta}; \underline{y}) &= ln\left(\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n}(y_i-\mu_i)^2/2\sigma^2}\right) \\ &= -nln(\sqrt{2\pi}) - nln(\sigma) - \textstyle\sum_{i=1}^{n}(y_i-\mu_i)^2/2\sigma^2 \end{aligned}$$

□

A continuación se enuncia el teorema que proporciona los estimadores de máxima verosimilitud de los parámetros $\underline{\beta}$ y σ^2.

Teorema 1.13. *Si $\varepsilon_i \sim iidN(0, \sigma^2)$, $i = 1, 2, \ldots, n$, los estimadores de máxima verosimilitud de los parámetros $\underline{\beta}$ y σ^2 son $\widehat{\underline{\beta}} = (X'X)^{-1}X'\underline{Y}$ (los estimadores de mínimos cuadrados) y $\sigma^2 = n^{-1}SCE$, así, el valor máximo de $L(\underline{\theta}; \underline{y})$ en $\overline{\underline{\Theta}} = \mathbb{R}^{p+1} \times \mathbb{R}_+$ es:*

$$L\left(\widehat{\underline{\beta}}, n^{-1}SCE; \underline{y}\right) = \left(\frac{n}{2\pi SCE}\right)^{n/2} e^{-n/2}.$$

Demostración. En la regresión lineal múltiple se desconoce el parámetro $\underline{\theta} = (\beta_0, \beta_1, \ldots, \beta_p, \sigma^2)'$, luego el espacio paramétrico es $\overline{\underline{\Theta}} = \mathbb{R}^{p+1} \times \mathbb{R}_+$. Así, si $\underline{\theta} = (\beta_0, \beta_1, \cdots, \beta_p, \sigma^2)' \in \overline{\underline{\Theta}}$, usando el Lema 1.12 se tiene que:

$$\begin{aligned} l(\underline{\theta}; \underline{y}) &= -nln(\sqrt{2\pi}) - nln(\sigma) - \textstyle\sum_{i=1}^{n}(y_i-\mu_i)^2/2\sigma^2 \\ &= -nln(\sqrt{2\pi}) - nln((\sigma^2)^{1/2}) - (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})/2\sigma^2 \\ &= -nln(\sqrt{2\pi}) - nln(\sigma^2)/2 - (\underline{Y}'\underline{Y} - 2\underline{\beta}'X'\underline{Y} + \underline{\beta}'X'X\underline{\beta})/2\sigma^2, \end{aligned}$$

cuyas derivadas parciales con respecto a $\underline{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)'$ y σ^2 son:

$$\begin{aligned} \frac{\partial l(\underline{\theta};\underline{y})}{\partial \underline{\beta}} &= -(-X'\underline{Y} + X'X\underline{\beta})/\sigma^2 \\ \frac{\partial l(\underline{\theta};\underline{y})}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{(\underline{Y}'\underline{Y} - 2\underline{\beta}'X'\underline{Y} + \underline{\beta}'X'X\underline{\beta})}{2(\sigma^2)^2} \\ &= \frac{-n\sigma^2 + (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})}{2(\sigma^2)^2} \end{aligned}$$

Al igualar a cero cada una de las expresiones anteriores, se tiene:

$$X'\underline{Y} - X'X\underline{\beta} = 0 \;\Leftrightarrow\; X'X\underline{\beta} = X'\underline{Y}$$

$$-n\sigma^2 + (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta}) = 0 \;\Leftrightarrow\; \sigma^2 = n^{-1}(\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})$$

La primera ecuación matricial es igual a las ecuaciones normales que se enuncian en el Teorema 1.2, por lo tanto $(\beta_0, \beta_1, \ldots, \beta_p)' = \underline{\widehat{\beta}} = (X'X)^{-1}X'\underline{Y}$ y $\sigma^2 = n^{-1}\sum_{i=1}^{n}(y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \cdots + \widehat{\beta}_p x_{ip}))^2 = n^{-1}SCE$.

A continuación se muestra que el punto crítico obtenido anteriormente determina un máximo de la función de verosimilitud. Esto es, la matriz que contiene las segundas derivadas parciales es:

$$\mathcal{H}(\underline{\beta}, \sigma^2) = \frac{1}{\sigma^2}\begin{bmatrix} -X'X & \frac{-X'\underline{Y}+X'X\underline{\beta}}{\sigma^2} \\ \frac{-X'\underline{Y}+X'X\underline{\beta}}{\sigma^2} & \frac{n}{2\sigma^2} - \frac{(\underline{Y}-X\underline{\beta})'(\underline{Y}-X\underline{\beta})}{(\sigma^2)^2} \end{bmatrix}$$

Al evaluar $\mathcal{H}(\underline{\beta}, \sigma^2)$ en $\underline{\beta} = \underline{\widehat{\beta}}$ y $\sigma^2 = n^{-1}SCE$, se tiene:

$$\begin{aligned}
&\mathcal{H}(\underline{\widehat{\beta}}, n^{-1}SCE) \\
&= \frac{n}{SCE}\begin{bmatrix} -X'X & \frac{-X'\underline{Y}+X'X\underline{\widehat{\beta}}}{n^{-1}SCE} \\ \frac{-X'\underline{Y}+X'X\underline{\widehat{\beta}}}{n^{-1}SCE} & \frac{n}{2n^{-1}SCE} - \frac{(\underline{Y}-X\underline{\widehat{\beta}})'(\underline{Y}-X\underline{\widehat{\beta}})}{(n^{-1}SCE)^2} \end{bmatrix} \\
&= \frac{n}{SCE}\begin{bmatrix} -X'X & \frac{-X'\underline{Y}+X'\underline{Y}}{n^{-1}SCE} \\ \frac{-X'\underline{Y}+X'\underline{Y}}{n^{-1}SCE} & \frac{n^2}{2SCE} - \frac{n^2 SCE}{(SCE)^2} \end{bmatrix} \\
&= \frac{n}{SCE}\begin{bmatrix} -X'X & \underline{0}' \\ \underline{0} & \frac{n^2}{2SCE} - \frac{2n^2}{2SCE} \end{bmatrix} = -\frac{n}{SCE}\begin{bmatrix} X'X & \underline{0}' \\ \underline{0} & \frac{n^2}{2SCE} \end{bmatrix}.
\end{aligned}$$

Se observa que $-\frac{SCE}{n}\mathcal{H}(\underline{\widehat{\beta}}, n^{-1}SCE) = -\frac{SCE}{n}\mathcal{H}(\underline{\widehat{\beta}}, n^{-1}SCE)'$, además que $X'X$ es definida positiva (pues $(X'X)' = X'X$ y dado $\underline{w} \in \mathbb{R}^{p+1} - \{\underline{0}\}$, se cumple que $\underline{w}'(X'X)\underline{w} = (X\underline{w})'(X\underline{w}) > 0$) y $\frac{n^2}{2SCE} > 0$, luego $-\frac{SCE}{n}\mathcal{H}(\underline{\widehat{\beta}}, n^{-1}SCE)$ es definida positiva. Por lo tanto $\mathcal{H}(\underline{\widehat{\beta}}, n^{-1}SCE)$ es definida negativa.

Como $\mathcal{H}(\underline{\widehat{\beta}}, n^{-1}SCE)$ es definida negativa, se tiene que $\underline{\beta} = \underline{\widehat{\beta}}$ y $\sigma^2 = n^{-1}SCE$

maximizan $L(\underline{\theta};\underline{y})$ en $\overline{\underline{\Theta}}$. Además, se tiene que:

$$\begin{aligned}\text{máx}_{\underline{\theta}\in\overline{\underline{\Theta}}}\{L(\underline{\theta};\underline{y})\} &= L\left(\widehat{\beta}_0,\widehat{\beta}_1,\ldots,\widehat{\beta}_p,n^{-1}SCE;\underline{y}\right)\\ &= \left(\sqrt{\tfrac{1}{n}SCE}\sqrt{2\pi}\right)^{-n}e^{-\sum_{i=1}^{n}(y_i-\sum_{j=0}^{p}\widehat{\beta}_j x_{ij})^2/2\frac{1}{n}SCE}\\ &= \left(\tfrac{n}{2\pi SCE}\right)^{n/2}e^{-nSCE/2SCE} = \left(\tfrac{n}{2\pi SCE}\right)^{n/2}e^{-n/2}\end{aligned}$$

□

1.6. Propiedades de los estimadores de mínimos cuadrados

Antes de enunciar las propiedades de los estimadores de mínimos cuadrados, se mencionarán algunas definiciones y propiedades (Véase [11]).

Definición 1.14. *El vector $E(\underline{X}) = \underline{\mu}$ es llamado la **esperanza** o la **media poblacional** del vector aleatorio $\underline{X}_{p\times 1}$ con función de densidad de probabilidad $f(\underline{x})$, donde $\mu_i = \int x_i f(\underline{x})d\underline{x}$, $i = 1,\cdots,p$.*

Definición 1.15. *La matriz $E((\underline{X}-\underline{\mu})(\underline{X}-\underline{\mu})') = \Sigma = Var(\underline{X})$ es llamada la **matriz de covarianza de** $\underline{X}$, también conocida como la **matriz de varianza-covarianza** o **matriz de dispersión**.*

Definición 1.16. *Se puede definir la **covarianza** entre dos vectores, $\underline{X}_{p\times 1}$ y $\underline{Y}_{q\times 1}$, dada por la matriz de tamaño $p\times q$ como sigue $Cov(\underline{X},\underline{Y}) = E((\underline{X}-\underline{\mu})(\underline{Y}-\underline{\nu})')$ donde $E(\underline{X}) = \underline{\mu}$ y $E(\underline{Y}) = \underline{\nu}$.*

Definición 1.17. *$\underline{X}$ tiene una distribución normal $p-$variada si y sólo si $\underline{a}'\underline{X}$ es normal univariada para todo $p-$vector $\underline{a}$ fijo.*

Para permitir el caso $\underline{a} = \underline{0}$, se considera a la constante como forma degenerada de la distribución normal.

A continuación se mencionan algunas propiedades de la media poblacional y la covarianza.

Teorema 1.18. *Sean $\underline{X}_{p\times 1}$, $\underline{Y}_{q\times 1}$ vectores aleatorios con media $E(\underline{X}) = \underline{\mu}$, $E(\underline{Y}) = \underline{\nu}$ y matriz de covarianza $Var(\underline{X}) = \Sigma = (\sigma_{ij})$; además, sean $A_{q\times p}$, $B_{p\times p}$, $C_{c\times p}$, $D_{d\times q}$, $\underline{a}_{1\times r}$, $\underline{b}_{q\times 1}$ y $\underline{c}_{p\times 1}$ matrices constantes*

i) *La media poblacional poseé la propiedad de linealidad:* $E(A\underline{X} + \underline{b}) = AE(\underline{X}) + \underline{b}$

ii) $E(A\underline{X}\ \underline{a}) = AE(\underline{X})\underline{a}$

iii) $E(\underline{X}'B\underline{X}) = E(\underline{X}')BE(\underline{X}) + tr(BVar(\underline{X}))$

iv) $\sigma_{ij} = Cov(X_i, X_j)$, $i \neq j$ *y* $\sigma_{ii} = Var(X_i)$

v) $\Sigma = E(\underline{X}\ \underline{X}') - \underline{\mu}\ \underline{\mu}'$

vi) $Var(A\underline{X}) = AVar(\underline{X})A'$

vii) $Var(\underline{X} + \underline{c}) = Var(\underline{X})$

viii) *El elemento* (i, j) *de la matriz* $Cov(\underline{X}, \underline{Y})$ *es igual a* $Cov(X_i, Y_j)$

ix) $Cov(C\underline{X}, D\underline{Y}) = CCov(\underline{X}, \underline{Y})D'$

x) $Cov(\underline{X} + \underline{c}, \underline{Y} + \underline{b}) = Cov(\underline{X}, \underline{Y})$

Demostración. Sólo se hará la prueba de la propiedad iii)

El elemento $(1, j)$ de la matriz $(\underline{X}'B)_{1\times p}$ es el producto punto de $\underline{X}'$ con la columna j de B, es decir, $\sum_{i=1}^{p} X_i b_{ij}$. Luego $((\underline{X}'B)\underline{X})_{1\times 1}$ es igual a $\sum_{j=1}^{p}((\sum_{i=1}^{p} X_i b_{ij})(X_j))$. Entonces,

$$\begin{aligned} E(\underline{X}'B\underline{X}) &= E(\textstyle\sum_{j=1}^{p}((\sum_{i=1}^{p} X_i b_{ij})(X_j))) = E(\sum_{j=1}^{p}\sum_{i=1}^{p}(X_i b_{ij} X_j)) \\ &= \textstyle\sum_{j=1}^{p}\sum_{i=1}^{p} E(b_{ij} X_i X_j)) = \sum_{j=1}^{p}\sum_{i=1}^{p} b_{ij} E(X_i X_j)). \end{aligned}$$

Ahora, como $E(X_i X_j) = Cov(X_i, X_j) + E(X_i)E(X_j)$, se tiene que:

$$\begin{aligned} &E(\underline{X}'B\underline{X}) = \textstyle\sum_{j=1}^{p}\sum_{i=1}^{p}(b_{ij}(Cov(X_i, X_j) + E(X_i)E(X_j))) \\ &= \textstyle\sum_{j=1}^{p}\sum_{i=1}^{p}(b_{ij}Cov(X_i, X_j)) + \sum_{j=1}^{p}\sum_{i=1}^{p}(b_{ij}E(X_i)E(X_j)) \\ &= \textstyle\sum_{j=1}^{p}(\sum_{i=1}^{p}(b_{ij}Cov(X_i, X_j))) + \sum_{j=1}^{p}((\sum_{i=1}^{p} E(X_i)b_{ij})E(X_j)). \end{aligned}$$

El elemento (j, l) de la matriz ΣB es el producto punto de la fila j de Σ con la columna l de B, es decir, $\sum_{i=1}^{p}(Cov(X_j, X_i)b_{il}) = \sum_{i=1}^{p}(b_{il}Cov(X_i, X_j))$. Luego $\sum_{i=1}^{p}(b_{ij}$

$Cov(X_i, X_j))$ es el elemento (j, j) de la matriz ΣB. Por lo tanto el primer término de la última expresión es la suma de los elementos de la diagonal de ΣB, es decir, la traza de esa matriz. Análogamente, como se calculó $\underline{X}'B\underline{X} = \sum_{j=1}^{p}((\sum_{i=1}^{p} X_i b_{ij})(X_j))$, se puede obtener $E(\underline{X}')BE(\underline{X}) = \sum_{j=1}^{p}((\sum_{i=1}^{p} E(X_i)b_{ij})E(X_j))$. Por lo tanto:

$$E(\underline{X}'B\underline{X}) = Tr(\Sigma B) + E(\underline{X}')BE(\underline{X}) = E(\underline{X}')BE(\underline{X}) + Tr(B\Sigma).$$

□

Teorema 1.19. *Si en el modelo de regresión lineal múltiple* $\varepsilon_i \sim iidN(0, \sigma^2)$, $i = 1, 2, \ldots, n$, *entonces el vector aleatorio* $\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \sim N_n(\mu_{\underline{Y}}, \Sigma_{\underline{Y}})$ *con* $\mu_{\underline{Y}} = E(\underline{Y}) = X\underline{\beta}$ *y* $\Sigma_{\underline{Y}} = Var(\underline{Y}) = \sigma^2 I_{n\times n}$.

Demostración. Sea $\underline{a}_{n\times 1}$ un vector constante, luego $\underline{a}'\underline{Y} = \sum_{k=1}^{n}(a_k Y_k)$ es una combinación lineal de variables con distribución normal, por lo tanto se distribuye normalmente. Luego $\underline{Y}$ tiene una distribución normal n-variada.

Usando la propiedad I) del Teorema 1.18 y la suposición de los errores $E(\varepsilon_i) = 0$, $i = 1, \ldots, n$, se tiene $E(\underline{Y}) = E(X\underline{\beta} + \underline{\varepsilon}) = X\underline{\beta} + E(\underline{\varepsilon}) = X\underline{\beta}$.

Usando la propiedad IV) del Teorema 1.18, además de que Y_i y Y_j, $i \neq j$, son variables aleatorias independientes.

$$\Sigma_{\underline{Y}} = \begin{bmatrix} Var(Y_1) & Cov(Y_1, Y_2) & \cdots & Cov(Y_1, Y_n) \\ Cov(Y_1, Y_2) & Var(Y_2) & \cdots & Cov(Y_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_1, Y_n) & Cov(Y_2, Y_n) & \cdots & Var(Y_n) \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

□

En seguida se enuncian 3 teoremas que se usarán más adelante.

Teorema 1.20. *Sea* $\underline{Y}_{n\times 1}$ *un vector aleatorio con distribución* $N_n(\underline{\mu}, I)$. *Si se define* U *por* $U = \underline{Y}'\underline{Y}$, *entonces* U *tiene f.d.p. dada por:*

$$f_U(u; n, \lambda) = \begin{cases} \sum_{j=0}^{\infty}\left(\frac{e^{-\lambda}\lambda^j}{j!}\right)\left(\frac{u^{(n+2j-2)/2}e^{-u/2}}{\Gamma\left(\frac{n+2j}{2}\right)2^{j+(n/2)}}\right) & \text{para } u > 0 \\ 0 & \text{para } u \leqslant 0, \end{cases} \tag{1.7}$$

donde $\lambda = \frac{1}{2}\underline{\mu}'\underline{\mu}$ *(por tanto* $\lambda \geqslant 0$*), y se define* $\lambda^j = 1$ *cuando* $\lambda = 0$, $j = 0$.

La demostración se puede ver en [7], págs. 125-126.

La distribución definida por la f.d.p. dada en (1.7) es llamada la distribución chi-cuadrada no central. La cantidad n (un entero positivo) representa los grados de libertad y λ es llamado el parámetro de no centralidad. Se usará el símbolo $\chi^2_{(n;\lambda)}$ para denotar la f.d.p. Claramente $\lambda = 0$ si y sólo si $\mu = 0$ y en ese caso la f.d.p. se simplifica a la chi-cuadrada "central", en cuyo caso se denotará como $\chi^2_{(n)}$.

La demostración de los siguientes teoremas puede verse en [7], págs. 135-139:

Teorema 1.21. *Sea $\underline{Y}_{n\times 1}$ un vector aleatorio con distribución $N_n(\underline{\mu}, \Sigma)$, donde Σ tiene rango n. La forma cuadrática $U = \underline{Y}'A\underline{Y}$ tiene distribución $\chi^2_{(p;\lambda)}$, donde $\lambda = \frac{1}{2}\underline{\mu}'A\underline{\mu}$ si y sólo si alguna de las siguientes condiciones se cumple:*

i) *$A\Sigma$ es una matriz idempotente de rango p.*

ii) *ΣA es una matriz idempotente de rango p.*

iii) *Σ es una inversa condicional de A ($A\Sigma A = A$) y A tiene rango p.*

Teorema 1.22. *Sea $\underline{Y}_{n\times 1}$ un vector aleatorio con distribución $N_n(\underline{\mu}, I)$ y sea $B_{q\times n}$ una matriz. El vector $B\underline{Y}_{q\times 1}$ es independiente de la forma cuadrática $\underline{Y}'A\underline{Y}$ si $BA = 0$.*

Teorema 1.23. *Sea $\underline{Y}_{n\times 1}$ un vector aleatorio con distribución $N_n(\underline{\mu}, \Sigma)$, donde Σ tiene rango n. Si $A\Sigma B = 0$, entonces las formas cuadráticas $\underline{Y}'A\underline{Y}$ y $\underline{Y}'B\underline{Y}$ son independientes.*

Las propiedades de los estimadores de mínimos cuadrados se enuncian en el siguiente teorema.

Teorema 1.24. *Los estimadores de mínimos cuadrados en la regresión lineal múltiple cumplen lo siguiente:*

i) *$E(\underline{\widehat{\beta}}) = \underline{\beta}$.*

ii) *$\Sigma_{\underline{\widehat{\beta}}} = Cov(\underline{\widehat{\beta}}, \underline{\widehat{\beta}}) = \sigma^2(X'X)^{-1}$.*

iii) *Un estimador insesgado de σ^2 es $S^2 = \frac{SCE}{n-(p+1)}$.*

Si además, las $\varepsilon_i \sim iidN(0, \sigma^2)$, $i = 1, 2, \ldots, n$:

iv) *Cada $\widehat{\beta}_i$ está distribuida normalmente.*

v) *La variable aleatoria $\frac{(n-(p+1))S^2}{\sigma^2}$ tiene una distribución χ^2 con $n-(p+1)$ grados de libertad.*

vi) *Los estadísticos S^2 y $\underline{\widehat{\beta}}$ son independientes.*

Demostración.

i) $E(\underline{\widehat{\beta}}) = E((X'X)^{-1}X'\underline{Y}) = (X'X)^{-1}X'E(\underline{Y}) = (X'X)^{-1}X'X\underline{\beta} = \underline{\beta}$.

ii) Usando la propiedad v) del Teorema 1.18, se tiene:

$$\begin{aligned}\Sigma_{\underline{\widehat{\beta}}} &= E(\underline{\widehat{\beta}}\,\underline{\widehat{\beta}}') - \underline{\mu}_{\underline{\widehat{\beta}}}\underline{\mu}'_{\underline{\widehat{\beta}}} = E((X'X)^{-1}X'\underline{Y}\;\underline{Y}'X(X'X)^{-1}) - \underline{\beta}\;\underline{\beta}' \\ &= (X'X)^{-1}X'E(\underline{Y}\;\underline{Y}')X(X'X)^{-1} - \underline{\beta}\;\underline{\beta}'.\end{aligned}$$

De la propiedad v) del Teorema 1.18 también se tiene: $E(\underline{Y}\;\underline{Y}') = \Sigma_{\underline{Y}} + \underline{\mu}_{\underline{Y}}\;\underline{\mu}'_{\underline{Y}} = \sigma^2 I_{n\times n} + (X\underline{\beta})(X\underline{\beta})'$. Luego,

$$\begin{aligned}\Sigma_{\underline{\widehat{\beta}}} &= (X'X)^{-1}X'(\sigma^2 I_{n\times n} + (X\underline{\beta})(X\underline{\beta})')X(X'X)^{-1} - \underline{\beta}\;\underline{\beta}' \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} + (X'X)^{-1}X'X\underline{\beta}\;\underline{\beta}'X'X(X'X)^{-1} - \underline{\beta}\;\underline{\beta}' \\ &= \sigma^2(X'X)^{-1} + \underline{\beta}\;\underline{\beta}' - \underline{\beta}\;\underline{\beta}' \\ &= \sigma^2(X'X)^{-1}.\end{aligned}$$

Esta igualdad es equivalente a:

- $Var\left(\widehat{\beta}_i\right) = \sigma^2_{\widehat{\beta}_i} = c_{ii}\sigma^2$, donde c_{ii} es el elemento en la fila i y columna i de $(X'X)^{-1}$.
- $Cov\left(\widehat{\beta}_i, \widehat{\beta}_j\right) = c_{ij}\sigma^2$, donde c_{ij} es el elemento en la fila i y columna j de $(X'X)^{-1}$.

iii) Usando la propiedad iii) del Teorema 1.18, se tiene:

$$\begin{aligned}&E(SCE) = E(\underline{Y}'(I - X(X'X)^{-1}X')\underline{Y}) \\ &= E(\underline{Y}')(I - X(X'X)^{-1}X')E(\underline{Y}) + tr((I - X(X'X)^{-1}X')Var(\underline{Y})) \\ &= E(\underline{Y}')(X\underline{\beta} - X(X'X)^{-1}X'X\underline{\beta}) + \sigma^2 tr(I_{n\times n} - X(X'X)^{-1}X') \\ &= E(\underline{Y}')(X\underline{\beta} - X\underline{\beta}) + \sigma^2(tr(I_{n\times n}) - tr(X(X'X)^{-1}X')) \\ &= \sigma^2(n - tr((X'X)^{-1}X'X)) = \sigma^2(n - tr(I_{(p+1)\times(p+1)}) \\ &= \sigma^2(n-(p+1)),\end{aligned}$$

de donde

$$E(S^2) = E\left(\frac{SCE}{n-(p+1)}\right) = \frac{E(SCE)}{n-(p+1)} = \sigma^2.$$

iv) Como $\underline{\widehat{\beta}} = (X'X)^{-1}X'\underline{Y}$, $\widehat{\beta}_i$ es el producto punto de la fila i de la matriz de elementos de valor real $(X'X)^{-1}X'_{(p+1)\times n}$ con $\underline{Y}_{n\times 1}$, es decir, $\widehat{\beta}_i = \sum_{j=1}^{n} d_{ij}Y_j$, donde d_{ij} es el elemento (i,j) de la matriz $(X'X)^{-1}X'$, entonces $\widehat{\beta}_i$ es una combinación lineal de variables aleatorias normales, por lo tanto $\widehat{\beta}_i$ está distribuida normalmente.

v) Nótese que se cumple la siguiente igualdad:

$$\frac{(n-(p+1))S^2}{\sigma^2} = \frac{(n-(p+1))}{\sigma^2}\frac{SCE}{n-(p+1)} = SCE/\sigma^2 = \underline{Y}'A\underline{Y}, \tag{1.8}$$

donde $A = (\sigma^2)^{-1}(I - X(X'X)^{-1}X')$.

A continuación se verá que:

$$A\Sigma_{\underline{Y}} = (\sigma^2)^{-1}(I_{n\times n} - X(X'X)^{-1}X')\sigma^2 I_{n\times n} = I_{n\times n} - X(X'X)^{-1}X'$$

es una matriz idempotente,

$$\begin{aligned}
&(I_{n\times n} - X(X'X)^{-1}X')^2 \\
&= I_{n\times n}(I_{n\times n} - X(X'X)^{-1}X') - X(X'X)^{-1}X'(I_{n\times n} - X(X'X)^{-1}X') \\
&= I_{n\times n} - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X'X(X'X)^{-1}X' \\
&= I_{n\times n} - 2X(X'X)^{-1}X' + X(X'X)^{-1}X' = I_{n\times n} - X(X'X)^{-1}X',
\end{aligned}$$

por el Teorema B.3 se sigue que:

$$\begin{aligned}
rang(A) &= tr(A) = tr(I_{n\times n} - X(X'X)^{-1}X') \\
&= n - tr(X(X'X)^{-1}X') = n - tr((X'X)^{-1}X'X) \\
&= n - tr(I_{p+1\times p+1}) = n-(p+1).
\end{aligned}$$

Del Teorema 1.19 se tiene que $\underline{Y} \sim N_n(X\underline{\beta}, \sigma^2 I_{n\times n})$, usando el Teorema 1.21 se sigue que la forma cuadrática $SCE/\sigma^2 = \underline{Y}'A\underline{Y}$ está distribuida como $\chi^2_{(n-(p+1),\lambda)}$, donde

$$\begin{aligned}
\lambda &= 2^{-1}(X\underline{\beta})'(\sigma^2)^{-1}(I_{n\times n} - X(X'X)^{-1}X')X\underline{\beta} \\
&= (2\sigma^2)^{-1}(X\underline{\beta})'(X\underline{\beta} - X\underline{\beta}) = 0.
\end{aligned}$$

Por lo tanto $SCE/\sigma^2 \sim \chi^2_{(n-(p+1))}$.

VI) Como $\underline{Y} \sim N_n(X\underline{\beta}, \sigma^2 I)$, entonces, $\underline{Y}/\sigma \sim N_n(\sigma^{-1}X\underline{\beta}, I)$ de la Ecuación (1.8) se sigue que $S^2 = (\underline{Y}/\sigma)'A(\underline{Y}/\sigma)$, con $A = \sigma^2(n-(p+1))^{-1}(I - X(X'X)^{-1}X')$; por otro lado, $\widehat{\underline{\beta}} = B\underline{Y}$ con $B = (X'X)^{-1}X'$. Pero,

$$\begin{aligned} BA &= (X'X)^{-1}X'\sigma^2(n-(p+1))^{-1}(I - X(X'X)^{-1}X') \\ &= \sigma^2(n-(p+1))^{-1}(X'X)^{-1}X'(I - X(X'X)^{-1}X') \\ &= \sigma^2(n-(p+1))^{-1}((X'X)^{-1}X' - (X'X)^{-1}X') \\ &= 0_{(p+1)\times n} \end{aligned}$$

y por el Teorema 1.22 se sigue que S^2 y $\widehat{\underline{\beta}}$ son independientes. □

Los siguientes corolarios son consecuencias del Teorema 1.24.

Corolario 1.25. *La distribución del i-ésimo estimador de mínimos cuadrados $\widehat{\beta}_i$ es $N(\beta_i, \sigma^2_{\widehat{\beta}_i})$, donde $\sigma^2_{\widehat{\beta}_i} = c_{ii}\sigma^2$ y c_{ii} es el elemento ii-ésimo de la matriz $(X'X)^{-1}$, $i = 0, 1, \ldots, p$.*

Demostración. El resultado se sigue de I), II) y IV) del Teorema 1.24. □

Corolario 1.26. $\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \ldots + \widehat{\beta}_p x_{ip}$, $i = 1, \ldots, n$ *tiene una distribución normal.*

Demostración. $\widehat{Y}_i$ es una variable aleatoria normal, ya que es una combinación lineal de variables aleatorias normales, por el Corolario 1.25. □

El estimador insesgado de la varianza $\sigma^2_{\widehat{\beta}_i}$ de cada coeficiente de mínimos cuadrados $\widehat{\beta}_i$ es $\widehat{\sigma}^2_{\widehat{\beta}_i} = c_{ii}S^2$, donde c_{ii} es el elemento en la fila i y columna i de $(X'X)^{-1}$.

Corolario 1.27. *Cuando $\varepsilon_i \sim iidN(0, \sigma^2)$, $i = 1, 2, \ldots, n$ en el modelo de regresión lineal múltiple, la cantidad $\frac{\widehat{\beta}_i - \beta_i}{\widehat{\sigma}_{\widehat{\beta}_i}}$ tiene una distribución t de Student con $n-(p+1)$ grados de libertad, esto es: $\frac{\widehat{\beta}_i - \beta_i}{\widehat{\sigma}_{\widehat{\beta}_i}} \sim t_{n-(p+1)}$.*

Demostración.

$$\frac{\widehat{\beta}_i - \beta_i}{\widehat{\sigma}_{\widehat{\beta}_i}} = \frac{\frac{\widehat{\beta}_i - \beta_i}{\sigma}}{\sqrt{\frac{\widehat{\sigma}^2_{\widehat{\beta}_i}}{\sigma^2}}} = \frac{\frac{\widehat{\beta}_i - \beta_i}{\sigma}}{\sqrt{\frac{c_{ii}S^2}{\sigma^2}}} = \frac{\frac{\widehat{\beta}_i - \beta_i}{\sigma\sqrt{c_{ii}}}}{\sqrt{\frac{1}{(n-(p+1))}\frac{(n-(p+1))S^2}{\sigma^2}}}$$

Como $\frac{\widehat{\beta}_i - \beta_i}{\sigma\sqrt{c_{ii}}} \sim N(0,1)$ y $\frac{(n-(p+1))S^2}{\sigma^2} \sim \chi^2_{(n-(p+1))}$, además son independientes dado que $\widehat{\beta}_i$ y S^2 son independientes, siguiéndose que $\frac{\widehat{\beta}_i - \beta_i}{\widehat{\sigma}_{\widehat{\beta}_i}} \sim t_{n-(p+1)}$. □

Con esta variable aleatoria pivotal se calculan los intervalos de confianza y bajo H_0 se utiliza como un estadístico para realizar pruebas de hipótesis sobre los valores de β_i.

1.7. Coeficiente de determinación

Un estadístico de la bondad del ajuste representa una cantidad que mide qué tan bien un modelo explica un conjunto específico de datos.

Tanto la recta de mínimos cuadrados como la recta $Y = \overline{Y}$ se pueden usar para pronosticar. Cuando se utiliza $Y = \overline{Y}$, los errores de predicción son $Y_i - \overline{Y}$, y la suma de los errores pronosticados al cuadrado será $\sum_{i=1}^{n} (Y_i - \overline{Y})^2$. Si se utiliza la recta de mínimos cuadrados, el error pronosticado será $Y_i - \widehat{Y}_i$, y la suma de los errores pronosticados al cuadrado es $\sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$. La fuerza de la relación lineal se mide al calcular la reducción obtenida en la suma de los errores pronosticados al cuadrado con $\widehat{Y}_i$, en lugar de $\overline{Y}$, esto es, la diferencia $\sum_{i=1}^{n} (Y_i - \overline{Y})^2 - \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$. Cuanto mayor sea la diferencia, más fuerte será la agrupación de puntos alrededor de la recta de mínimos cuadrados. Por lo tanto, $\sum_{i=1}^{n} (Y_i - \overline{Y})^2 - \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$ puede usarse como un estadístico de la bondad de ajuste.

Sin embargo, existe un problema al utilizar

$$\sum_{i=1}^{n} (Y_i - \overline{Y})^2 - \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$$

como un estadístico de la bondad del ajuste. Esta cantidad tiene unidades cuadradas de Y. No se podría usar este estadístico para comparar la bondad del ajuste de dos modelos que ajusten diferentes conjuntos de datos, puesto que las unidades serían diferentes. Se necesita utilizar un estadístico de la bondad del ajuste que sea un número puro para que se pueda medir la bondad del ajuste en una escala absoluta. Considerando esta situación, se da la siguiente definición.

Definición 1.28. *El estadístico de bondad del ajuste en la regresión múltiple representa una cantidad que se denota mediante R^2, denominado* ***coeficiente de determinación****. Éste es:*

$$R^2 = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2 - \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2} = \frac{STC - SCE}{STC} = \frac{SCR}{STC}.$$

El coeficiente de determinación es la reducción obtenida en la suma de los errores pronosticados al cuadrado al utilizar $\widehat{Y}_i$ en lugar $\overline{Y}$, expresado como una fracción de la suma de los errores pronosticados al cuadrado $\sum_{i=1}^{n}(Y_i - \overline{Y})^2$, obtenidos al usar $\overline{Y}$.

Puesto que la suma total de cuadrados de las desviaciones ($STC = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$) es exactamente la varianza muestral de los Y_i sin dividir entre $n-1$, la cantidad STC proporciona una medida de la variación total entre los valores Y, pasando por alto las x_1, x_2, ..., x_p. De manera alternativa, la SCE mide la variación en los valores Y que permanecen sin explicación después de usar las x_1, x_2, ..., x_p para ajustar el modelo de regresión lineal múltiple. Entonces, $\frac{SCE}{STC}$ da la proporción de la variación total en las Y_i que no explica el modelo de regresión lineal. Así, $R^2 = 1 - \frac{SCE}{STC}$ se puede interpretar como la *proporción de la variación total en las* Y_i*, explicada por las covariables* $x_1, x_2, \ldots, x_p$ *en un modelo de regresión lineal múltiple* o *proporción de la varianza en* Y *explicada por la regresión*.

Como $0 \leqslant \frac{SCR}{STC} = R^2 = 1 - \frac{SCE}{STC} \leqslant 1$. Si el modelo de regresión lineal múltiple se ajusta bien a los datos, las diferencias entre los valores observados y los pronosticados son pequeñas, lo cual lleva a un valor pequeño para la SCE, por lo que el R^2 es cercano a 1. De igual manera, si el modelo de regresión se ajusta mal, la SCE será grande, en este caso el R^2 es cercano a 0.

1.8. Prueba de hipótesis

En la regresión lineal simple casi siempre se prueba la hipótesis nula $\beta_1 = 0$; si ésta no se rechaza, el modelo lineal no será útil. La hipótesis nula análoga en la regresión múltiple es: $H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$, ésta es una hipótesis muy fuerte, establece que ninguna de las variables independientes tiene relación lineal con la variable dependiente. En la práctica, los datos por lo general proporcionan evidencia suficiente para rechazar esta hipótesis. A continuación se construye la prueba de razón de verosimilitudes para contrastar el juego de hipótesis $H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$ vs $H_1 : Al\ menos\ un\ \beta_i \neq 0$. Por lo que se analiza la función de verosimilitud y los puntos donde alcanza su máximo dependiendo del espacio donde ésta esté definida.

Teorema 1.29. *El estadístico de prueba para la hipótesis* $H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$ *es:*

$$F = \frac{SCR/p}{SCE/(n-p-1)},$$

y su distribución nula (bajo la hipótesis nula) es $F_{p,n-(p+1)}$*. (Obsérvese que el denominador del estadístico* F *es* S^2*, un estimador de* σ^2*).*

Demostración. En la regresión lineal múltiple se desconoce el parámetro $\underline{\theta} = (\beta_0, \beta_1, \ldots, \beta_p, \sigma^2)'$, luego el espacio paramétrico es $\overline{\underline{\Theta}} = \mathbb{R}^{p+1} \times \mathbb{R}_+$. Considérese la hipótesis $H_0 : \underline{\theta} \in \overline{\underline{\Theta}}_0 = \{(\beta_0, \beta_1, \ldots, \beta_p, \sigma^2)' \in \overline{\underline{\Theta}} \mid \beta_1 = \ldots = \beta_p = 0\}$ en contraste con la hipótesis $H_1 : \underline{\theta} \in \overline{\underline{\Theta}}_1 = \overline{\underline{\Theta}} - \overline{\underline{\Theta}}_0$; H_0 y H_1 son hipótesis compuestas.

A continuación se obtiene la prueba de razón de verosimilitudes generalizada, para lo cual, en principio, se obtendrán los estimadores de M. V. restringidos al subconjunto determinado en la hipótesis nula.

Sea $\underline{\theta}_0 = (\beta_0, 0, \cdots, 0, \sigma^2)' \in \overline{\underline{\Theta}}_0$; usando el Lema 1.12 se tiene que $l(\underline{\theta}_0; \underline{y}) = -nln(\sqrt{2\pi}) - nln((\sigma^2)^{1/2}) - \sum_{i=1}^n (y_i - \beta_0)^2/2\sigma^2$, cuyas parciales con respecto a β_0 y σ^2 son:

$$\begin{array}{rcl} \frac{\partial l(\underline{\theta}_0;\underline{y})}{\partial \beta_0} & = & \sum_{i=1}^n (y_i - \beta_0)/\sigma^2 \\ \frac{\partial l(\underline{\theta}_0;\underline{y})}{\partial \sigma^2} & = & -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (y_i-\beta_0)^2}{2(\sigma^2)^2} \quad = \quad \frac{-n\sigma^2+\sum_{i=1}^n (y_i-\beta_0)^2}{2(\sigma^2)^2} \end{array}$$

Al igualar estas derivadas parciales a cero, se tiene que:

$$\begin{array}{lll} \sum_{i=1}^n (y_i - \beta_0) = 0 & \Leftrightarrow & \beta_0 = n^{-1} \sum_{i=1}^n y_i = \overline{y} \\ -n\sigma^2 + \sum_{i=1}^n (y_i - \beta_0)^2 = 0 & \Leftrightarrow & \sigma^2 = n^{-1} \sum_{i=1}^n (y_i - \beta_0)^2 \\ & \Leftrightarrow & \sigma^2 = n^{-1} \sum_{i=1}^n (y_i - \overline{y})^2 \\ & \Leftrightarrow & \sigma^2 = n^{-1} STC. \end{array}$$

La matriz que contiene las segundas derivadas parciales es:

$$\mathcal{H}(\beta_0, \sigma^2) = -\frac{1}{\sigma^2} \begin{bmatrix} n & \frac{\sum_{i=1}^n (y_i-\beta_0)}{\sigma^2} \\ \frac{\sum_{i=1}^n (y_i-\beta_0)}{\sigma^2} & \frac{\sum_{i=1}^n (y_i-\beta_0)^2}{(\sigma^2)^2} - \frac{n}{2\sigma^2} \end{bmatrix}.$$

Al evaluar $\mathcal{H}(\beta_0, \sigma^2)$ en $(\overline{y}, n^{-1}STC)$, se obtiene:

$$\mathcal{H}(\overline{y}, n^{-1}STC) = -\frac{n}{STC}\begin{bmatrix} n & \frac{\sum_{i=1}^{n}(y_i-\overline{y})}{n^{-1}STC} \\ \frac{\sum_{i=1}^{n}(y_i-\overline{y})}{n^{-1}STC} & \frac{\sum_{i=1}^{n}(y_i-\overline{y})^2}{(n^{-1}STC)^2} - \frac{n}{2n^{-1}STC} \end{bmatrix}$$

$$= -\frac{n}{STC}\begin{bmatrix} n & 0 \\ 0 & \frac{n^2 STC}{STC^2} - \frac{n^2}{2STC} \end{bmatrix} = -\frac{n}{STC}\begin{bmatrix} n & 0 \\ 0 & \frac{n^2}{2STC} \end{bmatrix}.$$

Como $n > 0$ y el $Det\left(-n^{-1}STC\mathcal{H}\left(\overline{y}, n^{-1}STC\right)\right) > 0$, entonces "la matriz de las segundas derivadas evaluadas en $\overline{y}$ y $n^{-1}STC$" se define negativa. Así que $(\overline{y}, 0, \ldots, 0, n^{-1}STC)$ maximiza a la f. v. $L(\underline{\theta}; \underline{y})$ en $\overline{\underline{\Theta}}_0$. Además se tiene que:

$$\begin{aligned} \text{máx}_{\underline{\theta}\in\overline{\underline{\Theta}}_0}\{L(\underline{\theta}; \underline{y})\} &= L(\overline{y}, 0, \ldots, 0, n^{-1}STC; \underline{y}) \\ &= \left(\tfrac{n}{2\pi STC}\right)^{n/2} e^{-n\sum_{i=1}^{n}(y_i-\overline{y})^2/2STC} \\ &= \left(\tfrac{n}{2\pi STC}\right)^{n/2} e^{-nSTC/2STC} = \left(\tfrac{n}{2\pi STC}\right)^{n/2} e^{-n/2}. \end{aligned}$$

Por el Teorema 1.13, se tiene que bajo $\overline{\underline{\Theta}}$, los estimadores de máxima verosimilitud de los parámetros $\underline{\beta}$ y σ^2 son $\widehat{\underline{\beta}} = (X'X)^{-1}X'\underline{Y}$ (los estimadores de mínimos cuadrados) y $\sigma^2 = n^{-1}SCE$, de forma que la razón de verosimilitudes, usando los resultados anteriores, es:

$$\lambda(\underline{y}) = \frac{\text{máx}_{\underline{\theta}\in\overline{\underline{\Theta}}_0}\{L(\underline{\theta}; \underline{y})\}}{\text{máx}_{\underline{\theta}\in\overline{\underline{\Theta}}}\{L(\underline{\theta}; \underline{y})\}} = \frac{\left(\frac{n}{2\pi STC}\right)^{n/2} e^{-n/2}}{\left(\frac{n}{2\pi SCE}\right)^{n/2} e^{-n/2}} = \left(\frac{STC}{SCE}\right)^{-n/2}$$

La prueba que es la más potente de tamaño α ($\alpha \in [0, 1]$) es igual a 1 si y sólo si

$$\begin{aligned} \left(\tfrac{STC}{SCE}\right)^{-n/2} \leqslant k_1 &\Leftrightarrow \tfrac{STC}{SCE} \geqslant k_2 \\ &\Leftrightarrow \tfrac{n-(p+1)}{p}\left(\tfrac{STC}{SCE} - \tfrac{SCE}{SCE}\right) \geqslant \tfrac{n-(p+1)}{p}(k_2 - 1) \\ &\Leftrightarrow \tfrac{SCR/p}{SCE/(n-(p+1))} \geqslant k. \end{aligned}$$

Por lo tanto la función de prueba es:

$$\phi(\underline{y}) = \begin{cases} 1, & \text{si } \frac{SCR/p}{SCE/(n-(p+1))} \geqslant k \\ 0, & \text{si } \frac{SCR/p}{SCE/(n-(p+1))} < k, \end{cases}$$

donde k es tal que

$$E_{\underline{\theta}_0}\{\phi(\underline{Y})\} = \int_{\gamma} \phi(\underline{y}) f_{\underline{Y}}(\underline{y}; \underline{\theta}_0) d\underline{y} = \alpha,$$

donde γ es el soporte del vector aleatorio $\underline{Y}$. Como

$$E_{\underline{\theta}_0}\{\phi(\underline{Y})\} = P(\phi(\underline{Y}) = 1 \mid \underline{\theta}_0) = P\left(\frac{SCR/p}{SCE/(n-(p+1))} \geqslant k \mid \underline{\theta}_0\right),$$

lo siguiente es encontrar la distribución de $\frac{SCR/p}{SCE/(n-(p+1))}$ bajo H_0.

De la Definición 1.8 se tiene que $SCR/\sigma^2 = \underline{Y}' A \underline{Y}$, donde $A = (\sigma^2)^{-1}\left(X(X'X)^{-1}X' - n^{-1}\underline{1}\underline{1}'\right)$. Además se tiene que

$$A\Sigma_{\underline{Y}} = (\sigma^2)^{-1}(X(X'X)^{-1}X' - n^{-1}\underline{1}\underline{1}')\sigma^2 I_{n\times n} = (X(X'X)^{-1}X' - n^{-1}\underline{1}\underline{1}')$$

es una matriz idempotente. Un caso más general se puede consultar en [7], págs. 185-191, que en la notación de la referencia se tiene $X^- = (X'X)^{-1}X'$, $B = XG^-$, $B^- = (B'B)^{-1}B'$, $G^- = G'(GG')^{-1}$,

$$H_{p\times p+1} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \text{ y } G_{1\times p+1} = (1, 0, 0, \cdots, 0).$$

Por el Teorema B.3 se sigue que:

$$\begin{aligned} rang(A\Sigma_{\underline{Y}}) &= tr(A\Sigma_{\underline{Y}}) = tr(X(X'X)^{-1}X' - n^{-1}\underline{1}\underline{1}') \\ &= tr(X(X'X)^{-1}X') - tr(n^{-1}\underline{1}\underline{1}') \\ &= tr((X'X)^{-1}X'X) - n^{-1}tr(\underline{1}\underline{1}') \\ &= tr(I_{p+1\times p+1}) - n^{-1}n \\ &= p + 1 - 1 \\ &= p. \end{aligned}$$

Bajo H_0, por el Teorema 1.19 se tiene que $\underline{Y} \sim N_n(X\underline{\beta_0}, \sigma^2 I_{n\times n})$, donde $\underline{\beta_0} = (\beta_0, 0, \ldots, 0)'$. Usando el Teorema 1.21, se tiene que la forma cuadrática $SCR/\sigma^2 = \underline{Y}' A \underline{Y}$ está distribuida como $\chi^2_{(p,\lambda)}$, donde:

$$\begin{aligned}
\lambda &= 2^{-1}(X\underline{\beta_0})'(\sigma^2)^{-1}(X(X'X)^{-1}X' - n^{-1}\underline{11}')X\underline{\beta_0} \\
&= (2\sigma^2)^{-1}\underline{\beta_0}'X'(X(X'X)^{-1}X' - n^{-1}\underline{11}')X\underline{\beta_0} \\
&= (2\sigma^2)^{-1}(\underline{\beta_0}'X' - \underline{\beta}'X'n^{-1}\underline{11}')X\underline{\beta_0} \\
&= (2\sigma^2)^{-1}\underline{\beta_0}'X'(I - n^{-1}\underline{11}')X\underline{\beta_0} \\
&= -(2n\sigma^2)^{-1}\underline{\beta_0}'X'(-nI + \underline{11}')X\underline{\beta_0}.
\end{aligned}$$

Recordemos que la primera columna de la matriz X está formada por 1's, por lo que: $X\underline{\beta_0} = X(\beta_0, 0, \ldots, 0)' = (x_{10}\beta_0, x_{20}\beta_0, \ldots, x_{n0}\beta_0)' = \beta_0(1, \ldots, 1)'$. Luego,

$$\begin{aligned}
\lambda &= -\beta_0^2(2n\sigma^2)^{-1}(1, \ldots, 1)(-nI + \underline{11}')(1, \ldots, 1)' \\
&= -\beta_0^2(2n\sigma^2)^{-1}(1, \ldots, 1)\begin{bmatrix} 1-n & 1 & \ldots & 1 \\ 1 & 1-n & \ldots & 1 \\ \vdots & \vdots & \ddots & \ldots \\ 1 & 1 & \ldots & 1-n \end{bmatrix}_{n\times n}(1, \ldots, 1)' \\
&= -\beta_0^2(2n\sigma^2)^{-1}(0, \ldots, 0)(1, \ldots, 1)' \\
&= 0.
\end{aligned}$$

Por lo tanto se sigue que $SCR/\sigma^2 \sim \chi^2_{(p)}$ bajo H_0. Por v) del Teorema 1.24, se tiene que $\frac{(n-(p+1))S^2}{\sigma^2} = SCE/\sigma^2 \sim \chi^2_{(n-(p+1))}$. Por lo tanto, bajo H_0, se cumple que:

$$\frac{SCR/p}{SCE/(n-(p+1))} = \frac{\frac{SCR/\sigma^2}{p}}{\frac{SCE/\sigma^2}{(n-(p+1))}} \sim F_{p,n-(p+1)}.$$

□

1.9. Inferencias respecto a funciones lineales de los parámetros del modelo

Supongamos que se desea hacer inferencia acerca de la función lineal:

$$a_0\beta_0 + a_1\beta_1 + a_2\beta_2 + \ldots + a_p\beta_p,$$

donde a_0, a_1, a_2, …, a_p son constantes conocidas (algunas de las cuales pueden ser iguales a cero). Definiendo el vector $\underline{a}_{(p+1)\times 1} = (a_0, a_1, a_2, \ldots, a_p)'$, se deduce que una

combinación lineal de las β_0, β_1, β_2, ..., β_p correspondiente a a_0, a_1, a_2, ..., a_p se puede expresar como:

$$\underline{a}'\underline{\beta} = a_0\beta_0 + a_1\beta_1 + a_2\beta_2 + \ldots + a_p\beta_p.$$

De aquí en adelante se hace referencia a estas combinaciones lineales en su forma matricial.

Teorema 1.30. *En el modelo de regresión lineal múltiple tal que* $\varepsilon_i \sim iidN(0,\sigma^2)$, $i = 1,\ldots,n$, *un estimador insesgado de* $\underline{a}'\underline{\beta}$ *es* $\widehat{\underline{a}'\underline{\beta}} = a_0\widehat{\beta_0} + a_1\widehat{\beta_1} + a_2\widehat{\beta_2} + \ldots + a_p\widehat{\beta_p} = \underline{a}'\widehat{\underline{\beta}}$, *el cual cumple que* $\widehat{\underline{a}'\underline{\beta}} \sim N(\underline{a}'\underline{\beta}, [\underline{a}'(X'X)^{-1}\underline{a}]\sigma^2)$.

Demostración. Por la propiedad IV del Teorema 1.24, se tiene que $\widehat{\beta_0}$, $\widehat{\beta_1}$, $\widehat{\beta_2}$, ..., $\widehat{\beta_p}$ están distribuidas normalmente, luego la combinación lineal $\underline{a}'\widehat{\underline{\beta}}$ también se distribuye normalmente.

Usando la propiedad de linealidad de la esperanza (Propiedad I del Teorema 1.18), se tiene que $E(\underline{a}'\widehat{\underline{\beta}}) = \underline{a}'E(\widehat{\underline{\beta}}) = \underline{a}'\underline{\beta}$, es decir, $\widehat{\underline{a}'\underline{\beta}}$ es un estimador insesgado de $\underline{a}'\underline{\beta}$.

De la propiedad VI del Teorema 1.18 se sigue que

$$Var(\underline{a}'\widehat{\underline{\beta}}) = \underline{a}'Var(\widehat{\underline{\beta}})\underline{a} = \underline{a}'\sigma^2(X'X)^{-1}\underline{a} = \sigma^2\underline{a}'(X'X)^{-1}\underline{a},$$

con lo cual se concluye la demostración del teorema. □

Una consecuencia del Teorema 1.30 es que $\frac{\underline{a}'\widehat{\underline{\beta}}-\underline{a}'\underline{\beta}}{[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}\sigma} \sim N(0,1)$, y que al sustituir σ por su estimador S, se distribuye como una t de Student, como se muestra a continuación.

Teorema 1.31. *Si* $\varepsilon_i \sim iidN(0,\sigma^2)$, $i = 1,\ldots,n$, *en el modelo de regresión lineal múltiple se cumple que:* $\frac{\underline{a}'\widehat{\underline{\beta}}-\underline{a}'\underline{\beta}}{S\sqrt{\underline{a}'(X'X)^{-1}\underline{a}}}$ *tiene una distribución* t *de Student con* $(n-(p+1))$ *grados de libertad.*

Demostración.

$$\frac{\underline{a}'\widehat{\underline{\beta}}-\underline{a}'\underline{\beta}}{S\sqrt{\underline{a}'(X'X)^{-1}\underline{a}}} = \frac{\frac{\underline{a}'\widehat{\underline{\beta}}-\underline{a}'\underline{\beta}}{[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}\sigma}}{\sqrt{\frac{1}{(n-(p+1))}\frac{(n-(p+1))S^2}{\sigma^2}}}$$

Como $\frac{\underline{a}'\widehat{\underline{\beta}}-\underline{a}'\underline{\beta}}{[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}\sigma} \sim N(0,1)$ y $\frac{(n-(p+1))S^2}{\sigma^2} \sim \chi^2_{(n-(p+1))}$, además de ser independientes dado que $\widehat{\underline{\beta}}$ y S^2 son independientes, se sigue que:

$$\frac{\underline{a}'\widehat{\underline{\beta}} - \underline{a}'\underline{\beta}}{S\sqrt{\underline{a}'(X'X)^{-1}\underline{a}}} \sim t_{n-(p+1)}.$$

□

Teorema 1.32. *Si $\varepsilon_i \sim iidN(0,\sigma^2), i = 1,\ldots,n$, en el modelo de regresión lineal múltiple se cumple que, para contrastar los siguientes juegos de hipótesis:*

i) $H_0 : \underline{a}'\underline{\beta} = (\underline{a}'\underline{\beta})_0$ *vs* $H_1 : \underline{a}'\underline{\beta} > (\underline{a}'\underline{\beta})_0$

ii) $H_0 : \underline{a}'\underline{\beta} = (\underline{a}'\underline{\beta})_0$ *vs* $H_1 : \underline{a}'\underline{\beta} < (\underline{a}'\underline{\beta})_0$

iii) $H_0 : \underline{a}'\underline{\beta} = (\underline{a}'\underline{\beta})_0$ *vs* $H_1 : \underline{a}'\underline{\beta} \neq (\underline{a}'\underline{\beta})_0$

con un nivel de significancia α, se puede usar como estadístico de prueba a $T = \frac{\underline{a}'\widehat{\underline{\beta}}-(\underline{a}'\underline{\beta})_0}{S[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}}$, teniendo, respectivamente, las siguientes reglas de decisión:

$$\textit{Región de rechazo:} \begin{cases} t \geqslant t_{n-(p+1);\alpha} \\ t \leqslant t_{n-(p+1);\alpha} \\ |t| \geqslant t_{n-(p+1);\alpha/2}. \end{cases}$$

Demostración. Dado que $\varepsilon_i \sim iidN(0,\sigma^2)$, $i = 1,\ldots,n$, en el modelo de regresión lineal múltiple se tiene que $\underline{Y} \sim N(X\underline{\beta},\sigma^2 I)$ y se desconoce el parámetro $\underline{\theta} = (\underline{\beta},\sigma^2)$, luego el espacio paramétrico es $\overline{\underline{\Theta}} = \mathbb{R}^{p+1} \times \mathbb{R}_+$. Sea $H_0 : \underline{\theta} \in \overline{\underline{\Theta}}_0 = \{(\underline{\beta},\sigma^2) \in \overline{\underline{\Theta}} \mid \underline{a}'\underline{\beta} = (\underline{a}'\underline{\beta})_0\}$ la hipótesis nula en contraste con $H_1 : \underline{\theta} \in \overline{\underline{\Theta}} - \overline{\underline{\Theta}}_0$.

Primero se hallarán los estimadores de máxima verosimilitud bajo H_0; para ello se sustituirá la restricción $\underline{a}'\underline{\beta} = (\underline{a}'\underline{\beta})_0$ en el modelo $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$, lo cual proporciona un nuevo modelo reducido por la hipótesis. Para efectuarlo, sea $G_{p\times p+1}$ una matriz de rango igual a $rang(G) = p$ tal que $\underline{a}'G' = \underline{0}$ (G es una completación ortogonal de $\underline{a}'$). La matriz

$$\begin{bmatrix} \underline{a}' \\ G \end{bmatrix}_{p+1\times p+1}$$

es de rango $p+1$, entonces existe su inversa.

Dado que $rang(\underline{a}'_{1\times p+1}) = 1$ y $rang(G_{p\times p+1}) = p$, por el Teorema B.10 se sigue que sus inversas generalizadas son $\underline{a}'^{-} = \underline{a}(\underline{a}'\underline{a})^{-1}$ y $G^{-} = G'(GG')^{-1}$, respectivamente. Usando esto se tiene que:

$$\begin{bmatrix} \underline{a}' \\ G \end{bmatrix}^{-1}_{p+1\times p+1} = \left[\underline{a}'^{-}, G^{-}\right],$$

ya que:

$$\begin{bmatrix} \underline{a}' \\ G \end{bmatrix} \left[\underline{a}'^{-}, G^{-}\right] = \begin{bmatrix} \underline{a}'\underline{a}'^{-} & \underline{a}'G^{-} \\ G\underline{a}'^{-} & GG^{-} \end{bmatrix} = \begin{bmatrix} 1 & \underline{a}'G'(GG')^{-1} \\ G\underline{a}(\underline{a}'\underline{a})^{-1} & I_{p\times p} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \underline{0}(GG')^{-1} \\ \underline{0}'(\underline{a}'\underline{a})^{-1} & I_{p\times p} \end{bmatrix} = I_{p+1\times p+1}.$$

Nótese que $\left[\underline{a}'^{-}, G^{-}\right] \begin{bmatrix} \underline{a}' \\ G \end{bmatrix} = \underline{a}'^{-}\underline{a}' + G^{-}G$. Luego el modelo de regresión lineal múltiple se puede escribir como sigue:

$$\begin{aligned} \underline{Y} &= X\underline{\beta} + \underline{\varepsilon} = X\left[\underline{a}'^{-}, G^{-}\right] \begin{bmatrix} \underline{a}' \\ G \end{bmatrix} \underline{\beta} + \underline{\varepsilon} = X(\underline{a}'^{-}\underline{a}' + G^{-}G)\underline{\beta} + \underline{\varepsilon} \\ &= X\underline{a}'^{-}\underline{a}'\underline{\beta} + XG^{-}G\underline{\beta} + \underline{\varepsilon}. \end{aligned} \tag{1.9}$$

Como se está bajo la hipótesis H_0 de la Ecuación (1.9), se sigue que:

$$\underline{Y} = X\underline{a}'^{-}(\underline{a}'\underline{\beta})_0 + XG^{-}G\underline{\beta} + \underline{\varepsilon}, \tag{1.10}$$

considerando a $\underline{Z} = \underline{Y} - X\underline{a}'^{-}(\underline{a}'\underline{\beta})_0$, $X_{\underline{Z}} = XG^{-}$ y $\underline{\beta}_{\underline{Z}} = G\underline{\beta}$, de la Ecuación (1.10) se sigue que:

$$\underline{Z} = X_{\underline{Z}}\underline{\beta}_{\underline{Z}} + \underline{\varepsilon}, \tag{1.11}$$

donde $\underline{Z}_{n\times 1}$ es un vector aleatorio observable, $(\underline{\beta}_{\underline{Z}})_{p\times 1}$ es un vector de parámetros desconocidos que puede ser cualquier vector de $\mathbb{R}^p$ y $(X_{\underline{Z}})_{n\times p}$ es una matriz no aleatoria observable de rango $rang(X_{\underline{Z}}) = p$, ya que se supuso que $(X'X)^{-1}$ existe, es decir, $rang(X'X) = p+1$, también se cumple que $rang(X'X) = rang(X_{n\times p+1})$, así, por el Teorema B.10, se tiene que $X^{-} = (X'X)^{-1}X'$ y $X^{-}X = I$. Usando el Teorema B.9, se sigue que $p = rang(G) = rang(G^{-}) = rang(X^{-}XG^{-}) \leqslant rang(XG^{-}) \leqslant rang(G^{-}) = p$, por

lo que $rang(X_{\underline{Z}}) = p$. Se concluye que la Ecuación (1.11) es un modelo de regresión lineal múltiple con la restricción de H_0.

El modelo proporcionado por la Ecuación (1.11) cumple que $\varepsilon_i \sim iidN(0,\, \sigma^2)$, $i = 1, 2, \ldots, n$, por el Teorema 1.13, los estimadores de máxima verosimilitud de los parámetros $\underline{\beta}_{\underline{Z}}$ y $\sigma^2_{\underline{Z}}$ son $\widehat{\underline{\beta}}_{\underline{Z}} = (X'_{\underline{Z}} X_{\underline{Z}})^{-1} X'_{\underline{Z}} \underline{Z}$ y $\sigma^2_{\underline{Z}} = \frac{1}{n} SCE_{\underline{Z}} = \frac{1}{n} \underline{Z}'(I - X_{\underline{Z}}(X'_{\underline{Z}} X_{\underline{Z}})^{-1} X'_{\underline{Z}})\underline{Z}$; así, el valor máximo de $L(\underline{\theta}; \underline{z})$ en $\mathbb{R}^p \times \mathbb{R}_+$ y por lo tanto de $L(\underline{\theta}; \underline{y})$ en $\overline{\underline{\Theta}}_0$ es:

$$L\left(\widehat{\underline{\beta}}_{\underline{Z}}, n^{-1} SCE_{\underline{Z}}; \underline{z}\right) = \left(\frac{n}{2\pi SCE_{\underline{Z}}}\right)^{n/2} e^{-n/2}.$$

Por el Teorema 1.13, se tiene que, bajo $\overline{\underline{\Theta}}$, los estimadores de máxima verosimilitud de los parámetros $\underline{\beta}$ y σ^2 son $\widehat{\underline{\beta}} = (X'X)^{-1} X' \underline{Y}$ y $\sigma^2 = n^{-1} SCE = \underline{Y}'(I - X(X'X)^{-1}X')\underline{Y}$; así, el valor máximo de $L(\underline{\theta}; \underline{y})$ en $\overline{\underline{\Theta}}$ es:

$$L\left(\widehat{\underline{\beta}}, n^{-1} SCE; \underline{y}\right) = \left(\frac{n}{2\pi SCE}\right)^{n/2} e^{-n/2}.$$

La razón de verosimilitudes, usando los resultados anteriores es:

$$\lambda(\underline{y}) = \frac{\text{máx}_{\underline{\theta} \in \overline{\underline{\Theta}}_0} \{L(\underline{\theta}; \underline{y})\}}{\text{máx}_{\underline{\theta} \in \overline{\underline{\Theta}}} \{L(\underline{\theta}; \underline{y})\}} = \frac{\left(\frac{n}{2\pi SCE_{\underline{Z}}}\right)^{n/2} e^{-n/2}}{\left(\frac{n}{2\pi SCE}\right)^{n/2} e^{-n/2}} = \left(\frac{SCE_{\underline{Z}}}{SCE}\right)^{-n/2}$$

la prueba que es la más potente de tamaño α ($\alpha \in [0, 1]$) es igual a 1 si y sólo si

$$0 \leqslant \left(\tfrac{SCE_{\underline{Z}}}{SCE}\right)^{-n/2} \leqslant k_1 \Leftrightarrow k_2 \leqslant \tfrac{SCE_{\underline{Z}}}{SCE}$$
$$\Leftrightarrow (n - (p+1))(k_2 - 1) \leqslant (n - (p+1)) \tfrac{SCE_{\underline{Z}} - SCE}{SCE}, \text{ si } n \geqslant (p+1),$$

esto es,

$$\phi(\underline{y}) = \begin{cases} 1, & \text{si } (n - (p+1)) \frac{SCE_{\underline{Z}} - SCE}{SCE} \geqslant k \\ \\ 0, & \text{si } (n - (p+1)) \frac{SCE_{\underline{Z}} - SCE}{SCE} < k, \end{cases}$$

donde k es tal que:

$$E_{\underline{\theta}_0}\{\phi(\underline{Y})\} = \int_{\gamma} \phi(\underline{y}) f_{\underline{Y}}(\underline{y}, \underline{\theta}_0) d\underline{y} = \alpha,$$

donde γ es el soporte del vector aleatorio $\underline{Y}$. Como

$$E_{\underline{\theta}_0}\{\phi(\underline{Y})\} = P(\phi(\underline{Y}) = 1 \mid \underline{\theta}_0) = P\left(k \geqslant (n - (p+1)) \frac{SCE_{\underline{Z}} - SCE}{SCE} | \, \underline{\theta}_0\right),$$

lo siguiente es encontrar la distribución de $(n-(p+1))\frac{SCE_{\underline{Z}}-SCE}{SCE_{\underline{Z}}}$ bajo H_0. Para hallar dicha distribución, nótese que:

$$\begin{aligned}
&SCE = \underline{Y}'(I - X(X'X)^{-1}X')\underline{Y} \\
&= (\underline{Z} + X\underline{a}'^{-}(\underline{a}'\underline{\beta})_0)'(I - X(X'X)^{-1}X')(\underline{Z} + X\underline{a}'^{-}(\underline{a}'\underline{\beta})_0) \\
&= (\underline{Z} + X\underline{a}'^{-}(\underline{a}'\underline{\beta})_0)'((\underline{Z} + X\underline{a}'^{-}(\underline{a}'\underline{\beta})_0) - X(X'X)^{-1}X'(\underline{Z} + X\underline{a}'^{-}(\underline{a}'\underline{\beta})_0)) \\
&= (\underline{Z} + X\underline{a}'^{-}(\underline{a}'\underline{\beta})_0)'(\underline{Z} - X(X'X)^{-1}X'\underline{Z}) \\
&= (\underline{Z}' + (\underline{a}'\underline{\beta})_0(\underline{a}'^{-})'X')(I - X(X'X)^{-1}X')\underline{Z} \\
&= ((\underline{Z}' + (\underline{a}'\underline{\beta})_0(\underline{a}'^{-})'X') - (\underline{Z}' + (\underline{a}'\underline{\beta})_0(\underline{a}'^{-})'X')X(X'X)^{-1}X')\underline{Z} \\
&= (\underline{Z}' - \underline{Z}'X(X'X)^{-1}X')\underline{Z} \\
&= \underline{Z}'(I - X(X'X)^{-1}X')\underline{Z},
\end{aligned}$$

luego,

$$\begin{aligned}
&(\sigma^2)^{-1}(SCE_{\underline{Z}} - SCE) \\
&= (\sigma^2)^{-1}(\underline{Z}'(I - X_{\underline{Z}}(X'_{\underline{Z}}X_{\underline{Z}})^{-1}X'_{\underline{Z}})\underline{Z} - \underline{Z}'(I - X(X'X)^{-1}X')\underline{Z}) \\
&= (\sigma^2)^{-1}\underline{Z}'(X(X'X)^{-1}X' - X_{\underline{Z}}(X'_{\underline{Z}}X_{\underline{Z}})^{-1}X'_{\underline{Z}})\underline{Z} \\
&= (\sigma^2)^{-1}\underline{Z}'(XX^{-} - X_{\underline{Z}}X_{\underline{Z}}^{-})\underline{Z} \\
&= \underline{Z}'A\underline{Z},
\end{aligned}$$

donde $A = (\sigma^2)^{-1}(XX^{-} - X_{\underline{Z}}X_{\underline{Z}}^{-})$.

Por el Teorema 1.19, se tiene que $\underline{Y} \sim N_n(X\underline{\beta}, \sigma^2 I_{n\times n})$; sin embargo, bajo la hipótesis nula de la Ecuación 1.10, se sigue que $E(\underline{Y}) = X\underline{a}'^{-}(\underline{a}'\underline{\beta})_0 + XG^{-}G\underline{\beta}$, luego se tiene que $\underline{Z} = \underline{Y} - X\underline{a}'^{-}(\underline{a}'\underline{\beta})_0 \sim N_n(XG^{-}G\underline{\beta}, \sigma^2 I_{n\times n})$.

La matriz $A\Sigma_{\underline{Z}} = (\sigma^2)^{-1}(XX^{-} - X_{\underline{Z}}X_{\underline{Z}}^{-})\sigma^2 I_{n\times n} = XX^{-} - X_{\underline{Z}}X_{\underline{Z}}^{-}$ es una matriz idempotente. Un caso más general se puede consultar en [7], págs. 185-189. Por el Teorema B.3, se sigue que:

$$\begin{aligned}
rang(A\Sigma_{\underline{Z}}) &= tr(A\Sigma_{\underline{Z}}) = tr(XX^{-} - X_{\underline{Z}}X_{\underline{Z}}^{-}) \\
&= tr(XX^{-}) - tr(X_{\underline{Z}}X_{\underline{Z}}^{-}) \\
&= tr(X(X'X)^{-1}X') - tr(X_{\underline{Z}}(X'_{\underline{Z}}X_{\underline{Z}})^{-1}X'_{\underline{Z}}) \\
&= tr((X'X)^{-1}X'X) - tr((X'_{\underline{Z}}X_{\underline{Z}})^{-1}X'_{\underline{Z}}X_{\underline{Z}}) \\
&= tr(I_{p+1\times p+1}) - tr(I_{p\times p}) = p + 1 - p = 1.
\end{aligned}$$

Así, bajo la hipótesis nula, usando el Teorema 1.21 se tiene que la forma cuadrática $(SCE_{\underline{Z}} - SCE)/\sigma^2 = \underline{Z}'A\underline{Z}$ está distribuida como $\chi^2_{(1;\lambda)}$, donde:

$$\begin{aligned}
\lambda &= 2^{-1}(XG^{-}G\underline{\beta})'((\sigma^2)^{-1}(XX^{-}-X_{\underline{Z}}X_{\underline{Z}}^{-}))(XG^{-}G\underline{\beta}) \\
&= (2\sigma^2)^{-1}(G^{-}G\underline{\beta})'X'(X(X'X)^{-1}X'-X_{\underline{Z}}X_{\underline{Z}}^{-})XG^{-}G\underline{\beta} \\
&= (2\sigma^2)^{-1}(G^{-}G\underline{\beta})'X'(I-X_{\underline{Z}}X_{\underline{Z}}^{-})XG^{-}G\underline{\beta} \\
&= (2\sigma^2)^{-1}(G^{-}G\underline{\beta})'X'(I-X_{\underline{Z}}X_{\underline{Z}}^{-})X_{\underline{Z}}G\underline{\beta} \\
&= (2\sigma^2)^{-1}(G^{-}G\underline{\beta})'X'(X_{\underline{Z}}-X_{\underline{Z}}X_{\underline{Z}}^{-}X_{\underline{Z}})G\underline{\beta} \\
&= (2\sigma^2)^{-1}(G^{-}G\underline{\beta})'X'(X_{\underline{Z}}-X_{\underline{Z}})G\underline{\beta} \\
&= 0.
\end{aligned}$$

Por lo tanto se sigue que $(SCE_{\underline{Z}}-SCE)/\sigma^2 \sim \chi^2_{(1)}$, bajo H_0. Además, como la Ecuación (1.11) es un modelo de regresión lineal múltiple con la restricción de H_0, por v) del Teorema 1.24 se tiene que $SCE_{\underline{Z}}/\sigma^2 \sim \chi^2_{(n-p)}$.

Dado que $\underline{Z}=\underline{Y}-X\underline{a}'^{-}(\underline{a}'\underline{\beta})_0 \sim N_n(X\underline{\beta}-X\underline{a}'^{-}(\underline{a}'\underline{\beta})_0, \Sigma_{\underline{Z}})$, donde $\Sigma_{\underline{Z}}=\sigma^2 I_{n\times n}$ tiene rango n, por el Teorema 1.23 se tiene que $(SCE_{\underline{Z}}-SCE)/\sigma^2=\underline{Z}'A\underline{Z}$ y $SCE/\sigma^2=\underline{Z}'B\underline{Z}$, donde $A=(\sigma^2)^{-1}(XX^{-}-X_{\underline{Z}}X_{\underline{Z}}^{-})$ y $B=(\sigma^2)^{-1}(I-XX^{-})$ son formas cuadráticas independientes, ya que:

$$\begin{aligned}
B\Sigma_{\underline{Z}}A &= (\sigma^2)^{-1}(I-XX^{-})(\sigma^2)^{-1}I(\sigma^2)^{-1}(XX^{-}-X_{\underline{Z}}X_{\underline{Z}}^{-}) \\
&= (\sigma^2)^{-3}(I-XX^{-})(XX^{-}-X_{\underline{Z}}X_{\underline{Z}}^{-}) \\
&= (\sigma^2)^{-3}(XX^{-}-XX^{-}XX^{-}-X_{\underline{Z}}X_{\underline{Z}}^{-}+XX^{-}X_{\underline{Z}}X_{\underline{Z}}^{-}) \\
&= (\sigma^2)^{-3}(XX^{-}-XX^{-}-X_{\underline{Z}}X_{\underline{Z}}^{-}+XX^{-}X_{\underline{Z}}X_{\underline{Z}}^{-}) \\
&= (\sigma^2)^{-3}(-X_{\underline{Z}}X_{\underline{Z}}^{-}+XX^{-}X_{\underline{Z}}X_{\underline{Z}}^{-}).
\end{aligned}$$

Además, $XX^{-}X_{\underline{Z}}X_{\underline{Z}}^{-}=X(X'X)^{-1}X'XG^{-}(XG^{-})^{-}=XG^{-}(XG^{-})^{-}=X_{\underline{Z}}X_{\underline{Z}}^{-}$, luego $B\Sigma_{\underline{Z}}A=(\sigma^2)^{-3}(-X_{\underline{Z}}X_{\underline{Z}}^{-}+X_{\underline{Z}}X_{\underline{Z}}^{-})=0$.

Por lo tanto, bajo H_0 se cumple que $(n-(p+1))\frac{SCE_{\underline{Z}}-SCE}{SCE} = \frac{\frac{(SCE_{\underline{Z}}-SCE)/\sigma^2}{1}}{\frac{SCE/\sigma^2}{n-(p+1)}} \sim F_{1,n-(p+1)}$.

Lo siguiente será mostrar la equivalencia de:

$$(\underline{a}'\widehat{\underline{\beta}}-(\underline{a}'\underline{\beta})_0)'(\underline{a}'(X'X)^{-1}\underline{a})^{-1}(\underline{a}'\widehat{\underline{\beta}}-(\underline{a}'\underline{\beta})_0) \tag{1.12}$$

con:

$$\underline{Z}'(I-X_{\underline{Z}}X_{\underline{Z}}^{-})\underline{Z}-\underline{Y}'(I-XX^{-})\underline{Y}=\underline{Z}'(XX^{-}-X_{\underline{Z}}X_{\underline{Z}}^{-})\underline{Z}, \tag{1.13}$$

para lo cual se debe notar que:

- $X^{-}X = I$;
- $(\underline{a}'X^{-})(X(\underline{a}')^{-}) = \underline{a}'(X^{-}X)(\underline{a}')^{-} = \underline{a}'(\underline{a}')^{-} = 1$;
- $(\underline{a}'X^{-}\underline{Y} - (\underline{a}'\underline{\beta})_0) = \underline{a}'X^{-}(\underline{Y} - X(\underline{a}')^{-}(\underline{a}'\underline{\beta})_0)$;
- $(\underline{a}'X^{-})_{1\times n}$ cumple que $1 = rang(\underline{a}') = rang(\underline{a}'X^{-}X) \leqslant rang(\underline{a}'X^{-}) \leqslant rang(\underline{a}') = 1$, por lo que el $rang(\underline{a}'X^{-}) = 1$, luego:

$$\begin{aligned}(\underline{a}'X^{-})^{-} &= (\underline{a}'X^{-})'(\underline{a}'X^{-}(\underline{a}'X^{-})')^{-1}\\ &= (\underline{a}'(X'X)^{-1}X')'(\underline{a}'(X'X)^{-1}X'(\underline{a}'(X'X)^{-1}X')')^{-1}\\ &= X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}X'X(X'X)^{-1}\underline{a})^{-1}\\ &= X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1};\end{aligned}$$

- $(\underline{a}'(X'X)^{-1}\underline{a})^{-1} = (\underline{a}'X^{-})^{'-}(\underline{a}'X^{-})^{-}$, ya que:

$$\begin{aligned}&(\underline{a}'X^{-})^{'-}(\underline{a}'X^{-})^{-} = (\underline{a}'X^{-})^{-'}(\underline{a}'X^{-})^{-}\\ &= (X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1})'X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\\ &= ((\underline{a}'(X'X)^{-1}\underline{a})^{-1})'\underline{a}'(X'X)^{-1}X'X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\\ &= ((\underline{a}'(X'X)^{-1}\underline{a})^{-1})'\underline{a}'(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\\ &= (\underline{a}'(X'X)^{-1}\underline{a})^{-1};\end{aligned}$$

- $\underline{a}'\widehat{\underline{\beta}} = \underline{a}'X^{-}\underline{Y}$;
- $\underline{a}'\widehat{\underline{\beta}} - (\underline{a}'\underline{\beta})_0 = \underline{a}'X^{-}(\underline{Y} - X(\underline{a}')^{-}(\underline{a}'\underline{\beta})_0)$;
- $\underline{Z} = \underline{Y} - X(\underline{a}')^{-}(\underline{a}'\underline{\beta})_0$, bajo H_0.

Sustituyendo en la Ecuación (1.12), se sigue que:

$$\begin{aligned}&(\underline{a}'\widehat{\underline{\beta}} - (\underline{a}'\underline{\beta})_0)'(\underline{a}'(X'X)^{-1}\underline{a})^{-1}(\underline{a}'\widehat{\underline{\beta}} - (\underline{a}'\underline{\beta})_0)\\ &= (\underline{a}'X^{-}(\underline{Y} - X(\underline{a}')^{-}(\underline{a}'\underline{\beta})_0))'(\underline{a}'X^{-})^{'-}(\underline{a}'X^{-})^{-}(\underline{a}'X^{-}(\underline{Y} - X(\underline{a}')^{-}(\underline{a}'\underline{\beta})_0))\\ &= (\underline{a}'X^{-}\underline{Z})'(\underline{a}'X^{-})^{'-}(\underline{a}'X^{-})^{-}(\underline{a}'X^{-}\underline{Z})\\ &= \underline{Z}'(\underline{a}'X^{-})'(\underline{a}'X^{-})^{-'}(\underline{a}'X^{-})^{-}\underline{a}'X^{-}\underline{Z}\\ &= \underline{Z}'(\underline{a}'X^{-})'(\underline{a}'X^{-})^{-'}(\underline{a}'X^{-})^{-}\underline{a}'X^{-}\underline{Z}\\ &= \underline{Z}'((\underline{a}'X^{-})^{-}\underline{a}'X^{-})'(\underline{a}'X^{-})^{-}\underline{a}'X^{-}\underline{Z};\end{aligned}$$

también se cumple:

$$(\underline{a}'X^-)^-\underline{a}'X^- = X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X',$$

por lo que:

$$\begin{aligned}
&((\underline{a}'X^-)^-\underline{a}'X^-)'(\underline{a}'X^-)^-\underline{a}'X^- \\
&= X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X'X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X' \\
&= X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X' \\
&= X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X' \\
&= (\underline{a}'X^-)^-\underline{a}'X^-.
\end{aligned}$$

Usando esto último se llega a la siguiente igualdad:

$$(\underline{a}'\underline{\widehat{\beta}} - (\underline{a}'\underline{\beta})_0)'(\underline{a}'(X'X)^{-1}\underline{a})^{-1}(\underline{a}'\underline{\widehat{\beta}} - (\underline{a}'\underline{\beta})_0) = \underline{Z}'(\underline{a}'X^-)^-\underline{a}'X^-\underline{Z},$$

lo cual implica que la Ecuación (1.12) es equivalente a la Ecuación (1.13) si $XX^- - X_{\underline{Z}}X_{\underline{Z}}^- = (\underline{a}'X^-)^-\underline{a}'X^-$ ó $XX^- = (\underline{a}'X^-)^-\underline{a}'X^- + X_{\underline{Z}}X_{\underline{Z}}^-$. Para ello considérese $N_{n\times n} = (\underline{a}'X^-)^-\underline{a}'X^- + X_{\underline{Z}}X_{\underline{Z}}^-$ y nótese que:

- $(XG^-)(XG^-)^- = X_{\underline{Z}}X_{\underline{Z}}^- = X_{\underline{Z}}(X_{\underline{Z}}'X_{\underline{Z}})^{-1}X_{\underline{Z}}'$ es una matriz simétrica e idempotente, y por el Teorema B.3 se sigue que:

$$\begin{aligned}
rang((XG^-)(XG^-)^-) &= tr((XG^-)(XG^-)^-) = tr(X_{\underline{Z}}(X_{\underline{Z}}'X_{\underline{Z}})^{-1}X_{\underline{Z}}') \\
&= tr((X_{\underline{Z}}'X_{\underline{Z}})^{-1}X_{\underline{Z}}'X_{\underline{Z}}) = tr(I_{p\times p}) = p;
\end{aligned}$$

- $(\underline{a}'X^-)^-\underline{a}'X^- = X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X'$ es una matriz simétrica e idempotente y por el Teorema B.3 se sigue

$$\begin{aligned}
&rang((\underline{a}'X^-)^-\underline{a}'X^-) = tr(X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X') \\
&= tr((\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X'X(X'X)^{-1}\underline{a}) \\
&= tr((\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}\underline{a}) = tr(1) = 1
\end{aligned}$$

- $$\begin{aligned}
&(\underline{a}'X^-)^-\underline{a}'X^-(XG^-)(XG^-)^- \\
&= X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X'(XG^-)(XG^-)^- \\
&= X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'G^-(XG^-)^- \\
&= X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underbrace{\underline{a}'G'}_{0}(GG')^{-1}(XG^-)^- = 0
\end{aligned}$$

$$
\begin{aligned}
&(XG^-)(XG^-)^-(\underline{a}'X^-)^-\underline{a}'X^- \\
&= (XG^-)(XG^-)^-X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X' \\
&= (XG^-)((XG^-)'XG^-)^{-1}(XG^-)'X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X' \\
&= (XG^-)((XG^-)'XG^-)^{-1}(G^-)'X'X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X' \\
&= (XG^-)((XG^-)'XG^-)^{-1}(G^-)'\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X' \\
&= (XG^-)((XG^-)'XG^-)^{-1}(\underbrace{\underline{a}'G^-}_{0})'(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X' = 0.
\end{aligned}
$$

Esto implica que $N = (\underline{a}'X^-)^-\underline{a}'X^- + X_{\underline{Z}}X_{\underline{Z}}^-$ es simétrica e idempotente de rango $rang(N) = p+1$, ya que:

$$
\begin{aligned}
&N^2 = ((\underline{a}'X^-)^-\underline{a}'X^- + X_{\underline{Z}}X_{\underline{Z}}^-)^2 \\
&= ((\underline{a}'X^-)^-\underline{a}'X^-)^2 + (\underline{a}'X^-)^-\underline{a}'X^-X_{\underline{Z}}X_{\underline{Z}}^- + X_{\underline{Z}}X_{\underline{Z}}^-(\underline{a}'X^-)^-\underline{a}'X^- + (X_{\underline{Z}}X_{\underline{Z}}^-)^2 \\
&= (\underline{a}'X^-)^-\underline{a}'X^- + X_{\underline{Z}}X_{\underline{Z}}^- = N
\end{aligned}
$$

y

$$rang(N) = tr((\underline{a}'X^-)^-\underline{a}'X^- + X_{\underline{Z}}X_{\underline{Z}}^-) = p+1.$$

Ahora defínase $A_{p+1\times n} = (X'X)^{-1}\underline{a}(\underline{a}'X^-)'^- + G^-(XG^-)^-$, esta matriz cumple lo siguiente:

- $$
\begin{aligned}
XA &= X((X'X)^{-1}\underline{a}(\underline{a}'X^-)'^- + G^-(XG^-)^-) \\
&= X(X'X)^{-1}\underline{a}(\underline{a}'X^-)^{-'} + XG^-(XG^-)^- \\
&= X(X'X)^{-1}\underline{a}(X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1})' + XG^-(XG^-)^- \\
&= X(X'X)^{-1}\underline{a}(\underline{a}'(X'X)^{-1}\underline{a})^{-1}\underline{a}'(X'X)^{-1}X' + XG^-(XG^-)^- \\
&= (\underline{a}'X^-)^-\underline{a}'X^- + XG^-(XG^-)^- = N
\end{aligned}
$$
- $XAX = X$. Un caso más general se puede consultar en [7], págs. 223-224.

Todo esto implica que:

$$
\begin{aligned}
XX^- &= (XX^-)' = (XAXX^-)' = ((XA)(XX^-))' = (XX^-)'(XA)' \\
&= (XX^-)N' = XX^-N = XX^-XA = XA = N,
\end{aligned}
$$

por lo tanto se logra probar la equivalencia entre las Ecuaciones (1.12) y (1.13)

$$(\underline{a}'\widehat{\underline{\beta}} - (\underline{a}'\underline{\beta})_0)'(\underline{a}'(X'X)^{-1}\underline{a})^{-1}(\underline{a}'\widehat{\underline{\beta}} - (\underline{a}'\underline{\beta})_0) = \underline{Z}'(I - X_{\underline{Z}}X_{\underline{Z}}^-)\underline{Z} - \underline{Y}'(I - XX^-)\underline{Y}, \quad (1.14)$$

Regresando a la prueba de hipótesis y usando este resultado, se tiene que bajo H_0:

$$\begin{aligned} \frac{\frac{(SCE_{\underline{Z}}-SCE)/\sigma^2}{1}}{\frac{SCE/\sigma^2}{n-(p+1)}} &= \frac{((\underline{a}'\widehat{\underline{\beta}}-(\underline{a}'\underline{\beta})_0)'(\underline{a}'(X'X)^{-1}\underline{a})^{-1}(\underline{a}'\widehat{\underline{\beta}}-(\underline{a}'\underline{\beta})_0))}{\frac{SCE}{n-(p+1)}} \\ &= \frac{(\underline{a}'\widehat{\underline{\beta}}-(\underline{a}'\underline{\beta})_0)^2}{S^2(\underline{a}'(X'X)^{-1}\underline{a})} \sim F_{1,n-(p+1)} \end{aligned} \tag{1.15}$$

Además se cumple que $F_{1,n-(p+1)} = t^2_{n-(p+1)}$. Por último, se hallan las regiones de rechazo para las hipótesis contrastadas, para lo cual se observa que:

$$\begin{aligned} \alpha &= P\left(k \geqslant (n-(p+1))\frac{SCE_{\underline{Z}}-SCE}{SCE_{\underline{Z}}} |\ \underline{\theta}_0\right) = P\left(k \geqslant \frac{(\underline{a}'\widehat{\underline{\beta}}-(\underline{a}'\underline{\beta})_0)^2}{S^2(\underline{a}'(X'X)^{-1}\underline{a})} |\ \underline{\theta}_0\right) \\ &= P(k \geqslant F_{1,n-(p+1)}) = P(k \geqslant t^2_{n-(p+1)}) = P(k^{1/2} \geqslant |t_{n-(p+1)}|) \\ &= P(k^{1/2} \geqslant |\frac{\underline{a}'\widehat{\underline{\beta}}-(\underline{a}'\underline{\beta})_0}{S[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}}|), \end{aligned}$$

luego las diferentes pruebas de hipótesis tienen las siguientes regiones de rechazo:

i) $H_0 : \underline{a}'\underline{\beta} = (\underline{a}'\underline{\beta})_0$ vs $H_1 : \underline{a}'\underline{\beta} > (\underline{a}'\underline{\beta})_0$
$t_{n-(p+1);\alpha} \leqslant \frac{\underline{a}'\widehat{\underline{\beta}}-(\underline{a}'\underline{\beta})_0}{S[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}}$

ii) $H_0 : \underline{a}'\underline{\beta} = (\underline{a}'\underline{\beta})_0$ vs $H_1 : \underline{a}'\underline{\beta} < (\underline{a}'\underline{\beta})_0$
$t_{n-(p+1);\alpha} \geqslant \frac{\underline{a}'\widehat{\underline{\beta}}-(\underline{a}'\underline{\beta})_0}{S[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}}$

iii) $H_0 : \underline{a}'\underline{\beta} = (\underline{a}'\underline{\beta})_0$ vs $H_1 : \underline{a}'\underline{\beta} \neq (\underline{a}'\underline{\beta})_0$
$t_{n-(p+1);\alpha/2} \leqslant |\frac{\underline{a}'\widehat{\underline{\beta}}-(\underline{a}'\underline{\beta})_0}{S[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}}|$

□

El correspondiente intervalo de confianza $100(1-\alpha)\,\%$ para $\underline{a}'\underline{\beta}$ se da en el siguiente teorema.

Teorema 1.33. *Un intervalo de confianza* $100(1-\alpha)\,\%$ *para* $\underline{a}'\underline{\beta}$ *es:*

$$\left(\underline{a}'\widehat{\underline{\beta}} - t_{\alpha/2}S\sqrt{\underline{a}'(X'X)^{-1}\underline{a}}, \underline{a}'\widehat{\underline{\beta}} + t_{\alpha/2}S\sqrt{\underline{a}'(X'X)^{-1}\underline{a}}\right).$$

Demostración. Para hallar el intervalo de confianza se usará el *método del pivote*. Como $\frac{\underline{a}'\widehat{\underline{\beta}}-\underline{a}'\underline{\beta}}{S[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}}$ es una función de las medidas muestrales y el parámetro $\underline{a}'\underline{\beta}$, donde $\underline{a}'\underline{\beta}$ es la única cantidad desconocida, y tiene una distribución de probabilidad que no depende de $\underline{a}'\underline{\beta}$, entonces se puede usar como cantidad pivote. Sean $a, b \in \mathbb{R}$ tales que:

$$P\left(a < \frac{\underline{a}'\widehat{\underline{\beta}} - \underline{a}'\underline{\beta}}{S[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}} < b\right) = 1-\alpha.$$

Entonces los valores que optimizan $P\left(a < \frac{\underline{a}'\widehat{\underline{\beta}}-\underline{a}'\underline{\beta}}{S[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}} < b\right)$ son $a = -t_{\alpha/2}$ y $b = t_{\alpha/2}$, por lo que:

$$\begin{aligned} 1-\alpha &= P\left(-t_{\alpha/2} < \frac{\underline{a}'\widehat{\underline{\beta}}-\underline{a}'\underline{\beta}}{S[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}} < t_{\alpha/2}\right) \\ &= P\left(-t_{\alpha/2}S[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2} < \underline{a}'\widehat{\underline{\beta}} - \underline{a}'\underline{\beta} < t_{\alpha/2}S[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}\right) \\ &= P\left(\underline{a}'\widehat{\underline{\beta}} - t_{\alpha/2}S[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2} < \underline{a}'\underline{\beta} < \underline{a}'\widehat{\underline{\beta}} + t_{\alpha/2}S[\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}\right) \end{aligned}$$

Por lo tanto los límites de confianza inferior y superior respectivamente son $\underline{a}'\widehat{\underline{\beta}} \pm t_{\alpha/2}S\sqrt{\underline{a}'(X'X)^{-1}\underline{a}}$. □

Aunque por lo general no se considera una β_i individual como una combinación lineal de β_0, β_1, ..., β_p, si se escoge:

$$a_j = \begin{cases} 1, \text{ si } j = i \\ 0, \text{ si } j \neq i, \end{cases}$$

entonces $\beta_i = \underline{a}'\underline{\beta}$ para esta selección de $\underline{a}$, y se pueden usar los Teoremas 1.32 y Teorema 1.33 para hacer inferencias acerca de una β_i individual [16].

Una aplicación útil de las técnicas de prueba de hipótesis e intervalo de confianza que se acaban de presentar consiste en usarlas en el problema de calcular el valor medio de Y para valores fijos de las variables independientes x_1, x_2, ..., x_p, en particular si $x_i = x_i^*$ para $i = 1, 2, \ldots, p$ entonces $E(Y) = \beta_0 + \beta_1 x_1^* + \ldots + \beta_p x_p^*$.

Obsérvese que $E(Y)$ es un caso especial de $a_0\beta_0 + a_1\beta_1 + \ldots + a_p\beta_p = \underline{a}'\underline{\beta}$ con $a_0 = 1$ y $a_i = x_i^*$, para $i = 1, 2, \ldots, p$. Entonces, la inferencia alrededor de $E(Y)$ cuando $x_i = x_i^*$, para $i = 1, 2, \ldots, p$ se puede hacer usando los Teoremas 1.32 y 1.33 desarrollados para combinaciones lineales generales de las β.

1.10. Predicción de un valor particular de Y mediante regresión múltiple

Supongamos que se ha ajustado el modelo de regresión lineal múltiple $Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon$ y que interesa predecir el valor de Y^* cuando $x_i = x_i^*$, para

$i = 1, 2, \ldots, p$. Se predice el valor de $Y^* = \beta_0 + \beta_1 x_1^* + \ldots + \beta_p x_p^* + \varepsilon = \underline{a}'\underline{\beta} + \varepsilon$ con $\widehat{Y}^* = \widehat{\beta}_0 + \widehat{\beta}_1 x_1^* + \ldots + \widehat{\beta}_p x_p^* = \underline{a}'\underline{\widehat{\beta}}$, donde $\underline{a} = (1, x_1^*, \ldots, x_p^*)'$. El estudio y análisis se centrará en la diferencia entre la variable Y^* y el valor predicho: $Y^* - \widehat{Y}^* = \underline{a}'(\underline{\beta} - \underline{\widehat{\beta}}) + \varepsilon$.

Teorema 1.34. *En el modelo de regresión lineal múltiple bajo el supuesto $\varepsilon_i \sim iidN(0, \sigma^2)$, $i = 1, \ldots, n$, la diferencia entre la variable Y^* y su valor predicho satisface que $Y^* - \widehat{Y}^* \sim N(0, \sigma^2[1 + \underline{a}'(X'X)^{-1}\underline{a}])$.*

Demostración. Como Y^* e $\widehat{Y}^*$ están distribuidas normalmente, el error está distribuido normalmente y se tiene que:

$$E(Y^* - \widehat{Y}^*) = E(\underline{a}'(\underline{\beta} - \underline{\widehat{\beta}}) + \epsilon) = \underline{a}' E(\underline{\beta} - \underline{\widehat{\beta}}) + E(\epsilon) = 0,$$

y también que:

$$\begin{aligned} Var(Y^* - \widehat{Y}^*) &= Var(\underline{a}'(\underline{\beta} - \underline{\widehat{\beta}}) + \epsilon) = Var(-\underline{a}'\underline{\widehat{\beta}} + \epsilon) \\ &= Var(-\underline{a}'(X'X)^{-1}X'\underline{Y} + \epsilon) \\ &= Var(-\underline{a}'(X'X)^{-1}X'(X\underline{\beta} + \underline{\epsilon}) + \epsilon) \\ &= Var(-\underline{a}'(X'X)^{-1}X'\underline{\epsilon} + \epsilon) \\ &= \underline{a}'(X'X)^{-1}X' Var(\underline{\epsilon}) X(X'X)^{-1}\underline{a} + \sigma^2 \\ &= \sigma^2(\underline{a}'(X'X)^{-1}\underline{a} + 1). \end{aligned}$$

□

De este teorema se sigue que $Z = \frac{Y^* - \widehat{Y}^*}{\sigma[1 + \underline{a}'(X'X)^{-1}\underline{a}]^{1/2}}$ tiene una distribución normal estándar. Además, si σ es sustituido por S, se puede demostrar el siguiente teorema.

Teorema 1.35. *El modelo de regresión lineal múltiple bajo el supuesto $\varepsilon_i \sim iidN(0, \sigma^2)$, $i = 1, \ldots, n$ cumple que $T = \frac{Y^* - \widehat{Y}^*}{S[1 + \underline{a}'(X'X)^{-1}\underline{a}]^{1/2}}$ tiene una distribución t de Student con $[n - (p + 1)]$ grados de libertad.*

Demostración.

$$\frac{Y^* - \widehat{Y}^*}{S[1 + \underline{a}'(X'X)^{-1}\underline{a}]^{1/2}} = \frac{\frac{Y^* - \widehat{Y}^*}{\sigma[1 + \underline{a}'(X'X)^{-1}\underline{a}]^{1/2}}}{\sqrt{\frac{1}{(n-(p+1))}\frac{(n-(p+1))S^2}{\sigma^2}}}.$$

Como $\frac{Y^*-\widehat{Y}^*}{\sigma[1+\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}} \sim N(0,1)$, $\frac{(n-(p+1))S^2}{\sigma^2} \sim \chi^2_{(n-(p+1))}$ y son independientes, puesto que Y^* es una v.a. independiente de la m.a. con la cual se calcula S^2. Por lo tanto $T \sim t_{(n-(p+1))}$. □

Se obtiene el siguiente intervalo de predicción $100(1-\alpha)\,\%$ para Y^*.

Teorema 1.36. *Un intervalo de predicción* $100(1-\alpha)\,\%$ *para* Y^* *cuando* $x_1 = x_1^*, x_2 = x_2^*, \ldots, x_p = x_p^*$ *tiene límites de confianza inferior y superior* $\underline{a}'\widehat{\underline{\beta}} \pm t_{\alpha/2} S\sqrt{1+\underline{a}'(X'X)^{-1}\underline{a}}$ *respectivamente, donde* $\underline{a} = (1, x_1^*, \ldots, x_p^*)'$.

Demostración. Como $T = \frac{Y^*-\widehat{Y}^*}{S[1+\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}}$ es una función de las medidas muestrales y el parámetro Y^*, donde Y^* es la única cantidad desconocida, y tiene una distribución de probabilidad que no depende de Y^*, entonces se puede usar como cantidad pivote. Sean $a, b \in \mathbb{R}$ tales que:

$$P\left(a < \frac{Y^*-\widehat{Y}^*}{S[1+\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}} < b\right) = 1-\alpha,$$

entonces $a = -t_{\alpha/2}$ y $b = t_{\alpha/2}$, por lo que:

$$\begin{aligned}
1-\alpha &= P\left(-t_{\alpha/2} < \tfrac{Y^*-\widehat{Y}^*}{S[1+\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}} < t_{\alpha/2}\right) \\
&= P\left(-t_{\alpha/2} S[1+\underline{a}'(X'X)^{-1}\underline{a}]^{1/2} < Y^* - \underline{a}'\widehat{\underline{\beta}} < t_{\alpha/2} S[1+\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}\right) \\
&= P\left(\underline{a}'\widehat{\underline{\beta}} - t_{\frac{\alpha}{2}} S[1+\underline{a}'(X'X)^{-1}\underline{a}]^{1/2} < Y^* < \underline{a}'\widehat{\underline{\beta}} + t_{\frac{\alpha}{2}} S[1+\underline{a}'(X'X)^{-1}\underline{a}]^{1/2}\right).
\end{aligned}$$

Por lo tanto, los límites de confianza inferior y superior respectivamente son $\underline{a}'\widehat{\underline{\beta}} \pm t_{\alpha/2} S\sqrt{1+\underline{a}'(X'X)^{-1}\underline{a}}$. □

1.11. Comprobación de supuestos en la regresión múltiple

En la regresión múltiple, como en la regresión lineal simple, es importante probar la validez de los supuestos para los errores en modelos lineales. Los diagnósticos para estos supuestos empleados en el caso de la regresión lineal simple también son útiles en la regresión múltiple. Estos diagnósticos son las gráficas de residuos frente a los

valores ajustados, las de probabilidad normal de residuos y las de residuos frente al orden en que se hacen las observaciones. También es una buena idea hacer gráficas de residuos comparadas con cada una de las variables independientes. Si las gráficas de los residuos indican incumplimiento de los supuestos, es posible intentar arreglar el problema al transformar las variables.

Si la suposición de normalidad se cumple, el histograma de los residuos tendría una forma razonable de campana y sería razonablemente simétrico con respecto a cero. Otra manera de comprobar la suposición de normalidad es construir una **gráfica normal** de los residuos. Para elaborar una gráfica normal, primero se acomodan los residuos en orden ascendente. Los residuos así ordenados se denotan como $e_{(1)}$, $e_{(2)}$, ..., $e_{(n)}$. En la gráfica normal tradicional se grafica $e_{(i)}$ en el eje horizontal contra un punto $z_{(i)}$ que va en el eje vertical. En este caso, $z_{(i)}$ se define como el punto en el eje horizontal bajo la curva normal estándar de modo que el área bajo la curva a la izquierda de $z_{(i)}$ es $(3i-1)/(3n+1)$. Por lo tanto $z_{(i)}$ es el punto normal que tiene un área de $(3i-1)(3n+1)$ bajo la curva normal estándar a la izquierda [3].

Una gráfica equivalente se obtiene graficando el porcentaje $p_{(i)}$ del área bajo la curva normal estándar a la izquierda de $z_{(i)}$ en el eje vertical. Es importante resaltar que la escala en el eje vertical no tiene la separación usual entre porcentajes. La separación refleja la distancia entre el puntaje z que corresponde a los porcentajes en la distribución normal estándar.

Se puede demostrar que si se cumple la suposición de normalidad, entonces el valor esperado del i-ésimo residuo ordenado $e_{(i)}$ es proporcional a $z_{(i)}$. Por lo tanto, una gráfica de los valores $e_{(i)}$ en la escala horizontal contra los valores $z_{(i)}$ en la escala vertical (lo que es lo mismo, los valores $e_{(i)}$ en la escala horizontal contra los valores $p_{(i)}$ en la escala vertical) debe tener el aspecto de una recta. Es decir, si la suposición de normalidad se mantiene, entonces la gráfica normal debe parecer una línea recta. Una gráfica normal que no tiene la apariencia de una recta (lo que es una decisión subjetiva) quiere decir que está transgrediendo la suposición de normalidad.

Es importante considerar que, con frecuencia, las transgresiones de las suposiciones de la varianza constante y la forma funcional correcta generan un histograma de los

residuos no normal y pueden ocasionar que la gráfica normal tenga una apariencia curva. Por esta razón, es una buena idea usar gráficas de residuos para verificar la varianza no constante y la forma funcional incorrecta antes de dar una conclusión final con respecto a la suposición de normalidad.

1.12. Colinealidad

Ajustar modelos por separado para cada variable no es lo mismo que ajustar el modelo multivariado.

Cuando un modelo no tiene un valor alto de coeficiente de determinación R^2, es porque que hay otros factores importantes que afectan la variable dependiente y que no se han incluido en el modelo.

Definición 1.37. *Cuando dos variables independientes están fuertemente correlacionadas, la regresión múltiple no puede determinar cuál es la importante. En este caso se dice que las variables son* ***colineales****. Cuando dos variables están correlacionadas, su diagrama de dispersión es casi una línea recta. También a veces se utiliza la palabra* ***multicolinealidad****. Cuando la colinealidad está presente, a veces se dice que el conjunto de variables independientes está* ***mal condicionado****.*

Nótese que existe multicolinealidad si la dimensión del espacio en el que se encuentran las variables predictoras es menor que el número de éstas. La existencia de la multicolinealidad afecta en gran medida las interpretaciones de cualquier modelo ajustado de regresión [8].

Cuando hay una fuerte relación lineal entre dos variables independientes, se puede sospechar que Y podría tener realmente relación con sólo una de estas variables, con lo demás en confusión.

En general, no hay mucho que se pueda hacer cuando las variables son colineales. La única manera de arreglar la situación es reunir más datos, incluyendo algunos valores para las variables independientes que no están en la misma línea recta. Entonces la regresión múltiple será capaz de determinar cuáles de las variables son realmente importantes.

1.13. Selección de modelos

Hay muchas situaciones en las que se han medido bastantes variables independientes, por lo que frecuentemente es necesario decidir cuáles de ellas son importantes y deben aparecer en el modelo que explica el comportamiento de la variable dependiente del fenómeno en estudio. Esto es, se requiere determinar si es posible eliminar variables sin que el modelo pierda eficiencia. Esto lleva al problema de **selección de modelos**, el cual es frecuentemente difícil. Se establecerán algunos principios básicos.

La buena selección de modelos se basa en un principio básico conocido como navaja de Occam. Este principio se enuncia de la siguiente manera:

La navaja de Occam: El mejor modelo científico es el modelo más simple que explica los hechos observados.

En términos de modelos lineales, la navaja de Occam implica el **principio de parsimonia:** Un modelo debe contener el menor número de variables necesario para ajustar los datos.

Existen algunas excepciones al principio de parsimonia:

1. Un modelo lineal siempre debe contener un intercepto, a menos que una teoría física indique otra cosa.

2. Si una potencia x^n de una variable se incluye en un modelo, también estarán incluidas todas las potencias inferiores $x, x^2, \cdots, x^{n-1}$, a menos que una teoría física indique lo contrario.

3. Si un producto $x_i x_j$ de dos variables está incluido en un modelo, entonces las variables x_i y x_j también deben estar incluidas por separado, a menos que una teoría física indique algo distinto.

Los modelos que sólo contienen las variables necesarias para ajustar los datos se llaman **parsimoniosos**.

1.13.1. Determinación de la eliminabilidad de variables de un modelo

Con frecuencia ocurre que se ha formado un modelo con muchas variables independientes y se desea determinar si un subconjunto en particular de ellas se puede eliminar del modelo sin reducir significativamente la precisión de éste. Para ser más específico, supongamos que se conoce que el modelo:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon$$

es correcto en lo que respecta a representar la relación verdadera entre las variables x e Y. Se llamará a éste el **modelo "completo"**. Se desea probar la hipótesis nula $H_0 : \beta_{k+1} = \cdots = \beta_p = 0$. Si H_0 no se rechaza, el modelo permanecerá correcto si se eliminan las variables $x_{k+1}, \cdots, x_p$, por lo que se puede reemplazar el modelo completo con el siguiente **modelo reducido**:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon.$$

Cada uno de estos modelos tiene una suma de los cuadrados del error (SCE), las cuales se denotarán por SCE_c y SCE_r para el modelo completo y reducido respectivamente.

Teorema 1.38. *Si bajo el supuesto sobre los errores, $\varepsilon_i \sim iidN(0, \sigma^2)$, $i = 1, \ldots, n$, el modelo completo $Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon$ es correcto, entonces el estadístico de prueba para la hipótesis $H_0 : \beta_{k+1} = \cdots = \beta_p = 0$ es $f = \frac{(SCE_r - SCE_c)/(p-k)}{SCE_c/(n-(p+1))}$. El estadístico f tiene distribución nula $F_{p-k,n-(p+1)}$.*

Demostración. En la regresión lineal múltiple se desconoce el parámetro $\underline{\theta} = (\beta_0, \beta_1, \ldots, \beta_{k-1}, \beta_k, \beta_{k+1}, \ldots, \beta_p, \sigma^2)'$, luego el espacio paramétrico es $\overline{\underline{\Theta}} = \mathbb{R}^{p+1} \times \mathbb{R}_+$. Considérese la hipótesis $H_0 : \underline{\theta} \in \overline{\underline{\Theta}}_0 = \{(\beta_0, \beta_1, \ldots, \beta_{k-1}, \beta_k, \beta_{k+1}, \ldots, \beta_p, \sigma^2)' \in \overline{\underline{\Theta}} | \beta_{k+1} = \ldots = \beta_p = 0\}$ en contraste con la hipótesis $H_1 : \underline{\theta} \in \overline{\underline{\Theta}}_1 = \overline{\underline{\Theta}} - \overline{\underline{\Theta}}_0$; H_0 y H_1 son hipótesis compuestas.

Sea $\underline{\theta}_0 = (\beta_0, \beta_1, \cdots, \beta_{k-1}, \beta_k, 0, \ldots, 0, \sigma^2)' \in \overline{\underline{\Theta}}_0$; usando el Lema 1.12 se tiene que:

$$\begin{aligned}
&l(\underline{\theta}_0; \underline{y}) \\
&= -nln(\sqrt{2\pi}) - nln((\sigma^2)^{1/2}) - \textstyle\sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}))^2 / 2\sigma^2 \\
&= -nln(\sqrt{2\pi}) - nln((\sigma^2)^{1/2}) - (\underline{Y} - X_r \underline{\beta_r})'(\underline{Y} - X_r \underline{\beta_r}) / 2\sigma^2 \\
&= -nln(\sqrt{2\pi}) - nln((\sigma^2)^{1/2}) - (\underline{Y}'\underline{Y} - 2\underline{\beta_r}' X_r' \underline{Y} + \underline{\beta_r}' X_r' X_r \underline{\beta_r}) / 2\sigma^2,
\end{aligned}$$

donde $X_r = (X_{(0)}, X_{(1)}, \ldots, X_{(k)})$ con $X_{(i)}$ igual a la i-ésima columna de X y $\underline{\beta_r} = (\beta_0, \beta_1, \ldots, \beta_k)'$ son las matrices correspondientes al modelo reducido. Las derivadas parciales de $l(\underline{\theta}_0; \underline{y})$ con respecto a $\underline{\beta_r}$ y σ^2 son:

$$\begin{aligned}
\tfrac{\partial l(\underline{\theta}_0; \underline{y})}{\partial \underline{\beta_r}} &= -(-X_r'\underline{Y} + X_r' X_r \underline{\beta_r}) / \sigma^2 \\
\tfrac{\partial l(\underline{\theta}_0; \underline{y})}{\partial \sigma^2} &= -\tfrac{n}{2\sigma^2} + \tfrac{(\underline{Y}'\underline{Y} - 2\underline{\beta_r}' X_r' \underline{Y} + \underline{\beta_r}' X_r' X_r \underline{\beta_r})}{2(\sigma^2)^2} = \tfrac{-n\sigma^2 + (\underline{Y} - X_r \underline{\beta_r})'(\underline{Y} - X_r \underline{\beta_r})}{2(\sigma^2)^2}.
\end{aligned}$$

Al igualar a cero cada una de las expresiones anteriores, se tiene:

$$\begin{aligned}
X_r'\underline{Y} - X_r' X_r \underline{\beta_r} = 0 \quad &\Leftrightarrow \quad X_r' X_r \underline{\beta_r} = X_r'\underline{Y} \\
-n\sigma^2 + (\underline{Y} - X_r \underline{\beta_r})'(\underline{Y} - X_r \underline{\beta_r}) = 0 \quad &\Leftrightarrow \quad \sigma^2 = n^{-1}(\underline{Y} - X_r \underline{\beta_r})'(\underline{Y} - X_r \underline{\beta_r}).
\end{aligned}$$

La primera ecuación matricial es igual a las ecuaciones normales que se enuncian en el Teorema 1.2, pero para el modelo reducido $Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon$, por lo tanto $\widehat{\underline{\beta_r}} = (X_r' X_r)^{-1} X_r' \underline{Y}$ y $\sigma^2 = n^{-1} SCE_r$.

La matriz que contiene las segundas derivadas parciales es:

$$\mathcal{H}(\underline{\beta_r}, \sigma^2) = \tfrac{1}{\sigma^2} \begin{bmatrix} -X_r' X_r & \tfrac{-X_r'\underline{Y} + X_r' X_r \underline{\beta_r}}{\sigma^2} \\ \tfrac{-X_r'\underline{Y} + X_r' X_r \underline{\beta_r}}{\sigma^2} & \tfrac{n}{2\sigma^2} - \tfrac{(\underline{Y} - X_r \underline{\beta_r})'(\underline{Y} - X_r \underline{\beta_r})}{(\sigma^2)^2} \end{bmatrix}$$

Al evaluar $\mathcal{H}(\underline{\beta_r}, \sigma^2)$ en $\underline{\beta_r} = \widehat{\underline{\beta_r}}$ y $\sigma^2 = n^{-1} SCE_r$ se tiene:

$$\mathcal{H}(\underline{\widehat{\beta}_r}, n^{-1}SCE_r)$$
$$= \frac{1}{n^{-1}SCE_r}\begin{bmatrix} -X_r'X_r & \frac{-X_r'\underline{Y}+X_r'X_r\underline{\widehat{\beta}_r}}{n^{-1}SCE_r} \\ \frac{-X_r'\underline{Y}+X_r'X_r\underline{\widehat{\beta}_r}}{n^{-1}SCE_r} & \frac{n}{2n^{-1}SCE_r} - \frac{(\underline{Y}-X_r\underline{\widehat{\beta}_r})'(\underline{Y}-X_r\underline{\widehat{\beta}_r})}{(n^{-1}SCE_r)^2} \end{bmatrix}$$

$$= \frac{n}{SCE_r}\begin{bmatrix} -X_r'X_r & \frac{-X_r'\underline{Y}+X_r'\underline{Y}}{n^{-1}SCE_r} \\ \frac{-X_r'\underline{Y}+X_r'\underline{Y}}{n^{-1}SCE_r} & \frac{n^2}{2SCE_r} - \frac{n^2SCE_r}{(SCE_r)^2} \end{bmatrix}$$

$$= \frac{n}{SCE_r}\begin{bmatrix} -X_r'X_r & \underline{0}' \\ \underline{0} & \frac{n^2}{2SCE_r} - \frac{2n^2}{2SCE_r} \end{bmatrix} = -\frac{n}{SCE_r}\begin{bmatrix} X_r'X_r & \underline{0}' \\ \underline{0} & \frac{n^2}{2SCE_r} \end{bmatrix}.$$

Se observa que $-\frac{SCE_r}{n}\mathcal{H}(\underline{\widehat{\beta}_r}, n^{-1}SCE_r) = -\frac{SCE_r}{n}\mathcal{H}(\underline{\widehat{\beta}_r}, n^{-1}SCE_r)'$, además que $X_r'X_r$ es definida positiva, pues $(X_r'X_r)' = X_r'X_r$ y dado $\underline{w} \in \mathbb{R}^{p+1} - \{\underline{0}\}$ se cumple que $\underline{w}'(X_r'X_r)\underline{w} = (X_r\underline{w})'(X_r\underline{w}) > 0$, y $\frac{n^2}{2SCE_r} > 0$, luego $-\frac{SCE_r}{n}\mathcal{H}(\underline{\widehat{\beta}_r}, n^{-1}SCE_r)$ es definida positiva. Por lo tanto $\mathcal{H}(\underline{\widehat{\beta}_r}, n^{-1}SCE_r)$ es definida negativa.

Como $\mathcal{H}(\underline{\widehat{\beta}_r}, n^{-1}SCE_r)$ es definida negativa, se tiene que $\underline{\beta} = (\underline{\widehat{\beta}_r}, 0, \ldots, 0)$ y $\sigma^2 = n^{-1}SCE_r$ maximizan $L(\underline{\theta}; \underline{y})$ en $\overline{\underline{\Theta}}_0$. Por lo que el valor máximo para la función de verosimilitud en $\overline{\underline{\Theta}}_0$ es:

$$\begin{aligned} \text{máx}_{\underline{\theta}\in\overline{\underline{\Theta}}_0}\{L(\underline{\theta}; \underline{y})\} &= L\left(\underline{\widehat{\beta}_r}, 0, \ldots, 0, \tfrac{1}{n}SCE_r; \underline{y}\right) \\ &= \left(\tfrac{n}{2\pi SCE_r}\right)^{n/2} e^{-nSCE_r/2SCE_r} = \left(\tfrac{n}{2\pi SCE_r}\right)^{n/2} e^{-n/2}. \end{aligned}$$

Por otro lado, por el Teorema 1.13, se tiene que bajo $\overline{\underline{\Theta}}$, los estimadores de máxima verosimilitud de los parámetros $\underline{\beta}$ y σ^2 son $\underline{\widehat{\beta}} = (X'X)^{-1}X'\underline{Y}$ y $\widehat{\sigma}^2 = n^{-1}\sum_{i=1}^n(y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \cdots + \widehat{\beta}_p x_{ip}))^2 = n^{-1}SCE_c$, correspondientes al modelo completo $Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \epsilon$. Por lo tanto la razón de verosimilitudes es:

$$\lambda(\underline{y}) = \frac{\text{máx}_{\underline{\theta}\in\overline{\underline{\Theta}}_0}\{L(\underline{\theta}; \underline{y})\}}{\text{máx}_{\underline{\theta}\in\overline{\underline{\Theta}}}\{L(\underline{\theta}; \underline{y})\}} = \frac{\left(\frac{n}{2\pi SCE_r}\right)^{n/2} e^{-n/2}}{\left(\frac{n}{2\pi SCE_c}\right)^{n/2} e^{-n/2}} = \left(\frac{SCE_r}{SCE_c}\right)^{-n/2}.$$

Luego la prueba más potente de tamaño α ($\alpha \in [0,1]$) es igual a 1 si y sólo si:

$$\begin{aligned}\left(\frac{SCE_r}{SCE_c}\right)^{-n/2} \leqslant k_1 \quad &\Leftrightarrow \frac{SCE_r}{SCE_c} \geqslant k_2 \\ &\Leftrightarrow \frac{n-(p+1)}{(p-K)}\left(\frac{SCE_r}{SCE_c} - \frac{SCE_c}{SCE_c}\right) \geqslant \frac{n-(p+1)}{(p-K)}(k_2-1) \\ &\Leftrightarrow \frac{(SCE_r - SCE_c)/(p-K)}{SCE_c/(n-(p+1))} \geqslant k\end{aligned}$$

Por lo tanto, la función de prueba es de la forma:

$$\phi(\underline{y}) = \begin{cases} 1, & \text{si } \frac{(SCE_r - SCE_c)/(p-K)}{SCE_c/(n-(p+1))} \geqslant k \\ \\ 0, & \text{si } \frac{(SCE_r - SCE_c)/(p-K)}{SCE_c/(n-(p+1))} < k, \end{cases}$$

donde k es tal que:

$$E_{\underline{\theta}_0}\{\phi(\underline{Y})\} = \int_{\gamma} \phi(\underline{y}) f_{\underline{Y}}(\underline{y}, \underline{\theta}_0) d\underline{y} = \alpha,$$

donde γ representa el soporte del vector aleatorio $\underline{Y}$. Como

$$E_{\underline{\theta}_0}\{\phi(\underline{Y})\} = P(\phi(\underline{Y}) = 1 \mid \underline{\theta}_0) = P\left(\frac{(SCE_r - SCE_c)/(p-K)}{SCE_c/(n-(p+1))} \geqslant k \mid \underline{\theta}_0\right),$$

lo siguiente es encontrar la distribución de $\frac{(SCE_r - SCE_c)/(p-K)}{SCE_c/(n-(p+1))}$ bajo H_0.

Del Teorema 1.9 se sigue que $SCE_r = \underline{Y}'(I - X_r(X_r'X_r)^{-1}X_r')\underline{Y}$ y $SCE_c = \underline{Y}'(I - X(X'X)^{-1}X')\underline{Y}$, por lo tanto $SCE_r - SCE_c = \underline{Y}'A\underline{Y}$, donde $A = X(X'X)^{-1}X' - X_r(X_r'X_r)^{-1}X_r'$. Además, se tiene que:

$$\begin{aligned} A\Sigma_{\underline{Y}} &= (\sigma^2)^{-1}(X(X'X)^{-1}X' - X_r(X_r'X_r)^{-1}X_r')\sigma^2 I_{n\times n} \\ &= X(X'X)^{-1}X' - X_r(X_r'X_r)^{-1}X_r' \end{aligned}$$

es una matriz idempotente. Un caso más general se puede consultar en [7], págs. 185-191, que en la notación de la referencia se tiene $X^- = (X'X)^{-1}X'$, $B = X_r = XG^-$, $B^- = (X_r'X_r)^{-1}X_r'$, $G^- = G'(GG')^{-1}$,

$$H_{p-k\times p+1} = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

y

$$G_{k+1\times p+1} = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \ddots & \vdots & \\ 0 & 0 & \cdots & 1 & \cdots & 0 \end{bmatrix}.$$

Por el Teorema B.3 se sigue que:

$$\begin{aligned} rang(A\Sigma_{\underline{Y}}) &= tr(A\Sigma_{\underline{Y}}) = tr(X(X'X)^{-1}X' - X_r(X_r'X_r)^{-1}X_r') \\ &= tr(X(X'X)^{-1}X') - tr(X_r(X_r'X_r)^{-1}X_r') \\ &= tr((X'X)^{-1}X'X) - tr((X_r'X_r)^{-1}X_r'X_r) \\ &= tr(I_{p+1\times p+1}) - tr(I_{k+1\times k+1}) \\ &= p+1-(k+1) = p-k. \end{aligned}$$

Bajo H_0, por el Teorema 1.19 se tiene que $\underline{Y} \sim N_n(X\underline{\beta_0}, \sigma^2 I_{n\times n})$, donde $\underline{\beta_0} = (\beta_0, \beta_1, \ldots, \beta_k, 0, \ldots, 0)'$. Haciendo operaciones algebraicas se obtiene que $X\underline{\beta_0} = X_r\underline{\beta_r}$. Usando el Teorema 1.21 se tiene que la forma cuadrática $(SCE_r - SCE_c)/\sigma^2 = \underline{Y}'A\underline{Y}$ está distribuida como $\chi^2_{(p-k,\lambda)}$, donde:

$$\begin{aligned} \lambda &= 2^{-1}(X\underline{\beta_0})'(\sigma^2)^{-1}(X(X'X)^{-1}X' - X_r(X_r'X_r)^{-1}X_r')X\underline{\beta_0} \\ &= (2\sigma^2)^{-1}(X\underline{\beta_0})'X(X'X)^{-1}X'X\underline{\beta_0} - 2^{-1}(X\underline{\beta_0})'X_r(X_r'X_r)^{-1}X_r'X_r\underline{\beta_r} \\ &= (2\sigma^2)^{-1}(X\underline{\beta_0})'X\underline{\beta_0} - 2^{-1}(X\underline{\beta_0})'X_r\underline{\beta_r} \\ &= (2\sigma^2)(X\underline{\beta_0})'X\underline{\beta_0} - 2^{-1}(X\underline{\beta_0})'X\underline{\beta_0} \\ &= 0. \end{aligned}$$

Por lo tanto, bajo H_0, $(SCE_r - SCE_c)/\sigma^2 \sim \chi^2_{(p-k)}$. Como el modelo completo es correcto, se cumple que $SCE_c \sim \chi^2_{(n-(p+1))}$. Por lo tanto, bajo H_0

$$\frac{(SCE_r - SCE_c)/(p-K)}{SCE_c/(n-(p+1))} = \frac{\frac{(SCE_r - SCE_c)/\sigma^2}{p-k}}{\frac{SCE_c/\sigma^2}{n-(p+1)}} \sim F_{p-k,n-(p+1)}$$

□

El método recién descrito es muy útil en la práctica para desarrollar modelos parsimoniosos, al eliminar las variables no necesarias. Sin embargo, las condiciones en las

que ésto es formalmente válido rara vez se encuentran en la práctica. Primero, es raro el caso en el cual el modelo completo es correcto; habrá cantidades no aleatorias que afectan el valor de la variable dependiente y que no se consideran para las variables independientes. Segundo, para que el método sea formalmente válido, el subconjunto de variables que se eliminará debe determinarse independientemente de los datos. Éste por lo general no es el caso. Más a menudo, cuando un modelo grande se ajusta, algunas de las variables se ve que tienen p-valores grandes, y se utiliza la prueba F para determinar si se deben eliminar del modelo. Como se ha dicho, ésta es una técnica útil en la práctica, pero, de la misma manera que la mayoría de los métodos de la selección de modelos, debe verse como una herramienta informal, no como un procedimiento riguroso basado en la teoría.

Si se encuentra que 0.05 $< p <$ 0.10, de acuerdo con la regla general del 5 %, si $p >$ 0.05, el modelo reducido es creíble, pero sólo apenas es cierto. Más que establecer un modelo apenas creíble, es inteligente ir más lejos para buscar un modelo ligeramente menos reducido que contenga un P-valor mayor.

Si se encuentra nuevamente que 0.05 $< p <$ 0.10, el modelo reducido apenas es creíble como el mejor.

Se ha descrito un método informal para encontrar un buen modelo parsimonioso. Es importante darse cuenta que este procedimiento informal podría haberse realizado de manera diferente, con elecciones diferentes de las variables eliminadas e incluidas en el modelo. Se podría haber tenido un modelo final diferente que pudiera haber sido tan bueno como el que se encontró. Con frecuencia, en la práctica hay muchos modelos que ajustan los datos casi igualmente bien; no hay un único modelo “correcto”.

1.13.2. Regresión con los mejores subconjuntos

Con frecuencia los métodos de selección de modelos son informales, pero adecuados. Sin embargo, hay algunas herramientas que pueden tornar el proceso más sistemático. Una de ellas es la **regresión con los mejores subconjuntos**. El concepto es muy simple. Supongamos que hay p variables independientes $x_1, x_2, \cdots, x_p$, que están dispo-

nibles para incluirse en el modelo. Supongamos que se desea encontrar un buen modelo que contenga exactamente cuatro variables independientes. Se puede ajustar cada modelo potencial que contenga cuatro de las variables, y ordenarlos según su bondad de ajuste, midiendo el coeficiente de determinación R^2. El subconjunto de cuatro variables que produce el valor mayor de R^2 es el "mejor" subconjunto de tamaño 4. Se puede repetir el proceso para otros tamaños de subconjuntos, para hallar los mejores subconjuntos de tamaño $1, 2, \cdots, p$. Después se pueden examinar estos mejores subconjuntos para ver cuál proporciona un buen ajuste en tanto continúe siendo parsimonioso.

El valor de R^2 para el mejor subconjunto aumenta cuando el número de variables lo hace. Es un hecho matemático que el mejor subconjunto de $k+1$ variables de k siempre tendrá por lo menos un R^2 tan grande como el mejor subconjunto de variables k

También es importante darse cuenta de que el valor R^2 es una cantidad aleatoria, que depende de los datos. Si el proceso se repitiera y se obtuvieran nuevos datos, los valores de R^2 para los diferentes modelos serían algo diferentes, y posiblemente habría modelos diferentes que serían "mejores". Por esta razón no se debe usar este procedimiento, o cualquier otro, para elegir sólo un modelo. En lugar de ello se debe considerar que habrá muchos modelos que ajusten los datos casi igualmente bien.

Sin embargo, existen métodos que se han desarrollado para elegir un sólo modelo, presumiblemente el "mejor de los mejores". Aquí se describen dos de ellos. Si sólo se elige el modelo con el valor más alto de R^2, siempre se eligirá el que contiene todas las variables, ya que el valor de R^2 aumenta necesariamente cuando el número de variables en el modelo se incrementa. Los métodos de selección de un modelo implican estadísticos que ajustan el valor de R^2 para eliminar esta característica.

El primero de estos estadísticos considerados para efectuar una selección es el denominado R^2 **ajustado**, cuya definición se da a continuación.

Definición 1.39. *Sea n el número de observaciones y p el de variables independientes en el modelo. El R^2 ajustado se define como:*

$$R^2 ajustado = R^2 - \left(\frac{p}{n-p-1}\right)(1-R^2). \tag{1.16}$$

El estadístico R^2 ajustado es siempre menor que R^2, ya que a este último se le resta una cantidad positiva, como se indica en (1.16). Conforme el número de variables k aumenta, R^2 aumentará, pero la cantidad restada de éste también lo hará. El valor de k para el cual el valor de R^2 ajustado es un máximo se puede usar para determinar el número de variables en el modelo, y el mejor subconjunto de ese tamaño que se puede elegir como modelo.

Otro estadístico comúnmente usado es C_p **de Mallows**

Definición 1.40. *Sea n el número de observaciones, p el número total de variables independientes en consideración y k el de variables independientes en un subconjunto. Como antes, sea SCE_c la suma de los cuadrados del error para el modelo completo que contiene todas las variables p, y sea SCE_r la suma de los cuadrados del error para el modelo que contiene solamente el subconjunto de variables k. El C_p de Mallows se define como:*

$$C_p = \frac{(n-p-1)SCE_r}{SCE_c} - (n-2k-2).$$

Para modelos que contienen tantas variables independientes como las realmente necesarias, se supone que el valor de C_p es casi igual al número de variables, incluyendo el intercepto, en el modelo. Para elegir sólo un modelo, se puede elegir éste ya sea con el valor mínimo de C_p, o se puede escoger el modelo en el que el valor de C_p está más cerca del número de variables independientes en el modelo [12].

En la práctica no hay razón clara para preferir el modelo elegido con R^2 ajustado o el elegido con el C_p de Mallows.

1.13.3. Regresión paso a paso (*stepwise*)

La **regresión paso a paso** es quizás la técnica de selección de modelos más ampliamente usada. Su ventaja principal sobre la regresión con los mejores subconjuntos es que es menos intensa computacionalmente, por lo que se puede utilizar en situaciones donde hay un número muy grande de variables independientes candidatas y demasiados subconjuntos posibles para revisar cada uno. La versión de la regresión paso a paso

que se describirá está basada en los p-valores de los estadísticos t para las variables independientes. Una versión equivalente tiene como sustento el estadístico F (que es el cuadrado del estadístico t). Antes de operar el algoritmo, el usuario elige dos p-valores de umbral, α_{dentro} y α_{fuera} con $\alpha_{dentro} \leqslant \alpha_{fuera}$. Los pasos para la regresión paso a paso son:

1. **Selección hacia adelante:** Se selecciona la variable independiente con el p-valor más pequeño, suponiendo que se satisface $p < \alpha_{dentro}$. Esta variable se introduce en el modelo, creando un modelo de una sola variable independiente. Esta variable se denotará por x_1.

2. También en un paso de selección hacia adelante, se revisan una a una las variables restantes como candidatas para la segunda variable en el modelo. La que tenga el p-valor más pequeño se agrega al modelo, suponiendo nuevamente que $p < \alpha_{dentro}$.

3. **Eliminación hacia atrás:** Es posible que al haber agregado la segunda variable al modelo se provoque un aumento en el p-valor de la primera variable. La primera variable se elimina del modelo si su p-valor es mayor que α_{fuera}.

4. El algoritmo continúa alternando los pasos de selección hacia adelante con los de eliminación hacia atrás: en cada paso de selección hacia adelante se agrega la variable con el p-valor más pequeño si $p < \alpha_{dentro}$, y en cada paso de eliminación hacia atrás se elimina la variable con el p-valor más grande si $p > \alpha_{fuera}$. El algoritmo se termina cuando ninguna variable satisface los criterios para ser agregada o eliminada del modelo.

Una debilidad de todos los procedimientos automáticos de selección de variables, incluyendo la regresión paso a paso y la regresión con los mejores subconjuntos, es que operan sólo con base en la bondad del ajuste, y pueden no considerar las relaciones entre variables independientes, que son importantes.

1.13.4. Problemas en la selección de modelos

Cuando se construye un modelo para pronosticar el valor de una variable dependiente, parecería razonable tratar de empezar con tantas variables candidatas independientes como sea posible, para que el procedimiento de selección de modelos tenga un número muy grande de modelos donde elegir. Desgraciadamente, lo anterior no es una buena idea, como se mostrará a continuación.

Un coeficiente de correlación se puede calcular entre cualesquiera dos variables. A veces, dos variables sin ninguna relación real estarán correlacionadas fuertemente sólo por azar. Por ejemplo, George Undy Yule observó que la tasa de natalidad anual en Gran Bretaña estaba casi perfectamente correlacionada ($r =$-0.98) con la producción anual de hierro en lingotes en Estados Unidos durante 1875-1920, aunque nadie sugeriría tratar de pronosticar alguna de estas variables en términos de la otra. Ello ilustra una dificultad que comparten todos los procedimientos de selección de modelos. Lo más probable es que algunas de las variables independientes que se proporcionan sean mejores candidatas solamente por azar y presenten correlaciones sin sentido con la variable dependiente [12].

¿Como se puede determinar qué variables, si las hay, en el modelo seleccionado están muy relacionadas con la variable dependiente, y cuáles se seleccionaron solamente por azar? Los métodos estadísticos no son de mucha ayuda aquí. El método más confiable es repetir el experimento, reuniendo más datos de la variable dependiente y de las variables independientes que fueron seleccionadas por el modelo. Entonces las variables independientes sugeridas por el procedimiento de selección se pueden ajustar a la variable dependiente utilizando los nuevos datos. Si alguna de estas variables se ajusta bien a los nuevos datos, la evidencia de una relación real será más convincente.

Capítulo 2

Análisis de componentes principales (ACP)

En la práctica, lo más habitual es tomar el mayor número posible de variables, puesto que los estudios suelen requerirlo. Esto conlleva inconvenientes: es difícil visualizar relaciones entre las variables y pueden presentarse fuertes correlaciones entre ellas, ya que si se toman demasiadas variables (cosa que en general sucede cuando no se sabe demasiado sobre los datos), lo normal es que estén relacionadas o que midan lo mismo bajo distintos puntos de vista [14].

Los métodos multivariados son extraordinariamente útiles para grandes conjuntos de datos complicados y complejos de una gran cantidad de variables medidas en números grandes de unidades experimentales. La importancia y utilidad de los métodos multivariados aumentan al incrementarse el número de variables que se están midiendo y el número de unidades experimentales que se están evaluando [8].

A menudo, el objetivo primario de los análisis multivariados es resumir grandes cantidades de datos por medio de relativamente pocos parámetros. El tema subyacente de muchas técnicas multivariadas es la simplificación.

Muchas técnicas multivariadas tienden a ser de naturaleza exploratoria en lugar de confirmatoria, es decir, tienden a propiciar la generación de hipótesis más que a comprobarlas. Los métodos estadísticos tradicionales suelen exigir que un investigador establezca algunas hipótesis, reúna algunos datos y, a continuación, use estos datos para

rechazar o no rechazar esas hipótesis. Una situación alternativa que se presenta frecuentemente es el caso en el cual un investigador dispone de una gran cantidad de datos y se pregunta si pudiera haber una información valiosa en ellos. Las técnicas multivariadas suelen ser útiles para examinar los datos en un intento por saber si hay información que sea valiosa en esos datos [8].

Estos análisis multivariados presentan dos técnicas,

1. Las *técnicas dirigidas por las unidades experimentales*, que se interesan principalmente en las relaciones que podrían existir entre las unidades experimentales que se están midiendo. Algunos ejemplos de este tipo de técnica se encuentran en el análisis por agrupación y el análisis mutivariado de la varianza (MANOVA: *multivariate analysis of variance*).

2. Las *técnicas dirigidas por las variables*, que se enfocan principalmente en las relaciones que podrían existir entre las variables respuesta que se están midiendo.

   Algunos ejemplos de este tipo de técnica se encuentran en el análisis de regresión y el análisis de componentes principales (ACP) [8].

El análisis de componentes principales fue desarrollado por Hotelling (1933), a partir de los planteamientos de Karl Pearson (1901) [11].

2.1. El análisis de componentes principales: campos de aplicación

El análisis de componentes principales (ACP) es una técnica multivariada que crea nuevas variables no correlacionadas llamadas *componentes principales* a partir de un conjunto de variables correlacionadas.

Hay muchas razones para considerar el análisis de componentes principales, algunas de ellas se describen a continuación:

2.1.1. Cribado de datos

El análisis de componentes principales es quizá el más útil para cribar datos multivariados. En casi todas las situaciones de análisis de datos, se puede recomendar el ACP como un primer paso. Se debe realizar sobre un conjunto de datos, antes de realizar cualquier clase de análisis multivariados. Los análisis de seguimiento sobre las componentes principales son útiles para comprobar las hipótesis que el investigador podría establecer acerca de un conjunto de datos multivariados y para identificar y localizar posibles datos con valores atípicos (*outliers*) en un conjunto.

Definición 2.1. *Las unidades experimentales cuyas variables medidas parecen incoherentes con las mediciones realizadas en las otras unidades experimentales suelen llamarse* ***datos con valores atípicos****.*

Las calificaciones de las componentes principales se pueden usar como entrada para programas de trazado de gráficas y situación de datos y, con frecuencia, un examen de las presentaciones gráficas resultantes revelarán las anormalidades, como los datos con valores atípicos en los datos que se está planeando analizar.

2.1.2. Validar hipótesis

A menudo, se requiere que se cumplan algunas hipótesis para que sean válidas ciertas clases de análisis estadísticos. Se pueden analizar por separado las mediciones de las componentes principales para ver si se cumplen las hipótesis relativas a la distribución, como la normalidad de las variables, con una gráfica de probabilidad normal, y la independencia de las unidades experimentales.

Si la población de la que se está tomando la muestra es normal multivariada, entonces cada una de las componentes principales tendrá una distribución normal univariada y una gráfica de dispersión por pares de las calificaciones de las dos primeras componentes principales debe dar por resultado que los datos caigan en una región de forma elíptica, con los ejes mayor y menor paralelos al primer y segundo componente principal, respectivamente.

2.1.3. Agrupación

El análisis de componentes principales también es útil cuando el investigador desee agrupar las unidades experimentales en subgrupos de tipos semejantes. Se puede usar para ayudar a formar agrupamientos de las unidades experimentales en subgrupos o para verificar los resultados de los programas de agrupación. En este caso, se usarían las calificaciones de las componentes principales como entrada para los programas de agrupación, lo que suele incrementar la eficacia de estos programas y reducir el costo de su uso. Además, pueden y siempre deben usarse las mediciones de las componentes principales para ayudar a validar los resultados de los programas de agrupación.

2.1.4. Regresión

La regresión múltiple puede ser peligrosa cuando las variables predictoras están intensamente correlacionadas de alguna manera. Esto se conoce como *multicolinealidad* entre las variables predictoras. El análisis de componentes principales ayudará a determinar si ocurre multicolinealidad entre las variables predictoras.

2.2. Objetivos del análisis de componentes principales

El ACP debe usarse principalmente como una técnica exploratoria, ayuda a los investigadores a que adquieran cierta percepción respecto a un conjunto de datos. A veces, un ACP ayuda a los investigadores a comprender mejor la estructura de correlación entre las respuestas y, en ocasiones, a generar hipótesis acerca de las variables o de los datos. Los objetivos principales de un ACP son:

2.2.1. Identificar nuevas variables no correlacionadas

Es importante resaltar que el concepto de mayor información se relaciona con el de mayor variabilidad o varianza. Cuanto mayor sea la variabilidad de los datos, se considera que existe mayor cantidad información [14].

A continuación se muestra la definición de combinación lineal estandarizada.

Definición 2.2. *Una combinación lineal* $\underline{l}'\underline{x}$ *es una combinación lineal estandarizada (CLE) si* $\sum l_i^2 = 1$.

El análisis de componentes principales busca las CLE no correlacionadas que tengan varianza máxima. El ACP siempre identificará nuevas variables; sin embargo, no garantiza que las nuevas variables sean significativas. Desafortunadamente, lo más frecuente es que no sean significativas; no obstante, las variables componentes principales serán útiles y el análisis también.

2.2.2. Descubrir la verdadera dimensionalidad de los datos

El análisis de componentes principales sirve para determinar la dimensionalidad real de los datos y, cuando esa dimensionalidad es menor que p, el número de variables originales se pueden reemplazar por un número menor de variables subyacentes sin que se pierda información. Esta menor cantidad de variables se usará en los siguientes análisis.

En general, el análisis de componentes principales busca unas cuantas combinaciones lineales con el fin de resumir los datos, perdiendo en el proceso la menor información posible. Este intento de reducir la dimensionalidad se describiría como “el resumen parsimonioso” de los datos [11]. Cuando las variables originales no están correlacionadas, no tiene sentido realizar un análisis de componentes principales [14].

2.3. Componentes principales poblacionales

El análisis de componentes principales pretende explicar la estructura de covarianza de un vector aleatorio buscando un nuevo sistema de ejes coordenados que indiquen las direcciones de mayor variabilidad, ya sea en una situación teórica con matriz de covarianza, o con una matriz de covarianza estimada a partir de una muestra [14]. Primero se estudiará el modelo teórico. Sea $\underline{X} = (X_1, \ldots, X_p)'$ un vector aleatorio con matriz de covarianza Σ.

2.3.1. Construcción sucesiva de las componentes principales

En la construcción sucesiva de las componentes principales, se obtendrá la primera componente principal, es decir, la CLE de $\underline{X}$ que tenga varianza máxima. Posteriormente se construirá la segunda componente principal que es la CLE de $\underline{X}$ que tiene varianza máxima restante y es no correlacionada con la primera. Por último se construirá la (j+1)-ésima componente principal, es decir, la CLE de $\underline{X}$ que tiene varianza máxima restante y es no correlacionada con las j primeras componentes principales.

Construcción de la primera componente principal

La primera componente principal se denotará por Y_1 y, como ya se mencionó, es la CLE de las variables originales de varianza máxima, es decir, $Y_1 = \underline{l_1}'\underline{X} = l_{11}X_1 + \ldots + l_{1p}X_p$ tiene varianza máxima, donde $\underline{l_1} = (l_{11}, \ldots, l_{1p})'$ cumple que $\underline{l_1}'\underline{l_1} = 1$. Se tiene que la varianza de Y_1 es:

$$Var(Y_1) = Var(\underline{l_1}'\underline{X}) = \underline{l_1}'Var(\underline{X})\underline{l_1} = \underline{l_1}'\Sigma\underline{l_1}.$$

Por lo tanto se resolverá el siguiente problema:

$$\text{máx}\{Var(Y_1)\} = \text{máx}_{R_1}\{\underline{l_1}'\Sigma\underline{l_1}\}, \text{ donde } R_1 = \{\underline{l_1} \in \mathbb{R}^p | \underline{l_1}'\underline{l_1} - 1 = 0\}.$$

El método habitual para maximizar una función de varias variables sujeta a restricciones es el método de los multiplicadores de Lagrange. En este caso el objetivo es maximizar una función sujeta a una restricción de igualdad, por lo que se construye la función lagrangiana de la siguiente forma:

$$\mathcal{L}_1(\underline{l_1}, \lambda) = \underline{l_1}'\Sigma\underline{l_1} - \lambda(\underline{l_1}'\underline{l_1} - 1);$$

para calcular los puntos críticos de $\mathcal{L}_1$, se calculan sus derivadas parciales con respecto a $\underline{l_1}$ y a λ, y se igualan a cero.

$$\frac{\partial \mathcal{L}_1(\underline{l_1}, \lambda)}{\partial \underline{l_1}} = 2\Sigma\underline{l_1} - 2\lambda I_{p\times p}\underline{l_1} = \underline{0} \quad \Leftrightarrow \quad (\Sigma - \lambda I_{p\times p})\underline{l_1} = \underline{0} \tag{2.1}$$

$$\frac{\partial \mathcal{L}_1(\underline{l_1}, \lambda)}{\partial \lambda} = -(\underline{l_1}'\underline{l_1} - 1) = 0 \quad \Leftrightarrow \quad \underline{l_1}'\underline{l_1} = 1 \tag{2.2}$$

El sistema homogéneo (2.1) siempre tiene solución, sin embargo, para que tenga solución distinta de $\underline{0}$, por el teorema de Roché-Frobenius (Teorema B.4), se debe cumplir que $rango(\Sigma - \lambda I_{p\times p}) < p$, lo cual es equivalente a que $Det(\Sigma - \lambda I_{p\times p}) = 0$ (Teorema B.5). Por lo que λ es un valor propio de Σ; además, de (2.1) se sigue que $\Sigma\underline{l_1} = \lambda\underline{l_1}$, lo que indica que $\underline{l_1}$ es un vector propio de norma 1 (por (2.2)) correspondiente a λ. Así, $(\underline{l_1}, \lambda)$, donde $\underline{l_1}$ es el vector propio unitario correspondiente al valor propio λ, es punto crítico de $\mathcal{L}_1$ y $\underline{l_1}$ es punto crítico de la $Var(Y_1)$.

A continuación se calcula la $Var(Y_1)$ considerando que su punto crítico $\underline{l_1}$ satisface $\Sigma\underline{l_1} = \lambda\underline{l_1}$ y $\underline{l_1}'\underline{l_1} = 1$, donde λ es su valor propio correspondiente:

$$Var(Y_1) = \underline{l_1}'\Sigma\underline{l_1} = \underline{l_1}'\lambda\underline{l_1} = \lambda\underline{l_1}'\underline{l_1} = \lambda.$$

Por otro lado, recordaremos que $\Sigma = E((\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})')$; se observa que $\Sigma = \Sigma'$; dado $\underline{w} \in \mathbb{R}^p - \{\underline{0}\}$, se sigue:

$$\begin{aligned} \underline{w}'\Sigma\underline{w} &= E(\underline{w}'(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})'\underline{w}) \\ &= E([\underline{w}'(\underline{X} - \underline{\mu})]_{1\times 1}[\underline{w}'(\underline{X} - \underline{\mu})]'_{1\times 1}) \\ &= E([\underline{w}'(\underline{X} - \underline{\mu})]^2) \geqslant 0, \end{aligned}$$

por lo que Σ es semidefinida positiva y es de orden p. Por el Teorema B.6 se sigue que los p valores propios de Σ son mayores o iguales a cero, es decir, $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p \geqslant 0$. Así, $Var(Y_1) = \lambda$ está bien definida.

Por último, como el objetivo es maximizar $Var(Y_1)$, se toma a λ como el valor propio máximo de Σ, λ_1, por lo que $\underline{l_1}$ es el vector propio unitario correspondiente a λ_1, que se denotará por $\underline{v_1}$. Por lo tanto, la primera componente principal está dada por $Y_1 = \underline{v_1}'\underline{X}$ y cumple que $Var(Y_1) = \lambda_1$.

Construcción de la segunda componente principal

La segunda componente principal se denotará por Y_2, y es la CLE de las variables originales que tenga varianza máxima restante y no esté correlacionada con la primera componente principal, es decir, $Y_2 = \underline{l_2}'\underline{X} = l_{21}X_1 + \ldots + l_{2p}X_p$ tiene varianza máxima y la $Cov(Y_2, Y_1) = 0$, donde $\underline{l_2} = (l_{21}, \ldots, l_{2p})'$ cumple que $\underline{l_2}'\underline{l_2} = 1$. Además se tiene que

la varianza de Y_2 es:

$$Var(Y_2) = Var(\underline{l_2}'\underline{X}) = \underline{l_2}'Var(\underline{X})\underline{l_2} = \underline{l_2}'\Sigma\underline{l_2}$$

y la covarianza de Y_2 y Y_1 es:

$$Cov(Y_2, Y_1) = Cov(\underline{l_2}'\underline{X}, \underline{v_1}'\underline{X}) = \underline{l_2}'Cov(\underline{X}, \underline{X})\underline{v_1} = \underline{l_2}'Var(\underline{X})\underline{v_1} = \underline{l_2}'\Sigma\underline{v_1}.$$

Nótese que se cumple que $\Sigma\underline{v_1} = \lambda_1\underline{v_1}$ con $\lambda_1 > 0$, dado que es la varianza de la primera componente principal que recoge la mayor variabilidad de los datos que tienen variabilidad positiva. Luego,

$$Cov(Y_2, Y_1) = \underline{l_2}'\Sigma\underline{v_1} = \underline{l_2}'\lambda_1\underline{v_1} = \lambda_1\underline{l_2}'\underline{v_1} = 0$$

implica que $\underline{l_2}'\underline{v_1} = 0$, es decir, que $\underline{l_2}$ y $\underline{v_1}$ son ortogonales.

Por lo tanto se resolverá el siguiente problema:

máx$\{Var(Y_2)\}$ = máx$_{R_2}\{\underline{l_2}'\Sigma\underline{l_2}\}$, donde $R_2 = \{\underline{l_2} \in \mathbb{R}^p | \underline{l_2}'\underline{l_2} - 1 = 0$ y $\underline{l_2}'\underline{v_1} = 0\}$.

La función lagrangiana de este problema es:

$$\mathcal{L}_2(\underline{l_2}, \lambda, \delta) = \underline{l_2}'\Sigma\underline{l_2} - \lambda(\underline{l_2}'\underline{l_2} - 1) - \delta\underline{l_2}'\underline{v_1},$$

para calcular los puntos críticos de $\mathcal{L}_2$, se calculan sus derivadas parciales con respecto a $\underline{l_2}$, λ y δ, y se igualan a cero.

$$\frac{\partial\mathcal{L}_2(\underline{l_2}, \lambda, \delta)}{\partial\underline{l_2}} = 2\Sigma\underline{l_2} - 2\lambda I_{p\times p}\underline{l_2} - \delta\underline{v_1} = \underline{0} \tag{2.3}$$

$$\frac{\partial\mathcal{L}_2(\underline{l_2}, \lambda, \delta)}{\partial\lambda} = -(\underline{l_2}'\underline{l_2} - 1) = 0 \Leftrightarrow \underline{l_2}'\underline{l_2} = 1 \tag{2.4}$$

$$\frac{\partial\mathcal{L}_2(\underline{l_2}, \lambda, \delta)}{\partial\delta} = -\underline{l_2}'\underline{v_1} = 0 \Leftrightarrow \underline{l_2}'\underline{v_1} = 0. \tag{2.5}$$

El sistema de ecuaciones formado por (2.3), (2.4) y (2.5) es no lineal. Sin embargo, si se multiplica (2.3) por $\underline{v_1}'$, se tiene que:

$$2\underline{v_1}'\Sigma\underline{l_2} - 2\lambda\underbrace{\underline{v_1}'\underline{l_2}}_{0} - \delta\underbrace{\underline{v_1}'\underline{v_1}}_{1} = \underline{v_1}'\underline{0} = 0; \Leftrightarrow \delta = 2\underline{v_1}'\Sigma\underline{l_2} = 2(Cov(Y_2, Y_1))' = 0$$

de (2.3) se sigue que:

$$2\Sigma\underline{l_2} - 2\lambda I_{p\times p}\underline{l_2} = \underline{0} \Leftrightarrow \Sigma\underline{l_2} - \lambda I_{p\times p}\underline{l_2} = \underline{0}.$$

De manera análoga a la construcción de la primera componente principal, se llega a que λ es un valor propio de Σ y $\underline{l_2}$ es un vector propio de norma 1 (por (2.4)) correspondiente a λ. Así, $(\underline{l_2}, \lambda, 0)$, donde $\underline{l_2}$ es el vector propio unitario correspondiente al valor propio λ, es punto crítico de $\mathcal{L}_2$, y $\underline{l_2}$ es punto crítico de la $Var(Y_2)$.

A continuación se calcula la $Var(Y_2)$ considerando que su punto crítico $\underline{l_2}$ satisface $\Sigma\underline{l_2} = \lambda\underline{l_2}$ y $\underline{l_2}'\underline{l_2} = 1$, donde λ es su valor propio correspondiente

$$Var(Y_2) = \underline{l_2}'\Sigma\underline{l_2} = \underline{l_2}'\lambda\underline{l_2} = \lambda\underline{l_2}'\underline{l_2} = \lambda.$$

Por último, como el objetivo es maximizar $Var(Y_2)$, si se toma a λ como el primer valor propio de Σ, $\underline{l_2}$ podría ser igual a $\underline{l_1}$ o no necesariamente son perpendiculares y podrían estar midiendo la misma información, por lo que se toma a λ como el segundo valor propio máximo de Σ, λ_2; $\underline{l_2}$ es un vector propio unitario correspondiente λ_2, que se denotará por $\underline{v_2}$; si $\lambda_1 = \lambda_2$, se escoge a $\underline{v_2}$ ortogonal a $\underline{v_1}$. Por lo tanto, la segunda componente principal está dada por $Y_2 = \underline{v_2}'\underline{X}$ y cumple que $Var(Y_2) = \lambda_2$.

Construcción de la (j+1)-ésima componente principal

La (j+1)-ésima componente principal se denotará por Y_{j+1}, es la CLE de las variables originales que tenga varianza máxima restante y no esté correlacionada con Y_1, Y_2, …, Y_j, es decir, $Y_{j+1} = \underline{l_{j+1}}'\underline{X} = l_{(j+1)1}X_1 + \ldots + l_{(j+1)p}X_p$ tiene varianza máxima y $Cov(Y_{j+1}, Y_i) = 0$, $i = 1, 2, \ldots, j$, donde $\underline{l_{j+1}} = (l_{(j+1)1}, \ldots, l_{(j+1)p})'$ cumple que $\underline{l_{j+1}}'\underline{l_{j+1}} = 1$. Además se tiene que la varianza de Y_{j+1} es:

$$Var(Y_{j+1}) = Var(\underline{l_{j+1}}'\underline{X}) = \underline{l_{j+1}}'Var(\underline{X})\underline{l_{j+1}} = \underline{l_{j+1}}'\Sigma\underline{l_{j+1}}$$

y la covarianza de Y_{j+1} y Y_i, $i = 1, 2, \ldots, j$ es:

$$\begin{aligned} Cov(Y_{j+1}, Y_i) &= Cov(\underline{l_{j+1}}'\underline{X}, \underline{v_i}'\underline{X}) = \underline{l_{j+1}}'Cov(\underline{X}, \underline{X})\underline{v_i} = \underline{l_{j+1}}'Var(\underline{X})\underline{v_i} \\ &= \underline{l_{j+1}}'\Sigma\underline{v_i}. \end{aligned}$$

Nótese que se cumple que $\Sigma\underline{v_i} = \lambda_i\underline{v_i}$ con $\lambda_i > 0$, dado que es la varianza de la i-ésima componente principal la que recoge la mayor variabilidad posible de los datos, y si $\lambda_i = 0$, ya se habría recuperado la variabilidad total de los datos en las primeras $i-1$ componentes principales y no se construirían más. Luego,

$$Cov(Y_{j+1}, Y_i) = \underline{l_{j+1}}'\Sigma\underline{v_i} = \underline{l_{j+1}}'\lambda_i\underline{v_i} = \lambda_i\underline{l_{j+1}}'\underline{v_i} = 0$$

implica que $\underline{l_{j+1}}'\underline{v_i} = 0$, es decir, que $\underline{l_{j+1}}$ y $\underline{v_i}$ son ortogonales para $i = 1, 2, \ldots, j$.

Por lo tanto se resolverá el siguiente problema:

$$\text{máx}\{Var(Y_{j+1})\} = \text{máx}_{R_{j+1}}\{\underline{l_{j+1}}'\Sigma\underline{l_{j+1}}\},$$

donde $R_{j+1} = \{\underline{l_{j+1}} \in \mathbb{R}^p | \underline{l_{j+1}}'\underline{l_{j+1}} - 1 = 0$ y $\underline{l_{j+1}}'\underline{v_i} = 0, i = 1, 2, \ldots, j\}$.

La función lagrangiana de este problema es:

$$\mathcal{L}_{j+1}(\underline{l_{j+1}}, \lambda, \delta_1, \ldots, \delta_j) = \underline{l_{j+1}}'\Sigma\underline{l_{j+1}} - \lambda(\underline{l_{j+1}}'\underline{l_{j+1}} - 1) - \sum_{i=1}^{j}\delta_i\underline{l_{j+1}}'\underline{v_i}.$$

Para calcular los puntos críticos de $\mathcal{L}_{j+1}$, se calculan sus derivadas parciales con respecto a $\underline{l_{j+1}}$, λ, δ_1, ..., δ_j y se igualan a cero.

$$\frac{\partial\mathcal{L}_{j+1}(\underline{l_{j+1}}, \lambda, \delta_1, \ldots, \delta_j)}{\partial\underline{l_{j+1}}} = 2\Sigma\underline{l_{j+1}} - 2\lambda I_{p\times p}\underline{l_{j+1}} - \sum_{i=1}^{j}\delta_i\underline{v_i} = \underline{0} \tag{2.6}$$

$$\frac{\partial\mathcal{L}_{j+1}(\underline{l_{j+1}}, \lambda, \delta_1, \ldots, \delta_j)}{\partial\lambda} = -(\underline{l_{j+1}}'\underline{l_{j+1}} - 1) = 0 \Leftrightarrow \underline{l_{j+1}}'\underline{l_{j+1}} = 1 \tag{2.7}$$

$$\frac{\partial\mathcal{L}_{j+1}(\underline{l_{j+1}}, \lambda, \delta_1, \ldots, \delta_j)}{\partial\delta_i} = -\underline{l_{j+1}}'\underline{v_i} = 0 \Leftrightarrow \underline{l_{j+1}}'\underline{v_i} = 0, i = 1, 2, \ldots, j. \tag{2.8}$$

El sistema de ecuaciones formado por (2.6), (2.7) y (2.8) es no lineal. Sin embargo, si se multiplica (2.6) sucesivamente por $\underline{v_1}'$, ..., $\underline{v_j}'$, los cuales se tomaron ortonormales, se tiene que:

$$2\underline{v_1}'\Sigma\underline{l_{j+1}} - 2\lambda\underbrace{\underline{v_1}'\underline{l_{j+1}}}_{0} - \delta_1\underbrace{\underline{v_1}'\underline{v_1}}_{1} - \textstyle\sum_{i\neq 1}\delta_i\underbrace{\underline{v_1}'\underline{v_i}}_{0} = \underline{v_1}'\underline{0} = 0$$

$$\vdots$$

$$2\underline{v_j}'\Sigma\underline{l_{j+1}} - 2\lambda\underbrace{\underline{v_j}'\underline{l_{j+1}}}_{0} - \delta_j\underbrace{\underline{v_j}'\underline{v_j}}_{1} - \textstyle\sum_{i\neq j}\delta_i\underbrace{\underline{v_j}'\underline{v_i}}_{0} = \underline{v_j}'\underline{0} = 0,$$

por lo tanto,

$$\delta_1 = 2\underline{v_1}'\Sigma\underline{l_{j+1}} = 2(Cov(Y_{j+1}, Y_1))' = 0$$
$$\vdots$$
$$\delta_j = 2\underline{v_j}'\Sigma\underline{l_{j+1}} = 2(Cov(Y_{j+1}, Y_j))' = 0,$$

de (2.6) se sigue que:

$$2\Sigma\underline{l_{j+1}} - 2\lambda I_{p\times p}\underline{l_{j+1}} = \underline{0} \Leftrightarrow \Sigma\underline{l_{j+1}} - \lambda I_{p\times p}\underline{l_{j+1}} = \underline{0}.$$

De manera análoga a la construcción de la primera y la segunda componente principal se llega a que λ es un valor propio de Σ, y $\underline{l_{j+1}}$ es un vector propio de norma 1 (por (2.7)) correspondiente a λ. Así, $(\underline{l_{j+1}}, \lambda, \underline{0})$, donde $\underline{l_{j+1}}$ es el vector propio unitario correspondiente al valor propio λ, es punto crítico de $\mathcal{L}_{j+1}$, y $\underline{l_{j+1}}$ es punto crítico de $Var(Y_{j+1})$.

A continuación se calcula la $Var(Y_{j+1})$ considerando que su punto crítico $\underline{l_{j+1}}$ satisface $\Sigma\underline{l_{j+1}} = \lambda\underline{l_{j+1}}$ y $\underline{l_{j+1}}'\underline{l_{j+1}} = 1$, donde λ es su valor propio correspondiente:

$$Var(Y_{j+1}) = \underline{l_{j+1}}'\Sigma\underline{l_{j+1}} = \underline{l_{j+1}}'\lambda\underline{l_{j+1}} = \lambda\underline{l_{j+1}}'\underline{l_{j+1}} = \lambda.$$

Por último, como el objetivo es maximizar $Var(Y_{j+1})$, se toma a λ como el (j+1)-ésimo valor propio más grande de Σ, λ_{j+1}; así, $\underline{l_{j+1}}$ es un vector propio unitario correspondiente a λ_{j+1}, que se denotará por $\underline{v_{j+1}}$, cuando $\lambda_{j-r} = \lambda_{j-r+1} = \ldots = \lambda_j = \lambda_{j+1}$; se toma a $\underline{v_{j+1}}$ ortogonal a $\underline{v_{j-r}}$, $\underline{v_{j-r+1}}$, …, $\underline{v_j}$. Por lo tanto, la (j+1)-ésima componente principal está dada por $Y_{j+1} = \underline{v_{j+1}}'\underline{X}$ y cumple que $Var(Y_{j+1}) = \lambda_{j+1}$.

Si se define Γ como la matriz que tiene a los vectores propios unitarios $\underline{v_1}$, $\underline{v_2}$, …, $\underline{v_p}$ correspondientes a λ_1, λ_2, …, λ_p como columnas ($\Gamma = (\underline{v_1}, \underline{v_2}, \ldots, \underline{v_p})$), el vector aleatorio de las componentes principales $\underline{Y} = (Y_1, Y_2, \ldots, Y_p)'$ se puede expresar como:

$$\underline{Y} = \Gamma'\underline{X}.$$

Por la forma de construir las componentes principales, se tiene que $Var(Y_i) = \lambda_i$, $i = 1, 2, \ldots, p$ y $Cov(Y_i, Y_j) = 0$, $i, j = 1, 2, \ldots, p$, $i \neq j$. Luego:

$$Var(\underline{Y}) = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \lambda_p \end{bmatrix} = \Lambda;$$

por otro lado,

$$\Lambda = Var(\underline{Y}) = Var(\Gamma'\underline{X}) = \Gamma' Var(\underline{X})\Gamma = \Gamma'\Sigma\Gamma;$$

como Γ es una matriz ortogonal, $\Gamma'\Gamma = I$, se tiene también que:

$$\Sigma = \Gamma\Lambda\Gamma'.$$

2.3.2. Construcción conjunta de las componentes principales

En la construcción conjunta de las componentes principales se obtendrán de manera conjunta las CLE que satisfacen los objetivos de las componentes principales.

Dado que Σ, la matriz de covarianza de $\underline{X}$, es simétrica, por el teorema de la descomposición espectral (Teorema B.7) se cumple que:

$$\Sigma = \Gamma\Lambda\Gamma',$$

donde Λ es una matriz diagonal de valores propios de Σ y Γ es una matriz ortogonal cuyas columnas son vectores propios estandarizados.

Como Γ es una matriz ortogonal, cumple $\Gamma'\Gamma = I$, lo cual equivale a que $\|\underline{\gamma_{(i)}}\|^2 = \underline{\gamma_{(i)}}'\underline{\gamma_{(i)}} = 1$, $i = 1, 2, \ldots, p$ y $\underline{\gamma_{(i)}}'\underline{\gamma_{(j)}} = 0$, $i \neq j$, $i, j = 1, 2, \ldots, p$; como $\underline{\gamma_{(i)}}$ representa la columna i de Γ, es decir, un vector propio de Σ, esto equivale a que los vectores propios son unitarios y ortogonales. Luego la combinación lineal formada por los vectores propios y el vector aleatorio $\underline{X}$, $\underline{\gamma_{(i)}}'\underline{X}$ es una CLE.

Se propone $\underline{Y} = (Y_1, \ldots, Y_p)' = \Gamma'\underline{X}$, cuya varianza está dada por:

$$Var(\underline{Y}) = Var(\Gamma'\underline{X}) = \Gamma' Var(\underline{X})\Gamma = \Gamma'\Sigma\Gamma = \Gamma'\Gamma\Lambda\Gamma'\Gamma = \Lambda,$$

por lo que $Var(Y_i) = \lambda_i$, $i = 1, \ldots, p$ y $Cov(Y_i, Y_j) = 0$, $i \neq j$, $i, j = 1, \ldots, p$, es decir, las Y_i no están correlacionadas.

Dado que Σ es semidefinida positiva y es de orden p, por el Teorema B.6 se sigue que los p valores propios de Σ son mayores o iguales a cero, es decir, $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p \geqslant 0$. Así, $Var(Y_i) = \lambda_i$ está bien definida.

Hasta el momento Y_1, ..., Y_p son CLE's no correlacionadas, sin embargo Y_1 no necesariamente tiene varianza máxima, por lo que se define:

$$Y_1 = \underline{v_1}'\underline{X},$$

donde $\underline{v_1}$ es un vector propio unitario de λ_1, el valor propio máximo. Luego Y_2 debe tener la mayor varianza restante, pero Y_2 y Y_1 no están correlacionadas, por lo que se define:

$$Y_2 = \underline{v_2}'\underline{X},$$

donde $\underline{v_2}$ es un vector unitario de λ_2, el segundo valor propio máximo. Y_{j+1} no está correlacionada con $Y_1, \ldots, Y_j$, si se define como:

$$Y_{j+1} = \underline{v_{j+1}}'\underline{X},$$

donde $\underline{v_{j+1}}$ es un vector propio unitario de λ_{j+1}, el (j+1)-ésimo valor propio máximo se está maximizando la varianza restante de las $Y_1, \ldots, Y_j$.

Por lo tanto, las componentes principales quedan definidas de manera conjunta como:

$$\underline{Y} = \Gamma'\underline{X},$$

donde Γ es la matriz que tiene a los vectores propios unitarios $\underline{v_1}, \underline{v_2}, \ldots, \underline{v_p}$ correspondientes a $\lambda_1, \lambda_2, \ldots, \lambda_p$ como columnas ($\Gamma = (\underline{v_1}, \underline{v_2}, \ldots, \underline{v_p})$).

2.3.3. Definición de las componentes principales poblacionales

Como se vio, para un vector aleatorio $\underline{X}$ con media $\underline{\mu}$ y varianza Σ, el vector aleatorio de sus componentes principales está dado por $\underline{Y} = \Gamma'\underline{X}$; sin embargo, la varianza de $\underline{X} - \underline{\mu}$ es $Var(\underline{X} - \underline{\mu}) = Var(\underline{X}) = \Sigma$, por tanto $\Gamma'(\underline{X} - \underline{\mu})$ también es un vector aleatorio que satisface los objetivo de las componentes principales pero centrando la distribución, por lo cual a $\underline{Y} = \Gamma'(\underline{X} - \underline{\mu})$ se le llamará el ACP centrado y a $\underline{Y} = \Gamma'\underline{X}$ se le llamará el ACP no centrado. A continuación se da la definición de la transformación de componentes principales

Definición 2.3. *Si $\underline{X}$ es un vector aleatorio con media $\underline{\mu}$ y matriz de covarianza Σ, entonces la transformación a componentes principales es la transformación:*

$$\underline{X} \to \underline{Y} = \Gamma'(\underline{X} - \underline{\mu}), \tag{2.9}$$

donde Γ es una matriz ortogonal cuyas columnas son vectores propios estandarizados de Σ; $\Gamma'\Sigma\Gamma = \Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_p)$ es diagonal de valores propios de Σ y $\lambda_1 \geqslant \lambda_2 \geqslant$

$\ldots \geqslant \lambda_p \geqslant 0$. *La positividad estricta de los valores propios* λ_i *está garantizada si* Σ *es definida positiva.*

El i-ésimo componente principal de $\underline{X}$ *se puede definir como el i-ésimo elemento del vector* $\underline{Y}$*, es decir, como:*

$$Y_i = \underline{\gamma_{(i)}}'(\underline{X} - \underline{\mu}), \tag{2.10}$$

donde $\gamma_{(i)}$ *es la i-ésima columna de* Γ*. La función* Y_p *se puede llamar la última componente principal de* $\underline{X}$*.*

Esta definición plantea las siguientes observaciones:

1. Si hay valores propios iguales, supongamos λ_k, ..., λ_{k+r}, los vectores propios $\underline{\gamma_{(k)}}$, ..., $\underline{\gamma_{(k+r)}}$ asociados no son únicos, por lo que las respectivas componentes principales no serán únicas [14].

2. Si el valor propio no se repite, supongamos λ_l, los vectores propios $\underline{\gamma_{(l)}}$ y $-\underline{\gamma_{(l)}}$ son unitarios y las SCL's formadas por ellos tienen la misma varianza, λ_l, por lo que la componente principal no es única.

Por lo tanto, las componentes principales no son únicas, lo que permite escoger a los $\underline{\gamma_{(i)}}$ como más convenga, por ejemplo, se puede escoger a $\underline{\gamma_{(i)}}$ de tal forma que $Y_i = \underline{\gamma_{(i)}}'\underline{X}$ sea más interpretable que las otras Y_i que pueden ser la i-ésima componente principal.

2.3.4. Calificaciones de las componentes principales

Por definición, las componentes principales están construidas con vectores propios ortogonales, lo cual permite colocar un sistema de ejes ortogonales dentro del espacio de componentes principales en el cual caen los datos.

Para usar las variables componentes principales en los análisis estadísticos subsiguientes, es necesario calcular las calificaciones de esas componentes (valores de las variables componentes principales) para cada unidad experimental en el conjunto de datos. Estas calificaciones proporcionan las ubicaciones de las observaciones en un conjunto de datos con respecto a sus ejes componentes principales.

Sea $\underline{x_r}$ el vector de variables medidas para la r-ésima unidad experimental. Entonces el valor (calificación) de la j-ésima variable componente principal para la r-ésima unidad experimental es: $y_{rj} = \underline{\gamma_{(j)}}(\underline{x}_r - \underline{\mu})$, para $j = 1, \ldots, p$ y $r = 1, \ldots, n$ [8].

2.3.5. Propiedades de las componentes principales

Las componentes principales cumplen los siguientes resultados:

Teorema 2.4. *Si $\underline{X} \sim (\underline{\mu}, \Sigma)$ y $\underline{Y}$ se define como en la Ecuación (2.9), entonces:*

i) $E(\underline{Y}) = \underline{0}$

ii) $Var(\underline{Y}) = \Lambda$

iii) $Var(Y_1) \geqslant Var(Y_2) \geqslant \ldots \geqslant Var(Y_p) \geqslant 0$

iv) $\sum_{i=1}^{p} Var(Y_i) = tr(\Sigma)$

v) $\prod_{i=1}^{p} Var(Y_i) = |\Sigma|$.

Demostración.

i) $E(\underline{Y}) = E(\Gamma'(\underline{X} - \underline{\mu})) = \Gamma' E(\underline{X} - \underline{\mu}) = \Gamma'(E(\underline{X}) - \underline{\mu}) = \Gamma'(\underline{\mu} - \underline{\mu}) = \underline{0}$

ii)$Var(\underline{Y}) = Var(\Gamma'(\underline{X} - \underline{\mu})) = \Gamma' Var(\underline{X} - \underline{\mu})\Gamma = \Gamma' Var(\underline{X})\Gamma = \Gamma'\Sigma\Gamma = \Lambda$. Lo que equivale a:

- $Var(Y_i) = \lambda_i$
- $Cov(Y_i, Y_j) = 0$ para $i \neq j$.

iii) De la Definición 2.3 se tiene que $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p \geqslant 0$; como $Var(Y_i) = \lambda_i$; se sigue que $Var(Y_1) \geqslant Var(Y_2) \geqslant \ldots \geqslant Var(Y_p) \geqslant 0$.

iv) De la Definición 2.10 se tiene $\Gamma'\Sigma\Gamma = \Lambda$; se sigue que $\Sigma = \Gamma\Lambda\Gamma'$; luego:

$$tr(\Sigma) = tr(\Gamma\Lambda\Gamma') = tr(\Lambda\Gamma'\Gamma) = tr(\Lambda) = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} Var(Y_i)$$

v) $|\Sigma| = |\Gamma\Lambda\Gamma'| = |\Gamma||\Lambda\Gamma'| = |\Lambda\Gamma'||\Gamma| = |\Lambda\Gamma'\Gamma| = |\Lambda| = \prod_{i=1}^{p} \lambda_i = \prod_{i=1}^{p} Var(Y_i)$ □

A continuación se presenta un resultado más significativo

Teorema 2.5. *Sea $\underline{X} \sim (\mu, \Sigma)$ un vector aleatorio de orden $p \times 1$; se tiene que:*

- *Ninguna CLE de $\underline{X}$ tiene varianza mayor que λ_1, la varianza de la primera componente principal.*
- *Ninguna CLE de $\underline{X}$ tiene varianza menor que λ_p, la varianza de la última componente principal.*

Demostración. Sea $\underline{a}'\underline{X}$ una CLE, es decir, $\underline{a}'\underline{a} = 1$. Se puede escribir:

$$\underline{a} = c_1\underline{\gamma_{(1)}} + \ldots + c_p\underline{\gamma_{(p)}} = \Gamma\underline{c}, \tag{2.11}$$

donde $\underline{c} = (c_1, \ldots, c_p)'$, $\Gamma = (\underline{\gamma_{(1)}}, \cdots, \underline{\gamma_{(p)}})$ y $\underline{\gamma_{(1)}}$, ...,$\underline{\gamma_{(p)}}$ son los vectores propios de Σ (cualquier vector se puede escribir de esta forma, ya que los vectores propios forman una base para $\mathbb{R}^p$). Ahora,

$$Var(\underline{a}'\underline{X}) = \underline{a}'Var(\underline{X})\underline{a} = \underline{a}'\Sigma\underline{a};$$

usando el teorema de la descomposición espectral (Teorema B.7), se cumple que $\Sigma = \Gamma\Lambda\Gamma'$, por lo que:

$$Var(\underline{a}'\underline{X}) = \underline{a}'\Sigma\underline{a} = \underline{c}'\Gamma'\Sigma\Gamma\underline{c} = \underline{c}'\underbrace{\Gamma'\Gamma}_{I}\Lambda\underbrace{\Gamma'\Gamma}_{I}\underline{c} = \underline{c}'\Lambda\underline{c} = \sum_{i=1}^{p}\lambda_i c_i^2. \tag{2.12}$$

Además,

$$1 = \underline{a}'\underline{a} = \underline{c}'\underbrace{\Gamma'\Gamma}_{I}\underline{c} = \underline{c}'\underline{c} = \sum_{i=1}^{p} c_i^2. \tag{2.13}$$

Dado que Σ es definida positiva y es de orden p, por el Teorema B.6 se sigue que los p valores propios de Σ son mayores a cero, es decir, $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p > 0$. Por lo tanto, de (2.12) y (2.13) se sigue que:$\lambda_1 = \lambda_1 \sum_{i=1}^{p} c_i^2 = \sum_{i=1}^{p} \lambda_1 c_i^2 \geqslant Var(\underline{a}'\underline{X}) \geqslant \sum_{i=1}^{p} \lambda_p c_i^2 = \lambda_p \sum_{i=1}^{p} c_i^2 = \lambda_p$. □

Para un vector aleatorio multinormal, el teorema anterior se puede interpretar geométricamente diciendo que el primer componente principal representa el eje mayor de la elipsoide de concentración y el último componente principal representa el eje menor de la elipsoide de concentración (véase Figura 2.1).

Las componentes intermedias tienen una propiedad de varianza máxima dada por el siguiente teorema:

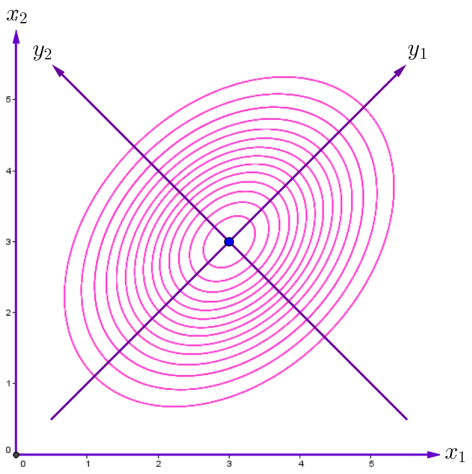

Figura 2.1: Las elipses de concentración para la distribución normal bivariada, se muestran las componentes principales y_1 e y_2, donde $\underline{\mu} = (3,3)'$, $\sigma_{11} = 3$, $\sigma_{12} = 1$ y $\sigma_{22} = 3$. Elaboración propia basada en [11]

Teorema 2.6. *Sea $\underline{X} \sim (\mu, \Sigma)$ un vector aleatorio de orden $p \times 1$; se tiene que si $\alpha = \underline{a}'\underline{X}$ es una CLE de $\underline{X}$ no correlacionada con las primeras k componentes principales de $\underline{X}$, entonces la varianza de α se maximizada cuando α es el (k+1)-ésimo componente principal de $\underline{X}$, $k < p$.*

Demostración. Se puede escribir $\underline{a} = c_1\underline{\gamma_{(1)}} + \ldots + c_p\underline{\gamma_{(p)}} = \Gamma\underline{c}$ (Ecuación (2.11)), donde $\underline{c} = (c_1, \ldots, c_p)'$; $\Gamma = (\underline{\gamma_{(1)}}, \cdots, \underline{\gamma_{(p)}})$, y $\underline{\gamma_{(1)}}$, ...,$\underline{\gamma_{(p)}}$ son los vectores propios correspondientes a los valores propios $\lambda_1, \ldots, \lambda_p$ de Σ, es decir, se cumple que $\Sigma\underline{\gamma_{(i)}} = \lambda_i\underline{\gamma_{(i)}}$. Dado que α no está correlacionado con $\underline{\gamma_{(i)}}'\underline{X}$, $i = 1, \ldots, k$,

$$\begin{aligned} 0 &= Cov(\underline{a}'\underline{X}, \underline{\gamma_{(i)}}'\underline{X}) = \underline{a}'Cov(\underline{X}, \underline{X})\underline{\gamma_{(i)}} = \underline{a}'Var(\underline{X})\underline{\gamma_{(i)}} = \underline{a}'\Sigma\underline{\gamma_{(i)}} \\ &= \underline{a}'\lambda_i\underline{\gamma_{(i)}} = \lambda_i\underline{a}'\underline{\gamma_{(i)}}; \end{aligned}$$

ya que Σ es definida positiva, se tiene que $\lambda_1 \geqslant \ldots \geqslant \lambda_p > 0$, por lo cual $\underline{a}'\underline{\gamma_{(i)}} = 0$, $i = 1, \ldots, k$, luego:

$$\begin{aligned} 0 &= \underline{a}'\underline{\gamma_{(i)}} = (c_1\underline{\gamma_{(1)}} + \ldots + c_i\underline{\gamma_{(i)}} + \ldots + c_p\underline{\gamma_{(p)}})'\underline{\gamma_{(i)}} \\ &= (c_1\underline{\gamma_{(1)}}' + \ldots + c_i\underline{\gamma_{(i)}}' + \ldots + c_p\underline{\gamma_{(p)}}')\underline{\gamma_{(i)}} \\ &= (c_1\underbrace{\underline{\gamma_{(1)}}'\underline{\gamma_{(i)}}}_{0} + \ldots + c_i\underbrace{\underline{\gamma_{(i)}}'\underline{\gamma_{(i)}}}_{1} + \ldots + c_p\underbrace{\underline{\gamma_{(p)}}'\underline{\gamma_{(i)}}}_{0}) = c_i \text{ para } i = 1, \ldots, k; \end{aligned}$$

se sigue que:

$$\begin{aligned} Var(\underline{a}'\underline{X}) &= \underline{a}'\Sigma\underline{a} = \underline{c}'\Gamma'\Sigma\Gamma\underline{c} = \underline{c}'\underbrace{\Gamma'\Gamma}_{I}\Lambda\underbrace{\Gamma'\Gamma}_{I}\underline{c} = \underline{c}'\Lambda\underline{c} = \textstyle\sum_{i=1}^{p}\lambda_i c_i^2 \\ &= \textstyle\sum_{i=k+1}^{p}\lambda_i c_i^2 \leqslant \sum_{i=k+1}^{p}\lambda_{k+1}c_i^2 = \lambda_{k+1}\sum_{i=k+1}^{p}c_i^2 = \lambda_{k+1} \end{aligned}$$

y la $Var(Y_{k+1}) = \lambda_{k+1}$, por lo tanto la varianza de α es maximizada cuando α es la (k+1)-ésima componente principal de $\underline{X}$. □

Las propiedades más importantes de las componentes principales se enunciaron en los Teoremas 2.4, 2.5, 2.6. Ahora se examinarán otras propiedades útiles.

a) *La varianza total y la varianza generalizada son invariantes respecto de las variables originales y respecto de las componentes principales.*

   La matriz de covarianza Σ es una posible generalización multivariada de la noción univariada de varianza, medida de la desviación respecto a la media. Sin embargo, algunas veces es conveniente tener sólo un número para medir la desviación multivariada. Dos medidas comunes son:

   - *la varianza generalizada*, $|\Sigma|$, y
   - *la varianza total*, $tr(\Sigma)$,

   que son invariantes respecto de las variables originales y respecto de las componentes principales, pues, por el Teorema 2.4, la $Var(\underline{Y}) = \Lambda$, $|\Sigma| = \prod_{i=1}^{p} Var(Y_i) = \prod_{i=1}^{p}\lambda_i = |\Lambda|$ y $tr(\Sigma) = \sum_{i=1}^{p} Var(Y_i) = \sum_{i=1}^{p}\lambda_i = tr(\Lambda)$.

b) Para medir la importancia de la j-ésima componente principal se considera la relación $\lambda_j/(\lambda_1 + \ldots + \lambda_p)$, para $j = 1, 2, \ldots, p$. Ésta mide la proporción de la variabilidad total en las variables originales explicada por la j-ésima componente principal [8].

c) *La suma de los k primeros valores propios dividida entre la suma de todos los valores propios, $(\lambda_1 + \ldots + \lambda_k)/(\lambda_1 + \ldots + \lambda_p)$, representa la "proporción de la variación total" explicada por los k primeros componentes principales.*

   La proporción de variación total, $(\lambda_1 + \ldots + \lambda_k)/(\lambda_1 + \ldots + \lambda_p)$, da una medida cuantitativa de la cantidad de información retenida en la reducción de dimensión p

a k [11]. En la práctica, si las componentes principales son tales que unas pocas explican un alto porcentaje de la varianza total, merece la pena sustituir el vector original $\underline{X}$ por dichas componentes principales [14].

d) *Las componentes principales de un vector aleatorio no son invariantes bajo escala.*

Una desventaja del análisis de componentes principales es que no es invariante bajo escala, es decir, si se tienen variables en cierta escala y se busca un análisis de componentes principales en otra escala (por ejemplo, si se tiene peso en libras, altura en pies y edad en años y se buscan componentes principales expresados en onzas, pulgadas y décadas), se sugieren los siguientes procesos:

a) multiplicar las variables por ciertos escalares que proporcionen el cambio de escala y después llevar a cabo un ACP

b) llevar a cabo un ACP y después multiplicar los elementos de las componentes por sus correspondientes escalares que proporcionan el cambio de escala;

sin embargo, estos procedimientos no siempre proporcionan los mismos resultados [11].

Algebraicamente, la falta de invarianza bajo escala se explica como sigue:

Sea Σ la matriz de covarianza del vector aleatorio $\underline{X}$. Si se divide la i-ésima variable por d_i, la matriz de covarianza de las nuevas variables es:

$$Var(D\underline{X}) = DVar(\underline{X})D' = D\Sigma D, \tag{2.14}$$

donde $D = diag(d_i^{-1})$. Sin embargo, si $\underline{\gamma}$ es un vector propio de Σ, $\Sigma\underline{\gamma} = \lambda\underline{\gamma}$, se sigue que $D\Sigma DD^{-1}\underline{\gamma} = \lambda D\underline{\gamma}$, entonces los vectores $D^{-1}\underline{\gamma}$ y $D\underline{\gamma}$ no son vectores propios de $D\Sigma D$, además se tiene que $D\underline{\gamma}\underline{X}$ es la variable resultante del cambio de escala de la componente principal, que es diferente a todas las componentes principales construidas a partir de $D\Sigma D$, por lo tanto las componentes principales no son invariantes bajo escala.

La falta de invarianza bajo escala implica sensibilidad a la forma en que se eligen las escalas. Dos posibles soluciones son:

a) Buscar las llamadas unidades "naturales", para asegurar que todas las variables medidas son del mismo tipo.

b) Estandarizar todas las variables para que tengan varianza igual a 1, y luego encontrar las componentes principales de la matriz de correlación en lugar de la matriz de covarianza. Esta opción es la que comúnmente se usa, a pesar de que complica las cuestiones de las pruebas de hipótesis [11].

En general las componentes principales se modifican por un cambio de escala de las variables, por lo que no son una característica única de los datos.

e) *Si la matriz de covarianza de $\underline{X}$ tiene rango $r < p$, entonces la varianza total de $\underline{X}$ se explicaría completamente por las primeras r componentes principales.*

Si Σ tiene rango r, entonces los últimos $(p-r)$ valores propios de Σ son cero, luego la proporción de la variación total explicada por las primeros r componentes principales es $\sum_{i=1}^{r} \lambda_i / \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{r} \lambda_r / \sum_{i=1}^{r} \lambda_r = 1$, es decir, las primeras r componentes principales explican la variación total de $\underline{X}$.

2.3.6. Vectores de carga de componentes

Supongamos que se desea usar componentes principales para identificar a los individuos con valores grandes en las variables originales. Un procedimiento consistiría en escoger a los individuos con calificaciones grandes en las componentes principales que cumplen que los signos en los elementos de cada vector propio, que corresponden a elementos que no están próximos a cero, y son en su mayoría positivos. Por lo tanto, los individuos con valores grandes en las variables originales también tendrán calificaciones grandes en las componentes principales elegidas, y viceversa.

Nótese que los vectores propios de la matriz de varianzas-covarianzas que se usan para definir componentes principales se normalizan para tener longitud 1, esto es, $\underline{\gamma_{(j)}}' \underline{\gamma_{(j)}} = 1$, para $j = 1, \ldots, p$. Puede resultar confuso cuando se trata de interpretar a las componentes principales mediante el *examen de los elementos en los vectores propios* que definen esas componentes. Un elemento grande en un vector propio puede serlo o

no en otro vector propio, es decir, los elementos dentro de un vector propio son comparables entre sí, pero los elementos en vectores propios diferentes no son comparables, en virtud de que los vectores propios se normalizan para tener longitud 1, lo cual requiere que la suma de los cuadrados de los elementos en cada vector debe ser igual a 1. Por consiguiente, cuantos más elementos haya en un solo vector propio que sean diferentes de cero, más pequeño debe ser cada elemento.

Para comparar entre vectores propios, muchos investigadores cambian la escala de los vectores propios multiplicando los elementos en cada vector por la raíz cuadrada de su valor propio correspondiente.

Definición 2.7. *Sea* $\underline{\beta_{(j)}} = \lambda_j^{1/2}\underline{\gamma_{(j)}}$*, para* $j = 1, \ldots, p$*. Estos nuevos vectores se llaman* vectores de carga de componentes *y los elementos del vector* $\underline{\beta_{(j)}}$ *reciben el nombre de* cargas de componentes*. En notación matricial se puede escribir* $B = (\underline{\beta_{(1)}}, \cdots, \underline{\beta_{(p)}}) = \Gamma\Lambda^{1/2}$*.*

Los $\underline{\beta_{(j)}}$ todavía son vectores propios de Σ, pero tienen longitudes iguales a $\|\underline{\beta_{(j)}}\| = \|\lambda_j^{1/2}\underline{\gamma_{(j)}}\| = |\lambda_j^{1/2}|\|\underline{\gamma_{(j)}}\| = \lambda_j^{1/2}$, para $j = 1, \ldots, p$, en lugar de longitud 1. Las cargas de componentes tienen una escala tal que permite comparar entre sí todos los elementos en todos los $\underline{\beta_{(j)}}$.

2.3.7. Covarianza y correlación entre las componentes principales y las variables originales

Teorema 2.8. *Sea* $\underline{Y}$ *el vector aleatorio de componentes principales del vector aleatorio* $\underline{X} \sim (\mu, \Sigma)$*. El i-ésimo elemento en* $\lambda_j\underline{\gamma_{(j)}}$ *da la covarianza entre la i-ésima variable original y la j-ésima componente principal.*

Demostración. Recordemos que $\underline{Y} = \Gamma'(\underline{X} - \underline{\mu})$ y $\Sigma = \Gamma\Lambda\Gamma'$

$$\begin{aligned} Cov(\underline{X}, \underline{Y}) &= Cov(\underline{X}, \Gamma'(\underline{X} - \underline{\mu})) = Cov(\underline{X}, \underline{X} - \underline{\mu})\Gamma = Cov(\underline{X}, \underline{X})\Gamma \\ &= Var(\underline{X})\Gamma = \Sigma\Gamma = \Gamma\Lambda\underbrace{\Gamma'\Gamma}_{I} = \Gamma\Lambda. \end{aligned}$$

La $Cov(X_i, Y_j)$ es el elemento (i, j) de $Cov(\underline{X}, \underline{Y}) = \Gamma\Lambda$, es decir, $\gamma_{ij}\lambda_j$, el i-ésimo elemento en $\lambda_j\underline{\gamma_{(j)}}$. □

En la mayoría de los casos no existen interpretaciones obvias para las componentes principales. Cualquiera de esas interpretaciones debe surgir de un *examen de vectores propios*. El teorema anterior, $\lambda_j\underline{\gamma_{(j)}}$, proporciona la covarianza de la j-ésima componente principal con las p variables originales, además, como $\underline{\gamma_{(j)}}$ es un múltiplo de $\lambda_j\underline{\gamma_{(j)}}$, se puede decir que las variables con fuertes relaciones con la j-ésima componente principal, $Y_j = \underline{\gamma_{(j)}}'\underline{X} = (\gamma_{1j}, \ldots, \gamma_{pj})\underline{X}$, son las que tienen elementos en $\underline{\gamma_{(j)}}$ que tienden a ser mayores en valor absoluto que los otros elementos de $\underline{\gamma_{(j)}}$, es decir, las X_i. tal que $|\gamma_{ij}|$ es un valor grande de $|\gamma_{1j}|$, ..., $|\gamma_{pj}|$.

Corolario 2.9. *Sea $\underline{Y}$ el vector aleatorio de componentes principales del vector aleatorio $\underline{X} \sim (\mu, \Sigma)$. La correlación de la i-ésima variable original y la j-ésima componente principal es $\rho_{X_iY_j} = \frac{\gamma_{ij}\sqrt{\lambda_j}}{\sqrt{\sigma_{ii}}}$ [14].*

Demostración. Se tiene que $Var(X_i) = \sigma_{ii}$ y $Var(Y_j) = \lambda_j$, así que:

$$\rho_{X_iY_j} = \frac{Cov(X_i, Y_j)}{\sqrt{Var(X_i)}\sqrt{Var(Y_j)}} = \frac{\gamma_{ij}\lambda_j}{\sqrt{\sigma_{ii}}\sqrt{\lambda_j}} = \frac{\gamma_{ij}\sqrt{\lambda_j}}{\sqrt{\sigma_{ii}}} = \frac{\sqrt{\lambda_j}}{\sqrt{\sigma_{ii}}}\gamma_{ij}$$

□

Éste es también un resultado importante, ya que permite medir la importancia de cada variable original, X_i, sobre cada componente principal Y_j. La correlación entre las variables originales y la j-ésima componente principal están dadas en el vector $(\rho_{X_1Y_j}, \ldots, \rho_{X_pY_j})' = \sqrt{\lambda_j}(\gamma_{1j}/\sqrt{\sigma_{11}}, \ldots, \gamma_{pj}/\sqrt{\sigma_{pp}})'$. A raíz de la anterior expresión, se deduce que cuanto mayor sea la i-ésima componente de $\underline{\gamma_{(j)}}$, $|\gamma_{ij}|$, mayor será la correlación entre X_i y Y_j [14].

Se puede decir que la proporción de variación de X_i "explicada" por Y_j es $\rho^2_{X_iY_j}$. Entonces, dado que los elementos de $\underline{Y}$ no están correlacionados, cualquier conjunto $\mathcal{G}$ de componentes explica una proporción:

$$\rho^2_{X_i\mathcal{G}} = \sum_{j\in\mathcal{G}} \rho^2_{X_iY_j} = \sum_{j\in\mathcal{G}} \left(\frac{\sqrt{\lambda_j}\gamma_{ij}}{\sqrt{\sigma_{ii}}}\right)^2 = \frac{1}{\sigma_{ii}}\sum_{j\in\mathcal{G}} \lambda_j\gamma^2_{ij} \tag{2.15}$$

de la variación de X_i. El denominador de esta expresión representa y explica la variación en X_i, y el numerador da la cantidad explicada por el conjunto $\mathcal{G}$. Cuando $\mathcal{G}$ contiene todas las componentes, el numerador es el elemento (i,i) de $\Gamma\Lambda\Gamma' = \Sigma$, que es σ_{ii}, así, la proporción (2.15) es 1.

Nótese que la parte de la *variación total* explicada por las componentes en $\mathcal{G}$ se puede expresar como la suma de todas las p variables de la proporción de *la variación en cada variable* explicada por las componentes en $\mathcal{G}$, donde cada variable es ponderada por su varianza, esto es,

$$\sum_{i=1}^{p} \sigma_{ii}\rho^2_{i\mathcal{G}} = \sum_{i=1}^{p}\sum_{j\in\mathcal{G}} \lambda_j\gamma^2_{ij} = \sum_{j\in\mathcal{G}}\left(\lambda_j \sum_{i=1}^{p}\gamma^2_{ij}\right) = \sum_{j\in\mathcal{G}}\left(\lambda_j \|\underline{\gamma_{(j)}}\|^2\right) = \sum_{j\in\mathcal{G}}\lambda_i.$$

2.4. Componentes principales poblacionales de la matriz de correlaciones

El análisis de componentes principales sólo es apropiado en aquellos casos donde todas las variables surgen “sobre un fundamento igual”, es decir:

1. *Todas las variables deben estar medidas en las mismas unidades o, por lo menos, en unidades comparables.*

   Si las variables respuesta no se miden en las mismas unidades, entonces cualquier cambio en la escala de medición en una o más de las variables tendrá un efecto sobre las componentes principales. Ese cambio de escala podría invertir los papeles de las variables importantes y las no importantes [8].

2. *Las variables deben tener varianzas con tamaños aproximadamente semejantes.*

   Si una de las variables tiene una varianza mucho mayor que las demás, dominará la primera componente principal, sin importar la estructura de las covarianzas de las variables y, en este caso, tiene poco sentido la realización de un ACP.

Cuando no parezca que las variables están ocurriendo sobre un fundamento igual, se aplicará el ACP a las calificaciones Z o a datos estandarizados, en lugar de aplicarlo a los valores de los datos en bruto.

Sea $\underline{X} = (X_1, \ldots, X_p)'$ un vector aleatorio con media $\underline{\mu} = (\mu_1, \ldots, \mu_p)'$ y matriz de varianzas-covarianzas Σ. Estandarizando el vector $\underline{X}$, se tiene:

$$\underline{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_p \end{bmatrix} = \begin{bmatrix} \sigma_{11}^{-1/2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{pp}^{-1/2} \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 \\ \vdots \\ X_p - \mu_p \end{bmatrix} = D^{-1/2}(\underline{X} - \underline{\mu}),$$

donde $D = diag(\sigma_{11}, \cdots, \sigma_{pp})$.

Teorema 2.10. *La i-ésima componente principal del vector $\underline{Z}$ con matriz de covarianzas P viene dada por $Y_i = \underline{\gamma_{(i)}^*}'\underline{Z} = \underline{\gamma_{(i)}^*}'D^{-1/2}(\underline{X} - \underline{\mu})$, $i = 1, \ldots, p$, siendo $\underline{\gamma_{(i)}^*}$ los vectores propios asociados a los valores propios λ_i^* de P, la matriz de correlaciones de $\underline{X}$, cumpliéndose la propiedad de que $\lambda_1^* \geqslant \lambda_2^* \geqslant \cdots \geqslant \lambda_p^* \geqslant 0$, y verificándose además que $\sum_{i=0}^p Var(Y_i) = \sum_{i=0}^p Var(Z_i) = p$.*

Demostración. La media y la matriz de covarianzas de $\underline{Z}$, respectivamente, son:

$$E(\underline{Z}) = E(D^{-1/2}(\underline{X} - \underline{\mu})) = D^{-1/2}E(\underline{X} - \underline{\mu}) = \underline{0}$$

$$\begin{aligned} P &= Var(\underline{Z}) = Var(D^{-1/2}(\underline{X} - \underline{\mu})) = D^{-1/2}Var(\underline{X} - \underline{\mu})D^{-1/2\,\prime} \\ &= D^{-1/2}Var(\underline{X})D^{-1/2} = D^{-1/2}\Sigma D^{-1/2}, \end{aligned}$$

la transformación de componentes principales es la transformación $\underline{Z} \to \underline{Y} = \Gamma^{*\,\prime}(\underline{Z} - \underline{0}) = \Gamma^{*\,\prime}\underline{Z}$, donde Γ^* es una matriz ortogonal cuyas columnas son vectores propios estandarizados de P; $\Gamma^{*\,\prime}P\Gamma^* = \Lambda^* = diag(\lambda_1^*, \lambda_2^*, \cdots, \lambda_p^*)$ es la matriz diagonal de valores propios de P y $\lambda_1^* \geqslant \lambda_2^* \geqslant \ldots \geqslant \lambda_p^* \geqslant 0$.

El i-ésimo componente principal de $\underline{Z}$ se define como el i-ésimo elemento del vector $\underline{Y}$, es decir, como $Y_i = \gamma_{(i)}^{*\,\prime}\underline{Z} = \underline{\gamma_{(i)}^*}'D^{-1/2}(\underline{X} - \underline{\mu})$, $i = 1, \ldots, p$, donde $\gamma_{(i)}^*$ es la i-ésima columna de Γ^*, es decir, el vector propio asociado al valor propio de P, λ_i^*.

La covarianza de las variables aleatorias Z_i y Z_j, $i, j = 1, \ldots, p$ es igual al coeficiente de correlación de X_i y X_j, ya que:

$$\begin{aligned} Cov(Z_i, Z_j) &= E((Z_i - \mu_{Z_i})(Z_j - \mu_{Z_j})) = E(Z_iZ_j) = E\left(\tfrac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}\tfrac{X_j - \mu_j}{\sqrt{\sigma_{jj}}}\right) \\ &= \tfrac{E((X_i - \mu_i)(X_j - \mu_j))}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} = \tfrac{Cov(X_i, X_j)}{\sqrt{Var(X_i)}\sqrt{Var(X_j)}} \\ &= \rho_{X_iX_j}, \end{aligned}$$

por lo tanto la matriz de covarianzas de $\underline{Z}$ coincide con la matriz de correlaciones de $\underline{X}$.

Por último, por el Teorema 2.4, se tiene que $\sum_{i=0}^{p} Var(Y_i) = \sum_{i=0}^{p} Var(Z_i)$ y se verifica que:

$$\sum_{i=0}^{p} Var(Z_i) = tr(P) = tr(D^{-1/2}\Sigma D^{-1/2}) = tr(\Sigma D^{-1/2} D^{-1/2}) = tr(\Sigma D^{-1})$$

el elemento (i, j) de ΣD^{-1} es $(\sigma_{i1}, \cdots, \sigma_{ij}, \cdots, \sigma_{ip})(0, \cdots, \sigma_{jj}^{-1}, \cdots, 0)' = \sigma_{ij}/\sigma_{jj}$ cuando $i = j$ es 1, por lo que $tr(\Sigma D^{-1}) = \sum_{i=1}^{p} 1 = p$. □

Aplicar el ACP a las calificaciones Z o a datos estandarizados es equivalente a aplicar el ACP a la matriz de correlación de las respuestas, en lugar de a la matriz de varianzas-covarianzas [8].

Los valores propios y vectores propios de P son diferentes a los de Σ y no existe simplificación sencilla para llevar uno de los conjuntos de valores hacia el otro.

2.4.1. Calificaciones de las componentes principales

Cuando se realiza el ACP sobre una matriz de correlaciones, las calificaciones de componentes principales deben calcularse a partir de los valores Z. En este caso, la calificación de la j-ésima componente principal para la r-ésima unidad experimental se define por:

$$y_{rj}^{*} = \underline{\gamma_{(j)}^{*}}' \underline{z_r},$$

para $j = 1, 2, \ldots, p$.

2.4.2. Vectores de correlaciones de componentes

Cuando se ha realizado el ACP sobre la matriz de correlaciones, se tiene la siguiente definición:

Definición 2.11. *Sea* $\underline{\beta_{(j)}^{*}} = \lambda_j^{*1/2} \underline{\gamma_{(j)}}^{*}$*, para* $j = 1, \ldots, p$*. Estos nuevos vectores se llaman* vectores de correlaciones de componentes *y los elementos del vector* $\underline{\beta_{(j)}^{*}}$ *reciben*

el nombre de correlaciones de componentes. *En notación matricial se puede escribir* $B^* = (\underline{\beta^*_{(1)}}, \cdots, \underline{\beta^*_{(p)}}) = \Gamma^* \Lambda^{*1/2}$.

2.4.3. Correlación entre las componentes principales y las variables originales

Teorema 2.12. *Los elementos de* $\underline{\beta^*_{(j)}}$ *dan las correlaciones entre las variables originales y la j-ésima variable componente principal.*

Demostración. Por el Corolario 2.9, se sigue que si $\underline{Y}$ es el vector aleatorio de componentes principales del vector aleatorio $\underline{Z} \sim (\underline{0}, P)$, la correlación de la i-ésima variable original y la j-ésima componente principal es $\rho_{Z_i Y_j} = \frac{\gamma^*_{ij}\sqrt{\lambda^*_j}}{\sqrt{Var(Z_i)}} = \frac{\gamma^*_{ij}\sqrt{\lambda^*_j}}{\sqrt{1}} = \gamma^*_{ij}\sqrt{\lambda^*_j}$.

El vector de las correlaciones entre las variables originales y la j-ésima variable componente principal es $(\rho_{Z_1 Y_j}, \cdots, \rho_{Z_p Y_j})' = (\gamma^*_{1j}\sqrt{\lambda^*_j}, \cdots, \gamma^*_{pj}\sqrt{\lambda^*_j})' = \sqrt{\lambda^*_j}(\gamma^*_{1j}, \cdots, \gamma^*_{pj})'$ $= \sqrt{\lambda^*_j}\underline{\gamma_{(j)}}^* = \underline{\beta^*_{(j)}}$. □

Sólo debe efectuarse ACP cuando el determinante de la matriz de correlaciones sea cercano a cero, porque esto indica que existen dependencias lineales entre las variables respuesta.

Elegir que se analice P en lugar de Σ significa decidir que todas las variables respuesta tienen igual importancia y no es correcto establecer esta hipótesis de manera arbitraria. Pueden existir situaciones en que las variables no están en unidades comparables y en las que el investigador preferiría no tratar todas las variables como si tuvieran igual importancia. Algunos paquetes para cálculos estadísticos permitirán la asignación de un peso a cada variable. Los datos podrían estandarizarse y, a continuación, otorgar pesos mayores a las variables que se piensa tienen más importancia.

2.5. Componentes principales muestrales

En el desarrollo precedente se supuso que se conocían tanto μ como Σ, lo cual no es frecuente, de modo que se requerirá una estimación de μ y Σ a partir de los datos de

la muestra.

Supongamos que se dispone de una muestra aleatoria de tamaño n de una población $\underline{X} = (X_1, \ldots, X_p)'$ con vector de medias $\underline{\mu}$ y matriz de covarianzas Σ desconocida, entonces la matriz de datos es $\mathbb{X} = (\underline{x_1}', \ldots, \underline{x_n}')'$, donde $\underline{x_i}' = (x_{i1}, \ldots, x_{ip})$, $i = 1, \ldots, n$. Los estimadores de $\underline{\mu}$ y Σ son $\overline{\underline{x}}$ y S, respectivamente.

Definición 2.13. *La media muestral de la i-ésima variable es* $\overline{x}_i = \frac{1}{n}\sum_{r=1}^{n} x_{ri}$ *y el vector que contiene a las medias muestrales de* $\underline{x_1}$*, ...,* $\underline{x_p}$ *es:*

$$\overline{\underline{x}} = (\overline{x}_1, \cdots, \overline{x}_p)' = \frac{1}{n}\sum_{r=1}^{n} \underline{x_r} = \frac{1}{n}\mathbb{X}'\underline{1},$$

denominado el vector de medias muestral o simplemente vector de medias.

Definición 2.14. *La matriz de covarianza muestral o simplemente la matriz de covarianza es la matriz:*

$$S = \frac{1}{n}\sum_{r=1}^{n}(\underline{x_r} - \overline{\underline{x}})(\underline{x_r} - \overline{\underline{x}})' = \frac{1}{n}\left(\mathbb{X}'\mathbb{X} - \frac{1}{n}\mathbb{X}'\underline{1}\underline{1}'\mathbb{X}\right) = \frac{1}{n}\mathbb{X}'H\mathbb{X},$$

donde $H = I - \frac{1}{n}\underline{1}\underline{1}'$.

En esta matriz, s_{ii} *es la varianza muestral de la i-ésima variable y* s_{ij} *es la covarianza muestral de las variables i-ésima y j-ésima:*

$$s_{ii} = n^{-1}\sum_{r=1}^{n}(x_{ri} - \overline{x}_i)^2,$$

$$s_{ij} = n^{-1}\sum_{r=1}^{n}(x_{ri} - \overline{x}_i)(x_{rj} - \overline{x}_j).$$

La matriz H es una matriz idempotente simétrica, pues $H' = (I - n^{-1}\underline{1}\underline{1}')' = I - n^{-1}\underline{1}\underline{1}' = H$ y $H^2 = (I - n^{-1}\underline{1}\underline{1}')(I - n^{-1}\underline{1}\underline{1}') = I - n^{-1}\underline{1}\underline{1}' - n^{-1}\underline{1}\underline{1}' + n^{-1}\underline{1}\underline{1}'n^{-1}\underline{1}\underline{1}' = I - 2n^{-1}\underline{1}\underline{1}' + n^{-2}\underline{1}\underline{1}'\underline{1}\underline{1}' = I - 2n^{-1}\underline{1}\underline{1}' + n^{-1}\underline{1}\underline{1}' = I - n^{-1}\underline{1}\underline{1}' = H$.

El objetivo, como en el caso teórico, es explicar el mayor porcentaje posible de variación de la muestra con combinaciones lineales no correlacionadas de las variables con varianzas máximas [14].

Una combinación lineal para la muestra $\underline{x_1}, \ldots, \underline{x_n}$ viene dada por:

$$\begin{aligned} l_{1i}x_{r1} + l_{2i}x_{r2} + \cdots + l_{pi}x_{rp} &= (l_{1i}, l_{2i}, \cdots, l_{pi})(x_{r1}, x_{r2}, \cdots, x_{rp})' \\ &= \underline{l_{(i)}}'\underline{x_r},\ r = 1, 2, \ldots, n. \end{aligned}$$

Las n observaciones de la variable definida como $\underline{l_{(i)}}'\underline{X}$ son:

$$\begin{bmatrix} \underline{l_{(i)}}'\underline{x_1} \\ \vdots \\ \underline{l_{(i)}}'\underline{x_n} \end{bmatrix} = \begin{bmatrix} \underline{x_1}'\underline{l_{(i)}} \\ \vdots \\ \underline{x_n}'\underline{l_{(i)}} \end{bmatrix} = \begin{bmatrix} \underline{x_1}' \\ \ldots \\ \underline{x_n}' \end{bmatrix} l_{(i)} = (\underline{x_1}', \ldots, \underline{x_n}')'\underline{l_{(i)}} = \mathbb{X}\underline{l_{(i)}}.$$

La media muestral y la varianza muestral de esta nueva variable son:

$$\overline{\mathbb{X}\underline{l_{(i)}}} = \frac{1}{n}(\mathbb{X}\underline{l_{(i)}})'\underline{1} = \frac{1}{n}\underline{l_{(i)}}'\mathbb{X}'\underline{1} = \underline{l_{(i)}}'\frac{1}{n}\mathbb{X}'\underline{1} = \underline{l_{(i)}}'\overline{\underline{x}}$$

$$S_{\mathbb{X}\underline{l_{(i)}}} = \frac{1}{n}(\mathbb{X}\underline{l_{(i)}})'H(\mathbb{X}\underline{l_{(i)}}) = \underline{l_{(i)}}'\left(\frac{1}{n}\mathbb{X}'H\mathbb{X}\right)\underline{l_{(i)}} = \underline{l_{(i)}}'S\underline{l_{(i)}},$$

la covarianza muestral de $\mathbb{X}\underline{l_{(i)}}$ y $\mathbb{X}\underline{l_{(j)}}$ es:

$$\begin{aligned} Cov(\mathbb{X}\underline{l_{(i)}}, \mathbb{X}\underline{l_{(j)}}) &= n^{-1}\textstyle\sum_{r=1}^{n}(\underline{l_{(i)}}'\underline{x_r} - \overline{\mathbb{X}\underline{l_{(i)}}})(\underline{l_{(j)}}'\underline{x_r} - \overline{\mathbb{X}\underline{l_{(j)}}}) \\ &= n^{-1}\textstyle\sum_{r=1}^{n}(\underline{l_{(i)}}'\underline{x_r} - \underline{l_{(i)}}'\overline{\underline{x}})(\underline{l_{(j)}}'\underline{x_r} - \underline{l_{(i)}}'\overline{\underline{x}})' \\ &= n^{-1}\textstyle\sum_{r=1}^{n}\underline{l_{(i)}}'(\underline{x_r} - \overline{\underline{x}})(\underline{x_r} - \overline{\underline{x}})'\underline{l_{(j)}} \\ &= \underline{l_{(i)}}'n^{-1}\textstyle\sum_{r=1}^{n}(\underline{x_r} - \overline{\underline{x}})(\underline{x_r} - \overline{\underline{x}})'\underline{l_{(j)}} \\ &= \underline{l_{(i)}}'S\underline{l_{(j)}}, \end{aligned}$$

donde $\overline{\underline{x}}$ y S son la media muestral y la varianza muestral de $\mathbb{X}$, respectivamente.

La *primera componente principal muestral* es la combinación lineal $\underline{l_{(1)}}'\underline{X}$, tal que al considerar sus n valores sobre la muestra, $\mathbb{X}\underline{l_{(1)}}$, éstos hacen máxima la varianza $S_{\mathbb{X}\underline{l_{(1)}}} = \underline{l_{(1)}}'S\underline{l_{(1)}}$ sujeto a la restricción $\underline{l_{(1)}}'\underline{l_{(1)}} = 1$.

La *segunda componente principal muestral* es la combinación lineal $\underline{l_{(2)}}'\underline{X}$, tal que al considerar sus n valores sobre la muestra, $\mathbb{X}\underline{l_{(2)}}$, éstos hacen máxima la varianza $S_{\mathbb{X}\underline{l_{(2)}}} = \underline{l_{(2)}}'S\underline{l_{(2)}}$ sujeto a la restricción $\underline{l_{(2)}}'\underline{l_{(2)}} = 1$ y que no esté correlacionada con la anterior, es decir, $Cov(\underline{l_{(1)}}'\underline{X}, \underline{l_{(2)}}'\underline{X}) = \underline{l_{(1)}}'S\underline{l_{(2)}} = 0$.

La *j+1-ésima componente principal muestral* es la combinación lineal $\underline{l_{(j+1)}}'\underline{X}$, tal que al considerar sus n valores sobre la muestra, $\mathbb{X}\underline{l_{(j+1)}}$, éstos hacen máxima la varianza $S_{\mathbb{X}\underline{l_{(j+1)}}} = \underline{l_{(j+1)}}'S\underline{l_{(j+1)}}$ sujeto a la restricción $\underline{l_{(j+1)}}'\underline{l_{(j+1)}} = 1$ y que no esté correlacionada con las anteriores, es decir, $Cov(\underline{l_{(i)}}'\underline{X}, \underline{l_{(j+1)}}'\underline{X}) = \underline{l_{(i)}}'S\underline{l_{(j+1)}} = 0$, $i = 1, \ldots, j$.

2.5.1. Construcción conjunta de las componentes principales muestrales

Las componentes principales muestrales se obtendrán de manera análoga a las componentes principales poblacionales.

Dado que S, la matriz de covarianza de $\mathbb{X}$, es simétrica, por el teorema de la descomposición espectral (Teorema B.7) se cumple que:

$$S = GLG',$$

donde L es una matriz diagonal de valores propios de S, y G es una matriz ortogonal cuyas columnas son vectores propios estandarizados.

Como G es una matriz ortogonal, cumple $G'G = I$, lo cual equivale a que $\|\underline{g_{(i)}}\| = \sqrt{\underline{g_{(i)}}'\underline{g_{(i)}}} = 1$, $i = 1, 2, \ldots, p$ y $\underline{g_{(i)}}'\underline{g_{(j)}} = 0$, $i \neq j$, $i, j = 1, 2, \ldots, p$; como $\underline{g_{(i)}}$ representa la columna i de G, es decir, un vector propio de S, equivale a que los vectores propios son unitarios y ortogonales. Luego las n observaciones de la variable definida como $\underline{g_{(i)}}'\underline{X}$ son $\mathbb{X}\underline{g_{(i)}}$.

Se propone:

$$\underline{Y} = \begin{bmatrix} \underline{g_{(1)}}'\underline{X} \\ \vdots \\ \underline{g_{(p)}}'\underline{X} \end{bmatrix} = \begin{bmatrix} \underline{g_{(1)}}' \\ \vdots \\ \underline{g_{(p)}}' \end{bmatrix} \underline{X} = G'\underline{X},$$

las n observaciones para el vector aleatorio $\underline{Y}$ son:

$$\mathbb{Y} = (\underline{\mathbb{Y}_{(1)}}, \cdots, \underline{\mathbb{Y}_{(p)}}) = (\mathbb{X}\underline{g_{(1)}}, \cdots, \mathbb{X}\underline{g_{(p)}}) = \mathbb{X}(\underline{g_{(1)}}, \cdots, \underline{g_{(p)}}) = \mathbb{X}G,$$

cuya varianza muestral está dada por:

$$\begin{aligned} S_{\mathbb{Y}} &= \tfrac{1}{n}\mathbb{Y}'H\mathbb{Y} = \tfrac{1}{n}(\mathbb{X}G)'H\mathbb{X}G = \tfrac{1}{n}G'\mathbb{X}'H\mathbb{X}G = G'\tfrac{1}{n}\mathbb{X}'H\mathbb{X}G = G'SG \\ &= G'GLG'G = L, \end{aligned}$$

por lo que la varianza muestral de $\mathbb{X}\underline{g_{(i)}}$ es $S_{\mathbb{Y}ii} = \widehat{\lambda}_i$, $i = 1, \ldots, p$, y la covarianza muestral de $\mathbb{X}\underline{g_{(i)}}$ y $\mathbb{X}\underline{g_{(j)}}$ es $S_{\mathbb{Y}ij} = 0$, $i \neq j$, $i, j = 1, \ldots, p$, es decir, $\mathbb{X}\underline{g_{(i)}}$ y $\mathbb{X}\underline{g_{(j)}}$ no están correlacionadas.

Dado que $S' = \frac{1}{n}\mathbb{X}'H\mathbb{X} = S$ y dado $\underline{w} \in \mathbb{R}^p - \{\underline{0}\}$, se sigue que:

$$\begin{aligned} \underline{w}'S\underline{w} &= \underline{w}'\tfrac{1}{n}\textstyle\sum_{r=1}^{n}(\underline{x_r} - \overline{\underline{x}})(\underline{x_r} - \overline{\underline{x}})'\underline{w} = \tfrac{1}{n}\textstyle\sum_{r=1}^{n}(\underline{w}'(\underline{x_r} - \overline{\underline{x}})(\underline{x_r} - \overline{\underline{x}})'\underline{w}) \\ &= \tfrac{1}{n}\textstyle\sum_{r=1}^{n}(\underline{w}'(\underline{x_r} - \overline{\underline{x}}))(\underline{w}'(\underline{x_r} - \overline{\underline{x}}))' = \tfrac{1}{n}\textstyle\sum_{r=1}^{n}(\underline{w}'(\underline{x_r} - \overline{\underline{x}}))^2 \geqslant 0, \end{aligned}$$

por lo que S es semidefinida positiva.

Dado que S es semidefinida positiva y es de orden p, por el Teorema B.6 se sigue que los p valores propios de S son mayores o iguales a cero, es decir, $\widehat{\lambda_1} \geqslant \widehat{\lambda_2} \geqslant \ldots \geqslant \widehat{\lambda_p} \geqslant 0$. Así, $S_{\mathbb{Y}ii} = \widehat{\lambda_i}$ está bien definida.

Hasta el momento la covarianza muestral de $\mathbb{X}\underline{g_{(1)}}$, ..., $\mathbb{X}\underline{g_{(p)}}$ es cero, por lo que no están correlacionadas; sin embargo, $\mathbb{X}\underline{g_{(1)}}$ no necesariamente tiene varianza muestral máxima, y se tiene que $S_{\mathbb{Y}11}$ es igual a un valor propio de S, por lo que se define:

$$\underline{\mathbb{Y}_{(1)}} = \mathbb{X}\widehat{\underline{\gamma_{(1)}}},$$

donde $\widehat{\underline{\gamma_{(1)}}}$ es un vector propio unitario de $\widehat{\lambda_1}$, el valor propio máximo, de este modo, $\mathbb{X}\underline{g_{(1)}}$ sí tiene varianza muestral máxima. Luego $\mathbb{X}\underline{g_{(2)}}$ debe tener la mayor varianza muestral restante; $S_{\mathbb{Y}22}$ también es igual a un valor propio de S, por lo que se define:

$$\underline{\mathbb{Y}_{(2)}} = \mathbb{X}\widehat{\underline{\gamma_{(2)}}},$$

donde $\widehat{\underline{\gamma_{(2)}}}$ es un vector propio unitario de $\widehat{\lambda_2}$, el segundo valor propio máximo. Por último, como $\mathbb{X}\underline{g_{(j+1)}}$ no está correlacionada con $\mathbb{X}\underline{g_{(1)}}$, ..., $\mathbb{X}\underline{g_{(j)}}$ y se desea maximizar la varianza muestral $S_{\mathbb{Y}j+1\ j+1}$ que es igual a un valor propio de S, si se define a:

$$\underline{\mathbb{Y}_{(j+1)}} = \mathbb{X}\widehat{\underline{\gamma_{(j+1)}}},$$

donde $\widehat{\underline{\gamma_{(j+1)}}}$ es un vector unitario de $\widehat{\lambda_{j+1}}$, el (j+1)-ésimo valor propio máximo, se está maximizando la varianza muestral restante de las $\underline{\mathbb{Y}_{(1)}}$, ..., $\underline{\mathbb{Y}_{(j)}}$.

Por lo tanto, las componentes principales muestrales quedan definidas de manera conjunta como:

$$\mathbb{Y} = (\underline{\mathbb{Y}_{(1)}}, \cdots, \underline{\mathbb{Y}_{(p)}}) = (\mathbb{X}\widehat{\underline{\gamma_{(1)}}}, \cdots, \mathbb{X}\widehat{\underline{\gamma_{(p)}}}) = \mathbb{X}(\widehat{\underline{\gamma_{(1)}}}, \cdots, \widehat{\underline{\gamma_{(p)}}}) = \mathbb{X}G,$$

donde G es la matriz que tiene a los vectores propios unitarios $\widehat{\underline{\gamma_{(1)}}}$, $\widehat{\underline{\gamma_{(2)}}}$, ..., $\widehat{\underline{\gamma_{(p)}}}$ correspondientes a $\widehat{\lambda_1}$, $\widehat{\lambda_2}$, ..., $\widehat{\lambda_p}$ como columnas ($G = (\widehat{\underline{\gamma_{(1)}}}, \widehat{\underline{\gamma_{(2)}}}, \ldots, \widehat{\underline{\gamma_{(p)}}})$).

2.5.2. Definición de las componentes principales muestrales

Definición 2.15. *Si $\mathbb{X} = (\underline{x_1}', \ldots, \underline{x_n}')'$ es una muestra aleatoria de tamaño n de una población $\underline{X} = (X_1, \ldots, X_p)'$ con vector de medias $\underline{\mu}$ y matriz de covarianzas Σ desconocida, y $\underline{\overline{x}}$ y S son el vector de medias muestral y la matriz de covarianza muestral, respectivamente, donde $\underline{x_i}' = (x_{i1}, \ldots, x_{ip})$, $i = 1, \ldots, n$, la transformación de componentes principales:*

$$\mathbb{X} \to \mathbb{Y} = (\mathbb{X} - \underline{1}\,\underline{\overline{x}}')G, \tag{2.16}$$

donde $\underline{1}_{p\times 1}$ es un vector de 1's; G es una matriz ortogonal cuyas columnas son vectores propios estandarizados de S; $G'SG = L = diag(\widehat{\lambda}_1, \widehat{\lambda}_2, \cdots, \widehat{\lambda}_p)$, donde los λ_i's, son los valores propios de S y $\widehat{\lambda}_1 \geqslant \widehat{\lambda}_2 \geqslant \ldots \geqslant \widehat{\lambda}_p \geqslant 0$. La positividad estricta de los valores propios $\widehat{\lambda}_i$ se garantiza si S es definida positiva.

La i-ésima componente principal muestral de $\mathbb{X}$ puede definirse como la i-ésima columna de $\mathbb{Y}$, es decir, como:

$$\underline{\mathbb{Y}_{(i)}} = (\mathbb{X} - \underline{1}\,\underline{\overline{x}}')\underline{\widehat{\gamma_{(i)}}}, \tag{2.17}$$

donde $\widehat{\gamma_{(i)}}$ es la i-ésima columna de G.

2.5.3. Calificaciones de las componentes muestrales

El r-ésimo elemento de $\mathbb{Y}_{(i)}$, $\mathbb{Y}_{ri}$, representa la puntuación-calificación (*score*) de la i-ésima componente principal en el r-ésimo individuo. En términos de los individuos, se puede escribir la transformación de componentes principales como: $\mathbb{Y}_{ri} = (\underline{x_r} - \underline{\overline{x}})'\widehat{\gamma_{(i)}} = \widehat{\gamma_{(i)}}'(\underline{x_r} - \underline{\overline{x}})$, para $i = 1, 2, \ldots, p$; $r = 1, 2, \ldots, n$.

Frecuentemente se omite el subíndice r, $\mathbb{Y}_i = g'_{(i)}(x - \overline{x})$ para enfatizar la transformación en lugar de su efecto sobre algún individuo particular.

2.5.4. Propiedades de las componentes principales muestrales

Las componentes principales muestrales satisfacen la extensión muestral directa de los Teoremas 2.4-2.6

Teorema 2.16. *Si $\mathbb{X}$ es una matriz de datos muestrales de una población $\underline{X} \sim (\underline{\mu}, \Sigma)$ e $\mathbb{Y}$ se define como en la Ecuación (2.16), entonces:*

i) $\overline{\mathbb{Y}} = \underline{0}$

ii) $S_{\mathbb{Y}} = L$

iii) $S_{\mathbb{Y}11} \geqslant S_{\mathbb{Y}22} \geqslant \ldots \geqslant S_{\mathbb{Y}pp} \geqslant 0$

iv) $\sum_{i=1}^{p} S_{\mathbb{Y}ii} = tr(S)$

v) $\prod_{i=1}^{p} S_{\mathbb{Y}ii} = |S|$

Demostración.

i) $\overline{\mathbb{Y}} = n^{-1}\mathbb{Y}'\underline{1} = n^{-1}((\mathbb{X} - \underline{1}\ \overline{\underline{x}}')G)'\underline{1} = G'n^{-1}(\mathbb{X} - \underline{1}\ \overline{\underline{x}}')'\underline{1} = G'\left(n^{-1}\mathbb{X}' - n^{-1}\overline{\underline{x}}\ \underline{1}'\right)\underline{1} = G'\left(n^{-1}\mathbb{X}'\underline{1} - n^{-1}\overline{\underline{x}}\ \underline{1}'\underline{1}\right) = G'(\overline{\underline{x}} - \overline{\underline{x}}) = \underline{0}.$

ii)

$$\begin{aligned} S_{\mathbb{Y}} &= n^{-1}\mathbb{Y}'H\mathbb{Y} \\ &= n^{-1}G'(\mathbb{X} - \underline{1}\ \overline{\underline{x}}')'H(\mathbb{X} - \underline{1}\ \overline{\underline{x}}')G \\ &= n^{-1}G'(\mathbb{X}' - \overline{\underline{x}}\ \underline{1}')(H\mathbb{X} - H\underline{1}\ \overline{\underline{x}}')G \\ &= G'n^{-1}(\mathbb{X}'H\mathbb{X} - \mathbb{X}'H\underline{1}\ \overline{\underline{x}}' - \overline{\underline{x}}\ \underline{1}'H\mathbb{X} + \overline{\underline{x}}\ \underline{1}'H\underline{1}\ \overline{\underline{x}}')G, \end{aligned}$$

además, como $\overline{\underline{x}} = n^{-1}\mathbb{X}'\underline{1}$ y $\overline{\underline{x}}' = n^{-1}\underline{1}'\mathbb{X}$,

$$\begin{aligned} S_{\mathbb{Y}} &= G'n^{-1}(\mathbb{X}'H\mathbb{X})G - G'n^{-1}(\mathbb{X}'H\underline{1}\ n^{-1}\underline{1}'\mathbb{X} + n^{-1}\mathbb{X}'\underline{1}\ \underline{1}'H\mathbb{X} \\ &\quad -n^{-1}\mathbb{X}'\underline{1}\ \underline{1}'H\underline{1}\ n^{-1}\underline{1}'\mathbb{X})G \\ &= G'(n^{-1}\mathbb{X}'H\mathbb{X})G - G'n^{-2}\mathbb{X}'(H\underline{1}\underline{1}'\mathbb{X} + \underline{1}\ \underline{1}'H\mathbb{X} - \underline{1}\ \underline{1}'H\underline{1}\ n^{-1}\underline{1}'\mathbb{X})G \\ &= G'SG - G'n^{-2}\mathbb{X}'(H\underline{1}\underline{1}' + \underline{1}\ \underline{1}'H - n^{-1}\underline{1}\ \underline{1}'H\underline{1}\underline{1}')\mathbb{X}G \\ &= G'SG - G'n^{-2}\mathbb{X}'(H\underline{1}\underline{1}' + (H\underline{1}\underline{1}')' - n^{-1}\underline{1}\ \underline{1}'H\underline{1}\underline{1}')\mathbb{X}G \end{aligned}$$

recordemos que $H = I - n^{-1}\underline{1}\underline{1}'$, así, $H\underline{1}\underline{1}' = (I - n^{-1}\underline{1}\underline{1}')\underline{1}\underline{1}' = \underline{1}\underline{1}' - \underline{1}\underline{1}' = 0$, luego:

$$S_{\mathbb{Y}} = G'SG + G'n^{-2}\mathbb{X}'(-0 - (0)' + n^{-1}\underline{1}\ \underline{1}'0)\mathbb{X}G = G'SG = G'GLG'G = L.$$

Esto equivale a:

- $S_{\mathbb{Y}ii} = \widehat{\lambda}_i$

- $S_{\mathbb{Y}ij} = 0$ para $i \neq j$.

III) De la Definición 2.15 se tiene que $\widehat{\lambda_1} \geqslant \widehat{\lambda_2} \geqslant \ldots \geqslant \widehat{\lambda_p} \geqslant 0$, además, como $S_{\mathbb{Y}ii} = \widehat{\lambda_i}$, se sigue que $S_{\mathbb{Y}11} \geqslant S_{\mathbb{Y}22} \geqslant \ldots \geqslant S_{\mathbb{Y}pp} \geqslant 0$.

IV) De la Definición 2.15 se tiene $G'SG = L$, de donde se sigue que $S = GLG'$, luego:

$$tr(S) = tr(GLG') = tr(LG'G) = tr(L) = \sum_{i=1}^{p} \widehat{\lambda_i} = \sum_{i=1}^{p} S_{\mathbb{Y}ii}.$$

V) $|S| = |GLG'| = |G||LG'| = |LG'||G| = |LG'G| = |L| = \prod_{i=1}^{p} \widehat{\lambda_i} = \prod_{i=1}^{p} S_{\mathbb{Y}ii}$. □

Se mostrará un resultado significativo:

Teorema 2.17. *Si $\mathbb{X}$ es una matriz de datos muestrales de una población $\underline{X} \sim (\underline{\mu}, \Sigma)$, se tiene que:*

1. *Ninguna CLE de las variables observadas tiene varianza muestral mayor que $\widehat{\lambda_1}$, la varianza muestral de la primera componente principal muestral.*
2. *Ninguna CLE de las variables observadas tiene varianza muestral menor $\widehat{\lambda_p}$, la varianza muestral de la última componente principal muestral.*

Demostración. Sea $\underline{a}'\underline{X}$ una CLE, es decir, $\underline{a}'\underline{a} = 1$. Se puede escribir a:

$$\underline{a} = c_1\widehat{\underline{\gamma_{(1)}}} + \ldots + c_p\widehat{\underline{\gamma_{(p)}}} = G\underline{c}, \tag{2.18}$$

donde $\underline{c} = (c_1, \ldots, c_p)'$, $G = (\widehat{\underline{\gamma_{(1)}}}, \cdots, \widehat{\underline{\gamma_{(p)}}})$ y $\widehat{\underline{\gamma_{(1)}}}$, ...,$\widehat{\underline{\gamma_{(p)}}}$ son los vectores propios de S (cualquier vector puede ser escrito de esta forma, ya que los vectores propios forman una base para $\mathbb{R}^p$). Las observaciones de la variable $\underline{a}'\underline{X}$ son $\mathbb{X}\underline{a}$, cuya varianza muestral es:

$$S_{\mathbb{X}\underline{a}} = \underline{a}'S\underline{a},$$

donde S es la matriz de covarianza muestral de $\mathbb{X}$. Usando el teorema de la descomposición espectral (Teorema B.7), se cumple que $S = GLG'$, por lo que:

$$S_{\mathbb{X}\underline{a}} = \underline{a}'S\underline{a} = \underline{c}'G'SG\underline{c} = \underline{c}'\underbrace{G'G}_{I}L\underbrace{G'G}_{I}\underline{c} = \underline{c}'L\underline{c} = \sum_{i=1}^{p} \widehat{\lambda_i}c_i^2, \tag{2.19}$$

Además,

$$1 = \underline{a}'\underline{a} = \underline{c}'\underbrace{G'G}_{I}\underline{c} = \underline{c}'\underline{c} = \sum_{i=1}^{p} c_i^2 \tag{2.20}$$

Dado que S es definida positiva y es de orden p, por el Teorema B.6 se sigue que los p valores propios de S son mayores a cero, es decir, $\widehat{\lambda}_1 \geqslant \widehat{\lambda}_2 \geqslant \ldots \geqslant \widehat{\lambda}_p > 0$. Por lo tanto, de (2.19) y (2.20) se sigue que:

$$\widehat{\lambda}_1 = \widehat{\lambda}_1 \sum_{i=1}^{p} c_i^2 = \sum_{i=1}^{p} \widehat{\lambda}_1 c_i^2 \geqslant S_{\mathbb{X}\underline{a}} \geqslant \sum_{i=1}^{p} \widehat{\lambda}_p c_i^2 = \widehat{\lambda}_p \sum_{i=1}^{p} c_i^2 = \widehat{\lambda}_p$$

□

Las componentes principales muestrales intermedias tienen una propiedad de varianza máxima dada por el siguiente teorema:

Teorema 2.18. *Si $\mathbb{X}$ es una matriz de datos muestrales de una población $\underline{X} \sim (\mu, S)$, se tiene que si $\alpha = \mathbb{X}\underline{a}$ es una CLE de $\mathbb{X}$ que no está correlacionada con las primeras k componentes principales muestrales de $\mathbb{X}$, entonces la varianza muestral de α es maximizada cuando α es la (k+1)-ésima componente principal muestral de $\mathbb{X}$, $k < p$.*

Demostración. Se puede escribir a $\underline{a} = c_1\widehat{\underline{\gamma_{(1)}}} + \ldots + c_p\widehat{\underline{\gamma_{(p)}}} = G\underline{c}$ (Ecuación (2.18)), donde $\underline{c} = (c_1, \ldots, c_p)'$, $G = (\widehat{\underline{\gamma_{(1)}}}, \cdots, \widehat{\underline{\gamma_{(p)}}})$ y $\widehat{\underline{\gamma_{(1)}}}$, $\ldots$,$\widehat{\underline{\gamma_{(p)}}}$ son los vectores propios correspondientes a los valores propios $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_p$ de S, la matriz de covarianza muestral de $\mathbb{X}$, es decir, se cumple que $S\widehat{\underline{\gamma_{(i)}}} = \widehat{\lambda}_i\widehat{\underline{\gamma_{(i)}}}$. Dado que α no está correlacionada con $\widehat{\underline{\gamma_{(i)}}}'\underline{X}$, $i = 1, \ldots, k$,

$$0 = Cov(\mathbb{X}\underline{a}, \mathbb{X}\widehat{\underline{\gamma_{(i)}}}) = \underline{a}'S\widehat{\underline{\gamma_{(i)}}} = \underline{a}'\widehat{\lambda}_i\widehat{\underline{\gamma_{(i)}}} = \widehat{\lambda}_i\underline{a}'\widehat{\underline{\gamma_{(i)}}},$$

además, como S es definida positiva, se tiene que $\widehat{\lambda}_1 \geqslant \ldots \geqslant \widehat{\lambda}_p > 0$, por lo cual $\underline{a}'\widehat{\underline{\gamma_{(i)}}} = 0$, $i = 1, \ldots, k$, luego:

$$\begin{aligned}
0 &= \underline{a}'\widehat{\underline{\gamma_{(i)}}} = (c_1\widehat{\underline{\gamma_{(1)}}} + \ldots + c_i\widehat{\underline{\gamma_{(i)}}} + \ldots + c_p\widehat{\underline{\gamma_{(p)}}})'\widehat{\underline{\gamma_{(i)}}} \\
&= (c_1\widehat{\underline{\gamma_{(1)}}}' + \ldots + c_i\widehat{\underline{\gamma_{(i)}}}' + \ldots + c_p\widehat{\underline{\gamma_{(p)}}}')\widehat{\underline{\gamma_{(i)}}} \\
&= (c_1\underbrace{\widehat{\underline{\gamma_{(1)}}}'\widehat{\underline{\gamma_{(i)}}}}_{0} + \ldots + c_i\underbrace{\widehat{\underline{\gamma_{(i)}}}'\widehat{\underline{\gamma_{(i)}}}}_{1} + \ldots + c_p\underbrace{\widehat{\underline{\gamma_{(p)}}}'\widehat{\underline{\gamma_{(i)}}}}_{0}) = c_i \text{ para } i = 1, \ldots, k,
\end{aligned}$$

se sigue que:

$$\begin{aligned} S_{\mathbb{X}\underline{a}} &= \underline{a}'S\underline{a} = \underline{c}'G'SG\underline{c} = \underline{c}'\underbrace{G'G}_{I}L\underbrace{G'G}_{I}\underline{c} = \underline{c}'L\underline{c} = \textstyle\sum_{i=1}^{p}\widehat{\lambda}_i c_i^2 \\ &= \textstyle\sum_{i=k+1}^{p}\widehat{\lambda}_i c_i^2 \leqslant \sum_{i=k+1}^{p}\widehat{\lambda_{k+1}}c_i^2 = \widehat{\lambda_{k+1}}\sum_{i=k+1}^{p}c_i^2 = \widehat{\lambda_{k+1}}, \end{aligned}$$

y la $S_{\mathbb{X}\underline{l}_{(k+1)}} = \widehat{\lambda_{k+1}}$, por lo tanto la varianza muestral de α es maximizada cuando α es la (k+1)-ésima componente principal muestral de $\mathbb{X}$. □

2.5.5. Covarianza muestral y correlación muestral entre las componentes principales muestrales y las variables originales

Se tiene que $\mathbb{X} = (\underline{x_{(1)}}, \cdots, \underline{x_{(p)}})$, donde $\underline{x_{(i)}} = (x_{1i}, \cdots, x_{ni})' = \mathbb{X}\underline{e_i}$ son los valores observados en la muestra para la variable X_i, $\underline{e_i}_{p\times 1} = (0, \cdots, 0, \underbrace{1}_{i}, 0, \cdots, 0)'$, $\overline{\underline{x_i}} = n^{-1}(\mathbb{X}\underline{e_i})'\underline{1} = \underline{e_i}'n^{-1}\mathbb{X}'\underline{1} = \underline{e_i}'\underline{\overline{x}}$ y $x_{ri} = \underline{e_{(i)}}'\underline{x_r}$ es el valor de la i-ésima variable en el r-ésimo individuo.

Teorema 2.19. *Sea $\mathbb{Y}$ la matriz de componentes principales de una muestra $\mathbb{X}$ de una población con vector aleatorio $\underline{X} \sim (\mu, \Sigma)$. El i-ésimo elemento de $\widehat{\lambda_j}\widehat{\underline{\gamma_{(j)}}}$ da la covarianza muestral entre las observaciones de la i-ésima variable original y la j-ésima componente principal muestral.*

Demostración. Recordemos que $\underline{\mathbb{Y}_{(j)}} = (\mathbb{X} - \underline{1}\,\underline{\overline{x}}')\widehat{\underline{\gamma_{(j)}}}$, $\mathbb{Y}_{rj} = (\underline{x_r} - \underline{\overline{x}})'\widehat{\gamma_{(j)}} = \widehat{\gamma_{(j)}}'(\underline{x_r} - \underline{\overline{x}})$ y $S = GLG'$,

$$\begin{aligned} Cov(\underline{x_{(i)}}, \underline{\mathbb{Y}_{(j)}}) &= n^{-1}\textstyle\sum_{r=1}^{n}(x_{ri} - \overline{x}_i)(\mathbb{Y}_{rj} - \overline{\mathbb{Y}}_j) \\ &= n^{-1}\textstyle\sum_{r=1}^{n}((\underline{e_{(i)}}'\underline{x_r} - \underline{e_i}'\underline{\overline{x}})(\underline{x_r} - \underline{\overline{x}})'\widehat{\gamma_{(j)}}) \\ &= n^{-1}\textstyle\sum_{r=1}^{n}(\underline{e_{(i)}}'(\underline{x_r} - \underline{\overline{x}})(\underline{x_r} - \underline{\overline{x}})'\widehat{\gamma_{(j)}}) \\ &= n^{-1}\textstyle\sum_{r=1}^{n}(\underline{e_{(i)}}'(\underline{x_r} - \underline{\overline{x}})(\underline{x_r} - \underline{\overline{x}})'\widehat{\gamma_{(j)}}) \\ &= \underline{e_{(i)}}'n^{-1}\textstyle\sum_{r=1}^{n}((\underline{x_r} - \underline{\overline{x}})(\underline{x_r} - \underline{\overline{x}})')\widehat{\gamma_{(j)}} \\ &= \underline{e_{(i)}}'S\widehat{\underline{\gamma_{(j)}}}. \end{aligned}$$

Luego la covarianza muestral de $\underline{x_{(1)}}, \cdots, \underline{x_{(p)}}$ con $\underline{\mathbb{Y}_{(j)}}$ juntas son:

$$(\underline{e_{(1)}}'S\widehat{\underline{\gamma_{(j)}}}, \cdots, \underline{e_{(p)}}'S\widehat{\underline{\gamma_{(j)}}})' = (\underline{e_{(1)}}', \cdots, \underline{e_{(p)}}')'S\widehat{\underline{\gamma_{(j)}}} = S\widehat{\underline{\gamma_{(j)}}}$$

Así, las covarianzas muestral de las $x_{(i)}$'s con la $\mathbb{Y}_{(j)}$'s juntas son:

$$(S\widehat{\underline{\gamma_{(1)}}}, \cdots, S\widehat{\underline{\gamma_{(p)}}}) = S(\widehat{\underline{\gamma_{(1)}}}, \cdots, \widehat{\underline{\gamma_{(p)}}}) = SG = GL\underbrace{G'G}_{I} = GL.$$

Por lo tanto, la covarianza muestral de $\underline{x_{(i)}}$ con $\underline{\mathbb{Y}_{(j)}}$ es el elemento (i,j) de GL, es decir, $\widehat{\gamma_{ij}}\widehat{\lambda}_j$. □

Corolario 2.20. *Sea $\mathbb{Y}$ la matriz de componentes principales de una muestra $\mathbb{X}$ de una población con vector aleatorio $\underline{X} \sim (\mu, \Sigma)$. La correlación de las observaciones de la i-ésima variable original y la j-ésima componente principal muestral es $\rho_{x_{(i)}\mathbb{Y}_{(j)}} = \frac{\widehat{\gamma_{ij}}\sqrt{\widehat{\lambda_j}}}{\sqrt{s_{ii}}}$.*

Demostración. Se tiene que la varianza muestral de $\underline{x_{(i)}}$ es s_{ii} y la de $\underline{\mathbb{Y}_{(j)}}$ es $\widehat{\lambda}_j$, así que:

$$\rho_{x_{(i)}\mathbb{Y}_{(j)}} = \frac{Cov(x_{(i)}, \mathbb{Y}_{(j)})}{\sqrt{s_{ii}}\sqrt{\widehat{\lambda}_j}} = \frac{\widehat{\gamma_{ij}}\widehat{\lambda}_j}{\sqrt{s_{ii}}\sqrt{\widehat{\lambda}_j}} = \frac{\widehat{\gamma_{ij}}\sqrt{\widehat{\lambda}_j}}{\sqrt{s_{ii}}}$$

□

De manera análoga a las componentes principales poblacionales, en este caso se tiene *la varianza generalizada muestral*, $|S|$ y *la varianza total muestral*, $tr(S)$, las cuales son invariantes respecto de las observaciones de las variables originales y respecto de las componentes principales muestrales, pues por el Teorema 2.16 la $S_{\mathbb{Y}} = L$, $|S| = \prod_{i=1}^{p} s_{ii} = \prod_{i=1}^{p} \widehat{\lambda}_i = |L|$ y $tr(S) = \sum_{i=1}^{p} s_{ii} = \sum_{i=1}^{p} \widehat{\lambda}_i = tr(L)$.

2.6. Componentes principales muestrales de la matriz de correlaciones

Como en el caso teórico, se aplicará el ACP muestral a las calificaciones $\mathbb{Z}$ o a datos estandarizados, en lugar de aplicarlo a los valores de los datos en bruto.

Sea $\mathbb{X}$ una muestra de un vector aleatorio $\underline{X} \sim (\mu, \Sigma)$. Estandarizando los datos $\mathbb{X}$

se tiene:

$$\begin{aligned}\mathbb{Z} &= \left[\underline{\mathbb{Z}_{(1)}}, \cdots, \underline{\mathbb{Z}_{(p)}}\right] = \begin{bmatrix} x_{11}-\overline{x_1} & \cdots & x_{1p}-\overline{x_p} \\ \vdots & \ddots & \vdots \\ x_{n1}-\overline{x_1} & \cdots & x_{np}-\overline{x_p} \end{bmatrix} \begin{bmatrix} s_{\mathbb{Y}_{11}}^{-1/2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & s_{\mathbb{Y}_{pp}}^{-1/2} \end{bmatrix} \\ &= (\mathbb{X} - \underline{1}\,\underline{\overline{x}}')D^{-1/2},\end{aligned}$$

donde $D = diag(s_{\mathbb{Y}_{11}}, \cdots, s_{\mathbb{Y}_{pp}})$.

Teorema 2.21. *La i-ésima componente principal de $\mathbb{Z}$ con matriz de covarianzas R viene dada por $\underline{\mathbb{Y}_{(i)}} = \mathbb{Z}\underline{\widehat{\gamma^*_{(i)}}} = (\mathbb{X} - \underline{1}\,\underline{\overline{x}}')D^{-1/2}\underline{\widehat{\gamma^*_{(i)}}}$, $i = 1, \ldots, p$, siendo $\underline{\widehat{\gamma^*_{(i)}}}$ los vectores propios asociados a los valores propios $\widehat{\lambda}^*_i$ de R, la matriz de correlaciones de $\mathbb{X}$, cumpliéndose la propiedad de que $\widehat{\lambda}^*_1 \geqslant \widehat{\lambda}^*_2 \geqslant \cdots \geqslant \widehat{\lambda}^*_p \geqslant 0$, y verificándose además que $\sum_{i=0}^{p} s_{\mathbb{Y}_{ii}} = \sum_{i=0}^{p} s_{\mathbb{Z}_{ii}} = p$.*

Demostración. Dado que $\mathbb{Z}$ son los datos estandarizados, su media muestral es $\underline{0}$ y su matriz de covarianzas muestrales es:

$$\begin{aligned}S_{\mathbb{Z}} &= \tfrac{1}{n}\mathbb{Z}'H\mathbb{Z} = \tfrac{1}{n}((\mathbb{X} - \underline{1}\,\underline{\overline{x}}')D^{-1/2})'H(\mathbb{X} - \underline{1}\,\underline{\overline{x}}')D^{-1/2} \\ &= D^{-1/2\,\prime}\tfrac{1}{n}(\mathbb{X} - \underline{1}\,\underline{\overline{x}}')'H(\mathbb{X} - \underline{1}\,\underline{\overline{x}}')D^{-1/2} = D^{-1/2}\,SD^{-1/2}\end{aligned}$$

la transformación de componentes principales muestrales es la transformación $\mathbb{Z} \to \mathbb{Y} = (\mathbb{Z} - \underline{0})G^* = \mathbb{Z}G^*$, donde G^* es una matriz ortogonal cuyas columnas son los vectores propios estandarizados de R, $G^{*\,\prime}RG^* = L^* = diag(\widehat{\lambda}^*_1, \widehat{\lambda}^*_2, \cdots, \widehat{\lambda}^*_p)$ es la matriz diagonal de los valores propios de R y $\widehat{\lambda}^*_1 \geqslant \widehat{\lambda}^*_2 \geqslant \ldots \geqslant \widehat{\lambda}^*_p \geqslant 0$.

La i-ésima componente principal muestral de $\mathbb{Z}$ se define como la i-ésima columna de la matriz $\mathbb{Y}$, es decir, $\underline{\mathbb{Y}_{(i)}} = \underline{\mathbb{Z}}\gamma^*_{(i)} = (\mathbb{X} - \underline{1}\,\underline{\overline{x}}')D^{-1/2}\underline{\gamma^*_{(i)}}$, $i = 1, \ldots, p$, donde $\gamma^*_{(i)}$ es la i-ésima columna de G^*, es decir, el vector propio asociado al valor propio $\widehat{\lambda}^*_i$, de R.

La covarianza muestral de las variables aleatorias $\mathbb{Z}_{(i)}$ y $\mathbb{Z}_{(j)}$, $i, j = 1, \ldots, p$ es igual al coeficiente de correlación muestral de $x_{(i)}$ y $x_{(j)}$, ya que:

$$\begin{aligned}Cov(\mathbb{Z}_{(i)}, \mathbb{Z}_{(j)}) &= n^{-1}\textstyle\sum_{r=1}^{n}(\mathbb{Z}_{ri} - \overline{\mathbb{Z}}_i)(\mathbb{Z}_{rj} - \overline{\mathbb{Z}}_j) = n^{-1}\sum_{r=1}^{n}\mathbb{Z}_{ri}\mathbb{Z}_{rj} \\ &= n^{-1}\textstyle\sum_{r=1}^{n}\frac{x_{ri}-\overline{x_i}}{\sqrt{s_{ii}}}\frac{x_{rj}-\overline{x_j}}{\sqrt{s_{jj}}} \\ &= \frac{n^{-1}\sum_{r=1}^{n}(x_{ri}-\overline{x}_i)(x_{rj}-\overline{x}_j)}{\sqrt{s_{ii}}\sqrt{s_{jj}}} \\ &= \rho_{x_{(i)}x_{(j)}},\end{aligned}$$

por lo tanto la matriz de covarianzas de $\mathbb{Z}$, R, coincide con su matriz de correlaciones muestrales de $\mathbb{X}$.

Por último, por el Teorema 2.16, se tiene que: $\sum_{i=0}^{p} s_{\mathbb{Y}_{ii}} = \sum_{i=0}^{p} s_{\mathbb{Z}_{ii}}$ y se verifica que:

$$\sum_{i=0}^{p} s_{\mathbb{Z}_{ii}} = tr(R) = tr(D^{-1/2}SD^{-1/2}) = tr(SD^{-1/2}D^{-1/2}) = tr(SD^{-1}),$$

el elemento (i,j) de SD^{-1} es $(s_{i1}, \cdots, s_{ij}, \cdots, s_{ip})(0, \cdots, s_{jj}^{-1}, \cdots, 0)' = s_{ij}/s_{jj}$, cuando $i = j$ es 1, por lo que $tr(SD^{-1}) = \sum_{i=1}^{p} 1 = p$. □

2.6.1. Calificaciones de las componentes principales muestrales de la matriz de correlación

Si el análisis se realiza sobre R, las calificaciones de componentes principales se calculan a partir de los valores $\mathbb{Z}$, según la fórmula:

$$\mathbb{Y}^*_{rj} = \widehat{\underline{\gamma_{(j)}}^*}' \mathbb{Z}_r$$

para $j = 1, 2, \ldots, p$; $r = 1, 2, \ldots, n$ [8].

2.6.2. Correlación entre las componentes principales muestrales y las variables originales

Teorema 2.22. *Sea $\mathbb{Y}$ la matriz de componentes principales de $\mathbb{Z}$. Los elementos de $\underline{\beta^*_{(j)}}$ dan las correlaciones entre las variables originales y la j-ésima variable componente principal.*

Demostración. Por el Corolario 2.20 se sigue que si $\mathbb{Y}$ es la matriz de componentes principales muestrales de $\mathbb{Z}$, la correlación de la i-ésima variable original y la j-ésima componente principal muestral es:

$$\rho_{\underline{\mathbb{Z}_{(i)}}, \underline{\mathbb{Y}_{(j)}}} = \frac{\widehat{\gamma^*_{ij}}\sqrt{\widehat{\lambda^*_j}}}{\sqrt{s_{\mathbb{Z}_{ii}}}} = \frac{\widehat{\gamma^*_{ij}}\sqrt{\widehat{\lambda^*_j}}}{\sqrt{1}} = \sqrt{\widehat{\lambda^*_j}}\widehat{\gamma^*_{ij}}.$$

El vector de las correlaciones muestrales entre las variables originales y la j-ésima variable componente principal muestral es $(\rho_{\underline{\mathbb{Z}_{(1)}},\underline{\mathbb{Y}_{(j)}}},\cdots,\rho_{\underline{\mathbb{Z}_{(p)}},\underline{\mathbb{Y}_{(j)}}})' = (\sqrt{\widehat{\lambda}_j^*}\widehat{\gamma_{1j}^*},\cdots,\sqrt{\widehat{\lambda}_j^*}\widehat{\gamma_{pj}^*})' = \sqrt{\widehat{\lambda}_j^*}(\widehat{\gamma_{1j}^*},\cdots,\widehat{\gamma_{pj}^*})' = \sqrt{\widehat{\lambda}_j^*}\widehat{\underline{\gamma_{(j)}}^*}$. □

2.7. Determinación del número de componentes principales

Cuando se lleva a cabo un ACP, se necesita determinar la dimensionalidad real del espacio en el que caen los datos es decir, el número de componentes principales con varianzas mayores que cero. Si varios de los valores propios de S (o R) son cero o están suficientemente cercanos a cero, entonces la dimensionalidad real de los datos es la del número de valores propios diferentes de cero [8]. En la práctica se usan habitualmente ciertos criterios para la elección del número de componentes, según se esté trabajando con la matriz de covarianzas muestrales (S) o con la matriz de correlaciones muestrales (R) [14].

2.7.1. Actuación con la matriz de covarianzas muestrales

Existen dos métodos para ayudarse a elegir el número de componentes principales a usar cuando se está aplicando el ACP a S. Estos métodos se basan en los valores propios de S. Supóngase que d es la dimensión del espacio en el cual se encuentran en realidad los datos al leer acerca de estos dos métodos [8].

Supongamos que se desea tomar en cuenta $\gamma 100\,\%$ de la variabilidad total en las variables originales. En uno de los métodos para estimar d se considera $\sum_{i=1}^{k}\lambda_i/\sum_{i=1}^{p}\lambda_i$ para valores de $k = 1, 2, \ldots, p$. Entonces d se estima por el menor de los valores de k en el que, por primera vez, sobrepasa γ. Este porcentaje varía habitualmente según el campo en que se esté trabajando, puesto que en un estudio social, se usaría un porcentaje alrededor del 60 %, mientras que en un estudio médico o científico, este porcentaje sería próximo al 80 % [14]. Para datos obtenidos en laboratorio, suele ser bastante fácil

explicar más de $95\,\%$ de la variabilidad total con sólo dos o tres componentes principales. Por otra parte, para datos del "tipo de personas", es posible que se requieran cinco o seis componentes principales para explicar del 70 al $75\,\%$ de la variación total [8].

Por desgracia, cuantas más componentes principales se requieran, menos útil se vuelve cada una de ellas.

Criterio de Cattell (1966) o "Gráfica de sedimentación"

Un segundo método para estimar d utiliza una gráfica de sedimentación de los valores propios. Una gráfica de sedimentación se construye al situar el valor de cada valor propio contra su orden. Es decir, se sitúan las parejas $(1, \widehat{\lambda_1})$, $(2, \widehat{\lambda_2})$,. . .,$(p, \widehat{\lambda_p})$. Este diagrama frecuentemente indica con claridad dónde terminan los valores propios "grandes" y empiezan los valores propios "pequeños". Cuando los puntos de la gráfica tienden a nivelarse, estos valores propios suelen estar suficientemente cercanos a cero como para que puedan ignorarse. Es probable que los más pequeños estén midiendo nada más que el ruido aleatorio y éste no se debe tratar de interpretar. Por tanto, por este método se supone que la dimensionalidad del espacio de datos es la que corresponde al valor propio grande más pequeño.

En la Figura 2.2 se muestra un ejemplo de una gráfica de sedimentación. Esta gráfica de sedimentación sugeriría que la dimensionalidad real del espacio en el que se encuentran los datos es tres y, como consecuencia, el número apropiado de componentes principales que tienen que usarse también es de tres [8].

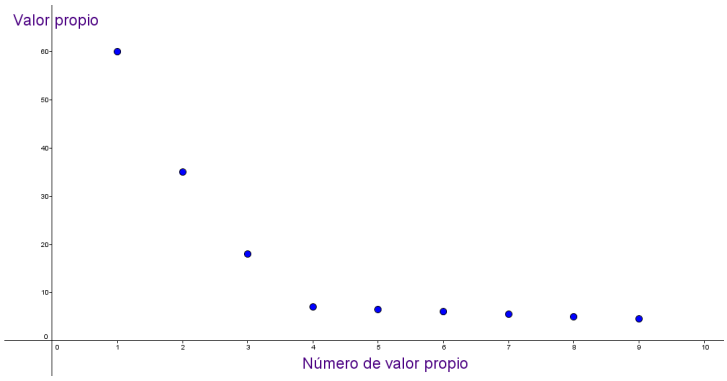

Figura 2.2: Una gráfica de sedimentación

Criterio de la raíz característica mayor que $\lambda^* = \frac{tr(S)}{p}$

Se buscan valores propios que sean mayores que la media de las varianzas, es decir, $\lambda^* = \frac{tr(S)}{p}$, y se estima que la dimensionalidad del espacio muestral debe ser el número de valores propios que sean mayores que λ^*. Estudios de Montecarlo consideran usar como punto de corte $0.7\lambda^*$.

2.7.2. Actuación con la matriz de correlaciones muestrales

Los dos primeros métodos descritos para determinar la dimensionalidad del espacio en el cual en realidad se encuentran los datos estandarizados también se pueden aplicar cuando se realiza un ACP sobre una matriz de correlación, R [8].

Criterio de Kaiser (1958) o criterio de la raíz característica mayor que 1

En este método se cuentan los valores propios que sean mayores que 1 y se estima que la dimensionalidad del espacio muestral es el del número de valores propios que sean mayores que 1. La razón para comparar los valores propios con 1 es que, cuando se realiza el análisis sobre datos estandarizados (es decir, la matriz de correlaciones), la varianza de cada variable estandarizada es igual a 1. La creencia es que si una componente principal no puede explicar más variación que una sola variable por sí misma, entonces es probable que no sea importante, por lo que frecuentemente se ignoran componentes cuyos valores propios son menores que 1. Nunca debe considerarse la comparación de los valores propios con 1 cuando se analizan los datos en bruto o, lo que es equivalente, la matriz de varianzas-covarianzas de la muestra [8]. Estudios de Montecarlo han probado que es más correcto el punto de corte $\lambda^* = 0.7$ [14].

Criterio de Horn (1965).

Se representan los valores propios de las componentes principales igual que en la "Prueba de la gráfica de sedimentación". Por otra parte, se consideran k conjuntos de una normal p-variada, todos de tamaño n, de los cuales se conoce la estructura de correlación. Se generan estas k muestras, se calculan los "valores propios medios" (media

aritmética de los valores propios de los k casos) y se van representando uno a uno. El criterio consiste en quedarse con las componentes principales anteriores al punto de cruce [14].

Un investigador puede emplear simultáneamente los métodos descritos; en todos los casos, la decisión de cuántas componentes principales considerar es subjetiva.

Se ha sugerido que cuando $p \leqslant 20$, el criterio de Kaiser tiende a incluir muy pocas componentes. Del mismo modo, el criterio de Cattel incluye muchas componentes. Por lo tanto, en la práctica frecuentemente se usa un acuerdo mutuo.

Mediante estos criterios, el ACP permite usar un número reducido de variables en los análisis subsecuentes, en general, no se pueden emplear estos criterios para eliminar variables originales debido a que se necesitan todas las variables originales para calificar o evaluar las variables componentes principales de cada uno de los individuos en un conjunto de datos.

2.8. Representación gráfica de las componentes principales

La reducción en la dimensión permitida por el análisis de componentes principales se puede usar gráficamente. Así, si las primeras dos componentes explican la mayoría de la varianza, entonces un diagrama de dispersión que muestre la distribución de los objetos en esas dos dimensiones suele dar una clara indicación de la distribución en general de los datos.

Además de mostrar la distribución de los *objetos* en las dos primeras componentes principales, también se pueden representar las *variables* en un diagrama similar, usando las coordenadas de las correlaciones [11].

2.9. Propiedades muestrales de las componentes principales

El análisis de componentes principales es una técnica matemática que no requiere la suposición de normalidad multivariante de los datos, aunque en el análisis de componentes principales paramétrico que se llevará a cabo, el vector aleatorio $\underline{X} = (X_1, \ldots, X_n)'$ se supondrá que es modelado al realizar inferencia por una distribución normal p-dimensional.

2.9.1. Estimación de máxima verosimilitud para datos normales

Cuando se tienen muestras pequeñas, la distribución de los valores propios y vectores propios de una matriz de covarianza S es extremadamente complicada, incluso cuando la correlación original desaparece. Una razón es que los valores propios son funciones no racionales de los elementos de S. Sin embargo, se conocen algunos resultados para muestras grandes (resultados límite) y muchas propiedades útiles de las componentes principales de la muestra de datos normales se derivan de los siguientes resultados de máxima verosimilitud [11].

Teorema 2.23. *Para datos normales cuando los valores propios de Σ son distintos, los valores propios y las componentes principales muestrales son los estimadores de máxima verosimilitud de los parámetros poblacionales correspondientes.*

La demostración puede consultarse en [11], pág. 229.

Cuando los valores propios de Σ no son distintos, el Teorema 2.23 no se cumple (esto puede ser fácilmente confirmado en el caso $\Sigma = \sigma^2 I$). Por tanto, cuando los valores propios de Σ no son distintos, las componentes principales muestrales no son los estimadores de máxima verosimilitud de sus equivalentes poblacionales. Sin embargo, en este caso puede usarse el siguiente resultado, que es una modificación del anterior.

Teorema 2.24. *Para datos normales, cuando $k > 1$ valores propios de Σ no son distintos, pero toman el valor común $\overline{\lambda}$, entonces:*

a) el estimador de máxima verosimilitud de $\overline{\lambda}$ *es* $\overline{\overline{\lambda}}$*, la media aritmética de los correspondientes valores propios muestrales, y*

b) los vectores propios muestrales correspondientes al valor propio repetido son los estimadores de máxima verosimilitud a pesar de que no son los únicos estimadores de máxima verosimilitud.

La demostración puede consultarse en [1], y en [10], pág. 439.

Aunque no se demuestra el teorema anterior, se puede notar que a) parece natural a partir del Teorema 2.23, ya que si λ_i es un valor propio diferente, entonces el estimador de máxima verosimilitud de λ_i es $\widehat{\lambda}_i$, y si los últimos k valores propios son iguales a $\overline{\lambda}$, entonces:

$$tr(\Sigma) = \lambda_1 + \ldots + \lambda_{p-k} + k\overline{\lambda}$$

y

$$tr(S) = \widehat{\lambda}_1 + \ldots + \widehat{\lambda_{p-k}} + k\overline{\overline{\lambda}}.$$

La parte b) del Teorema 2.24 también parece natural del Teorema 2.23.

2.9.2. Distribuciones asintóticas para datos normales

Para muestras grandes, el siguiente resultado asintótico, basado en el teorema del límite central, proporciona distribuciones útiles para los valores propios y vectores propios de la matriz de covarianza muestral.

Teorema 2.25. *Sea* Σ *una matriz definida positiva, con valores propios distintos. Si* $\underline{X}$ *se distribuye* $N_p(\underline{\mu}, \Sigma)$*, y* $\mathbb{X}_{n\times p}$ *es la matriz de la muestra, considerando la descomposición espectral* $\Sigma = \Gamma\Lambda\Gamma'$ *y* $S_u = (n-1)^{-1}nS = GLG'$*. Sean* $\widehat{\underline{\lambda}} = (\widehat{\lambda}_1, \widehat{\lambda}_2, \cdots, \widehat{\lambda}_p)'$ *y* $(\widehat{\underline{\gamma_{(1)}}}, \widehat{\underline{\gamma_{(2)}}}, \cdots, \widehat{\underline{\gamma_{(p)}}})$ *los valores propios y vectores propios de* S_u*, y análogamente* $\underline{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_p)'$ *y* $(\underline{\gamma_{(1)}}, \underline{\gamma_{(2)}}, \cdots, \underline{\gamma_{(p)}})$ *los valores propios y vectores propios de* Σ*. Entonces se tienen las siguientes distribuciones asintóticas cuando* $n \to \infty$*.*

1. $\widehat{\underline{\lambda}} \sim N_p(\underline{\lambda}, 2\Lambda^2/(n-1))$*; esto es, los valores propios de* S_u *son asintóticamente normales, insesgados, e independientes, con* $\widehat{\lambda}_i$ *teniendo varianza igual a* $2\lambda_i^2/(n-1)$*.*

2. $\widehat{\underline{\gamma_{(i)}}} \sim N_p(\underline{\gamma_{(i)}}, V_i/(n-1))$, *donde,*

$$V_i = \lambda_i \sum_{j \neq i} \frac{\lambda_j}{(\lambda_j - \lambda_i)^2} \underline{\gamma_{(j)}}\, \underline{\gamma_{(j)}}';$$

es decir, los vectores propios de S_u son asintóticamente normales e insesgados y con matriz de covarianza asintótica, $V_i/(n-1)$, dada anteriormente.

3. *La covarianza entre el r-ésimo elemento de $\widehat{\underline{\gamma_{(i)}}}$ y el s-ésimo elemento de $\widehat{\underline{\gamma_{(j)}}}$ es $-\lambda_i\lambda_j\gamma_{rj}\gamma_{si}/(n-1)(\lambda_i - \lambda_j)^2$.*

4. *Los elementos de $\widehat{\underline{\lambda}}$ son asintóticamente independientes de los elementos de G.*

Nótese que asintóticamente no hay diferencia si se trabaja con S o con S_u. La demostración de 1. puede consultarse en [11], pág. 231; también puede consultarse en [1].

Corolario 2.26. *Sean $\widehat{\lambda}_1, \widehat{\lambda}_2, \ldots, \widehat{\lambda}_p$ los valores propios de S para una muestra de tamaño n. Asintóticamente $\widehat{\lambda}_i \sim N(\lambda_i, 2\lambda_i^2/n)$. Esto permite establecer intervalos de confianza al $100(1-\alpha)\,\%$ con límites de confianza inferior y superior dados por $\widehat{\lambda}_i/(1 \pm z_{\alpha/2}\sqrt{2/n})$ [14].*

Demostración. De 1. del Teorema 2.25 se tiene que asintóticamente $\widehat{\lambda}_i \sim N(\lambda_i, 2\lambda_i^2/n)$, luego $\frac{\widehat{\lambda}_i - \lambda_i}{\lambda_i\sqrt{2/n}}$ es una función de las medidas muestrales y el parámetro λ_i, donde λ_i es la única cantidad desconocida; además tiene una distribución de probabilidad $N(0,1)$ que no depende de λ_i, entonces se puede usar como cantidad pivote. Sean $a, b \in \mathbb{R}$ tales que:

$$P\left(a < \frac{\widehat{\lambda}_i - \lambda_i}{\lambda_i\sqrt{2/n}} < b\right) = 1 - \alpha.$$

Entonces los valores que optimizan $P\left(a < \frac{\widehat{\lambda}_i - \lambda_i}{\lambda_i\sqrt{2/n}} < b\right)$ son $a = -z_{\alpha/2}$ y $b = z_{\alpha/2}$, por lo que:

$$\begin{aligned}
1-\alpha &= P\left(-z_{\alpha/2} < \tfrac{\widehat{\lambda}_i - \lambda_i}{\lambda_i\sqrt{2/n}} < z_{\alpha/2}\right) \\
&= P\left(-z_{\alpha/2}\sqrt{2/n} < \widehat{\lambda}_i/\lambda_i - 1 < z_{\alpha/2}\sqrt{2/n}\right) \\
&= P\left(1/(1 + z_{\alpha/2}\sqrt{2/n}) < \lambda_i/\widehat{\lambda}_i < 1/(1 - z_{\alpha/2}\sqrt{2/n})\right) \\
&= P\left(\widehat{\lambda}_i/(1 + z_{\alpha/2}\sqrt{2/n}) < \lambda_i < \widehat{\lambda}_i/(1 - z_{\alpha/2}\sqrt{2/n})\right)
\end{aligned}$$

Por lo tanto los límites de confianza inferior y superior son $\widehat{\lambda}_i/(1 \pm z_{\alpha/2}\sqrt{2/n})$. □

Hay que tener cuidado con estos intervalos cuando un λ_i es muy grande y n no lo es, ya que se producen intervalos muy amplios, que pueden originar error. Se recomienda trabajar siempre que se pueda con la matriz de correlaciones R [14].

2.10. Pruebas de hipótesis sobre las componentes principales

Frecuentemente es útil tener un procedimiento para decidir si las primeras k componentes principales incluyen toda la variación importante en $\underline{X}$. Uno desearía ignorar las restantes $p-k$ componentes si sus correspondientes valores propios poblacionales fueran todos cero. Sin embargo, esto ocurre sólo si Σ tiene rango k, en tal caso S también debe tener rango k. Por lo tanto esta situación solo se encuentra esporádicamente en la práctica.

Una segunda alternativa es la prueba de hipótesis de que la proporción de varianza explicada por las últimas $p-k$ componentes es menor que cierto valor.

Una hipótesis más conveniente es la cuestión de que los últimos $p-k$ valores propios son iguales. Esto implicaría que la variación es igual en todas las direcciones del espacio generado por los últimos $p-k$ vectores propios de dimensión $p-k$. Ésta es la situación de variación isotrópica e implicaría que si una de esas componentes es descartada, entonces todas las últimas k componentes deberían ser descartadas.

En todas estas subsecciones se asume que se da una muestra aleatoria de datos normales de tamaño n.

2.10.1. La hipótesis de que $(\lambda_1 + \ldots + \lambda_k)/(\lambda_1 + \ldots + \lambda_p) = \psi$

Sean $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_p$ los valores propios de S, se denota al equivalente muestral de ψ con:

$$\widehat{\psi} = (\widehat{\lambda}_1 + \ldots + \widehat{\lambda}_k)/(\widehat{\lambda}_1 + \ldots + \widehat{\lambda}_p)$$

Teorema 2.27. *La distribución asintótica de* $\widehat{\psi}$ *es:*

$$N\left(\psi, \frac{2tr(\Sigma^2)}{tr(\Sigma)^2 n}(\psi^2 - 2\alpha\psi + \alpha)\right),$$

donde $\alpha = (\lambda_1^2 + \ldots + \lambda_k^2)/(\lambda_1^2 + \ldots + \lambda_p^2)$.

Demostración. Del Teorema 2.25 se sabe que $\widehat{\underline{\lambda}}$ es asintóticamente normal con media $\underline{\lambda}$ y matriz de covarianzas $2\Lambda^2/n$, usando el Teorema C.2, con $\underline{t} = \widehat{\underline{\lambda}} = (\widehat{\lambda_1}, \ldots, \widehat{\lambda_p})'$, $\mu = \underline{\lambda}$, $V = 2\Lambda^2$ y $f(\underline{\lambda}) = \psi = (\lambda_1 + \ldots + \lambda_k)/(\lambda_1 + \ldots + \lambda_p)$.

Cuando $i \leqslant k$:

$$\begin{aligned} \frac{\partial\psi}{\partial\lambda_i} &= \left(\sum_{i=1}^p \lambda_i\right)^{-1} - \left(\sum_{i=1}^p \lambda_i\right)^{-2} \sum_{i=1}^k \lambda_i \\ &= \left(\sum_{i=1}^p \lambda_i\right)^{-1} \left(1 - \sum_{i=1}^k \lambda_i / \sum_{i=1}^p \lambda_i\right) \\ &= (1-\psi)/tr(\Sigma) \end{aligned}$$

y si $k < i$:

$$\frac{\partial\psi}{\partial\lambda_i} = -\left(\sum_{i=1}^k \lambda_i\right)\left(\sum_{i=1}^p \lambda_i\right)^{-2} = -\psi/\sum_{i=1}^p \lambda_i = -\psi/tr(\Sigma),$$

luego:

$$d_i = \frac{\partial\psi}{\partial\lambda_i} = \begin{cases} (1-\psi)/tr(\Sigma) \text{ para } i \leqslant k \\ -\psi/tr(\Sigma) \text{ para } k < i, \end{cases}$$

entonces $\underline{D} = (d_1, \ldots, d_p)'$.

Se tiene que $f'(\underline{\mu}) = f'(\underline{\lambda}) = D$, por lo que f es diferenciable en $\underline{\mu}$, por el Teorema C.2 se sigue que $f(\underline{t}) = f(\widehat{\underline{\lambda}}) = (\widehat{\lambda_1} + \ldots + \widehat{\lambda_k})/(\widehat{\lambda_1} + \ldots + \widehat{\lambda_p}) = \widehat{\psi}$ es asintóticamente

normal con media $f(\underline{\mu}) = f(\underline{\lambda}) = \psi$ y con varianza:

$$\begin{aligned}
\underline{D}'V\underline{D}/n &= 2tr(\Sigma)^{-2}n^{-1}(1-\psi,\ldots,1-\psi,-\psi,\ldots,-\psi)\Lambda^2 \begin{bmatrix} 1-\psi \\ \vdots \\ 1-\psi \\ -\psi \\ \vdots \\ -\psi \end{bmatrix} \\
&= 2tr(\Sigma)^{-2}n^{-1}\left(\textstyle\sum_{i=1}^{k}(1-\psi)^2\lambda_i^2 + \sum_{i=k+1}^{p}\psi^2\lambda_i^2\right) \\
&= 2tr(\Sigma)^{-2}n^{-1}\left((1-\psi)^2\textstyle\sum_{i=1}^{k}\lambda_i^2 + \psi^2\sum_{i=k+1}^{p}\lambda_i^2\right) \\
&= 2tr(\Sigma)^{-2}n^{-1}\left((1-2\psi+\psi^2)\textstyle\sum_{i=1}^{k}\lambda_i^2 + \psi^2\sum_{i=k+1}^{p}\lambda_i^2\right) \\
&= 2tr(\Sigma)^{-2}n^{-1}\left((1-2\psi)\textstyle\sum_{i=1}^{k}\lambda_i^2 + \psi^2\sum_{i=1}^{p}\lambda_i^2\right).
\end{aligned}$$

Dado que $\Sigma = \Gamma\Lambda\Gamma'$, se tiene que $tr(\Sigma^2) = tr(\Gamma\Lambda\Gamma'\Gamma\Lambda\Gamma') = tr(\Gamma\Lambda^2\Gamma') = tr(\Lambda^2\Gamma'\Gamma) = tr(\Lambda^2) = \sum_{i=1}^{p}\lambda_i^2$. Considérese $\alpha = (\lambda_1^2+\cdots+\lambda_k^2)/(\lambda_1^2+\cdots+\lambda_p^2) = (\lambda_1^2+\cdots+\lambda_k^2)/tr(\Sigma^2)$,

$$\begin{aligned}
\underline{D}'V\underline{D}/n &= 2tr(\Sigma)^{-2}n^{-1}\left((1-2\psi)\alpha\alpha^{-1}\textstyle\sum_{i=1}^{k}\lambda_i^2 + \psi^2 tr(\Sigma^2)\right) \\
&= 2tr(\Sigma)^{-2}n^{-1}\left((1-2\psi)\alpha tr(\Sigma^2) + \psi^2 tr(\Sigma^2)\right) \\
&= \tfrac{2tr(\Sigma^2)}{tr(\Sigma)^2 n}(\psi^2 - 2\alpha\psi + \alpha)
\end{aligned}$$

□

Nótese que α puede ser interpretada como la proporción de varianza explicada por las primeras k componentes principales de una variable cuya matriz de covarianza es Σ^2. Dado que sólo se tiene interés en proporciones de valores propios, no hay diferencia si se trabaja con los valores propios de S o de $S_u = (n-1)^{-1}nS$.

El Teorema 2.27 se puede usar para obtener intervalos de confianza aproximados para ψ, si se estiman los parámetros usando los valores propios muestrales.

2.10.2. Hipótesis de que algunos valores propios son iguales

La prueba de hipótesis $\lambda_p = \lambda_{p-1}$ es significativa, pues prueba que la dispersión es isotrópica parcialmente en un subespacio bidimensional. Si se considera la hipótesis más general dada por $\lambda_p = \lambda_{p-1} = \cdots = \lambda_{k+1}$, esta prueba de isotropía es útil en la

determinación del número de componentes principales que se usarán para describir los datos. Supongamos que se decide incluir al menos k componentes, y se quiere decidir si se incluye otra o no. El no rechazo de la hipótesis nula implica que si se incluyen más de k componentes principales se deben incluir todas las p componentes principales, porque cada una de las componentes restantes contienen la misma cantidad de información. A menudo se lleva a cabo una secuencia de pruebas isotrópicas empezando con $k = 0$ e incrementando k hasta que la hipótesis nula no es rechazada.

Teorema 2.28. *Si $\underline{X}$ se distribuye $N_p(\underline{\mu}, \Sigma)$, y $\mathbb{X}_{n\times p}$ es la matriz de la muestra. La prueba de hipótesis que contrasta que los $p-k$ valores propios más pequeños de Σ son todos iguales es $H_0 : \lambda_{k+1} = \lambda_{k+2} = \ldots = \lambda_p$. Un estadístico de prueba es:*

$$-2log\lambda = n(p-k)log\left(\overline{\lambda}/g_0\right), \tag{2.21}$$

donde $\widehat{\lambda}_i$ es el i-ésimo valor propio de S, $\overline{\lambda} = \sum_{i=k+1}^{p} \widehat{\lambda}_i/(p-k)$ y $g_0 = \left(\prod_{i=k+1}^{p} \widehat{\lambda}_i\right)^{1/(p-k)}$. Dicho estadístico, bajo la hipótesis nula, sigue una distribución χ^2 con $\frac{1}{2}(p-k-1)(p-k+2)$ grados de libertad, asintóticamente. Así, la prueba de razón de verosimilitud (PRV) de tamaño α rechazará la hipótesis nula cuando $-2log\lambda < \chi^2_{1-\alpha;\frac{1}{2}(p-k-1)(p-k+2)}$.

Demostración. En la distribución normal multivariada se desconocen los parámetros $\underline{\mu}$ y Σ, por lo que el espacio paramétrico es $\overline{\underline{\Theta}} = \mathbb{R}^p \times \{\Sigma \in M_{p\times p}(\mathbb{R}_+) | \Sigma \text{ es simétrica}\}$, la hipótesis nula es $H_0 : \theta \in \overline{\underline{\Theta}}_1 = \{(\underline{\mu}, \Sigma) \in \overline{\underline{\Theta}} | \lambda_{k+1} = \lambda_{k+2} = \ldots = \lambda_p$, donde $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_k \geqslant \lambda_{k+1} \geqslant \ldots \geqslant \lambda_p$ son los valores propios de $\Sigma\}$ y $H_1 : \underline{\theta} \in \overline{\underline{\Theta}}_1 = \overline{\underline{\Theta}} - \overline{\underline{\Theta}}_0$.

Bajo la hipótesis nula, por el teorema de la descomposición espectral (Teorema B.7), se tiene que $\Sigma = \Gamma\Lambda\Gamma'$, donde Γ y $\Lambda = diag(\lambda_1, \ldots, \lambda_k, \lambda_{k+1}, \ldots, \lambda_p)$ son las matrices de vectores propios y valores propios de Σ y se cumple que $\lambda_{k+1} = \ldots = \lambda_p$. Además, por los Teoremas 2.23 y 2.24, se sigue que los estimadores de máxima verosimilitud de Γ y Λ son G y L_0, donde G es una matriz ortogonal cuyas columnas son vectores propios estandarizados de S, y $L_0 = diag(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_k, \overline{\lambda}, \ldots, \overline{\lambda})$, donde $\widehat{\lambda}_i$ es el i-ésimo valor propio de S y $\overline{\lambda} = \sum_{i=k+1}^{p} \widehat{\lambda}_i/(p-k)$. Por lo que el estimador de máxima verosimilitud para Σ es $\widehat{\Sigma}_0 = GL_0G'$ y el estimador para $\underline{\mu}$ es $\widehat{\underline{\mu}} = \overline{\underline{x}}$.

Por el Teorema C.9 se tiene que los estimadores de máxima verosimilitud bajo $\overline{\underline{\Theta}}$ son $\widehat{\underline{\mu}} = \overline{\underline{x}}$ y $\widehat{\Sigma} = S$. Por el teorema de la descomposición espectral (Teorema B.7) se

tiene que $S = GLG'$, donde G es una matriz ortogonal cuyas columnas son vectores propios estandarizados de S, y $L = diag(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_k, \widehat{\lambda_{k+1}}, \ldots, \widehat{\lambda}_p)$ tiene en su diagonal a los valores propios de S. Nótese que el estimador de máxima verosimilitud bajo H_0 y bajo H_1 para Γ es el mismo, la matriz de vectores propios estandarizados de S.

Usando el Teorema C.14, se sigue que la prueba de razón de verosimilitud para contrastar H_0 contra H_1 está dada por $-2log\lambda = np(a - logg - 1)$, donde a y g son la media aritmética y media geométrica de los valores propios de $\widehat{\Sigma}_0^{-1}S$, luego:

$$\begin{aligned}
\widehat{\Sigma}_0^{-1}S &= (GL_0G')^{-1}GLG' = G'^{-1}L_0^{-1}G^{-1}GLG' = GL_0^{-1}LG' \\
&= Gdiag(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_k, \overline{\lambda}, \ldots, \overline{\lambda})^{-1}diag(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_k, \widehat{\lambda_{k+1}}, \ldots, \widehat{\lambda}_p)G' \\
&= Gdiag(\widehat{\lambda}_1^{-1}, \ldots, \widehat{\lambda}_k^{-1}, \overline{\lambda}^{-1}, \ldots, \overline{\lambda}^{-1})diag(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_k, \widehat{\lambda_{k+1}}, \ldots, \widehat{\lambda}_p)G' \\
&= Gdiag(1, \ldots, 1, \widehat{\lambda_{k+1}}/\overline{\lambda}, \ldots, \widehat{\lambda}_p/\overline{\lambda})G',
\end{aligned}$$

por lo que $1, \ldots, 1, \widehat{\lambda_{k+1}}/\overline{\lambda}, \ldots, \widehat{\lambda}_p/\overline{\lambda}$ son los valores propios de $\widehat{\Sigma}_0^{-1}S$, así:

$$a = \left(k + \sum_{i=k+1}^{p} \widehat{\lambda}_i/\overline{\lambda}\right)/p = \left(k + (p-k)\overline{\lambda}/\overline{\lambda}\right)/p = 1$$

y

$$\begin{aligned}
g &= \sqrt[p]{\prod_{i=k+1}^{p}\left(\widehat{\lambda}_i/\overline{\lambda}\right)} = \left(\prod_{i=k+1}^{p}\widehat{\lambda}_i/\overline{\lambda}^{(p-k)}\right)^{1/p} = \left(g_0^{(p-k)}/\overline{\lambda}^{(p-k)}\right)^{1/p} \\
&= \left(g_0/\overline{\lambda}\right)^{(p-k)/p},
\end{aligned}$$

donde $g_0 = \left(\prod_{i=k+1}^{p}\widehat{\lambda}_i\right)^{1/(p-k)}$, por lo que:

$$\begin{aligned}
-2log\lambda &= np(a - logg - 1) = np\left(1 - log\left(\left(g_0/\overline{\lambda}\right)^{(p-k)/p}\right) - 1\right) \\
&= -np(p-k)p^{-1}log\left(g_0/\overline{\lambda}\right) = -n(p-k)log\left(g_0/\overline{\lambda}\right) \\
&= n(p-k)log\left(\overline{\lambda}/g_0\right).
\end{aligned}$$

Debido a que H_0 afecta el número de condiciones de ortogonalidad que Σ debe satisfacer, bajo H_0, Σ está determinada por k valores propios distintos y un valor propio común, y los k vectores propios cada uno con p componentes que corresponden a los valores propios distintos. Esto da $k+1+kp$ parámetros. Sin embargo los vectores propios están restringidos por $\frac{1}{2}k(k+1)$ condiciones de ortogonalidad, dejando $k+1+kp-\frac{1}{2}k(k+1)$ de parámetros libres. En $\overline{\Theta}$, el número de parámetros libres se obtiene poniendo $k = p-1$, concretamente $p+(p-1)p-\frac{1}{2}(p-1)p = p+\frac{1}{2}(p-1)p = p(1+\frac{1}{2}(p-1)) = \frac{1}{2}p(p+1)$,

por lo tanto, la variable aleatoria $-2log\lambda(X)$ bajo H_0 es asintóticamente distribuida como una variable aleatoria χ^2 con grados de libertad igual a la diferencia entre el número de parámetros independientes en $\overline{\Theta}$ y el número de parámetros en $\overline{\Theta}_0$, es decir:

$$\frac{1}{2}p(p+1) - (k+1+kp-\frac{1}{2}k(k+1)) = \frac{1}{2}(p^2+p-2k-2-2kp+k^2+k)$$
$$= \frac{1}{2}(p(p-k-1)-k(p-k-1)+2(p-k-1)) = \frac{1}{2}(p-k+2)(p-k-1).$$

La prueba de razón de verosimilitud (PRV) de tamaño α para contrastar H_0 contra H_1 tiene región de rechazo $R = \{X| -2log(\lambda(X)) < c\}$, donde c es determinada de manera que:

$$\begin{aligned} \alpha &= sup_{\underline{\theta}\in\overline{\Theta}_0} P_{\underline{\theta}}(X \in R) = P(-2log(\lambda(X)) < c|(\underline{\overline{x}}, \widehat{\Sigma}_0)) \\ &= P\left(\chi^2_{\frac{1}{2}(p-k+2)(p-k-1)} < c\right) = 1 - P\left(\chi^2_{\frac{1}{2}(p-k+2)(p-k-1)} \geqslant c\right) \end{aligned}$$

y se sigue que $1-\alpha = P\left(\chi^2_{\frac{1}{2}(p-k+2)(p-k-1)} \geqslant c\right)$, por lo tanto:

$$c = \chi^2_{1-\alpha;\frac{1}{2}(p-k+2)(p-k-1)}.$$

□

También, se obtiene una mejor aproximación chi-cuadrada si n es reemplazada por $n' = n - (2p+11)/6$ en la Ecuación (2.21) para obtener la aproximación de Bartlett:

$$\left(n - \frac{2p+11}{6}\right)(p-k)log\left(\frac{\overline{\lambda}}{g_0}\right) \sim \chi^2_{(p-k+2)(p-k-1)/2},$$

asintóticamente [11].

Nótese que:

1. Si $k = 0$, la hipótesis nula que se contrasta es $H_0 : \lambda_1 = \lambda_2 = \ldots = \lambda_p$, es decir, la igualdad de todos los valores propios.

2. La proporción $\overline{\lambda}/g_0$ es la misma si se trabaja con los valores propios de S o de S_u [11].

Teorema 2.29. *Si $\underline{X}$ se distribuye $N_p(\underline{\mu}, \Sigma)$, y $\mathbb{X}_{n\times p}$ es la matriz de la muestra. La prueba de hipótesis que contrasta que los $p-k$ valores propios más pequeños de Σ son todos*

iguales a un valor dado es $H_0 : \lambda_{k+1} = \lambda_{k+2} = \ldots = \lambda_p = \lambda$. *El estadístico de prueba es:*

$$-2log\lambda = n(p-k)(a_0/\lambda - 1 - log\,(g_0/\lambda)),$$

donde $\widehat{\lambda}_i$ *es el i-ésimo valor propio de* S, $a_0 = (p-k)^{-1}\sum_{i=k+1}^{p}\widehat{\lambda}_i$ *y* $g_0 = \left(\prod_{i=k+1}^{p}\widehat{\lambda}_i\right)^{1/(p-k)}$. *Dicho estadístico, bajo la hipótesis nula, sigue una distribución* χ^2 *con* $\frac{1}{2}(p-k-1)(p-k+2)$ *grados de libertad, asintóticamente. Así, la prueba de razón de verosimilitud (PRV) de tamaño* α *rechazará la hipótesis nula cuando* $-2log\lambda < \chi^2_{1-\alpha;\frac{1}{2}(p-k-1)(p-k+2)}$.

Demostración. En la distribución normal multivariada, el espacio paramétrico es $\overline{\underline{\Theta}} = \mathbb{R}^p \times \{\Sigma \in M_{p\times p}(\mathbb{R}_+)|\Sigma \text{ es simétrica}\}$, la hipótesis nula es $H_0 : \underline{\theta} \in \overline{\underline{\Theta}}_1 = \{(\underline{\mu}, \Sigma) \in \overline{\underline{\Theta}}|\lambda_{k+1} = \lambda_{k+2} = \ldots = \lambda_p = \lambda$, donde $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_k \geqslant \lambda_{k+1} \geqslant \ldots \geqslant \lambda_p$ son los valores propios de $\Sigma\}$ y $H_1 : \underline{\theta} \in \overline{\underline{\Theta}}_1 = \overline{\underline{\Theta}} - \overline{\underline{\Theta}}_0$.

Bajo la hipótesis nula, por el teorema de la descomposición espectral (Teorema B.7) se tiene que $\Sigma = \Gamma\Lambda\Gamma'$, donde Γ y $\Lambda = diag(\lambda_1, \ldots, \lambda_k, \lambda_{k+1}, \ldots, \lambda_p)$ son las matrices de vectores propios y valores propios de Σ y se cumple que $\lambda_{k+1} = \ldots = \lambda_p = \lambda$. Por el Teorema 2.23 se sigue que los estimadores de máxima verosimilitud de Γ y Λ son G y L_0, donde G es una matriz ortogonal cuyas columnas son vectores propios estandarizados de S, y $L_0 = diag(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_k, \lambda, \ldots, \lambda)$, donde $\widehat{\lambda}_i$ es el i-ésimo valor propio de S. Por lo que el estimador de máxima verosimilitud para Σ es $\widehat{\Sigma}_0 = GL_0G'$ y el estimador para $\underline{\mu}$ es $\widehat{\underline{\mu}} = \overline{\underline{x}}$.

Por el Teorema C.9 se tiene que los estimadores de máxima verosimilitud bajo $\overline{\underline{\Theta}}$ son $\widehat{\underline{\mu}} = \overline{\underline{x}}$ y $\widehat{\Sigma} = S$. Por el teorema de la descomposición espectral (Teorema B.7) se tiene que $S = GLG'$, donde G es una matriz ortogonal cuyas columnas son vectores propios estandarizados de S, y $L = diag(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_k, \widehat{\lambda_{k+1}}, \ldots, \widehat{\lambda}_p)$ tiene en su diagonal a los valores propios de S. Nótese que el estimador de máxima verosimilitud bajo H_0 y bajo H_1 para Γ es el mismo, la matriz de vectores propios estandarizados de S.

Usando el Teorema C.14 se sigue que la prueba de razón de verosimilitud para contrastar H_0 contra H_1 está dada por $-2log\lambda = np(a - logg - 1)$, donde a y g son la media

aritmética y media geométrica de los valores propios de $\widehat{\Sigma}_0^{-1}S$, luego:

$$\begin{aligned}
&\widehat{\Sigma}_0^{-1}S = (GL_0G')^{-1}GLG' = G'^{-1}L_0^{-1}G^{-1}GLG' = GL_0^{-1}LG' \\
&= Gdiag(\widehat{\lambda}_1,\ldots,\widehat{\lambda}_k,\lambda,\ldots,\lambda)^{-1}diag(\widehat{\lambda}_1,\ldots,\widehat{\lambda}_k,\widehat{\lambda_{k+1}},\ldots,\widehat{\lambda}_p)G' \\
&= Gdiag(\widehat{\lambda}_1^{-1},\ldots,\widehat{\lambda}_k^{-1},\lambda^{-1},\ldots,\lambda^{-1})diag(\widehat{\lambda}_1,\ldots,\widehat{\lambda}_k,\widehat{\lambda_{k+1}},\ldots,\widehat{\lambda}_p)G' \\
&= Gdiag(1,\ldots,1,\widehat{\lambda_{k+1}}/\lambda,\ldots,\widehat{\lambda}_p/\lambda)G',
\end{aligned}$$

por lo que $1,\ldots,1,\widehat{\lambda_{k+1}}/\lambda,\ldots,\widehat{\lambda}_p/\lambda$ son los valores propios de $\widehat{\Sigma}_0^{-1}S$, así:

$$a = \left(k + \sum_{i=k+1}^{p}\widehat{\lambda}_i/\lambda\right)/p = k/p + \frac{1}{\lambda p}\sum_{i=k+1}^{p}\widehat{\lambda}_i$$

y

$$\begin{aligned}
g &= \sqrt[p]{\prod_{i=k+1}^{p}\left(\widehat{\lambda}_i/\lambda\right)} = \left(\prod_{i=k+1}^{p}\widehat{\lambda}_i/\lambda^{(p-k)}\right)^{1/p} = \left(g_0^{(p-k)}/\lambda^{(p-k)}\right)^{1/p} \\
&= (g_0/\lambda)^{(p-k)/p}
\end{aligned}$$

donde $g_0 = \left(\prod_{i=k+1}^{p}\widehat{\lambda}_i\right)^{1/(p-k)}$, por lo que:

$$\begin{aligned}
-2log\lambda &= np(a - logg - 1) \\
&= np(k/p + (\lambda p)^{-1}\textstyle\sum_{i=k+1}^{p}\widehat{\lambda}_i - 1) - nplog\left((g_0/\lambda)^{(p-k)/p}\right) \\
&= n(k - p + \lambda^{-1}\textstyle\sum_{i=k+1}^{p}\widehat{\lambda}_i) - n(p-k)log\,(g_0/\lambda) \\
&= n(k - p + \lambda^{-1}(p-k)a_0) - n(p-k)log\,(g_0/\lambda) \\
&= n(p-k)(-1 + a_0/\lambda) - n(p-k)log\,(g_0/\lambda) \\
&= n(p-k)(a_0/\lambda - 1 - log\,(g_0/\lambda)),
\end{aligned}$$

donde $a_0 = (p-k)^{-1}\sum_{i=k+1}^{p}\widehat{\lambda}_i$.

De manera análoga al Teorema 2.28, se tiene que H_0 afecta el número de condiciones de ortogonalidad que Σ debe satisfacer, bajo H_0, Σ está determinada por k valores propios distintos y un valor propio común, y los k vectores propios cada uno con p componentes que corresponden a los valores propios distintos. Esto da $k+1+kp$ parámetros. Sin embargo los vectores propios están restringidos por $\frac{1}{2}k(k+1)$ condiciones de ortogonalidad, dejando $k+1+kp-\frac{1}{2}k(k+1)$ de parámetros libres. En $\overline{\underline{\Theta}}$ el número de parámetros libres es $\frac{1}{2}p(p+1)$, por lo tanto, la variable aleatoria $-2log\lambda(X)$ bajo H_0 es asintóticamente distribuida como una variable aleatoria χ^2 con grados de libertad igual a

la diferencia entre el número de parámetros independientes en $\overline{\underline{\Theta}}$ y el número en $\overline{\underline{\Theta}}_0$, es decir:

$$\frac{1}{2}p(p+1) - (k+1+kp-\frac{1}{2}k(k+1)) = \frac{1}{2}(p^2+p-2k-2-2kp+k^2+k)$$
$$= \frac{1}{2}(p(p-k-1)-k(p-k-1)+2(p-k-1)) = \frac{1}{2}(p-k+2)(p-k-1).$$

La prueba de razón de verosimilitud (PRV) de tamaño α para contrastar H_0 contra H_1 tiene región de rechazo $R = \{X| -2log(\lambda(X)) < c\}$, donde c es determinada de manera que:

$$\begin{aligned} \alpha &= sup_{\underline{\theta}\in\overline{\underline{\Theta}}_0} P_{\underline{\theta}}(X \in R) = P(-2log(\lambda(X)) < c|(\overline{\underline{x}}, \widehat{\Sigma}_0)) \\ &= P\left(\chi^2_{\frac{1}{2}(p-k+2)(p-k-1)} < c\right) = 1 - P\left(\chi^2_{\frac{1}{2}(p-k+2)(p-k-1)} \geqslant c\right) \end{aligned}$$

y se sigue que $1-\alpha = P\left(\chi^2_{\frac{1}{2}(p-k+2)(p-k-1)} \geqslant c\right)$, por lo tanto:

$$c = \chi^2_{1-\alpha;\frac{1}{2}(p-k+2)(p-k-1)}.$$

□

Teorema 2.30. *Si $\underline{X}$ se distribuye $N_p(\underline{\mu}, \Sigma)$, y $\mathbb{X}_{n\times p}$ es la matriz de la muestra. La prueba de hipótesis que contrasta que r valores propios de Σ son iguales a un valor es H_0 : $\lambda_{q+1} = \ldots = \lambda_{q+r} = \lambda$. El estadístico de prueba es:*

$$-2log\lambda = n(p-k)(a_0/\lambda - 1 - log\,(g_0/\lambda)),$$

donde $\widehat{\lambda}_i$ es el i-ésimo valor propio de S,

$$a_0 = r^{-1}\sum_{i=q+1}^{q+r}\widehat{\lambda}_i \text{ y } g_0 = \left(\prod_{i=q+1}^{q+r}\widehat{\lambda}_i\right)^{1/r}.$$

Dicho estadístico, bajo la hipótesis nula, sigue una distribución χ^2 con $\frac{1}{2}(r+2)(r-1)$ grados de libertad, asintóticamente. Así, la prueba de razón de verosimilitud (PRV) de tamaño α rechazará la hipótesis nula cuando $-2log\lambda < \chi^2_{1-\alpha;\frac{1}{2}(r+2)(r-1)}$.

Demostración. En la distribución normal multivariada, el espacio paramétrico es $\overline{\underline{\Theta}} = \mathbb{R}^p \times \{\Sigma \in M_{p\times p}(\mathbb{R}_+)|\Sigma \text{ es simétrica}\}$, la hipótesis nula es $H_0 : \underline{\theta} \in \overline{\underline{\Theta}}_1 = \{(\underline{\mu}, \Sigma) \in$

$\overline{\Theta}|\lambda_{q+1} = \ldots = \lambda_{q+r} = \lambda,$ donde $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_q \geqslant \lambda_{q+1} \geqslant \cdots \geqslant \lambda_{q+r} \geqslant \lambda_{q+r+1} \geqslant \ldots \geqslant \lambda_p$ son los valores propios de $\Sigma\}$ y $H_1 : \underline{\theta} \in \overline{\Theta}_1 = \overline{\Theta} - \overline{\Theta}_0$. Por el Teorema C.9 se tiene que sus estimadores de máxima verosimilitud son $\widehat{\underline{\mu}} = \overline{\underline{x}}$ y $\widehat{\Sigma} = S$ respectivamente.

Bajo la hipótesis nula, por el teorema de la descomposición espectral (Teorema B.7) se tiene que $\Sigma = \Gamma\Lambda\Gamma'$, donde Γ y $\Lambda = diag(\lambda_1, \ldots, \lambda_q, \lambda_{q+1}, \ldots, \lambda_{q+r}, \lambda_{q+r+1}, \ldots, \lambda_p)$ son las matrices de vectores propios y valores propios de Σ, y se cumple que $\lambda_{q+1} = \ldots = \lambda_{q+r} = \lambda$. Además, por el Teorema 2.23 se sigue que los estimadores de máxima verosimilitud de Γ y Λ son G y L_0, donde G es una matriz ortogonal cuyas columnas son vectores propios estandarizados de S, y $L_0 = diag(\widehat{\lambda_1}, \ldots, \widehat{\lambda_q}, \lambda, \ldots, \lambda, \widehat{\lambda_{q+r+1}}, \ldots, \widehat{\lambda_p})$, donde $\widehat{\lambda_i}$ es el i-ésimo valor propio de S. Por lo que el estimador de máxima verosimilitud para Σ es $\widehat{\Sigma}_0 = GL_0G'$ y el estimador para $\underline{\mu}$ es $\widehat{\underline{\mu}} = \overline{\underline{x}}$.

Los estimadores de máxima verosimilitud bajo $\overline{\Theta}$ son $\widehat{\underline{\mu}} = \overline{\underline{x}}$ y $\widehat{\Sigma} = S$. Por el teorema de la descomposición espectral (Teorema B.7) se tiene que $S = GLG'$, donde G es una matriz ortogonal cuyas columnas son vectores propios estandarizados de S, y $L = diag(\widehat{\lambda_1}, \ldots, \widehat{\lambda_k}, \widehat{\lambda_{k+1}}, \ldots, \widehat{\lambda_p})$ tiene en su diagonal a los valores propios de S. Nótese que el estimador de máxima verosimilitud bajo H_0 y bajo H_1 para Γ es el mismo, la matriz de vectores propios estandarizados de S.

Usando el Teorema C.14 se sigue que la prueba de razón de verosimilitud para contrastar H_0 contra H_1 está dada por $-2log\lambda = np(a - logg - 1)$, donde a y g son la media aritmética y media geométrica de los valores propios de $\widehat{\Sigma}_0^{-1}S$, luego:

$$\begin{aligned}
&\widehat{\Sigma}_0^{-1}S = (GL_0G')^{-1}GLG' = G'^{-1}L_0^{-1}G^{-1}GLG' = GL_0^{-1}LG' \\
&= Gdiag(\widehat{\lambda_1}, \ldots, \widehat{\lambda_q}, \lambda, \ldots, \lambda, \widehat{\lambda_{q+r+1}}, \ldots, \widehat{\lambda_p})^{-1}diag(\widehat{\lambda_1}, \ldots, \widehat{\lambda_q}, \widehat{\lambda_{q+1}}, \ldots, \\
&\quad \widehat{\lambda_{q+r}}, \widehat{\lambda_{q+r+1}}, \ldots, \widehat{\lambda_p})G' \\
&= Gdiag(1, \ldots, 1, \widehat{\lambda_{q+1}}/\lambda, \ldots, \widehat{\lambda_{q+r}}/\lambda, 1, \ldots, 1)G',
\end{aligned}$$

por lo que $1, \ldots, 1, \widehat{\lambda_{q+1}}/\lambda, \ldots, \widehat{\lambda_{q+r}}/\lambda, 1, \ldots, 1$ son los valores propios de $\widehat{\Sigma}_0^{-1}S$, así:

$$a = \left(q + \sum_{i=q+1}^{q+r} \widehat{\lambda}_i/\lambda + p - (q+r) \right) /p = (p-r)/p + \frac{1}{\lambda p} \sum_{i=q+1}^{q+r} \widehat{\lambda}_i$$

y

$$g = \sqrt[p]{\prod_{i=q+1}^{q+r} \left(\widehat{\lambda}_i/\lambda \right)} = \left(\prod_{i=q+1}^{q+r} \widehat{\lambda}_i/\lambda^r \right)^{1/p} = (g_0^r/\lambda^r)^{1/p} = (g_0/\lambda)^{r/p},$$

donde $g_0 = \left(\prod_{i=q+1}^{q+r} \widehat{\lambda}_i\right)^{1/r}$, por lo que:

$$\begin{aligned} -2log\lambda &= np(a - logg - 1) \\ &= np((p-r)/p + (\lambda p)^{-1} \sum_{i=q+1}^{q+r} \widehat{\lambda}_i - log\left((g_0/\lambda)^{r/p}\right) - 1) \\ &= n(p - r - p + \lambda^{-1} \sum_{i=q+1}^{q+r} \widehat{\lambda}_i - rlog\,(g_0/\lambda)) \\ &= n(-r + \lambda^{-1} r a_0 - rlog\,(g_0/\lambda)) \\ &= nr(a_0/\lambda - 1 - log\,(g_0/\lambda)), \end{aligned}$$

donde $a_0 = r^{-1} \sum_{i=q+1}^{q+r} \widehat{\lambda}_i$.

De manera análoga al Teorema 2.28, se tiene que H_0 afecta el número de condiciones de ortogonalidad que Σ debe satisfacer, bajo H_0; Σ está determinada por $p - r$ valores propios distintos y un valor propio común, y los $p - r$ vectores propios cada uno con p componentes que corresponden a los valores propios distintos. Esto da $(p-r) + 1 + (p-r)p$ parámetros. Sin embargo los vectores propios están restringidos por $\frac{1}{2}(p-r)((p-r)+1)$ condiciones de ortogonalidad, dejando $(p-r) + 1 + (p-r)p - \frac{1}{2}(p-r)((p-r)+1)$ de parámetros libres. En $\overline{\underline{\Theta}}$, el número de parámetros libres es $\frac{1}{2}p(p+1)$, por lo tanto, la variable aleatoria $-2log\lambda(X)$ bajo H_0 es asintóticamente distribuida como una variable aleatoria χ^2, con grados de libertad igual a la diferencia entre el número de parámetros independientes en $\overline{\underline{\Theta}}$ y el número en $\overline{\underline{\Theta}}_0$, es decir:

$$\begin{aligned} &\tfrac{1}{2}p(p+1) - ((p-r) + 1 + (p-r)p - \tfrac{1}{2}(p-r)((p-r)+1)) \\ &= \tfrac{1}{2}(p^2 + p - 2(p-r) - 2 - 2(p-r)p + (p-r)^2 + (p-r)) \\ &= \tfrac{1}{2}(p(p-(p-r)-1) - (p-r)(p-(p-r)-1) + 2(p-(p-r)-1)) \\ &= \tfrac{1}{2}(p-(p-r)+2)(p-(p-r)-1) \\ &= \tfrac{1}{2}(r+2)(r-1). \end{aligned}$$

La prueba de razón de verosimilitud (PRV) de tamaño α para contrastar H_0 contra H_1 tiene región de rechazo $R = \{X| - 2log(\lambda(X)) < c\}$, donde c puede obtenerse a partir de:

$$\begin{aligned} \alpha &= sup_{\underline{\theta} \in \overline{\underline{\Theta}}_0} P_{\underline{\theta}}(X \in R) = P(-2log(\lambda(X)) < c | (\underline{\overline{x}}, \widehat{\Sigma}_0)) \\ &= P\left(\chi^2_{\frac{1}{2}(r+2)(r-1)} < c\right) = 1 - P\left(\chi^2_{\frac{1}{2}(r+2)(r-1)} \geqslant c\right) \end{aligned}$$

y se sigue que $1 - \alpha = P\left(\chi^2_{\frac{1}{2}(r+2)(r-1)} \geqslant c\right)$, por lo tanto $c = \chi^2_{1-\alpha;\frac{1}{2}(r+2)(r-1)}$. □

2.10.3. Hipótesis concernientes a la matriz de correlación

El análisis de componentes principales transforma un conjunto de variables correlacionadas en un nuevo conjunto de variables no correlacionadas. Si las variables ya están casi no correlacionadas, entonces nada se ganaría al llevar a cabo un ACP. En este caso, la dimensionalidad real de los datos es igual al número de variables respuesta medidas y no es posible examinar los datos en un espacio con un número reducido de dimensiones.

Si se cree que los datos provienen de una distribución normal multivariada, puede probarse si las variables respuesta son independientes (es decir, no correlacionadas), antes de realizar un ACP. Esto se puede llevar a cabo al probar que $P = I$ o, lo que equivale, al probar que Σ es una matriz diagonal, también es equivalente a la hipótesis de que todos los valores propios de P son iguales, si no se puede rechazar $P = I$, no se debe realizar un ACP.

Teorema 2.31. *Si $\underline{X}$ se distribuye $N_p(\underline{\mu}, \Sigma)$, y $\mathbb{X}_{n\times p}$ es la matriz de la muestra. Un estadístico de prueba de relación de probabilidad para probar $H_0 : P = I$ se expresa por $-2log\lambda = -nlog(|R|)$. Para valores grandes de n, la prueba de razón de verosimilitud (PRV) de tamaño α rechaza H_0 si $-2log\lambda < \chi^2_{1-\alpha,p(p-1)/2}$.*

Demostración. En la distribución normal multivariada, el espacio paramétrico es $\overline{\underline{\Theta}} = \mathbb{R}^p \times \{\Sigma \in M_{p\times p}(\mathbb{R}_+ \cup \{0\}) | \Sigma \text{ es simétrica}\}$, la hipótesis nula $H_0 : P = I$ es equivalente a la hipótesis $H_0 : \underline{\theta} \in \overline{\underline{\Theta}}_0 = \{(\underline{\mu}, \Sigma) \in \overline{\underline{\Theta}} | \Sigma = diag(\sigma_1^2, \sigma_2^2, \cdots, \sigma_p^2)\}$ y $H_1 : \underline{\theta} \in \overline{\underline{\Theta}}_1 = \overline{\underline{\Theta}} - \overline{\underline{\Theta}}_0$.

Bajo la hipótesis nula, la media y la varianza de cada variable son estimadas por separado, así, el estimador de máxima verosimilitud para $\underline{\mu}$ es $\widehat{\underline{\mu}} = \overline{\underline{x}}$ y el estimador para Σ es $\widehat{\Sigma}_0 = diag(s_1^2, s_2^2, \ldots, s_p^2)$. Además, por el Teorema C.9 se tiene que sus estimadores de máxima verosimilitud bajo $\overline{\underline{\Theta}}$ son $\widehat{\underline{\mu}} = \overline{\underline{x}}$ y $\widehat{\Sigma} = S$.

Usando el Teorema C.14 se sigue que la prueba de razón de verosimilitud para contrastar H_0 contra H_1 está dada por $-2log\lambda = np(a - logg - 1)$, donde a y g son la media aritmética y media geométrica de los valores propios de $\widehat{\Sigma}_0^{-1}S$, nótese que $S = DRD$,

donde R es la matriz de correlación muestral y $D = diag(s_1, s_2, \ldots, s_p)$, así:

$$\widehat{\Sigma}_0^{-1} S = (D^2)^{-1} S = D^{-1} D^{-1} D R D = D^{-1} R D,$$

luego el polinomio característico de $\widehat{\Sigma}_0^{-1} S$ es:

$$|D^{-1} R D - \lambda I| = |D^{-1}(R - \lambda I) D| = |D^{-1}||R - \lambda I||D| = |R - \lambda I|,$$

por lo que los valores propios de $\widehat{\Sigma}_0^{-1} S$ son los mismo que los valores propios de R, que es simétrica y cumple que la suma de sus valores propios es igual a su traza y el producto de sus valores propios es igual a su determinante, así:

$$a = tr(R)/p = p/p = 1 \text{ y } g = \sqrt[p]{|R|},$$

por lo que:

$$-2log\lambda = np(1 - log(\sqrt[p]{|R|}) - 1) = np(-log(\sqrt[p]{|R|})) = -nlog(|R|).$$

Dado que bajo H_0, Σ está determinada por los p valores de su diagonal que son los parámetros libres y en $\overline{\Theta}$ el número de parámetros libres es $\frac{1}{2}p(p+1)$, la variable aleatoria $-2log\lambda(X)$ bajo H_0 es asintóticamente distribuida como una variable aleatoria χ^2 con grados de libertad igual a la diferencia entre el número de parámetros independientes en $\overline{\Theta}$ y el número en $\overline{\Theta}_0$, es decir:

$$\frac{1}{2}p(p+1) - p = \frac{1}{2}(p(p+1) - 2p) = \frac{1}{2}(p(p-1)).$$

La prueba de razón de verosimilitud (PRV) de tamaño α para contrastar H_0 contra H_1 tiene región de rechazo $R = \{X| - 2log(\lambda(X)) < c\}$, donde c es determinada de manera que:

$$\begin{aligned} \alpha &= sup_{\underline{\theta} \in \overline{\Theta}_0} P_{\underline{\theta}}(X \in R) = P(-2log(\lambda(X)) < c | (\underline{\overline{x}}, \widehat{\Sigma}_0)) \\ &= P\left(\chi^2_{\frac{1}{2}(p(p-1))} < c\right) = 1 - P\left(\chi^2_{\frac{1}{2}(p(p-1))} \geqslant c\right) \end{aligned}$$

y se sigue que $1 - \alpha = P\left(\chi^2_{\frac{1}{2}(p(p-1))} \geqslant c\right)$, por lo tanto $c = \chi^2_{1-\alpha;\frac{1}{2}(p(p-1))}$. □

Box (1949) mostró que la aproximación χ^2 mejora si n se reemplaza por $n' = n - (2p+11)/6$. Así, se puede usar el estadístico $-n'log|R| \sim \chi^2_{p(p-1)/2}$ asintóticamente [11].

2.11. Algunas aplicaciones del ACP

2.11.1. Descartando variables

Supongamos que se realiza un análisis de componentes principales sobre la matriz de correlación de p variables. Inicialmente, se asume que k variables han de ser retenidas donde k es conocida y $(p-k)$ variables son descartadas. Se consideran los vectores propios correspondientes a los valores propios más pequeños, y se rechaza la variable con el coeficiente máximo (valor absoluto). Entonces se considera el siguiente valor propio más pequeño. Este proceso continúa hasta que se han considerado los valores propios $(p-k)$ más pequeños [11].

En general se descarta la variable de coeficiente máximo en valor absoluto en la componente, y que no había sido descartada previamente.

Este principio es consistente con la noción de considerar una componente con valor propio pequeño como de menor importancia y, consecuentemente, la variable que domina debería ser menos importante o redundante. Sin embargo se puede lograr una elección de k como sigue. Sea λ_0 un valor crítico (umbral) de modo que los valores propios $\lambda \leqslant \lambda_0$ pueden ser considerados como con poca contribución a los datos. En la literatura se ha recomendado $\lambda_0 = 0.70$ [11].

Observaciones

1. En lugar de usar λ_0, se puede establecer a k igual al número de componentes necesarias para tener en cuenta más que alguna proporción, α_0, de la variación total. Se ha encontrado en ejemplos que λ_0 parece ser mejor que α_0 como criterio para decidir cúantas variables rechazar, y que $\alpha_0 = 0.80$ es el mejor valor. Usando cualquiera de los dos criterios, siempre deben retenerse al menos en cuatro variables.

2. Una variante de este método de descarte de variables está dado por el siguiente procedimiento iterativo. En cada etapa del procedimiento se rechaza a la variable asociada con el valor propio más pequeño y se realiza el análisis de componentes

principales con las variables restantes, y se continúa hasta que todos los valores propios son grandes.

3. La relación entre la "insignificancia" de $Y_i = \sum \widehat{\lambda}_i X_i$ y el rechazo de una variable particular está abierta a la crítica. Sin embargo, hay estudios que indican que diferentes métodos de rechazo de variables con el fin de explicar la variación dentro de $\underline{X}$ producen en la práctica casi los mismos resultados. Por supuesto, el método seleccionado para tratar los datos de uno depende del propósito del estudio [11].

2.11.2. Relaciones estructurales

Si algunos de los valores propios de Σ (la matriz de covarianzas) o de P (la matriz de correlaciones), o de ambas, son cero, entonces algunas de las variables originales son linealmente dependientes de las otras, ya que si el valor propio λ_i es igual a cero, se tiene que la $Var(Y_i) = \lambda_i = 0$ y $E(Y_i) = 0$, luego $1 = P(Y_i = E(Y_i)) = P(\sum_{j=1}^{p} \gamma_{ji}(X_j - E(X_j)) = 0)$, además, se tiene que $\|\underline{\gamma_{(i)}}\| = 1$, por lo que algunos γ_{ji} son distintos de cero, por lo tanto algunas de las variables originales son linealmente dependientes de las otras.

Supongamos que se ha determinado que un ACP tiene d componentes principales. Entonces hay $p - d$ valores propios que son iguales a cero. Cuando hay $p - d$ valores propios iguales a cero, sus vectores propios correspondientes definen $p - d$ restricciones lineales, linealmente independientes, sobre las variables respuesta.

Definición 2.32. *Los vectores propios correspondientes a los valores propios pequeños definen las* relaciones estructurales *entre las variables originales. Las nuevas variables creadas por estas restricciones lineales se llaman* variables de relaciones estructurales.

Las relaciones estructurales nunca son únicas cuando $p - d > 1$, de modo que nunca debe intentarse dar interpretaciones físicas a las relaciones estructurales en los casos donde $p - d > 1$.

Las ecuaciones de relaciones estructurales proporcionan información acerca de cómo están relacionadas las variables entre sí, son muy útiles para examinar las relaciones

de multicolinealidad entre un conjunto de variables predictoras en los problemas de regresión [8].

2.11.3. Regresión

En el Capítulo 1 se estudió el modelo de regresión lineal múltiple; si las variables independientes de este modelo son altamente dependientes entre ellas, los estimadores de los coeficientes de regresión son muy imprecisos. En esta situación es ventajoso no considerar algunas de las variables con el fin de aumentar la estabilidad de los coeficientes.

Si se realizan n observaciones sobre una variable dependiente y p variables independientes ($\underline{y}_{n\times 1}$ y $X_{n\times p}$) y si las observaciones están centradas de modo que $\overline{y} = \overline{x_i} = 0$, $i = 1, \ldots, p$, se tiene que $S_X = n^{-1}X'X - \overline{xx}' = n^{-1}X'X$.

Supongamos que un valor propio muestral $\widehat{\lambda}$ es próximo a cero y su correspondiente vector propio es $\underline{\widehat{\gamma}}$, entonces:

$$S_X\underline{\widehat{\gamma}} - \widehat{\lambda}\underline{\widehat{\gamma}} = \underline{0} \Rightarrow X'X\underline{\widehat{\gamma}} \approx \underline{0} \Rightarrow \underline{\widehat{\gamma}}'X'X\underline{\widehat{\gamma}} = (X\underline{\widehat{\gamma}})'X\underline{\widehat{\gamma}} = \|X\underline{\widehat{\gamma}}\|^2 \approx 0 \Rightarrow X\underline{\widehat{\gamma}} \approx \underline{0}.$$

Por otro lado, la ecuación de regresión es $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, donde la media de $\underline{\varepsilon}$ es $\underline{0}$. Además, si $W = XG$ denota la transformación de las componentes principales, la ecuación de regresión se puede escribir:

$$\underline{y} = W\underline{\alpha} + \underline{\varepsilon},$$

donde $\underline{\alpha} = G'\underline{\beta}$. Si hay un cierto número de valores propios cercanos a cero, $p - k$, entonces:

$$\begin{aligned} W = XG &= X(\underline{\widehat{\gamma}_1}, \underline{\widehat{\gamma}_2}, \cdots, \underline{\widehat{\gamma}_k}, \widehat{\underline{\gamma_{k+1}}}, \cdots, \underline{\widehat{\gamma}_p}) \\ &= (X\underline{\widehat{\gamma}_1}, X\underline{\widehat{\gamma}_2}, \cdots, X\underline{\widehat{\gamma}_k}, \widehat{X\underline{\gamma_{k+1}}}, \cdots, X\underline{\widehat{\gamma}_p}) \\ &\approx (X\underline{\widehat{\gamma}_1}, X\underline{\widehat{\gamma}_2}, \cdots, X\underline{\widehat{\gamma}_k}, \underline{0}, \cdots, \underline{0}). \end{aligned}$$

En este caso, el modelo de regresión lineal se puede volver a escribir en términos de los componentes principales $\underline{\widehat{\gamma}_1}$, $\underline{\widehat{\gamma}_2}$, $\cdots$, $\underline{\widehat{\gamma}_k}$, es decir, las k componentes principales no nulas, pues:

$$X\underline{\beta} = XGG'\underline{\beta} \approx (X\underline{\widehat{\gamma}_1}, X\underline{\widehat{\gamma}_2}, \cdots, X\underline{\widehat{\gamma}_k}, \underline{0}, \cdots, \underline{0})\underline{\alpha}.$$

Dado que las columnas de W son ortogonales, los estimadores por mínimos cuadrados $\widehat{\alpha}_i$ no se alteran si algunas de las columnas de W son eliminadas de la regresión. El vector completo de estimadores de mínimos cuadrados está dado por:

$$\underline{\widehat{\alpha}} = (W'W)^{-1}W'\underline{y} \text{ y } \underline{\widehat{\varepsilon}} = \underline{y} - W\underline{\widehat{\alpha}},$$

donde:

$$\underline{\widehat{\alpha}} \;=\; (G'X'XG)^{-1}W'\underline{y} = G'(X'X)^{-1}GW'\underline{y},$$

además, como se tiene que $S_X = n^{-1}X'X$ por lo que:

$$\begin{aligned}\underline{\widehat{\alpha}} &= G'n^{-1}(n^{-1}X'X)^{-1}GW'\underline{y} = G'n^{-1}S_X^{-1}GW'\underline{y} = G'n^{-1}(GLG')^{-1}GW'\underline{y} \\ &= G'n^{-1}GL^{-1}G'GW'\underline{y} = n^{-1}L^{-1}W'\underline{y},\end{aligned}$$

entonces:

$$\widehat{\alpha}_i = \begin{cases} n^{-1}\widehat{\lambda}_i^{-1}\underline{w_{(i)}}'\underline{y}, & \text{si } \lambda_i \neq 0 \\ 0, & \text{si } \lambda_i = 0, \end{cases}$$

donde $\widehat{\lambda}_i$ es el i-ésimo valor propio de la matriz de covarianza, $S_X = n^{-1}X'X$, $i = 1, \ldots, p$. Con esto se puede ver que las componentes principales con varianzas igual a cero no explican a la variable dependiente.

Usando los resultados enunciados en el Teorema 1.24, de i) se tiene que la esperanza de $\widehat{\alpha}_i$ es α_i, además, dado que $(W'W)^{-1} = n^{-1}L^{-1}$ de ii), se sigue que la varianza de $\widehat{\alpha}_i$ es $\sigma^2/(n\underline{\lambda_i})$.

Hay otra forma de elegir algunas componentes principales en el contexto de la regresión lineal múltiple: se toman las componentes con las correlaciones máximas con la variable dependiente debido a que el propósito en la regresión es explicar a la variable dependiente. Dado que $\overline{W} = n^{-1}W'\underline{1} = n^{-1}G'X'\underline{1} = \underline{0}$, se sigue que la covarianza entre W_i y $\underline{y}$ es:

$$Cov(W_i, \underline{y}) = n^{-1}\sum_{r=1}^{n} w_{ri}y_r = n^{-1}\underline{w_{(i)}}'\underline{y} = n^{-1}\underline{y}'\underline{w_{(i)}} = n^{-1}\underline{\widehat{\gamma}_{(i)}}'X'\underline{y},$$

además, cuando $\lambda_i \neq 0$, se tiene que $Cov(W_i, \underline{y}) = \widehat{\lambda}_i\widehat{\alpha}_i$.

La varianza muestral de W_i y de $\underline{y}$ son $S_{W_i} = n^{-1}\sum_{r=1}^{n} w_{ri}^2 = n^{-1}\underline{w_{(i)}}'\underline{w_{(i)}}$ y $S_{\underline{y}} = n^{-1}\sum_{r=1}^{n} y_r^2 = n^{-1}\underline{y}'\underline{y}$. Luego la correlación muestral de W_i y de $\underline{y}$ es:

$$\rho_{W_i\underline{y}} = \frac{Cov(W_i,\underline{y})}{\sqrt{S_{W_i}S_{\underline{y}}}} = \frac{\widehat{\lambda}_i\widehat{\alpha}_i}{\sqrt{n^{-1}\underline{w_{(i)}}'\underline{w_{(i)}}n^{-1}\underline{y}'\underline{y}}} = \frac{n\widehat{\lambda}_i\widehat{\alpha}_i}{\sqrt{\underline{w_{(i)}}'\underline{w_{(i)}}\underline{y}'\underline{y}}}, \lambda_i \neq 0.$$

Si la ecuación de regresión es $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, donde $\underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 I)$, por el Teorema 1.32 se tiene que bajo la hipótesis nula, $H_0 : \alpha_i = 0$,

$$\begin{aligned} T &= \frac{\underline{a}'\widehat{\underline{\alpha}}}{S[\underline{a}'(W'W)^{-1}\underline{a}]^{1/2}} = \frac{\widehat{\alpha_i}}{S[\underline{a}'n^{-1}L^{-1}\underline{a}]^{1/2}} = \frac{\widehat{\alpha_i}}{S[n^{-1}\widehat{\lambda_i}^{-1}]^{1/2}} \\ &= \frac{\widehat{\alpha_i}(n\widehat{\lambda_i})^{1/2}}{(\widehat{\varepsilon}'\widehat{\varepsilon}/(n-(p+1)))^{1/2}} \sim t_{n-(p+1)} \end{aligned}$$

donde $\underline{a}' = (0, 0, \ldots, \underbrace{1}_{\text{i-ésimo}}, \ldots, 0, 0)$. Si $\lambda_i \neq 0$ se sigue que:

$$T = \frac{n^{-1}\widehat{\lambda}_i^{-1}\underline{w_{(i)}}'\underline{y}(n\widehat{\lambda}_i)^{1/2}}{(\widehat{\varepsilon}'\widehat{\varepsilon}/(n-(p+1)))^{1/2}} = \frac{\underline{w_{(i)}}'\underline{y}}{(n\widehat{\lambda}_i\widehat{\varepsilon}'\widehat{\varepsilon}/(n-(p+1)))^{1/2}} \sim t_{n-(p+1)} \tag{2.22}$$

independientemente de los valores verdaderos de la otras α_j. Seleccionar aquellas componentes para las cuales el estadístico (2.22) es significante es una forma sencilla de elegir qué componentes retener.

Si las componentes principales tienen un significado intuitivo natural, tal vez la mejor manera de dejar expresada la ecuación de regresión es en términos de las componentes. En otros casos, es más conveniente transformar de nuevo las variables originales [11].

Capítulo 3

El problema de la deforestación en la región transfronteriza entre Tabasco y Chiapas

3.1. Los ecosistemas base del bienestar humano

En los últimos decenios, la agenda ambiental se ha convertido en una de las más importantes para los gobiernos de los países de todo el mundo, en gran medida debido a que el cambio ambiental global (calentamiento climático, pérdida de ecosistemas y diversidad biológica) y sus problemas derivados son actualmente uno de los mayores desafíos para la humanidad. El bienestar humano, el desarrollo social y económico sustentable, la erradicación de la pobreza y la desigualdad requieren todos de la gestión del capital natural de la nación (véase Figura 3.1), es decir, de la conservación de los ecosistemas y su diversidad biológica así como del manejo sustentable de los bienes y servicios que éstos ofrecen [24], [18].

Aunque el ser humano ha modificado los ecosistemas desde hace milenios, en los últimos 50 años las transformaciones han alcanzado una magnitud y velocidad sin precedentes. En ese lapso se perdió la mitad de cubierta forestal nativa del planeta; 35 % de la extensión de manglares fue devastada; 20 % de los arrecifes coralinos han sido

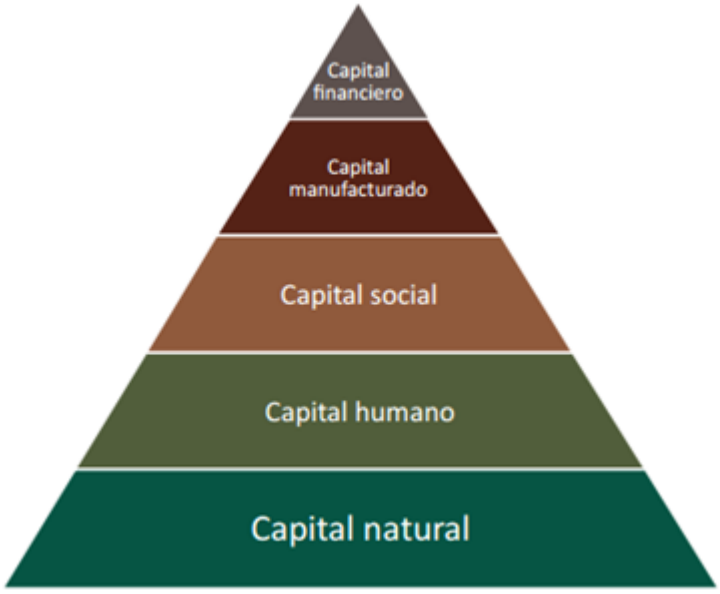

Figura 3.1: Los cinco tipos de capitales necesarios para el bienestar humano [18]

perturbados, y la demanda de agua se cuadruplicó. México no está exento de los efectos del cambio ambiental global y enfrenta una acelerada pérdida de ecosistemas y su degradación (con tasas de deforestación por encima de la media global) amenaza la disponibilidad de bienes y servicios que estos ofertan para el bienestar de la población; prácticamente todos los ecosistemas en el territorio nacional han sido afectados por las actividades humanas [18], [23].

3.2. Factores de cambio de los ecosistemas naturales

Comprender los factores y el proceso de cambio que inciden en los ecosistemas es sumamente importante para el diseño de instrumentos de intervención y estrategias de gestión que, por un lado, aseguren el acceso y disponibilidad a los bienes y servicios ecosistémicos necesarios para el bienestar humano y, por otro lado, garanticen su permanencia a largo plazo al reducir los impactos negativos sobre éstos. La interacción entre los ecosistemas naturales y los seres humanos es compleja y varía según el territorio y la escala; por un lado, las condiciones humanas cambiantes actúan como impulsores de cambios directos o indirectos en los ecosistemas y, al mismo tiempo, los cambios en

los ecosistemas provocan cambios en los niveles de bienestar humano [18].

El modelo base más utilizado para comprender la relación entre ecosistemas y bienestar humano es el propuesto por el grupo intergubernamental para la Evaluación de los Ecosistemas del Milenio (2003), cuyo énfasis se encuentra en la relación entre impulsores de cambio indirectos y directos que alteran los ecosistemas y sus servicios, y que a su vez impactan en las condiciones necesarias para el bienestar humano, el esquema que ejemplifica esta relación se puede ver en la Figura 3.2.

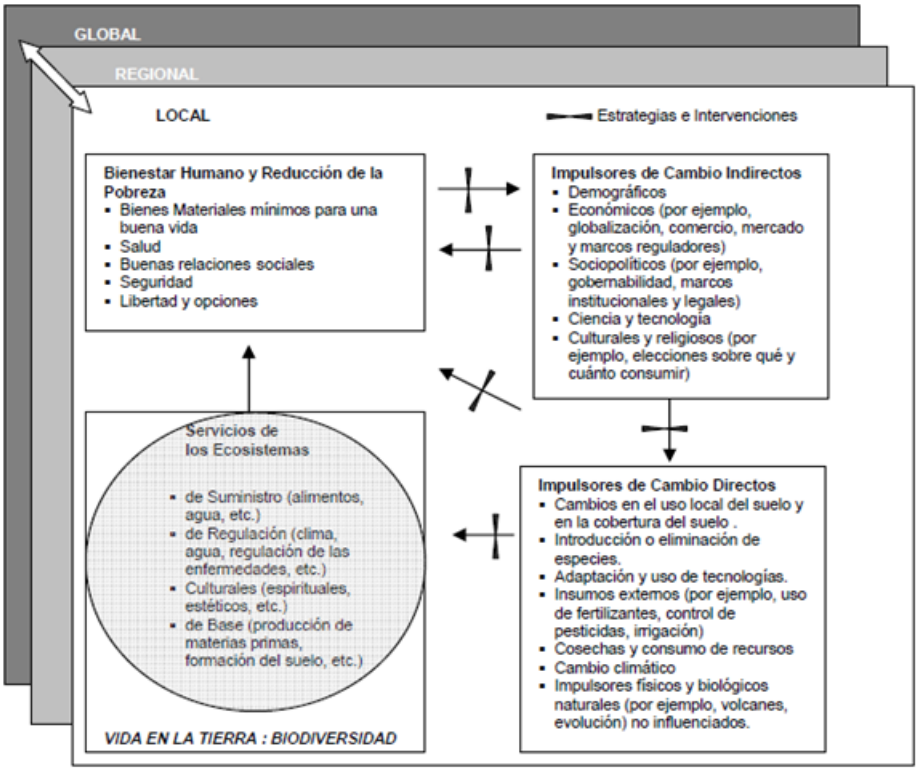

Figura 3.2: Relación entre ecosistemas naturales y bienestar humano [21].

Un impulsor es cualquier factor que altera algún aspecto de un ecosistema; los impulsores directos influyen directamente en el ecosistema, en su estructura y funciones, mientras que un indirecto puede actuar sobre uno o más impulsores directos. Sin embargo, estudios más recientes muestran que las interacciones entre los factores pueden ser más complejas, por ejemplo, dos o más factores directos pueden generar sinergias o un factor directo puede tener retroalimentaciones positivas o negativas con el ecosistema, como se observa en el Figura 3.3, propuesta por la CONABIO en su marco para

la evaluación del capital natural de México [18]. Un ejemplo de estas relaciones para el caso del país se observa en cómo la incidencia e intensidad de huracanes en el sureste mexicano (factor indirecto) interactúa con la deforestación antropogénica de manglares y la destrucción de arrecifes coralinos (factor directo), aumentando la magnitud y frecuencia de las inundaciones de la región y de los incendios forestales, alterando a su vez la estructura y diversidad de la vegetación (servicios y funciones ecosistémicas), que impactan en la seguridad de las localidades costeras [18].

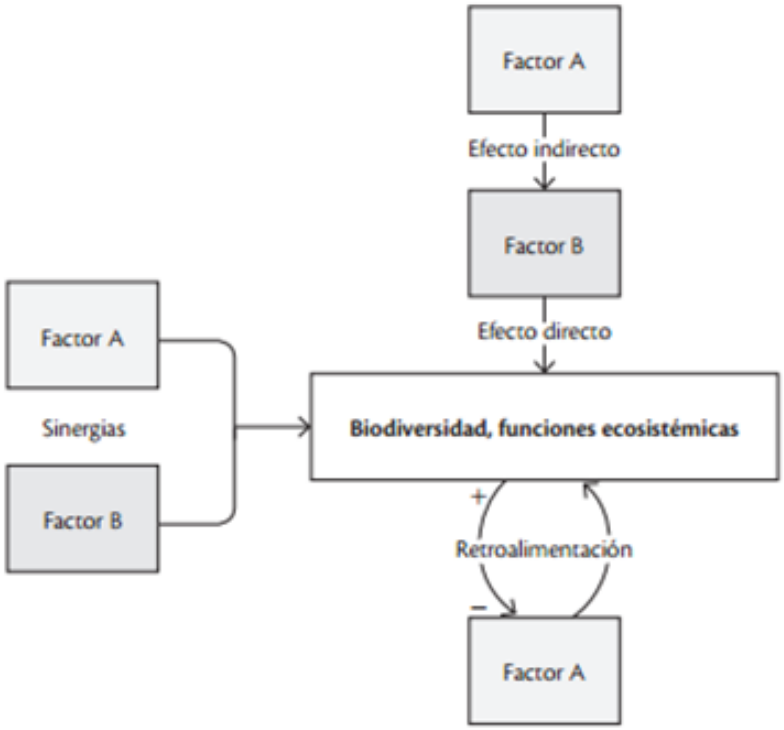

Figura 3.3: Relaciones entre los factores de cambio [18].

Adaptando el marco de Evaluación de los Ecosistemas del Milenio a la evaluación del capital natural en México, se propuso el siguiente esquema para ejemplificar las principales categorías de factores de cambio indirectos (aquí nombrados como de raíz o últimos) y de factores directos (o próximos) de la biodiversidad y los servicios ecosistémicos (véase Figura 3.4)

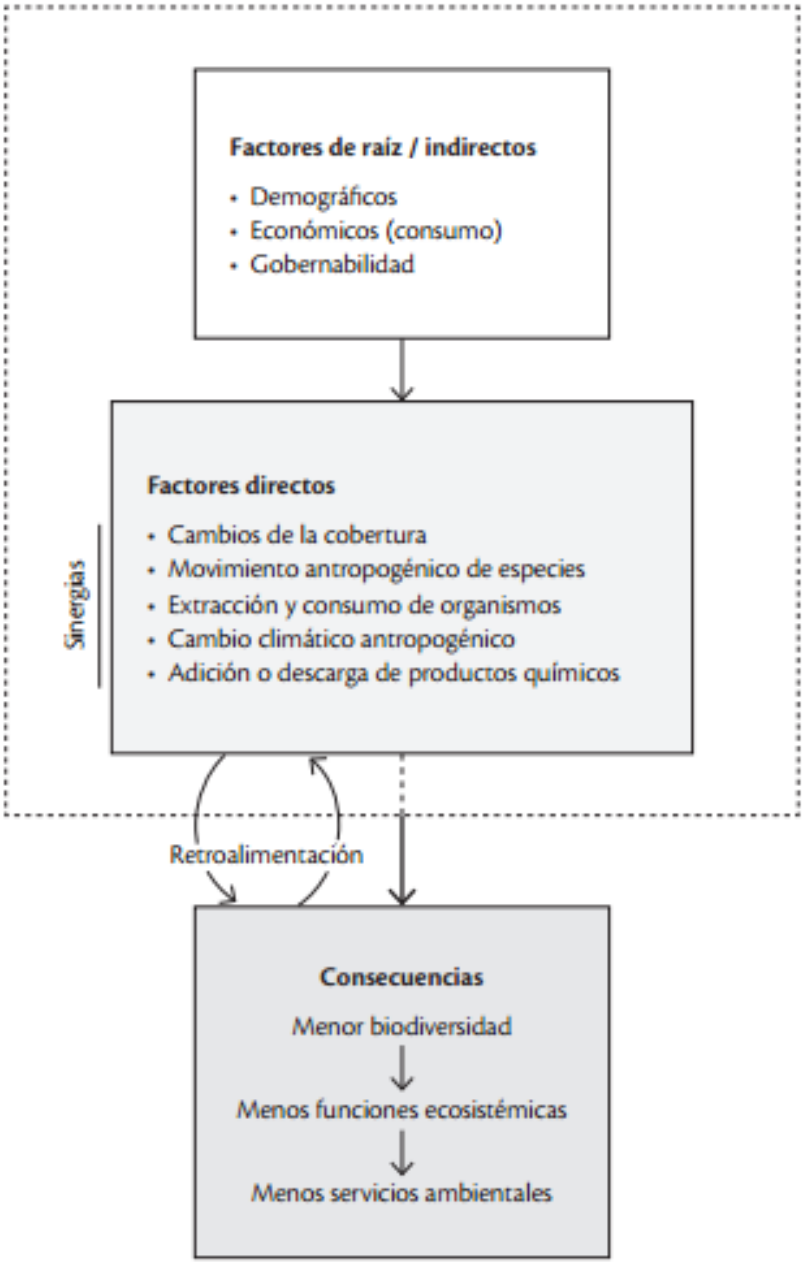

Figura 3.4: Los factores de cambio directos e indirectos de la biodiversidad y los ecosistemas [18].

3.2.1. Factores indirectos

Los factores indirectos (raíz o últimos) son las fuerzas de cambio que modifican los factores directos de cambio; su nivel de influencia se mide en función de los efectos en los factores directos, no en los ecosistemas. Varían según la escala de análisis y el

territorio que se estudia y también pueden generar sinergias entre ellos. Se resumen en cinco categorías [18]:

1. Demográficos: particularmente el tamaño de la población y la densidad poblacional, la edad, el género, la distribución espacial (rural-urbano), los movimientos migratorios y de colonización, que están asociados a otros factores, como el aumento en el consumo, o las políticas de reparto agrario y colonización de tierras “ociosas” (generalmente selvas y bosques).

2. De gobernabilidad: incluye la legislación, los planes de desarrollo y conjunto de políticas públicas de diferentes ámbitos como las de seguridad social, desarrollo urbano y planificación del uso del suelo, de reparto agrario, fomento agropecuario y forestal, desarrollo turístico y construcción de infraestructuras (transporte, petrolera, portuaria, comunicaciones). Las políticas públicas, en repetidas ocasiones, son contradictorias con la conservación de los ecosistemas y tienen un peso muy importante en la decisión de los actores para la conversión del uso del suelo [19].

3. Económicos: los modelos de desarrollo económico, el mercado, las opciones productivas, los precios de garantía de productos agroproductivos, el aumento en la demanda de insumos y bienes para la industria son factores que, combinados con las políticas públicas y otras condiciones sociales y políticas, juegan un papel importante en la toma de decisiones de actores directos e indirectos.

4. De ciencia y tecnología: como la inversión en tecnología, el conocimiento sobre uso y manejo de la biodiversidad (de los pueblos indígenas, por ejemplo) y la capacidad de adoptar tecnologías. En algunos casos la tecnología determina la magintud y velocidad de transformación de los ecosistemas; como el caso de los paquetes tecnológicos promovidos durante la “Revolución Verde” que han tenido un impacto negativo significativo en los ecosistemas.

5. Culturales: que abarcan desde las decisiones individuales sobre el consumo y la valoración de la biodiversidad hasta el conocimiento y uso de la biodiversidad; de culturas y pueblos indígenas que habitan los territorios nacionales.

3.2.2. Factores directos

Los factores directos son los procesos que modifican directamente los ecosistemas (en su estructura y funcionamiento) y la biodiversidad; son resultado de las actividades antropogénicas y son los causantes más inmediatos del deterioro ambiental. Entre ellos pueden generarse sinergias, retroalimentaciones positivas o negativas y todos pueden incidir simultáneamente en el ecosistema. Se clasifican en cinco categorías principales [21], [18]:

1. Cambio de uso del suelo y de la cobertura vegetal, que incluye procesos como la deforestación y sus consecuentes fragmentación y efecto de borde del hábitat, y los incendios forestales no naturales.

2. La sobreexplotación por la extracción y el consumo de organismos o parte de ellos, o por la extracción directa de otros recursos naturales (agua, minerales) de bienes ecosistémicos (tanto especies como recursos naturales).

3. El movimiento antropogénico de especies, como la introducción de especies exógenas al ecosistema natural (como las especies invasoras exóticas).

4. La contaminación por adición y descarga de sustancias químicas (como fertilizantes y pesticidas) que alteran los ciclos biogeoquímicos naturales

5. El cambio climático antropogénico con sus efectos en la temperatura y la precipitación, que impactan en la distribución de especies, la cobertura vegetal y la producción agropecuaria, entre otras.

3.2.3. El cambio de uso de suelo y alteración de la cobertura vegetal

El cambio en la cobertura vegetal (destrucción y transformación del hábitat), seguido por la sobreexplotación de recursos y la presencia de especies invasoras o de contaminantes ha constituido el factor de mayor impacto sobre la mayoría de los ecosistemas terrestres en el territorio mexicano. En 2002, la superficie de vegetación primaria era

solamente el 50 % de la superficie original, mientras que para los ecosistemas terrestres arbolados (selvas, bosques, manglares), el porcentaje fue de 38 %. La supresión de millones de hectáreas de vegetación natural constituye la mayor fuente de pérdida de biodiversidad (aumenta la extinción de especies) y de los ecosistemas, con consecuencias generalmente irreversibles. Afecta directamente a la provisión de bienes y servicios ecosistémicos; altera el clima local y regional, el ciclo hidrológico y los ciclos biogeoquímicos (incidencia en el cambio ambiental global), y aumenta la vulnerabilidad ante eventos hidrometeorológicos extremos. Desafortunadamente, pese a las medidas tomadas para frenar la deforestación, las tendencias y ritmos de transformación (aunque han disminuido) continúan con tendencias negativas para los ecosistemas terrestres y, en el caso del territorio mexicano, con impactos muy fuertes para las selvas (húmedas y secas) y para los bosques templados [23].

De todos los procesos de cambio de uso del suelo que impactan en la transformación de los ecosistemas terrestres del país, el más importante ha sido la expansión de la frontera agropecuaria, seguida de la construcción de infraestructuras (de transporte, comunicaciones, de riego, petrolera) y la urbanización [19]. La sustitución por pastizales para la actividad ganadera ha predominado en la zona de selvas húmedas, en tanto que la conversión a terrenos agrícolas ha sido más importante en las zonas de selvas secas, en lo matorrales xerófilos y en los bosques templados. En estos procesos, el papel del estado mexicano ha sido fundamental, pues, expresado en planes de desarrollo y políticas públicas, ha movilizado históricamente importantes recursos económicos y humanos y ha marcado la historia de diferentes regiones del país [23].

Comprender las fuerzas de cambio que han llevado a la transformación en la cobertura vegetal en los diferentes territorios locales y regionales del país no es tarea fácil, dada la complejidad de relaciones y la diversidad cultural y eco-geográfica, sin embargo, es fundamental no sólo para detener el proceso de degradación y deforestación de los ecosistemas, sino como base para la planeación de un desarrollo territorial auténticamente sustentable [19].

3.3. El caso de la región transfronteriza entre Tabasco y Chiapas

Características y problematización

Los Estados de Tabasco y Chiapas (Figura 3.5), en el sureste de México, comparten junto con Guatemala una de las regiones más importantes de Mesoamérica: la cuenca hidrográfica de los ríos Grijalva y Usumacinta, cuya riqueza biológica y cultural es invaluable. Por su posición geográfica, esta cuenca ha sido afectada históricamente por fenómenos meteorológicos que han ocasionado deslaves e inundaciones que impactan año con año a los asentamientos humanos y las actividades productivas. Es, además, una de las regiones más vulnerables ante los efectos de cambio climático, agravado por el deterioro de los ecosistemas y la modificación en la dinámica hidrológica de la cuenca [17].

La sección que corresponde a la frontera entre los Estados de Tabasco y Chiapas se encuentra dentro de la subdivisión denominada Parte baja de la cuenca del Río Grijalva, cuyos problemas principales son las constantes inundaciones en las zonas planas, debido a la alta precipitación pluvial de la región, a la deforestación de las partas altas y serranías y a las alteraciones de los cauces hidrológicos. Basta recordar las terribles consecuencias de las inundaciones que afectaron Villahermosa en el año 2009 y de las que cada año causan muertes humanas y pérdidas materiales para entender la necesidad de comprender la complejidad de relaciones que ocurren en este territorio [17].

Desafortunadamente, el sureste mexicano y en particular esta región, en los últimos años se ha visto sometida a transformaciones muy drásticas en sus ecosistemas forestales; si bien aún se encuentran remanentes de bosques, selvas, humedales y manglares, éstos cada vez ocupan una menor superficie y sobresalen cada vez más en el paisaje los pastizales inducidos y la agricultura, situación contraria al potencial productivo de la región y a la vocación del suelo, que es de conservación y aprovechamiento forestal sustentable [17].

Dada esta tendencia, los gobiernos de los Estado de Tabasco y Chiapas están coordinando un *Programa de Ordenamiento Ecológico Regional para los estados de Tabasco*

y Chiapas, cuyo eje central es la elaboración de un plan de manejo sustentable de los ecosistemas forestales presentes en la región transfronteriza a ambos estados, con dos objetivos centrales: a) detener la deforestación y degradación ambiental, y recuperar la cobertura forestal con el fin de frenar la erosión de suelo y aumentar la recarga de agua en las partes altas y disminuir el riesgo de inundación en las partes bajas, y b) mejorar las condiciones de la población local, al proponer estrategias de manejo sustentables que conserven y valoren los ecosistemas y la biodiversidad y al mismo tiempo generen ingresos productivos capaces de sostenerse a largo plazo y de esta manera contribuir a la disminución de la pobreza y desigualdad social [17].

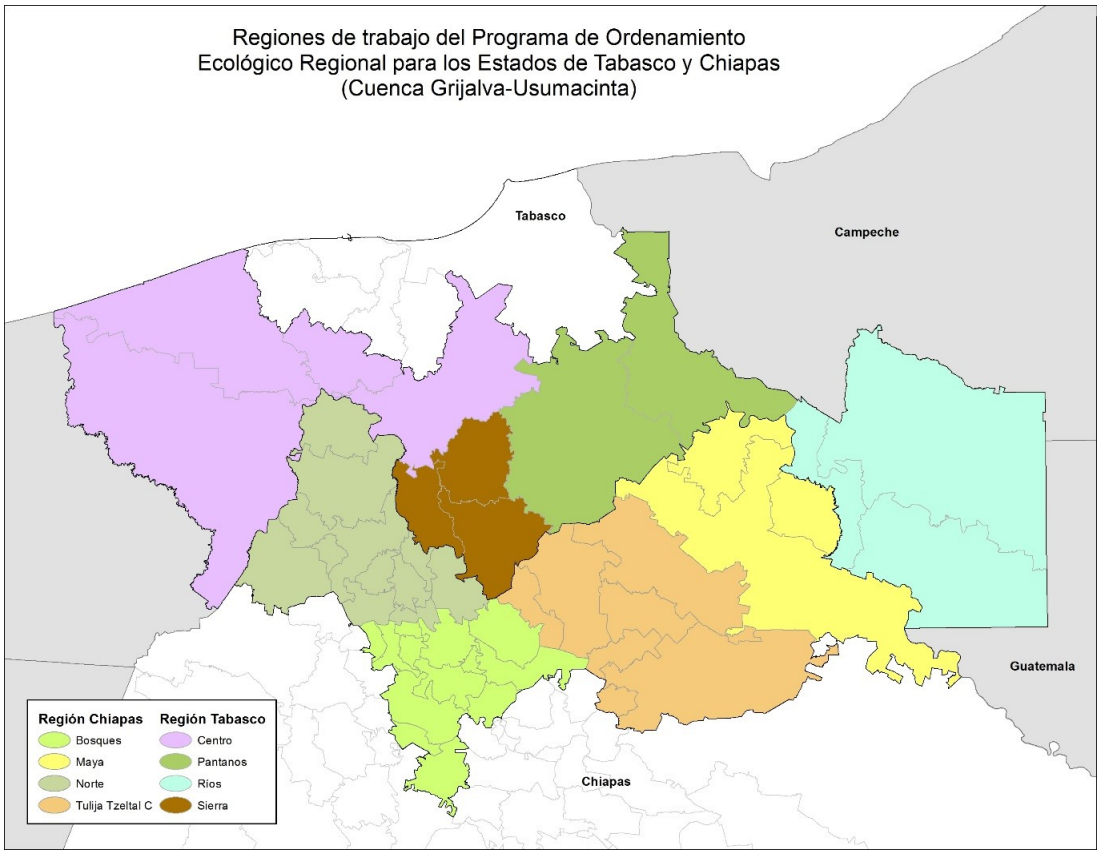

Figura 3.5: Regiones de trabajo del Programa de Ordenamiento Ecológico para los estados de Tabasco y Chiapas [17].

Delimitación de la región transfronteriza entre Tabasco y Chiapas

Si bien la extensión del territorio conocida como frontera Tabasco-Chiapas es flexible y responde a las diferentes dinámicas territoriales que se busquen analizar (por ejemplo,

la migración o la identidad cultural o las cuencas hidrológicas), para efectos del *Programa del Ordenamiento Ecológico*, los análisis se realizaron en 47 municipios, organizados en 8 regiones, correspondientes a los dos estados, como se observa en Tabla 3.1 [17].

Tabla 3.1. *Regiones y Municipios de trabajo de la región Tabasco-Chiapas.*

Región	***Municipios***
Bosques (Chiapas)	*Bochil, El Bosque, Huitiupán, Ixtapa, Jitotol, Pantepec, Pueblo Nuevo Solistahuacán, Rayón, Simojovel, Soyaló, Tapalapa, Tapilula, San Andrés Duraznal*
Maya (Chiapas)	*Catazajá, La Libertad y Palenque*
Norte (Chiapas)	*Amatán, Chapultenango, Ixhuatán, Ixtacomitán, Ixtapangajoya, Juárez, Ostuacán, Pichucalco, Reforma, Solosuchiapa, Sunuapa y Tecpatán*
Tulijá Tzeltal Chol (Chiapas)	*Chilón, Sabanilla, Salto de Agua, Sitalá, Tila, Tumbalá y Yajalón*
Centro (Tabasco)	*Cárdenas, Centro, Cunduacán y Huimanguillo*
Pantanos (Tabasco)	*Jonuta y Macuspana*
Ríos (Tabasco)	*Balancán, Emiliano Zapata y Tenosique*
Sierra (Tabasco)	*Jalapa, Tacotalpa y Teapa*

Fuente: Elaborado por M. Poisot-Cervantes con base en [17] (documento interno de trabajo sin publicar).

3.4. Principales factores de cambio de los ecosistemas forestales en la región transfronteriza entre Tabasco y Chiapas

Puesto que la conservación de los ecosistemas forestales de la región transfronteriza Tabasco-Chiapas es fundamental para el bienestar de la población local, comprender cuáles son las principales fuerzas de cambio (directas e indirectas) que afectan a estos ecosistemas es sumamente importante, sobre todo para la adecuada elaboración del

Ordenamiento Ecológico y de sus consecuentes programas que se aplicarán al territorio. La presente tesis pretende aportar a dicho análisis, en tanto busca determinar qué fuerzas de cambio inciden en mayor grado en la expansión de la frontera agropecuaria (principal factor de cambio directo de los ecosistemas forestales en la zona).

Si bien se han realizado diferentes estudios a nivel internacional y nacional para determinar las fuerzas principales que inciden en la deforestación, midiendo los efectos positivos o negativos de diferentes factores, tanto económicos (el precio de los productos agrícolas) como institucionales (régimen de propiedad de la tierra, la organización comunitaria agraria) y, más recientemente, de las políticas públicas (programas como PROCAMPO), la evidencia ha demostrado que no existe un solo factor o combinación de factores que permitan explicar todas las tendencias de cambio a nivel local o regional, puesto que además éstos cambian en el tiempo, por tanto los análisis regionales y locales no dejan de ser necesarios si se quiere contar con información oportuna y adaptada localmente [22].

Partiendo de la revisión y análisis propuestos por López [22], y de los análisis realizados sobre el estado de conservación del capital natural de México ([23], [18]) y el programa especial de gestión en zonas de alta biodiversidad [19], los factores indirectos seleccionados para analizar la región Tabasco-Chiapas fueron (Figura 3.6):

- El tamaño de la población, principalmente de la población joven, que combinado con otros factores, como la falta de empleos y el aumento en la brecha de la desigualdad social y económica, se ha convertido en un factor de presión para la deforestación, principalmente de los bosques y selvas anteriormente destinados a la conservación o de uso común en los ejidos y comunidades indígenas y también sobre las áreas naturales protegidas.

- El índice de marginación, como un indicador de la pobreza, condición social que amplía la presión sobre los ecosistemas y recursos naturales, sobre todo ante la falta de planeación para el manejo sustentable y de oportunidades de empleo e ingresos que valoren la biodiversidad.

- La superficie total de los municipios, como una limitante a la expansión agrope-

cuaria, ya sea por el límite del municipio o porque las mejores tierras ya han sido transformadas en las décadas pasadas.

- Los subsidios al campo, que incentivan directamente la transformación de ecosistemas naturales en pastizales o áreas de cultivos, generando contradicción con otras políticas públicas que buscan la conservación (como el pago por servicios ambientales). Particularmente del programa PROCAMPO por su amplia cobertura poblacional y temporal, y por ser el programa de SAGARPA con mayor presupuesto asignado.

Se analizaron dichas variables en su efecto sobre el total de hectáreas (por municipio) que fueron transformadas de ecosistemas forestales en cultivos agrícolas entre los años 1993-2002; 2002-2007, y 2007-2011 (que corresponden a las series II, III, IV y V de las cartas de uso de suelo y cobertura vegetal del INEGI), y cuyas estimaciones corresponden al análisis realizado por J. Galeana-Pizaña [20] para todos los municipios del país.

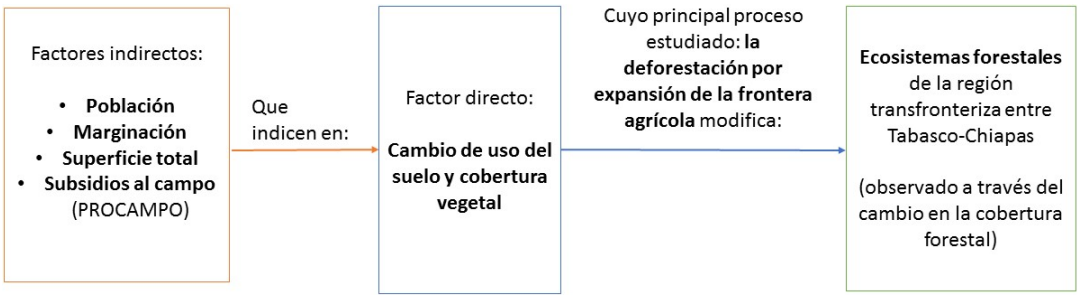

Figura 3.6: Modelo del problema planteado.

3.5. Descripción del análisis

El objetivo de este análisis es determinar si existe un modelo de regresión lineal múltiple que explique adecuadamente la relación entre la expansión de la frontera agrícola, como variable dependiente, y la superficie total; el *Programa de Apoyos Directos al Campo* (PROCAMPO); el total de población, y el índice de marginación, como variables independientes o regresoras, en tres periodos diferentes (1993-2002, 2003-2007, y

2008-2011). Para realizar los análisis estadísticos pertinentes, se utilizó el *software* R. Además, el proceso con el que se lleva a cabo el análisis en los tres periodos consta de:

- Análisis exploratorio

  Se realiza un Análisis de Componentes Principales, cuyo diagrama de dispersión de las dos primeras componentes principales muestra un municipio (municipio Centro) lejos de la dispersión de los demás municipios considerados, por lo cual no se considera en la siguiente fase del análisis; luego, se realiza otro Análisis de Componentes Principales a las puntuaciones de los 44 municipios restantes y se enuncian algunas interpretaciones.

- Análisis de regresión

  Primero se realizan diferentes análisis de regresión lineal simple, considerando a cada una de las variables independientes; luego se halla el mejor modelo mediante regresión paso a paso y, por último, se realiza un Análisis de regresión lineal múltiple mediante el Análisis de Componentes Principales.

El primer periodo que se analiza es 1993-2002; para ello se usan los datos de la Tabla A.2, que proporciona la superficie total (SupTotal); el promedio de productores beneficiados (PromProd); el monto del apoyo que reciben los habitantes (Monto); el promedio de productores beneficiados con más de 10 hectáreas (PromProd10); el monto del apoyo a productores con más de 10 hectáreas (Monto10); la población total (Pob); el índice de marginación (Marg), y la expansión de la frontera agrícola (Exp) de los municipios de interés (Tabla A.1) en este periodo. Dado que del municipio San Andrés Duraznal (SAD) no se cuenta con la información de PROCAMPO, no se incluirá en el análisis. El municipio Tecpatán tampoco se incluirá en el análisis debido a que se incorporó después al *"Programa de Ordenamiento Ecológico Regional"*.

Como el objetivo es determinar si existe una relación lineal entre la expansión de la frontera agrícola (Exp), la superficie total (SupTotal), PROCAMPO (PromProd, Monto, PromProd10 y Monto10), el total de población (Pob) y el índice de marginación (Marg), se considera la expansión de la frontera agrícola (Exp) como la variable dependiente y al resto como las variables independientes. De manera análoga se analizan los pe-

riodos 2003-2007 y 2008-2011, segundo y tercer periodo respectivamente. Los datos correspondientes a estos periodos se encuentran en las Tablas A.3 y A.4.

3.6. Análisis exploratorio

En el primer periodo, la dispersión de las variables independientes bajo estudio (Figura 3.7) medidas en los 45 municipios considerados muestra que existe una relación lineal entre las variables PromProd y Monto, y también entre PromProd10 y Monto10, lo cual se confirma al observar las correlaciones de estas variables (Tabla 3.2), pues sus correlaciones son 0.932 y 0.939 respectivamente; además se puede notar que existen altas correlaciones entre las variables independientes.

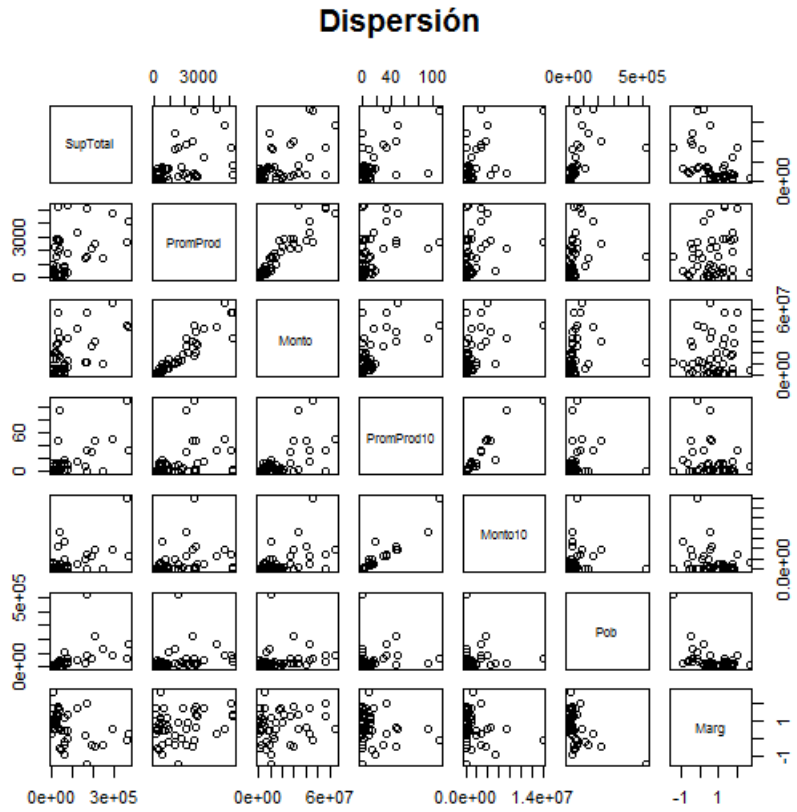

Figura 3.7: Dispersión de las variables independientes bajo estudio medidas en los 45 municipios considerados, en el primer periodo.

Tabla 3.2. *Correlaciones de las variables independientes medidas en los 45 municipios considerados, en el primer periodo.*

	SupTotal	*PromProd*	*Monto*	*PromProd10*	*Monto10*	*Pob*	*Marg*
SupTotal	*1.000*	*0.489*	*0.565*	*0.563*	*0.560*	*0.509*	*-0.386*
PromProd		*1.000*	*0.932*	*0.369*	*0.281*	*0.225*	*0.212*
Monto			*1.000*	*0.596*	*0.501*	*0.173*	*0.144*
PromProd10				*1.000*	*0.939*	*0.092*	*-0.165*
Monto10					*1.000*	*0.054*	*-0.230*
Pob						*1.000*	*-0.509*
Marg							*1.000*

Debido a la alta correlación que existe entre las variables independientes, como análisis exploratorio se realiza un análisis de componentes principales a los datos estandarizados de los 45 municipios bajo estudio, ya que las variables no están medidas en las mismas unidades ni en unidades comparables además de que no tienen varianzas semejantes (Tabla 3.3).

El análisis de componentes principales indica que las primeras tres componentes principales explican el 90.3 % de la variabilidad total (Tabla 3.4), además, su gráfica de sedimentación (Figura 3.8 a)) muestra que sólo estas componentes tienen valor propio mayor que 1, es decir, en un análisis posterior será suficiente usar las primeras tres componentes principales en lugar de las siete variables independientes, ya que existe multicolinealidad. El diagrama de dispersión de las dos primeras componentes principales (Figura 3.8 b)) muestra al municipio Centro (Cen) lejos de la dispersión del resto de municipios, quizás debido a que en Centro (Cen) se encuentra la ciudad de Villahermosa y su industria petroquímica, por lo que los principales procesos de deforestación se encuentran asociados a la urbanización y por lo tanto no se considera en el análisis posterior de este periodo.

Tabla 3.3. *Varianzas de las variables independientes bajo estudio medidas en los 45 municipios considerados, en el primer periodo.*

Variable	***Varianza***	***Variable***	***Varianza***
SupTotal	$8.604700 * 10^{9}$	*Monto10*	$6.380870 * 10^{12}$
PromProd	$2.333393 * 10^{6}$	*Pob*	$7.076719 * 10^{9}$
Monto	$3.016635 * 10^{14}$	*Marg*	$7.794386 * 10^{-1}$
PromProd10	$5.570225 * 10^{2}$		

Tabla 3.4. *Importancia de los componentes principales de las variables independientes bajo estudio medidas en los 45 municipios considerados, en el primer periodo, donde DE, PV y PA significan desviación estándar, proporción de varianza y proporción acumulada, respectivamente.*

	CP1	*CP2*	*CP3*	*CP4*	*CP5*	*CP6*	*CP7*
DE	*1.8583*	*1.2920*	*1.0946*	*0.55655*	*0.53283*	*0.23825*	*0.1694*
PV	*0.4933*	*0.2385*	*0.1712*	*0.04425*	*0.04056*	*0.00811*	*0.0041*
PA	*0.4933*	*0.7318*	*0.9030*	*0.94723*	*0.98779*	*0.99590*	*1.0000*

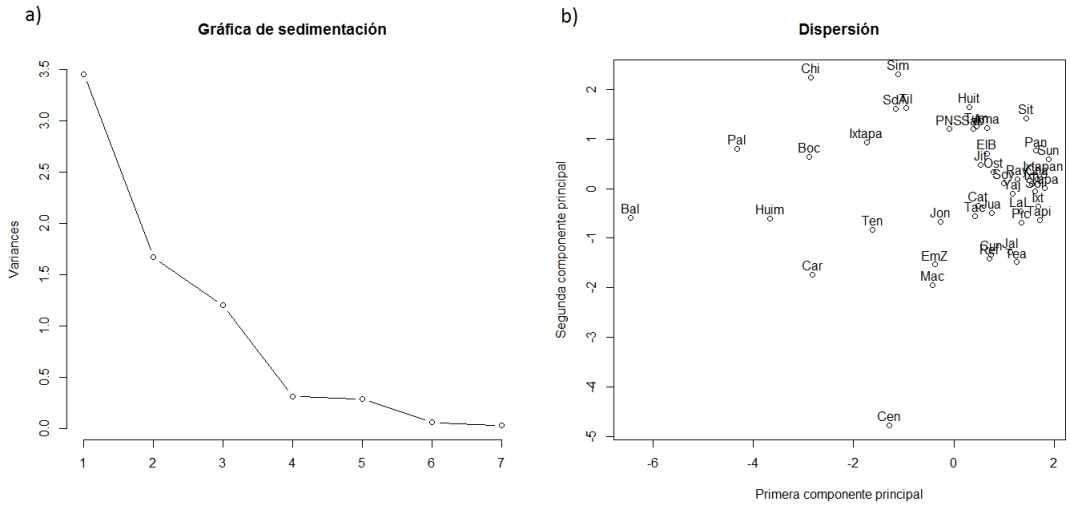

Figura 3.8: Gráficas del análisis de componentes principales de las variables independientes bajo estudio medidas en los 45 municipios, en el primer periodo: a) Gráfica de sedimentación y b) Dispersión de las dos primeras componentes principales.

El mismo procedimiento se llevó a cabo en el segundo y tercer periodo, lo que proporciono resultados similares, por lo que el municipio Centro (Cen) no se considera posteriormente. Las tablas y figuras correspondientes al segundo periodo son: Tablas 3.5, 3.6, 3.7 y Figuras 3.9, 3.10. Mientras que las tablas y figuras correspondientes al tercer periodo son: Tablas 3.8, 3.9, 3.10 y Figuras 3.11, 3.12.

Tablas y figuras correspondientes al segundo periodo

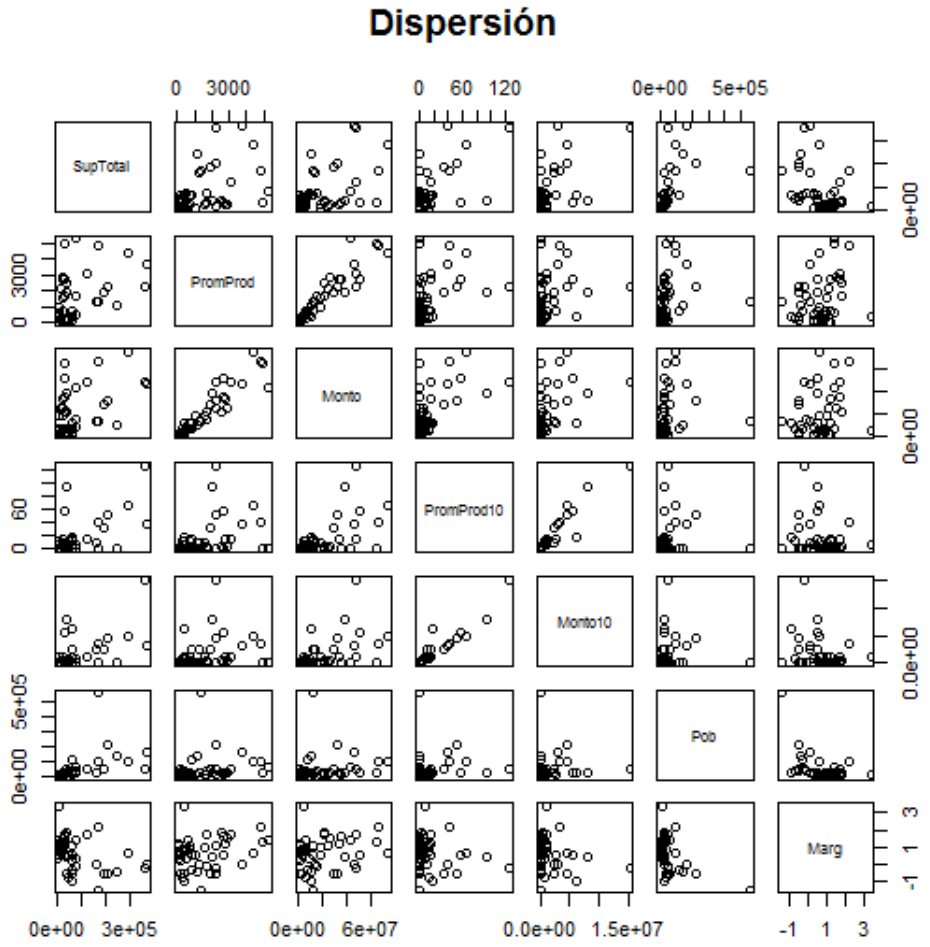

Figura 3.9: Dispersión de las variables independientes bajo estudio medidas en los 45 municipios considerados, en el segundo periodo.

Tabla 3.5. *Correlaciones de las variables independientes medidas en los 45 municipios considerados, en el segundo periodo.*

	SupTotal	*PromProd*	*Monto*	*PromProd10*	*Monto10*	*Pob*	*Marg*
SupTotal	*1.000*	*0.441*	*0.525*	*0.599*	*0.562*	*0.503*	*-0.373*
PromProd		*1.000*	*0.930*	*0.364*	*0.256*	*0.204*	*0.271*
Monto			*1.000*	*0.600*	*0.485*	*0.167*	*0.193*
PromProd10				*1.000*	*0.944*	*0.103*	*-0.155*
Monto10					*1.000*	*0.057*	*-0.217*
Pob						*1.000*	*-0.463*
Marg							*1.000*

Tabla 3.6. *Varianzas de las variables independientes bajo estudio medidas en los 45 municipios, en el segundo periodo.*

Variable	***Varianza***	***Variable***	***Varianza***
SupTotal	$8.604700 * 10^{9}$	*Monto10*	$7.976772 * 10^{12}$
PromProd	$2.069383 * 10^{6}$	*Pob*	$8.046871 * 10^{9}$
Monto	$4.067768 * 10^{14}$	*Marg*	$8.671147 * 10^{-1}$
PromProd10	$6.988500 * 10^{2}$		

Tabla 3.7. *Importancia de los componentes principales de las variables independientes bajo estudio medidas en los 45 municipios considerados, en el segundo periodo, donde DE, PV y PA significan desviación estándar, proporción de varianza y proporción acumulada, respectivamente.*

	CP1	*CP2*	*CP3*	*CP4*	*CP5*	*CP6*	*CP7*
DE	*1.8452*	*1.3087*	*1.0899*	*0.57281*	*0.5351*	*0.23134*	*0.16360*
PV	*0.4864*	*0.2447*	*0.1697*	*0.04687*	*0.0409*	*0.00765*	*0.00382*
PA	*0.4864*	*0.7311*	*0.9008*	*0.94763*	*0.9885*	*0.99618*	*1.00000*

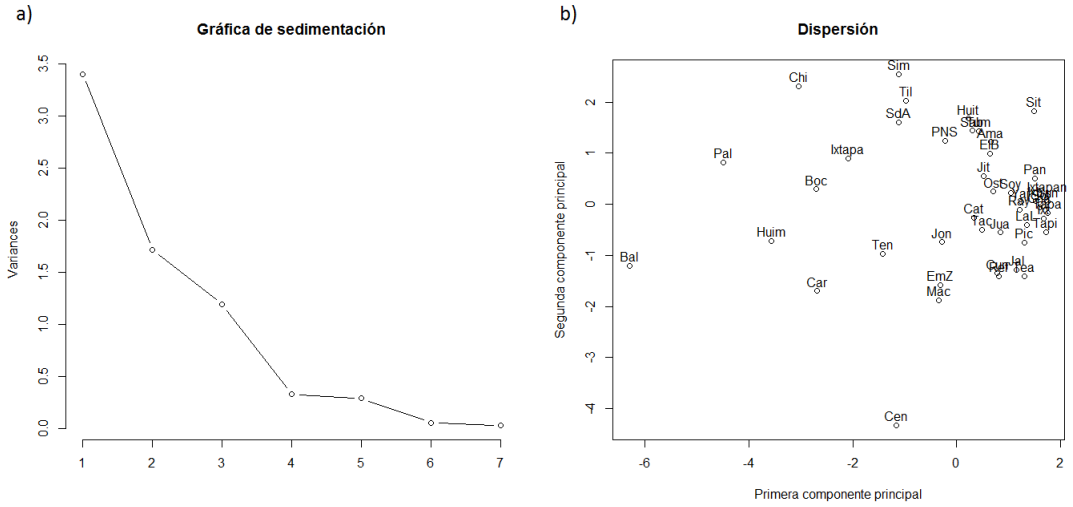

Figura 3.10: Gráficas del análisis de componentes principales de las variables independientes bajo estudio medidas en los 45 municipios, en el segundo periodo: a) Gráfica de sedimentación y b) Dispersión de las dos primeras componentes principales.

Tablas y figuras correspondientes al tercer periodo

Tabla 3.8. *Correlaciones de las variables independientes medidas en los 45 municipios considerados, en el tercer periodo.*

	SupTotal	*PromProd*	*Monto*	*PromProd10*	*Monto10*	*Pob*	*Marg*
SupTotal	*1.000*	*0.408*	*0.485*	*0.568*	*0.532*	*0.490*	*-0.331*
PromProd		*1.000*	*0.933*	*0.321*	*0.214*	*0.190*	*0.309*
Monto			*1.000*	*0.549*	*0.430*	*0.153*	*0.240*
PromProd10				*1.000*	*0.949*	*0.084*	*-0.094*
Monto10					*1.000*	*0.044*	*-0.164*
Pob						*1.000*	*-0.453*
Marg							*1.000*

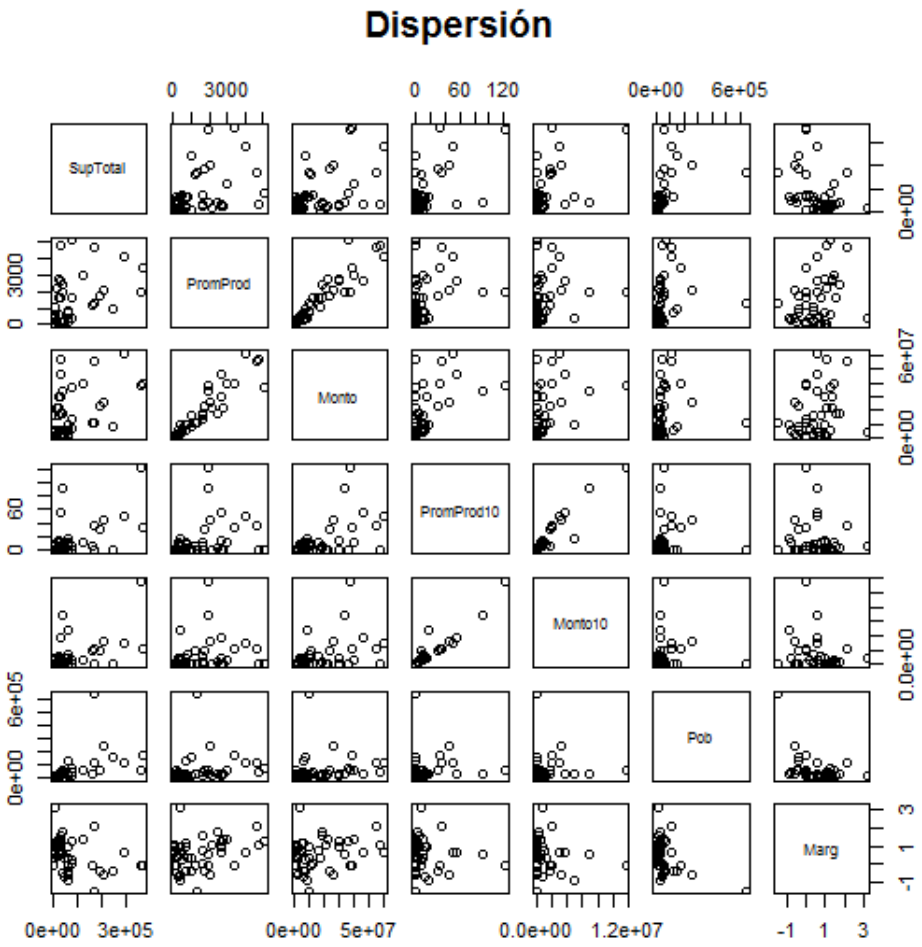

Figura 3.11: Dispersión de las variables independientes bajo estudio medidas en los 45 municipios considerados, en el tercer periodo.

Tabla 3.9. *Varianzas de las variables independientes bajo estudio medidas en los 45 municipios, en el tercer periodo.*

Variable	***Varianza***	***Variable***	***Varianza***
SupTotal	$8.604700 * 10^{9}$	*Monto10*	$4.674875 * 10^{12}$
PromProd	$1.907964 * 10^{6}$	*Pob*	$1.047662 * 10^{10}$
Monto	$2.884256 * 10^{14}$	*Marg*	$7.373943 * 10^{-1}$
PromProd10	$6.192855 * 10^{2}$		

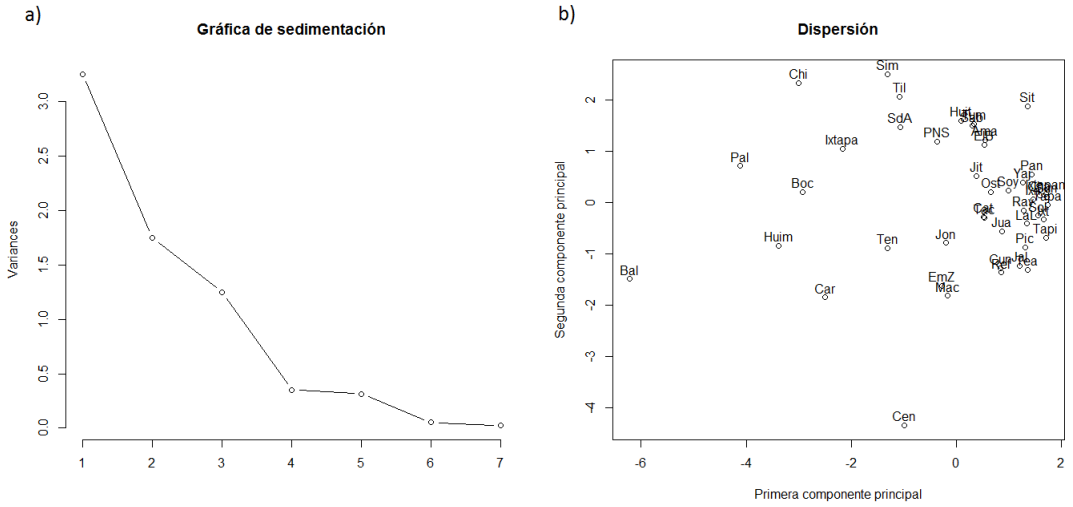

Figura 3.12: Gráficas del análisis de componentes principales de las variables independientes bajo estudio medidas en los 45 municipios, en el tercer periodo: a) Gráfica de sedimentación y b) Dispersión de las dos primeras componentes principales.

Tabla 3.10. *Importancia de los componentes principales de las variables independientes bajo estudio medidas en los 45 municipios considerados, en el tercer periodo, donde DE, PV y PA significan desviación estándar, proporción de varianza y proporción acumulada, respectivamente.*

	CP1	*CP2*	*CP3*	*CP4*	*CP5*	*CP6*	*CP7*
DE	*1.8040*	*1.3234*	*1.1184*	*0.59353*	*0.56068*	*0.22720*	*0.1587*
PV	*0.4649*	*0.2502*	*0.1787*	*0.05033*	*0.04491*	*0.00737*	*0.0036*
PA	*0.4649*	*0.7151*	*0.8938*	*0.94412*	*0.98903*	*0.99640*	*1.0000*

3.7. Análisis del primer periodo sin el municipio Centro

La dispersión de las variables independientes bajo estudio medidas en los 44 municipios (Figura 3.13) muestra que existe una relación lineal entre las variables PromProd y Monto, y también entre PromProd10 y Monto10, lo cual se confirma al observar las correlaciones de estas variables (Tabla 3.11), pues sus correlaciones son 0.933 y 0.938 respectivamente; además se puede notar que existen altas correlaciones entre las varia-

bles independientes, y que no existe una alta correlación lineal entre la variable Exp y el resto de variables, además de existir un punto atípico.

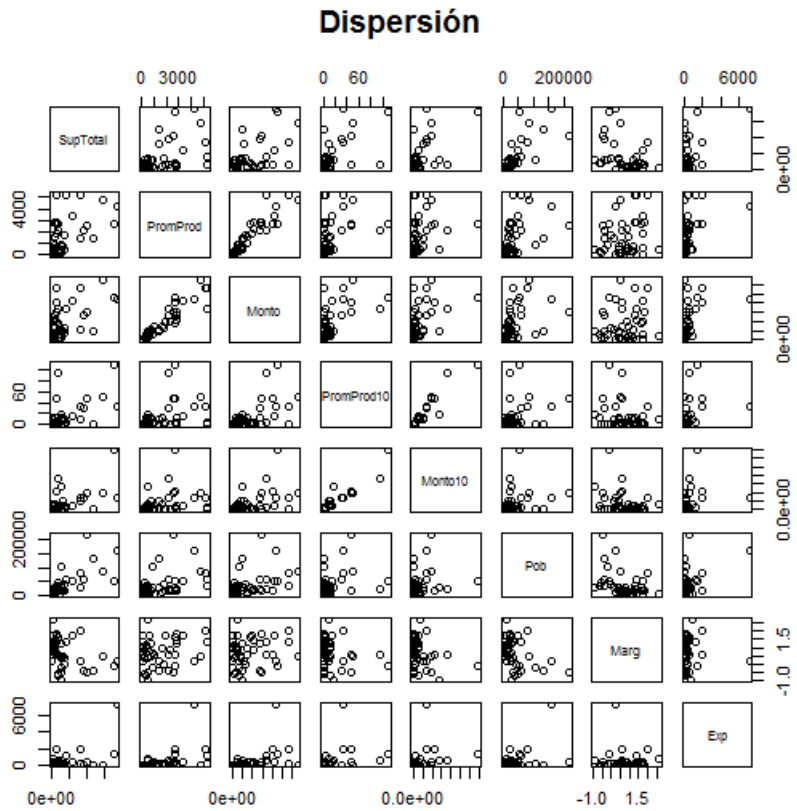

Figura 3.13: Dispersión de las variables bajo estudio medidas en los 44 municipios, en el primer periodo.

Tabla 3.11. *Correlaciones de las variables independientes medidas en los 44 municipios, en el primer periodo.*

	SupTotal	*PromProd*	*Monto*	*PromProd10*	*Monto10*	*Pob*	*Marg*
SupTotal	*1.000*	*0.495*	*0.581*	*0.586*	*0.582*	*0.736*	*-0.357*
PromProd		*1.000*	*0.933*	*0.370*	*0.281*	*0.443*	*0.227*
Monto			*1.000*	*0.594*	*0.499*	*0.427*	*0.134*
PromProd10				*1.000*	*0.938*	*0.325*	*-0.213*
Monto10					*1.000*	*0.242*	*-0.282*
Pob						*1.000*	*-0.399*
Marg							*1.000*

El análisis de componentes principales de los datos estandarizados de las variables independientes medidas en los 44 municipios; índica que las primeras tres componentes principales explican el 91.3 % de la variabilidad total (Tabla 3.12); su gráfica de sedimentación (Figura 3.14 a)) muestra que sólo estas componentes tienen valor propio mayor que 1, es decir, en el análisis posterior será suficiente usar las primeras tres componentes principales en lugar de las siete variables independientes, ya que existe multicolinealidad, pues al menos dos pares de variables independientes están altamente correlacionadas, como se observa en la Tabla 3.11. El diagrama de dispersión de las dos primeras componentes principales se muestra en la Figura 3.14 b).

Tabla 3.12. *Importancia de las componentes principales de las variables independientes bajo estudio medidas en los 44 municipios, en el primer periodo, donde DE, PV y PA significan desviación estándar, proporción de varianza y proporción acumulada, respectivamente.*

	CP1	*CP2*	*CP3*	*CP4*	*CP5*	*CP6*	*CP7*
DE	*1.9365*	*1.2418*	*1.0481*	*0.57735*	*0.44517*	*0.22994*	*0.15800*
PV	*0.5357*	*0.2203*	*0.1569*	*0.04762*	*0.02831*	*0.00755*	*0.00357*
PA	*0.5357*	*0.7560*	*0.9130*	*0.96057*	*0.98888*	*0.99643*	*1.00000*

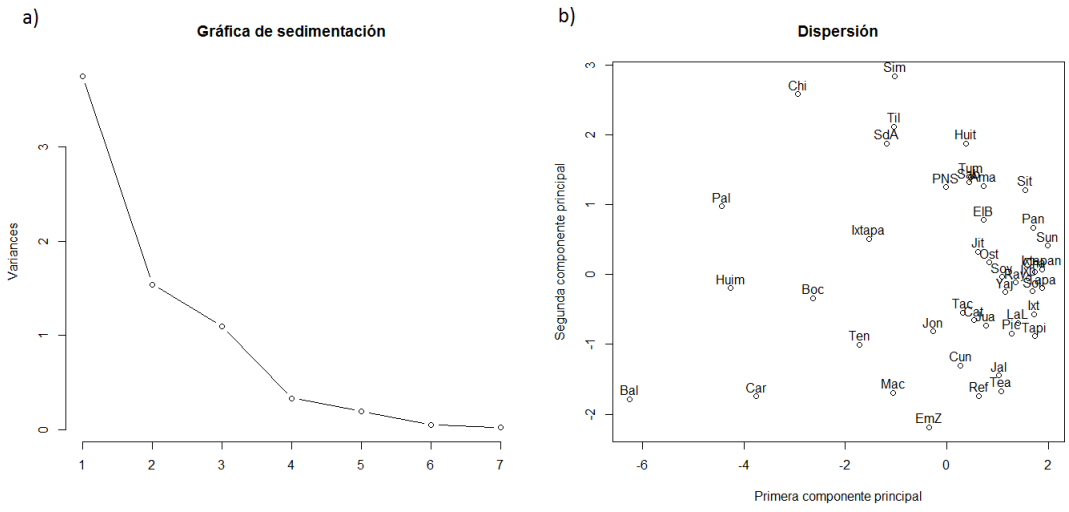

Figura 3.14: Gráficas del análisis de componentes principales de las variables independientes bajo estudio medidas en los 44 municipios, en el primer periodo: a) Gráfica de sedimentación y b) Dispersión de las dos primeras componentes principales.

Tabla 3.13. *Coeficientes de las combinaciones lineales de las variables independientes que definen a las siete componentes principales, en el primer periodo.*

	CP1	*CP2*	*CP3*	*CP4*	*CP5*	*CP6*	*CP7*
SupTotal	*-0.444*	*-0.161*	*-0.232*	*0.415*	*0.725*	*0.128*	*0.097*
PromProd	*-0.370*	*0.498*	*-0.204*	*-0.362*	*0.019*	*-0.373*	*0.550*
Monto	*-0.432*	*0.398*	*-0.015*	*-0.343*	*0.034*	*0.373*	*-0.630*
PromProd10	*-0.425*	*-0.123*	*0.480*	*0.104*	*-0.342*	*0.511*	*0.430*
Monto10	*-0.399*	*-0.201*	*0.534*	*0.066*	*-0.002*	*-0.654*	*-0.289*
Pob	*-0.350*	*-0.174*	*-0.594*	*0.332*	*-0.592*	*-0.121*	*-0.141*
Marg	*0.131*	*0.694*	*0.192*	*0.674*	*-0.075*	*-0.045*	*-0.053*

Como último paso del ACP, se le dará una interpretación a las primeras componentes principales. Como los coeficientes de la primera componente principal son en su mayoría negativos (Tabla 3.13), excepto Marg, que tiene un coeficiente cercano a cero, se tiene que esta componente es un promedio negativo ponderado, es decir:

- Los municipios que tienen los valores más pequeños en la primera componente principal son los que tienen valores grandes en alguna o algunas variables independientes, por ejemplo:
  - Balancán (Bal) es el segundo municipio más extenso (SupTotal), tiene más beneficiados con más de 10 hectáreas (PromProd10) y mayor monto asignado a estos productores (Monto10).
  - Palenque (Pal) es el tercer municipio más extenso (SupTotal), es el que tiene mayor monto asignado a productores en general (Monto) y el tercero con más productores con más de 10 hectáreas (PromProd10).
  - Huimanguillo (Huim) es el municipio más extenso (SupTotal) y el segundo con más población (Pob).
- Los municipios que tienen los valores máximos en la primera componente principal son los que tienen valores pequeños en las variables independientes, por ejemplo:

- Sunuapa (Sun) es de los municipios menos extensos, el segundo con menos productores apoyados (PromProd) y el que tiene menos población (Pob).
- Tapalapa (Tapa) es el segundo municipio menos extenso (SupTotal), el tercero con menos productores apoyados (PromProd) y el segundo con menos monto (Monto), además de carecer de apoyo a productores con más de 10 hectáreas (PromProd10, Monto10) y el segundo con menos población (Pob).
- Ixtapangajoya (Ixtapan) es de los municipios menos extensos (SupTotal), el cuarto municipio con menos productores y monto del apoyo (PromProd, Monto) y el tercer municipio con menos población (Pob).

Esto se puede visualizar en la Figura 3.15, que es la gráfica de las puntuaciones de la primera componente principal, y en la Tabla A.5, que indica el orden de los municipios en cada variable, en ella se puede observar que el municipio Balancán corresponde a un posible punto atípico.

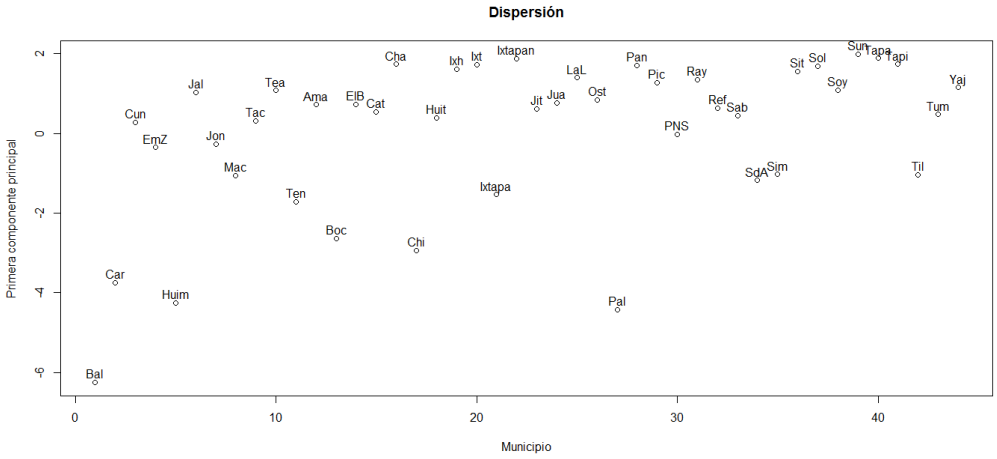

Figura 3.15: Dispersión de la primera componente principal frente municipio, en el primer periodo.

De la segunda componente principal, los coeficientes de PromProd, Monto y Marg son positivos y los de SupTotal, PromProd10, Monto10 y Pob son negativos (Tabla 3.13); así, se tiene que esta componente es un contraste de las variables PromProd, Monto y Marg contra SupTotal, PromProd10, Monto10 y Pob con mayor peso en PromProd, Monto y Marg, es decir:

- Los municipios que tienen los valores más altos en la segunda componente principal son los que tienen valores grandes en las variables PromProd, Monto y Marg, y no tienen valores tan grandes en SupTotal, PromProd10, Monto10 y Pob, por ejemplo:
  - Simojovel (Sim) es el segundo municipio con más productores apoyados y monto (PromProd y Monto) y es el número 12 en marginación (Marg).
  - Chilón (Chi) es el tercer municipio con más productores apoyados y monto (PromProd y Monto) y es el número segundo en marginación (Marg).
  - Tila (Til) es el municipio con más productores apoyados (PromProd) pero tiene el noveno lugar el monto del apoyo (Monto) y el décimo en marginación (Marg).
- Los municipios que tienen los valores más pequeños en la segunda componente principal son:
  - Emiliano Zapata (EmZ)
  - Balancán (Bal)
  - Cárdenas (Car).

Esto se observa en la Figura 3.16, que es la gráfica de las puntuaciones de la segunda componente principal, y en Tabla A.5, que indica el orden de los municipios en cada variable.

En la tercera componente principal, los coeficientes de PromProd10, Monto10 y Marg son positivos y los de SupTotal, PromProd, Monto y Pob son negativos (Tabla 3.13); así, se tiene que esta componente es un contraste de las variables PromProd10, Monto10 y Marg contra SupTotal, PromProd, Monto y Pob, con mayor peso en PromProd10, Monto10 y Pob, es decir:

- Los municipios que tienen los valores máximos en la tercera componente principal son los que tienen valores grandes en las variables PromProd10 y Monto10, que no tienen valores tan grandes en SupTotal, PromProd, Monto y Pob, por ejemplo:

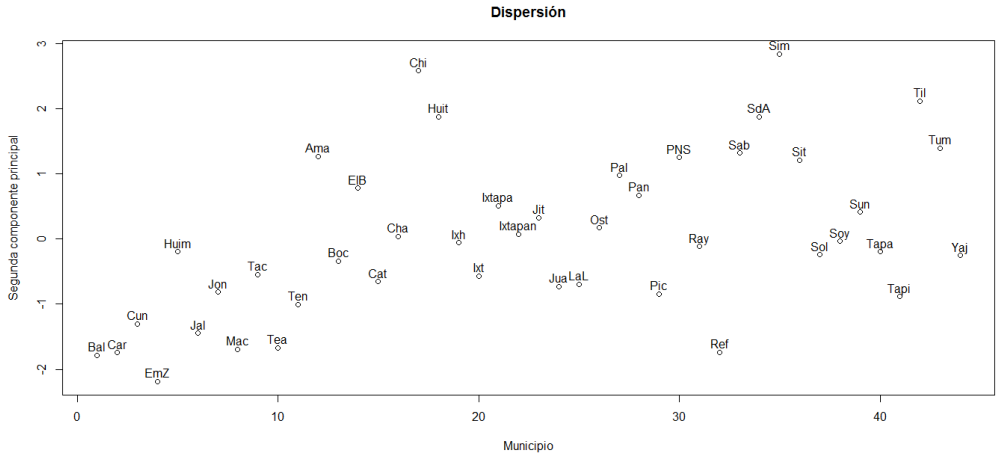

Figura 3.16: Dispersión la segunda componente principal frente municipio, en el primer periodo.

- Balancán (Bal), es el municipio con más productores con más de 10 hectáreas y mayor monto a estos productores (PromProd10 y Monto10).
- Bochil (Boc) es el segundo municipio con más productores con más de 10 hectáreas y mayor monto a estos productores (PromProd10 y Monto10).
- Ixtapa (Ixtapa) es el quinto municipio con más productores con más de 10 hectáreas y el cuarto con mayor monto a estos productores (PromProd10 y Monto10).

- Los municipios con los valores más pequeños en la tercera componente principal son:

  - Macuspana (Mac).
  - Huimanguillo (Huim).
  - Cárdenas (Car).

Esto se observa en la Figura 3.17, que es la gráfica de las puntuaciones de la tercera componente principal, y en la Tabla A.5, que indica el orden de los municipios en cada variable.

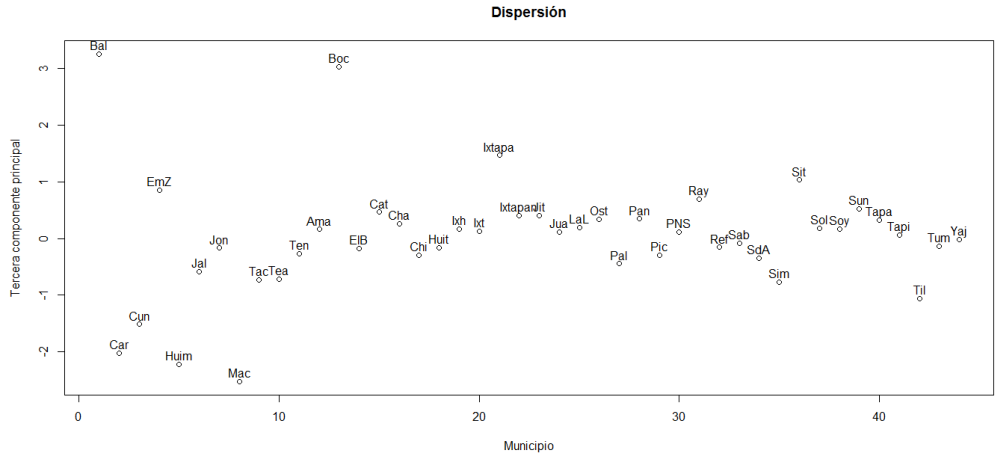

Figura 3.17: Dispersión de la tercera componente principal frente municipio, en el primer periodo.

3.7.1. Análisis de regresión del primer periodo

El primer paso de la regresión es realizar una regresión simple de la variable Exp con cada una de las variables independientes (SupTotal, PromProd, Monto, PromProd10, Monto10, Pob y Marg). Como resultado de estos modelos (Tabla 3.14) se obtuvo que las variables SupTotal, PromProd, Monto, PromProd10, Monto10, Pob sí son significativas, a diferencia de Marg, lo cual sugiere la conveniencia de realizar un análisis de regresión múltiple con todas las variables independientes, incluido índice de marginación, a pesar de que en este periodo no resultó significativo, ya que en el segundo periodo sí es significativo. El modelo que incluye a la superficie total (SupTotal) tiene el R^2 y R^2-ajustado máximos de la Tabla 3.14, sin embargo, estos son menores que 0.4 ya que la recta de regresión esta muy influenciada por el punto atípico que corresponde al municipio Huimanguillo (Figura 3.18).

Dado que existe alta correlación entre las variables originales (Tabla 3.11), un análisis de regresión usándolas a todas ellas como variables independientes conducirá a resultados inestables; sin embargo, una forma de realizar el análisis de interés es la regresión paso a paso de los datos estandarizados, que es un método de selección de modelos, o mediante un análisis de componentes principales de las variables independientes.

Tabla 3.14. *Resumen de los modelos en los cuales Exp es la variable dependiente con cada una de las variables independientes como predictor, en el primer periodo, donde: Estimación es el coeficiente estimado para el predictor en ese modelo, CS es el código significativo: 0 '***'0.001 '**'0.01 '*'0.05 '.'0.1 ' '1.*

Predictor (es)	*Estimación*	*Error estándar*	*Valor t*	$Pr(>\|t\|)$	*CS*	R^2	R^2 *ajustado*
SupTotal	*0.00654*	*0.0012*	*5.301*	$3.75*10^{-6}$	***	*0.3953*	*0.3812*
PromProd	*0.29964*	*0.0704*	*4.256*	*0.000111*	***	*0.2964*	*0.28*
Monto	$2.7*10^{-5}$	$6.42*10^{-6}$	*4.2*	*0.000132*	***	*0.2909*	*0.2744*
PromProd10	*20.003*	*6.131*	*3.262*	*0.00217*	**	*0.1984*	*0.1798*
Monto10	$1.77*10^{-4}$	$5.904*10^{-5}$	*2.999*	*0.0045*	**	*0.1729*	*0.1537*
Pob	*0.01203*	*0.0027*	*4.397*	$7.08*10^{-5}$	***	*0.3102*	*0.2942*
Intercepto	*465.51*	*234.30*	*1.987*	*0.0535*	.	*0.0003*	*-0.024*
Marg	*-22.85*	*214.81*	*-.11*	*0.9158*			

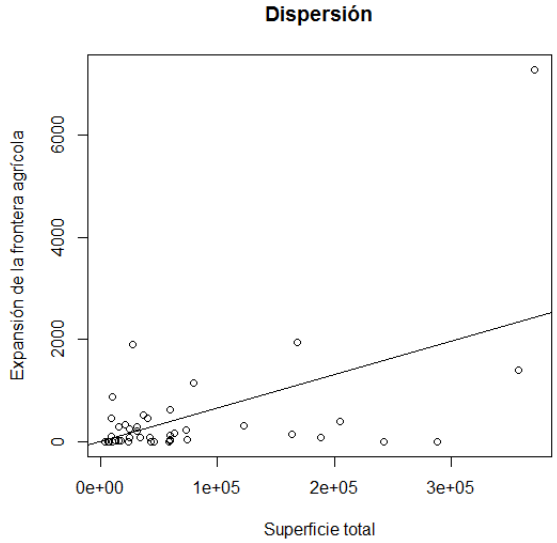

Figura 3.18: Diagrama de dispersión de la expansión de la frontera agrícola (Exp) contra la superficie total (SupTotal), con la recta ajustada por el modelo de regresión, en el primer periodo.

La regresión paso a paso proporciona el modelo que tiene como predictores a la superficie total (SupTotal) y al índice de marginación (Marg, a pesar de que cuando se considera sólo a la variable Marg en el modelo de regresión lineal simple, es la menos significativa) además de contener al intercepto (Tabla 3.15), sin embargo el índice de marginación y el intercepto no son significativos, por lo tanto elegimos al modelo que sólo incluye la superficie total (SupTotal) que está dado en la Tabla 3.14.

Tabla 3.15. *Resumen del mejor modelo elegido por regresión paso a paso, en el primer periodo, donde: Estimación es el coeficiente estimado para el predictor en ese modelo; CS es el código significativo: 0 '***'0.001 '**'0.01 '*'0.05 '.'0.1 ' '1.*

Predictor (es)	*Estimación*	*Error estándar*	*Valor t*	$Pr(>\|t\|)$	*CS*	R^2	R^2 *ajustado*
Intercepto	$2.9*10^{-16}$	$1.25*10^{-1}$	*0.000*	*1.000*		*0.3406*	*0.3084*
SupTotal	$6.25*10^{-1}$	$1.36*10^{-3}$	*4.6*	$4.02*10^{-5}$	***		
Marg	$2.07*10^{-1}$	$1.36*10^{-1}$	*1.522*	*0.136*			

Como tercer paso, se realiza la regresión con las primeras tres componentes principales. Como resultado se obtiene que sólo la primera componente principal es significativa para explicar a la variable Exp, el R^2 y el R^2-ajustado de este modelo son menores que 0.27, lo cual posiblemente se debe a que la recta de regresión esta muy influenciada por el punto atípico que corresponde a Huimanguillo (Figura 3.19).

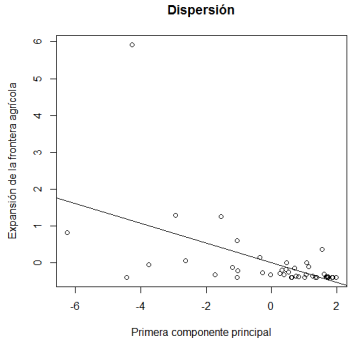

Figura 3.19: Diagrama de dispersión de la expansión de la frontera agrícola frente la primera componente principal con la recta ajustada por el modelo de regresión, en el primer periodo.

Tabla 3.16. *Resumen de los modelos en los cuales Exp es la variable dependiente con las tres primeras componentes principales como variables independientes, en el primer periodo, donde: CPi es la i-ésima componente principal, $i = 1, 2, 3$, Estimación es el coeficiente estimado para el predictor en ese modelo; CS es el código significativo: 0 '***'0.001 '**'0.01 '*'0.05 '.'0.1 ' '1.*

Predictor (es)	*Estimación*	*Error estándar*	*Valor t*	$Pr(>\|t\|)$	*CS*	R^2	R^2 *ajustado*
CP1	*-0.26767*	*0.06734*	*-3.98*	*0.000265*	***	*0.2687*	*0.2517*
Intercepto	$1.51*10^{-16}$	$1.30*10^{-1}$	*0.000*	*1.000*		*0.3067*	*0.2547*
CP1	$-2.7*10^{-1}$	$6.8*10^{-2}$	*-3.94*	*0.000321*	***		
CP2	$7.1*10^{-2}$	$1.06*10^{-1}$	*0.666*	*0.509019*			
CP3	$-1.7*10^{-1}$	$1.26*10^{-1}$	*-1.32*	*0.193582*			

Análisis de regresión sin el municipio Huimanguillo

Notar que en el primer periodo, Huimanguillo (Huim) tiene una expansión de frontera agrícola mayor que el resto de los municipios, por lo que no se considera para el Análisis de regresión ya que afecta de forma significativa los resultados, como se notó en los análisis anteriores.

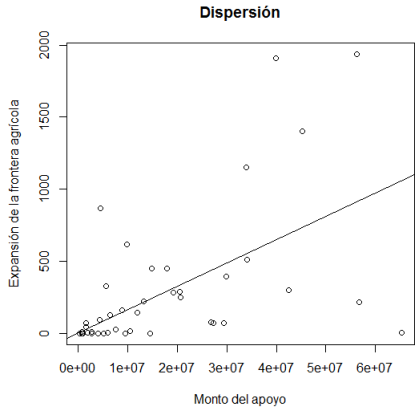

Figura 3.20: Diagrama de dispersión de expansión de la frontera agrícola (Exp) frente monto del apoyo a los productores (Monto), con la recta ajustada por el modelo de regresión, en el primer periodo (sin Huimanguillo).

El procedimiento que se lleva a cabo es el mismo que el del análisis que considera a Huimanguillo, sin embargo, en este caso todas las variables son significativas (Tabla 3.17), y en el análisis anterior el índice de marginación no es significativo. El modelo con R^2 y R^2-ajustado máximos es el que tiene a la variable monto asignado a productores (Monto, véase Figura 3.20).

Tabla 3.17. *Resumen de los modelos en los cuales Exp es la variable dependiente con cada una de las variables independientes como predictor, en el primer periodo, donde: Estimación es el coeficiente estimado para el predictor en ese modelo, CS es el código significativo: 0 '***'0.001 '**'0.01 '*'0.05 '.'0.1 ' '1.*

Predictor (es)	*Estimación*	*Error estándar*	*Valor t*	$Pr(>\|t\|)$	*CS*	R^2	R^2 *ajustado*
SupTotal	*0.00274*	*0.0007*	*3.923*	*0.00032*	***	*0.2681*	*0.2507*
PromProd	*0.17474*	*0.02897*	*6.032*	$3.58*10^{-7}$	***	*0.4642*	*0.4514*
Monto	$1.62*10^{-5}$	$2.56*10^{-6}$	*6.336*	$1.3*10^{-7}$	***	*0.4887*	*0.4766*
PromProd10	*13.094*	*2.418*	*5.414*	$2.75*10^{-6}$	***	*0.4111*	*0.3971*
Monto10	$1.27*10^{-4}$	$2.22*10^{-5}$	*5.717*	$1.01*10^{-6}$	***	*0.4376*	*0.4242*
Pob	*0.004737*	*0.001471*	*3.22*	*0.00247*	**	*0.198*	*0.1789*
Marg	*226.18*	*69.34*	*3.262*	*0.0022*	**	*0.2021*	*0.1831*

La regresión paso a paso proporciona el modelo que tiene como predictores la superficie total (SupTotal), las variables de PROCAMPO: PromProd, PromProd10, Monto10, la población (Pob) y el índice de marginación (Marg, Tabla 3.18), sin embargo, la población (Pob) y el índice de marginación no son significativos, por lo tanto elegimos al modelo que incluye la superficie total (SupTotal), promedio de productores (PromProd), promedio de productores con más de 10 hectáreas (PromProd10) y el monto asignado a productores con más de 10 hectáreas (Monto10):

$$\widehat{Exp} = -0.3298 SupTotal + 0.4897 PromProd + 0.6249 Monto10. \tag{3.1}$$

Esta ecuación esta dada para datos estandarizados. Las gráficas de diagnóstico (Fi-

gura 3.21) indican la falta de normalidad para los errores y la no independencia entre ellos, lo cual posiblemente se debe a la alta variabilidad que existe entre los valores de las variables en los municipios.

Tabla 3.18. *Resumen del mejor modelo elegido por regresión paso a paso, en el primer periodo, donde: Estimación es el coeficiente estimado para el predictor en ese modelo; CS es el código significativo: 0 '***'0.001 '**'0.01 '*'0.05 '.'0.1 ' '1.*

Predictor (es)	*Estimación*	*Error estándar*	*Valor t*	$Pr(>\|t\|)$	*CS*	R^2	R^2 *ajustado*
SupTotal	*-0.3298*	*0.1536*	*-2.147*	*0.03788*	*	*0.504*	*0.4668*
PromProd	*0.4897*	*0.1240*	*3.949*	*0.00031*	***		
Monto10	*0.6249*	*0.1433*	*4.359*	$8.88*10^{-5}$	***		
SupTotal	*-0.4406*	*0.2009*	*-2.193*	*0.03463*	*	*0.565*	*0.4945*
PromProd	*0.3974*	*0.1486*	*2.674*	*0.01109*	*		
PromProd10	*-0.5897*	*0.3476*	*-1.697*	*0.09818*	.		
Monto10	*1.2753*	*0.3713*	*3.434*	*0.00148*	**		
Pob	*0.2837*	*0.1772*	*1.601*	*0.11796*			
Marg	*0.2458*	*0.1486*	*1.655*	*0.10648*			

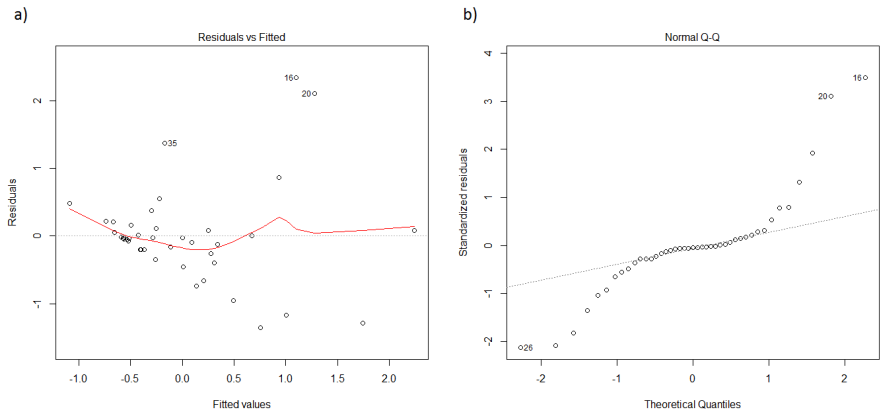

Figura 3.21: Gráficas de diagnósticos del análisis de regresión, en el primer periodo: a) Residuales frente valores ajustados y b) Gráfico Q-Q.

Por último, se realiza el análisis de regresión mediante las componentes principales. Como resultado se obtiene que las primeras tres componentes principales son significativas para explicar a la variable Exp (Tabla 3.19).

Tabla 3.19. *Resumen del modelo en el que Exp es la variable dependiente y las tres primeras componentes principales son las variables independientes, en el primer periodo, donde: CPi es la i-ésima componente principal,* $i = 1, 2, 3$*, Estimación es el coeficiente estimado para el predictor en ese modelo; CS es el código significativo: 0 '***'0.001 '**'0.01 '*'0.05 '.'0.1 ' '1.*

Predictor (es)	*Estimación*	*Error estándar*	*Valor t*	$Pr(>\|t\|)$	*CS*	R^2	R^2 *ajustado*
CP1	*-0.2753*	*0.06515*	*-4.23*	*0.000134*	***	*0.4299*	*0.3872*
CP2	*0.21195*	*0.09505*	*2.230*	*0.031429*	*		
CP3	*0.25902*	*0.11983*	*2.162*	*0.036690*	*		

El modelo de regresión lineal que se acepta es:

$$\widehat{Exp} = -0.2753CP1 + 0.21195CP2 + 0.25902CP3, \tag{3.2}$$

donde las componentes principales son las combinaciones lineales de las variables originales, dadas por:

$$\begin{aligned}
CP1 &= -0.444SupTotal - 0.37PromProd - 0.432Monto \\
&\quad -0.425PromProd10 - 0.399Monto10 - 0.35Pob + 0.131Marg, \\
CP2 &= -0.161SupTotal + 0.498PromProd + 0.398Monto \\
&\quad -0.123PromProd10 - 0.201Monto10 - 0.174Pob + 0.694Marg, \\
CP3 &= -0.232SupTotal - 0.204PromProd - 0.015Monto \\
&\quad +0.48PromProd10 + 0.534Monto10 - 0.594Pob + 0.192Marg.
\end{aligned}$$

Además se tiene que:

$$-0.2753CP1 + 0.21195CP2 + 0.25902CP3 = 0.028SupTotal + 0.155PromProd$$
$$+0.199Monto + 0.215PromProd10 + 0.206Monto10 - 0.095Pob + 0.161Marg.$$

Hay que señalar que esta ecuación considera a las variables estandarizadas y que sus coeficientes son en su mayoría positivos, excepto el de población (Pob), que tiene un coeficiente en valor absoluto pequeño, es decir, es de las variables que explican menos a la expansión de la frontera agrícola. Las gráficas de diagnóstico (Figura 3.22) indican la falta de normalidad para los errores y la no independencia entre ellos, lo cual posiblemente se debe a la alta variabilidad que existe entre los valores de las variables en los municipios.

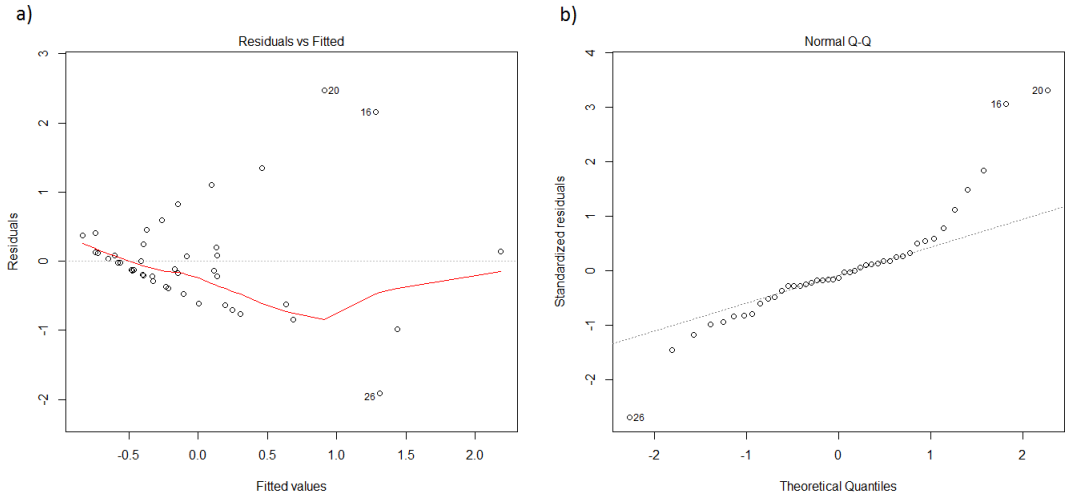

Figura 3.22: Gráficas de diagnósticos del análisis de regresión, en el primer periodo: a) Residuales frente valores ajustados y b) Gráfico Q-Q.

Resultados del primer periodo

En los municipios bajo estudio se observa que Huimanguillo tiene una expansión de frontera agrícola mayor que el resto de los municipios, por lo cual no se considera para el análisis de regresión. Al considerar a todas las variables se lograron dos modelos de regresión con problemas en los supuestos necesarios para poder considerar al modelo de regresión como aceptable. Aun con los problemas anteriores se trata de interpretar a los modelos obtenidos, esto es, el primer modelo está en términos de la superficie total (SupTotal), el promedio de productores (PromProd) y el monto del apoyo a productores con más de 10 hectáreas (Monto 10); el segundo modelo está en términos de todas

las variables pero tienen mayor influencias las variables de PROCAMPO (PromProd, Monto, PromProd10 y Monto10) y el índice de marginación (Marg). Además, el análisis de componentes principales arroja componentes con significados útiles.

3.8. Análisis del segundo periodo sin el municipio Centro

La dispersión de las variables independientes bajo estudio medidas en los 44 municipios (Figura 3.23) muestra que existe una relación lineal entre las variables PromProd y Monto, y entre PromProd10 y Monto10, como en el primer periodo, lo cual se confirma al observar las correlaciones de estas variables (Tabla 3.20): 0.931 y 0.943 respectivamente, además existen altas correlaciones entre las variables independientes.

Tabla 3.20. *Correlaciones de las variables independientes medidas en los 44 municipios, en el segundo periodo.*

	SupTotal	*PromProd*	*Monto*	*PromProd10*	*Monto10*	*Pob*	*Marg*
SupTotal	*1.000*	*0.450*	*0.541*	*0.621*	*0.584*	*0.737*	*-0.344*
PromProd		*1.000*	*0.931*	*0.364*	*0.255*	*0.434*	*0.282*
Monto			*1.000*	*0.598*	*0.483*	*0.425*	*0.186*
PromProd10				*1.000*	*0.943*	*0.346*	*-0.196*
Monto10					*1.000*	*0.250*	*-0.261*
Pob						*1.000*	*-0.348*
Marg							*1.000*

El análisis de componentes principales de los datos estandarizados de las variables independientes medidas en los 44 municipios estudiados, índica que las primeras tres componentes principales explican el 91.23 % de la variabilidad total (Tabla 3.21); su gráfica de sedimentación (Figura 3.24 a)) muestra que sólo estas componentes tienen valor propio mayor que 1, es decir, en el análisis posterior será suficiente usar las primeras tres componentes en lugar de las siete variables independientes. El diagrama de dispersión

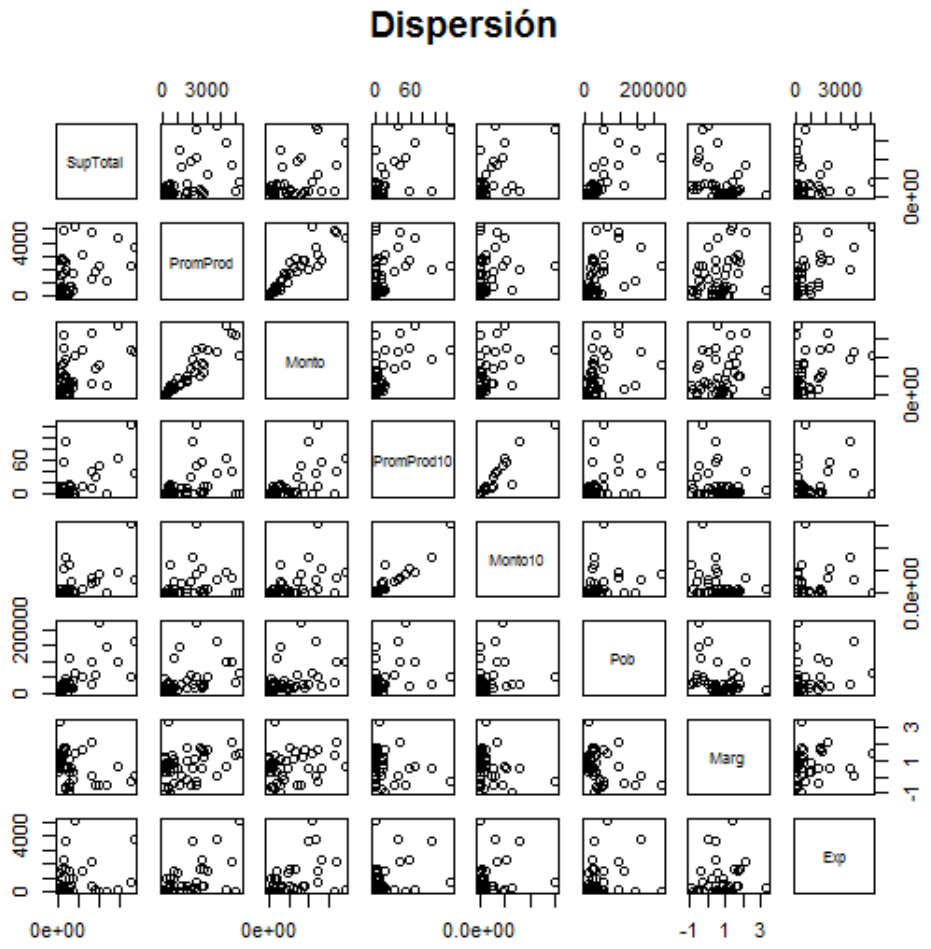

Figura 3.23: Dispersión de las variables bajo estudio medidas en los 44 municipios, en el segundo periodo.

de las dos primeras componentes principales se muestra en la Figura 3.24 b) y se puede notar que es similar al diagrama correspondiente al primer periodo (Figura 3.14 b)).

Tabla 3.21. *Importancia de las componentes principales de las variables independientes bajo estudio medidas en los 44 municipios, en el segundo periodo, donde DE, PV y PA significan desviación estándar, proporción de varianza y proporción acumulada, respectivamente.*

	CP1	*CP2*	*CP3*	*CP4*	*CP5*	*CP6*	*CP7*
DE	*1.9256*	*1.2696*	*1.0326*	*0.5904*	*0.4347*	*0.22825*	*0.15617*
PV	*0.5297*	*0.2303*	*0.1523*	*0.0498*	*0.0270*	*0.00744*	*0.00348*
PA	*0.5297*	*0.7600*	*0.9123*	*0.9621*	*0.9891*	*0.99652*	*1.00000*

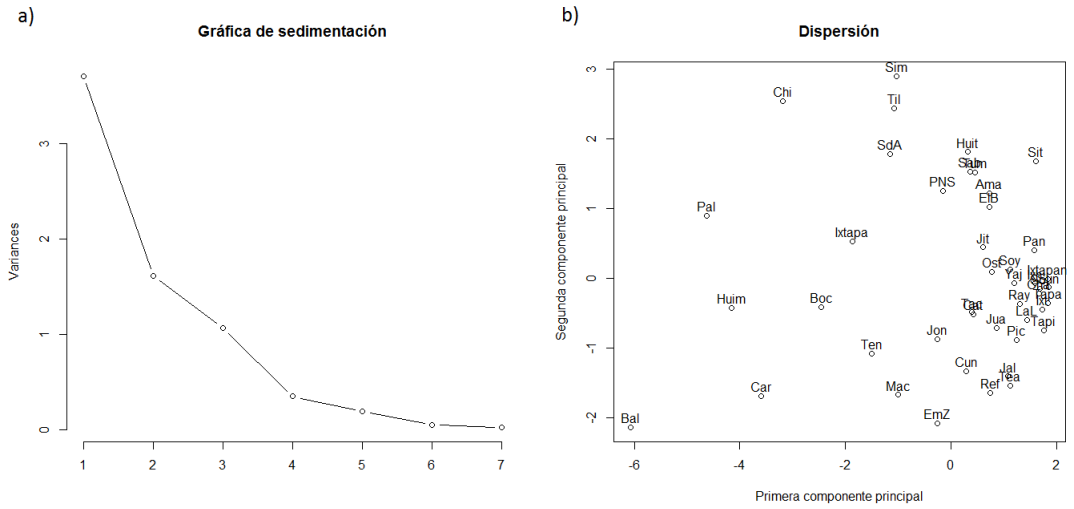

Figura 3.24: Gráficas del análisis de componentes principales de las variables independientes bajo estudio medidas en los 44 municipios, en el segundo periodo: a) Gráfica de sedimentación y b) Dispersión de las dos primeras componentes principales.

Tabla 3.22. *Coeficientes de las combinaciones lineales de las variables independientes que definen a las siete componentes principales, en el segundo periodo.*

	CP1	*CP2*	*CP3*	*CP4*	*CP5*	*CP6*	*CP7*
SupTotal	*-0.442*	*-0.187*	*-0.239*	*-0.386*	*-0.747*	*-0.068*	*0.037*
PromProd	*-0.363*	*0.504*	*-0.191*	*0.363*	*-0.048*	*0.396*	*0.534*
Monto	*-0.430*	*0.398*	*0.001*	*0.350*	*-0.021*	*-0.388*	*-0.620*
PromProd10	*-0.438*	*-0.139*	*0.456*	*-0.100*	*0.271*	*-0.521*	*0.476*
Monto10	*-0.401*	*-0.219*	*0.530*	*-0.063*	*0.083*	*0.640*	*-0.301*
Pob	*-0.352*	*-0.147*	*-0.613*	*-0.325*	*0.597*	*0.078*	*-0.103*
Marg	*0.105*	*0.681*	*0.207*	*-0.691*	*0.054*	*0.030*	*-0.045*

Como último paso del ACP, se interpretarán las primeras componentes principales. Como los coeficientes de la primera componente principal son en su mayoría negativos (Tabla 3.22), excepto Marg, que tiene un coeficiente cercano a cero, esta componente es un promedio negativo ponderado, como en el primer periodo, es decir:

- Los municipios con los valores más pequeños en la primera componente principal

son los que tienen valores grandes en alguna o algunas variables independientes, por ejemplo:

- Balancán (Bal) es el segundo municipio más extenso (SupTotal), con más beneficiados con más de 10 hectáreas (PromProd10) y mayor monto asignado a estos productores (Monto10).
- Palenque (Pal) es el tercer municipio más extenso (SupTotal), con mayor monto asignado a productores en general (Monto) y el tercero con más productores con más de 10 hectáreas (PromProd10).
- Huimanguillo (Huim) es el municipio más extenso (SupTotal) y el segundo con más población (Pob).

Nótese que los municipios enlistados son los mismos que en el primer periodo, y el motivo también es el mismo.

- Los municipios con los valores máximos en la primera componente principal son los que tienen valores pequeños en las variables independientes, por ejemplo:
  - Sunuapa (Sun) es de los municipios menos extensos, el segundo con menos productores apoyados (PromProd) y el de menor población (Pob).
  - Ixtapangajoya (Ixtapan) es uno de los municipio menos extensos (SupTotal), el cuarto municipio con menos productores apoyados (PromProd), y el tercero con menor monto del apoyo (Monto) y menor población (Pob).
  - Tapalapa (Tapa) es el segundo municipio menos extenso (SupTotal), el tercero con menos productores apoyados (PromProd) y el segundo con menos monto (Monto), menos apoyo a productores con más de 10 hectáreas (PromProd10, Monto10) y menos población (Pob).

  Nótese que son los mismos municipios que los del primer periodo.

Esto se observa en la Figura 3.25, que es la gráfica de las puntuaciones de la primera componente principal, y en la Tabla A.6, que indica el orden de los municipios en cada variable.

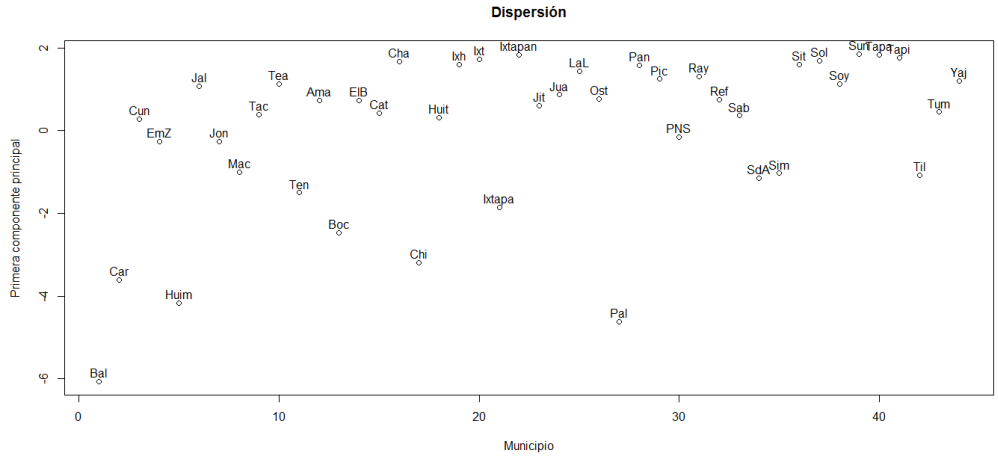

Figura 3.25: Dispersión de la primera componente principal frente municipio, en el segundo periodo.

De la segunda componente principal, los coeficientes de PromProd, Monto y Marg son positivos y los de SupTotal, PromProd10, Monto10 y Pob son negativos (Tabla 3.22), lo que supone un contraste de las variables PromProd, Monto y Marg contra SupTotal, PromProd10, Monto10 y Pob, con mayor peso en PromProd, Monto y Marg, es decir:

- Los municipios con los valores máximos en la segunda componente principal son los que tienen valores grandes en las variables PromProd, Monto y Marg, frente a valores menores en SupTotal, PromProd10, Monto10 y Pob, por ejemplo:
  - Simojovel (Sim) es el segundo municipio con más productores apoyados (PromProd), tercero con más monto (Monto) y el décimo en marginación (Marg).
  - Chilón (Chi) es el tercer municipio con más productores apoyados (PromProd), y segundo con más monto (Monto) y marginación (Marg).
  - Tila (Til) es el municipio con más productores apoyados (PromProd) pero el octavo en monto del apoyo y marginación (Monto y Marg).

  Nótese que son los mismos municipios que los del primer periodo.

- Los municipios con los valores más pequeños en la segunda componente principal son:
  - Balancán (Bal)

- Emiliano Zapata (EmZ)
- Cárdenas (Car).

Nótese que son los mismos municipios que los del primer periodo.

Esto se observa en la Figura 3.26, gráfica de las puntuaciones de la segunda componente principal, y en la Tabla A.6, que indica el orden de los municipios en cada variable.

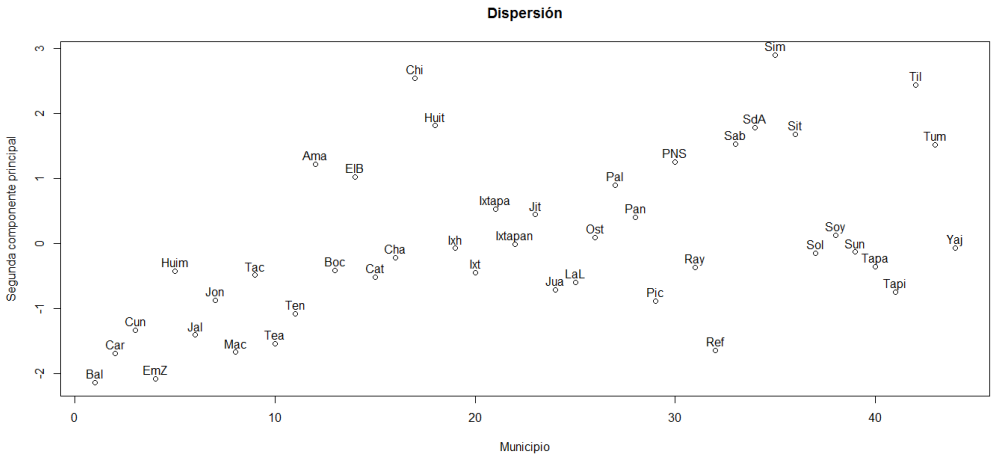

Figura 3.26: Dispersión de la segunda componente principal frente municipio, en el segundo periodo.

En la tercera componente principal, los coeficientes de Monto, PromProd10, Monto10 y Marg son positivos y los de SupTotal, PromProd y Pob son negativos; Monto es muy cercano a cero (Tabla 3.22), lo que contrasta las variables Monto, PromProd10, Monto10 y Marg contra SupTotal, PromProd y Pob. con mayor peso en PromProd10, Monto10 y Pob, es decir:

- Los municipios con los valores máximos en la tercera componente principal son los que tienen valores grandes en las variables PromProd10 y Monto10, y menores en SupTotal, PromProd y Pob, por ejemplo:
  - Balancán (Bal), el municipio con más productores con más de 10 hectáreas y mayor monto a estos productores (PromProd10 y Monto10), es el segundo municipio más extenso (SupTotal).

- Bochil (Boc), segundo municipio con más productores con más de 10 hectáreas y mayor monto a estos productores (PromProd10 y Monto10).
- Ixtapa (Ixtapa), cuarto municipio con más apoyo a productores con más de 10 hectáreas (PromProd10 y Monto10).

Nótese que son los mismos municipios que los del primer periodo.

- Los municipios con los valores más pequeños en la tercera componente principal son:
  - Macuspana (Mac).
  - Huimanguillo (Huim).
  - Cárdenas (Car).

  Nótese que son los mismos municipios que los del primer periodo.

Esto se observa en la Figura 3.27, gráfica de las puntuaciones de la tercera componente principal, y en la Tabla A.6, que indica el orden de los municipios en cada variable.

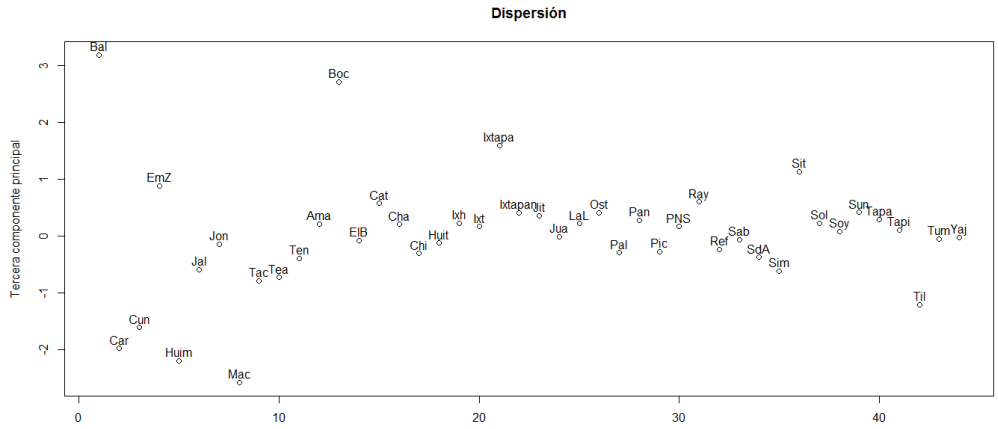

Figura 3.27: Dispersión de la tercera componente principal frente municipio, en el segundo periodo.

3.8.1. Análisis de regresión del segundo periodo

La regresión lineal simple de la variable Exp con cada una de las variables independientes (SupTotal, PromProd, Monto, PromProd10, Monto10, Pob y Marg), indica que

todas las variables independientes son significativas (Tabla 3.23), lo cual sugiere la conveniencia de realizar un análisis de regresión múltiple con todas las variables independientes. El modelo con R^2 y R^2-ajustado máximos es el que tiene a la variable promedio de productores (PromProd, véase Figura 3.28).

Tabla 3.23. *Resumen de los modelos en los que Exp es la variable dependiente, con cada una de las variables independientes como predictor, en el segundo periodo, donde: Estimación es el coeficiente estimado para el predictor en ese modelo; CS es el código significativo: 0 '***'0.001 '**'0.01 '*'0.05 '.'0.1 ' '1.*

Predictor (es)	*Estimación*	*Error estándar*	*Valor t*	$Pr(>\|t\|)$	*CS*	R^2	R^2 *ajustado*
SupTotal	*0.0055*	*0.0015*	*3.607*	*0.000801*	***	*0.2323*	*0.2144*
PromProd	*0.4612*	*0.06823*	*6.759*	$2.88*10^{-8}$	***	*0.5151*	*0.5039*
Monto	$3.1*10^{-5}$	$5.43*10^{-6}$	*5.71*	$9.64*10^{-7}$	***	*0.4313*	*0.418*
PromProd10	*21.452*	*5.876*	*3.651*	*0.000704*	***	*0.2366*	*0.2189*
Intercepto	*529.211*	*193.027*	*2.742*	*0.00894*	**	*0.0877*	*0.066*
PromProd10	*12.724*	*6.334*	*2.009*	*0.05100*	.		
Monto10	$1.75*10^{-4}$	$5.78*10^{-5}$	*3.028*	*0.00415*	**	*0.1758*	*0.1566*
Pob	*0.011787*	*0.002925*	*4.03*	*0.000224*	***	*0.2741*	*0.2572*
Intercepto	$4.615*10^2$	$2.22*10^2$	*2.079*	*0.0438*	*	*0.0727*	*0.0506*
Pob	$6.75*10^{-3}$	$3.718*10^{-3}$	*1.814*	*0.0768*	.		
Marg	*524.2*	*168.7*	*3.106*	*0.00335*	**	*0.1833*	*0.1643*

La regresión paso a paso proporciona el modelo que tiene como predictores al promedio de productores (PromProd), monto asignado (Monto) y el promedio de productores con más de 10 hectáreas (PromProd10), además de contener al intercepto (Tabla 3.24), sin embargo el intercepto no es significativo, por lo tanto elegimos al modelo que sólo incluye al promedio de productores (PromProd), monto asignado (Monto) y promedio de productores con más de 10 hectáreas (PromProd10), dado por (Tabla 3.24):

$$\widehat{Exp} = 1.834 PromProd - 1.592 Monto + 0.582 PromProd10. \tag{3.3}$$

Esta ecuación esta dada para datos estandarizados. Las gráficas de diagnóstico (Figura 3.29) indican la falta de normalidad para los errores y la no independencia entre ellos, lo cual posiblemente se debe a la alta variabilidad que existe entre los valores de las variables en los municipios.

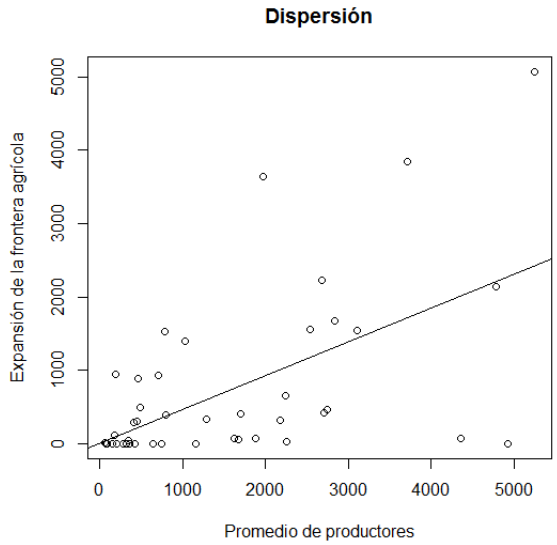

Figura 3.28: Diagrama de dispersión de la expansión de la frontera agrícola (Exp) frente promedio de productores (PromProd), con la recta ajustada por el modelo de regresión, en el segundo periodo.

Tabla 3.24. *Resumen del mejor modelo elegido por regresión paso a paso, en el segundo periodo, donde: Estimación es el coeficiente estimado para el predictor en ese modelo; CS es el código significativo: 0 '***'0.001 '**'0.01 '*'0.05 '.'0.1 ' '1.*

Predictor (es)	*Estimación*	*Error estándar*	*Valor t*	$Pr(>\|t\|)$	*CS*	R^2	R^2 *ajustado*
PromProd	*1.834*	*0.4140*	*4.428*	*6.9*10^{-5}*	***	*0.4689*	*0.4301*
Monto	*-1.592*	*0.4814*	*-3.3*	*0.00197*	**		
PromProd10	*0.582*	*0.1889*	*3.077*	*0.00371*	**		
Intercepto	*-3.4*10^{-16}*	*1.14*10^{-1}*	*0.000*	*1.00000*		*0.4689*	*0.4291*
PromProd	*1.834*	*4.19*10^{-1}*	*4.374*	*8.5*10^{-5}*	***		
Monto	*-1.592*	*4.87*10^{-1}*	*-3.3*	*0.00224*	**		
PromProd10	*5.82*10^{-1}*	*1.91*10^{-1}*	*3.040*	*0.00416*	**		

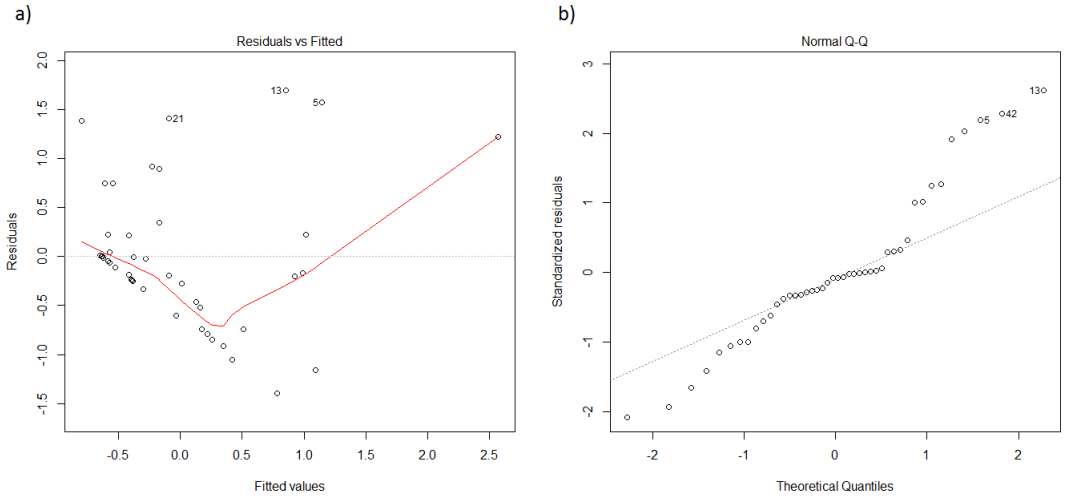

Figura 3.29: Gráficas de diagnósticos del análisis de regresión, en el segundo periodo: a) Residuales frente valores ajustados y b) Gráfico Q-Q.

El análisis de regresión de Exp con las primeras tres componentes principales indica que las dos primeras componentes principales son significativas para explicar a la variable Exp (Tabla 3.25).

Tabla 3.25. *Resumen de los modelo en los cuales Exp es la variable dependiente con las tres primeras componentes principales como variables independientes, en el segundo periodo, donde: CPi es la i-ésima componente principal, $i = 1, 2, 3$, Estimación es el coeficiente estimado para el predictor en ese modelo; CS es el código significativo: 0 '***'0.001 '**'0.01 '*'0.05 '.'0.1 ' '1.*

Predictor (es)	*Estimación*	*Error estándar*	*Valor t*	$Pr(>\|t\|)$	*CS*	R^2	R^2 *ajustado*
CP1	*-0.21441*	*0.06839*	*-3.14*	*0.00313*	**	*0.2716*	*0.2369*
CP2	*0.25051*	*0.10372*	*2.415*	*0.02016*	*		
Intercepto	$-1.8*10^{-16}$	$1.33*10^{-1}$	*0.000*	*1.0000*		*0.2735*	*0.219*
CP1	$-2.1*10^{-1}$	$7*10^{-2}$	*-3.06*	*0.0039*	**		
CP2	$2.5*10^{-1}$	$1.06*10^{-1}$	*2.360*	*0.0232*	*		
CP3	$-4.2*10^{-2}$	$1.31*10^{-1}$	*-0.32*	*0.7516*			

El modelo de regresión que se acepta es:

$$\widehat{Exp} = -0.21441CP1 + 0.25051CP2, \tag{3.4}$$

donde las primeras componentes principales están dadas por las siguientes combinaciones lineales de las variables originales:

$$\begin{aligned} CP1 &= -0.442SupTotal - 0.363PromProd - 0.43Monto - 0.438PromProd10 \\ &\quad -0.401Monto10 - 0.352Pob + 0.105Marg \\ CP2 &= -0.187SupTotal + 0.504PromProd + 0.398Monto - 0.139PromProd10 \\ &\quad -0.219Monto10 - 0.147Pob + 0.681Marg; \end{aligned}$$

además se tiene que:

$$-0.21441CP1 + 0.25051CP2 = 0.0481SupTotal + 0.2043PromProd$$
$$+0.1919Monto + 0.0589PromProd10 + 0.0312Monto10 + 0.0388Pob + 0.1481Marg.$$

Nuevamente esta ecuación considera las variables estandarizadas; todas las variables influyen positivamente, a diferencia del periodo anterior. En este análisis el promedio de productores apoyados (PromProd), su monto del apoyo (Monto) y la marginación (Marg) tienen los coeficientes máximos, es decir, contribuyen más que las demás variables a la expansión de la frontera agrícola. Las gráficas de diagnóstico (Figura 3.30) indican la falta de normalidad para los errores y la no independencia entre ellos, lo cual posiblemente se debe a la alta variabilidad que existe entre los valores de las variables en los municipios.

Resultados del segundo periodo

Al considerar a todas las variables se hallan dos modelos de regresión con problemas en los supuestos necesarios para poder considerar al modelo de regresión como aceptable. Aun con los problemas anteriores se trata de interpretar a los modelos obtenidos, esto es, en el primer modelo las variables que tienen más influencia son el promedio de productores beneficiados (PromProd), el monto del apoyo (Monto) y el promedio de productores con más de 10 hectáreas (PromProd10) mientras que en el segundo modelo las variables que tienen más influencia son el promedio de productores beneficiados

(PromProd), el promedio de productores con más de 10 hectáreas (PromProd10) y el índice de marginación (Marg). Además, el análisis de componentes principales arroja componentes con significados útiles similares a las del primer periodo.

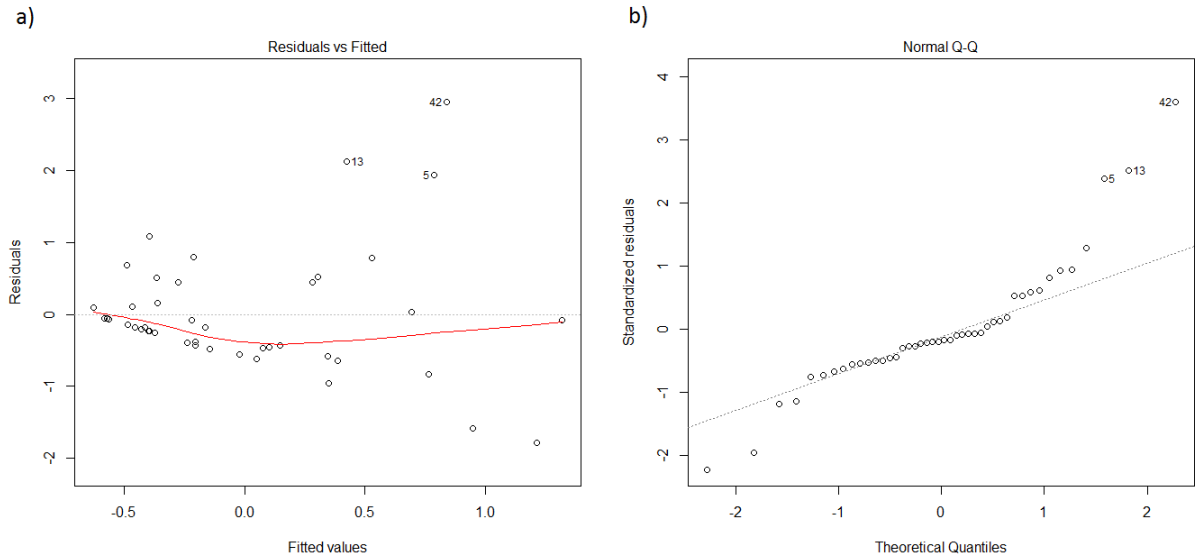

Figura 3.30: Gráficas de diagnósticos del análisis de regresión, en el segundo periodo: a) Residuales frente valores ajustados y b) Gráfico Q-Q.

3.9. Análisis del tercer periodo sin el municipio Centro

La dispersión de las variables independientes bajo estudio medidas en los 44 municipios (Figura 3.31) muestra que existe una relación lineal entre las variables PromProd y Monto, y también entre PromProd10 y Monto10, como en los primeros periodos, lo cual se confirma al observar las correlaciones de estas variables (Tabla 3.26), pues sus correlaciones son 0.933 y 0.949 respectivamente, además se observa que existen altas correlaciones entre las variables independientes.

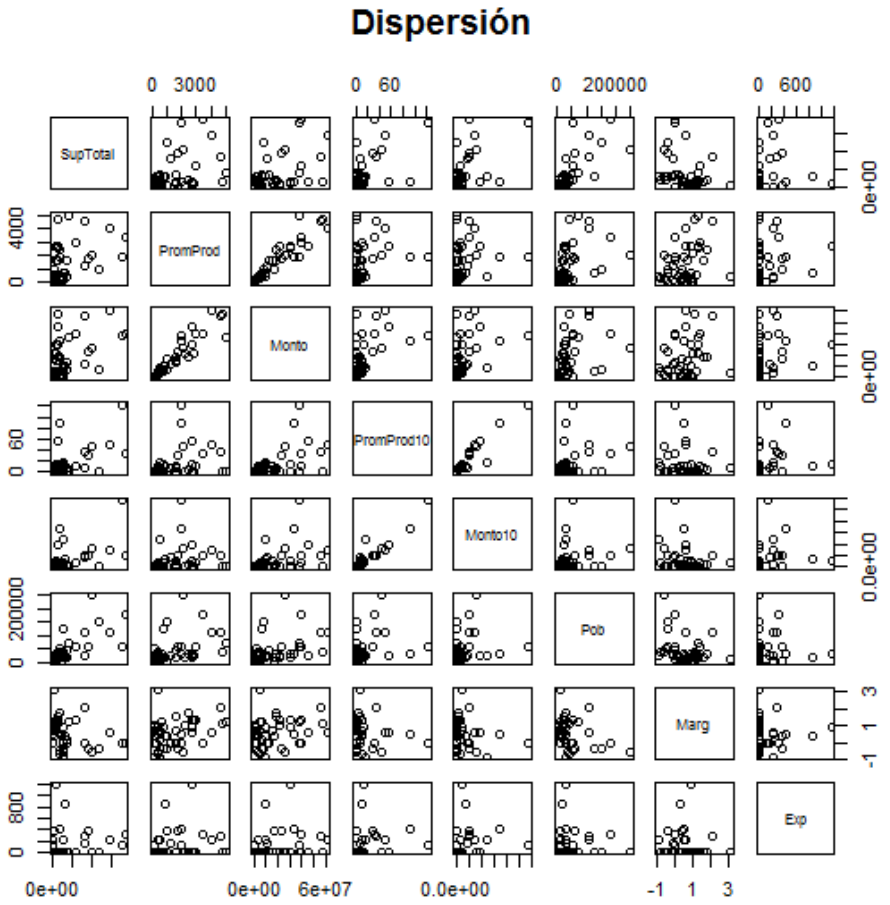

Figura 3.31: Dispersión de las variables bajo estudio medidas en los 44 municipios, en el tercer periodo.

Tabla 3.26. *Correlaciones de las variables independientes medidas en los 44 municipios, en el tercer periodo.*

	SupTotal	*PromProd*	*Monto*	*PromProd10*	*Monto10*	*Pob*	*Marg*
SupTotal	*1.000*	*0.416*	*0.501*	*0.589*	*0.553*	*0.725*	*-0.298*
PromProd		*1.000*	*0.933*	*0.320*	*0.213*	*0.421*	*0.323*
Monto			*1.000*	*0.547*	*0.428*	*0.412*	*0.235*
PromProd10				*1.000*	*0.949*	*0.312*	*-0.132*
Monto10					*1.000*	*0.227*	*-0.206*
Pob						*1.000*	*-0.308*
Marg							*1.000*

El siguiente paso es realizar un análisis de componentes principales de los datos estandarizados de las variables independientes medidas en los 44 municipios estudiados, el cual índica que las primeras tres componentes principales explican el 90.51 % de la variabilidad total (Tabla 3.27); su gráfica de sedimentación (Figura 3.32 a)) muestra que sólo estas componentes tienen valor propio mayor que 1, es decir, en el análisis posterior será suficiente usar las primeras tres componentes en lugar de las siete variables independientes. El diagrama de dispersión de las dos primeras componentes principales se muestra en la Figura 3.32 b), al compararlo con los periodos anteriores (Figura 3.14 b) y Figura 3.24 b)) se nota que en este periodo cambiaron de signo las puntuaciones de la segunda componente principal.

Tabla 3.27. *Importancia de las componentes principales de las variables independientes bajo estudio medidas en los 44 municipios, en el tercer periodo, donde DE, PV y PA significan desviación estándar, proporción de varianza y proporción acumulada, respectivamente.*

	CP1	*CP2*	*CP3*	*CP4*	*CP5*	*CP6*	*CP7*
DE	*1.8837*	*1.2874*	*1.0630*	*0.62117*	*0.45163*	*0.22640*	*0.15239*
PV	*0.5069*	*0.2368*	*0.1614*	*0.05512*	*0.02914*	*0.00732*	*0.00332*
PA	*0.5069*	*0.7437*	*0.9051*	*0.96022*	*0.98936*	*0.99668*	*1.00000*

Tabla 3.28. *Coeficientes de las combinaciones lineales de las variables independientes que definen a las siete componentes principales, en el tercer periodo.*

	CP1	*CP2*	*CP3*	*CP4*	*CP5*	*CP6*	*CP7*
SupTotal	*-0.442*	*0.202*	*-0.251*	*-0.382*	*-0.743*	*0.043*	*0.026*
PromProd	*-0.370*	*-0.502*	*-0.174*	*0.341*	*-0.043*	*-0.462*	*0.500*
Monto	*-0.435*	*-0.396*	*-0.008*	*0.369*	*-0.030*	*0.446*	*-0.565*
PromProd10	*-0.437*	*0.176*	*0.460*	*-0.072*	*0.235*	*0.479*	*0.525*
Monto10	*-0.399*	*0.258*	*0.516*	*-0.032*	*0.101*	*-0.595*	*-0.377*
Pob	*-0.354*	*0.135*	*-0.600*	*-0.331*	*0.614*	*-0.042*	*-0.095*
Marg	*0.062*	*-0.659*	*0.262*	*-0.697*	*0.051*	*-0.016*	*-0.060*

Por último, se interpretarán las primeras componentes principales. Como los coeficientes de la primera componente principal son en su mayoría negativos (Tabla 3.28),

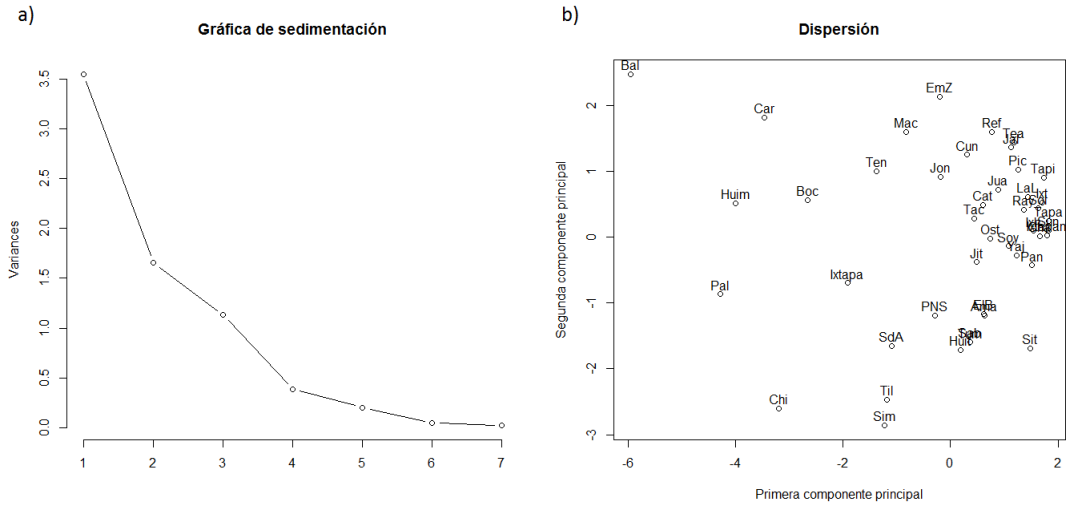

Figura 3.32: Gráficas del análisis de componentes principales de las variables independientes bajo estudio medidas en los 44 municipios, en el tercer periodo: a) Gráfica de sedimentación y b) Dispersión de las dos primeras componentes principales.

excepto Marg, que tiene un coeficiente cercano a cero, se tiene que esta componente es un promedio negativo ponderado correspondiente al tercer periodo, como en los dos periodos anteriores, es decir:

- Los municipios con los valores más pequeños en la primera componente principal son los que tienen valores grandes en alguna o algunas variables independientes, por ejemplo:
  - Balancán (Bal) es el segundo municipio más extenso (SupTotal), con más beneficiados con más de 10 hectáreas (PromProd10) y mayor monto del apoyo a estos productores (Monto10).
  - Palenque (Pal) es el tercer municipio más extenso (SupTotal), es el que tiene mayor monto del apoyo a productores en general (Monto), y el cuarto con más productores (PromProd) y más productores con más de 10 hectáreas (PromProd10).
  - Huimanguillo (Huim) es el municipio más extenso (SupTotal) y el segundo con más población (Pob).

Nótese que los municipios enlistados son los mismos que en el primer y segundo periodos, con los mismos motivos, a excepción de Palenque (Pal), donde son similares.

- Los municipios con los valores máximos en la primera componente principal son los que tienen valores pequeños en las variables independientes, por ejemplo:
  - Tapalapa (Tapa), es el segundo municipio menos extenso (SupTotal), tercero con menos productores apoyados (PromProd) y segundo con menos monto del apoyo (Monto), menos apoyo a productores con más de 10 hectáreas (PromProd10, Monto10) y menos población (Pob).
  - Sunuapa (Sun) es de los municipios menos extensos, el segundo con menos productores apoyados (PromProd) y el que tiene menos población (Pob).
  - Ixtapangajoya (Ixtapan) es uno de los municipios menos extensos (SupTotal), el cuarto con menos productores apoyados (PromProd), y el tercero con menor monto del apoyo (Monto) y menor población (Pob).

  Nótese que son los mismos municipios que los del primer y segundo periodo.

Esto se observa en la Figura 3.33, que es la gráfica de las puntuaciones de la primera componente principal, y en la Tabla A.7, que indica el orden de los municipios en cada variable.

De la segunda componente principal, los coeficientes de PromProd, Monto y Marg son negativos y los de SupTotal, PromProd10, Monto10 y Pob son positivos (Tabla 3.28); esta componente contrasta las variables PromProd, Monto y Marg contra SupTotal, PromProd10, Monto10 y Pob con mayor peso en PromProd, Monto y Marg; tiene los signos opuestos a los de la segunda componente principal de los primeros periodos, lo cual justifica que en la dispersión de las dos primeras componentes principales (Figura 3.32 b)) sus valores tengan signo contrario, es decir:

- Los municipios con los valores más pequeños en la segunda componente principal son los que tienen valores grandes en las variables PromProd, Monto y Marg, y no tienen valores tan grandes en SupTotal, PromProd10, Monto10 y Pob, por ejemplo:

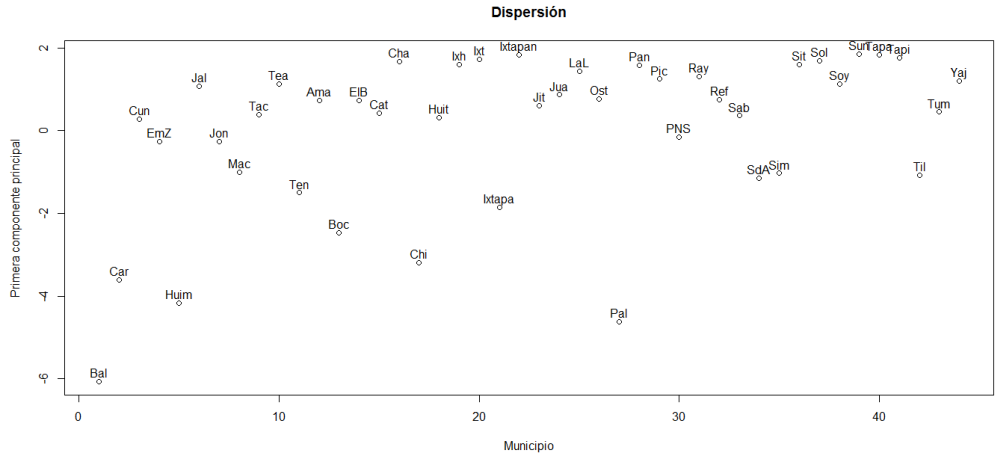

Figura 3.33: Dispersión de la primera componente principal frente municipio, en el tercer periodo.

- Simojovel (Sim) es el segundo municipio con más productores apoyados (PromProd) y más monto (Monto) y es el número 11 en marginación (Marg).
- Chilón (Chi) es el tercer municipio con más productores apoyados (PromProd) y con más monto (Monto) y segundo con más marginación (Marg).
- Tila (Til) es el municipio con más productores apoyados (PromProd) pero tiene el octavo lugar en monto del apoyo y marginación (Monto y Marg).

Nótese que son los mismos municipios de los dos periodos anteriores.

- Los municipios con los valores máximos en la segunda componente principal son:
  - Balancán (Bal)
  - Emiliano Zapata (EmZ)
  - Cárdenas (Car).

Nótese que son los mismos municipios que en los dos periodos anteriores.

Esto se observa en la Figura 3.34, que es la gráfica de las puntuaciones de la segunda componente principal, y en la Tabla A.7, que indica el orden de los municipios en cada variable.

En la tercera componente principal, los coeficientes de PromProd10, Monto10 y Marg son positivos y los de SupTotal, Monto, PromProd y Pob son negativos (Tabla 3.28); así,

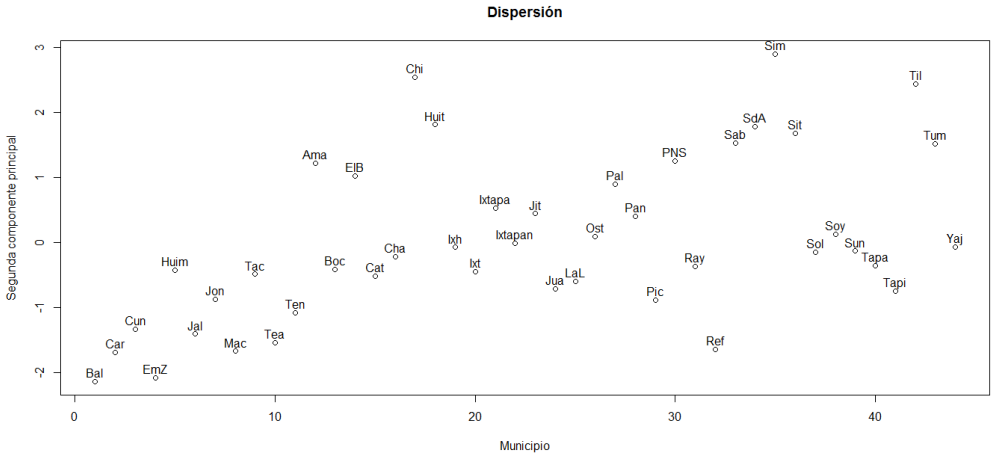

Figura 3.34: Dispersión de la segunda componente principal frente municipio, en el tercer periodo.

esta componente contrasta las variables PromProd10, Monto10 y Marg contra SupTotal, Monto, PromProd y Pob, con mayor peso en PromProd10, Monto10 y Pob, es decir:

- Los municipios con los valores máximos en la tercera componente principal son los que tienen valores grandes en las variables PromProd10 y Monto10, y no tienen valores tan grandes en SupTotal, PromProd y Pob, por ejemplo:
  - Balancán (Bal), que es el municipio con más productores con más de 10 hectáreas y mayor monto a estos productores (PromProd10 y Monto10), es el segundo municipio más extenso (SupTotal).
  - Bochil (Boc) es el segundo municipio con más productores con más de 10 hectáreas y mayor monto a estos productores (PromProd10 y Monto10).
  - Ixtapa (Ixtapa) es el cuarto municipio con más apoyo a productores y a productores con más de 10 hectáreas (PromProd y PromProd10) y el tercero con mayor monto a productores con más de 10 hectáreas (Monto10).

  Nótese que son los mismos municipios que los de los primeros periodos.
- Los municipios con los valores más pequeños en la tercera componente principal son:
  - Macuspana (Mac).

- Huimanguillo (Huim).
- Cárdenas (Car).

Nótese que son los mismos municipios que los primeros periodos.

Esto se observa en la Figura 3.35, que es la gráfica de las puntuaciones de la tercera componente principal, y en la Tabla A.7, que indica el orden de los municipios en cada variable.

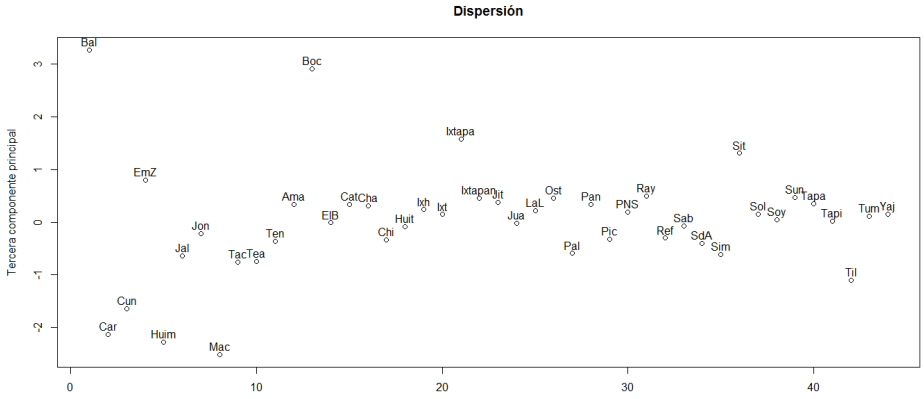

Figura 3.35: Dispersión de la tercera componente principal frente municipio, en el tercer periodo.

3.9.1. Análisis de regresión del tercer periodo

Para realizar el análisis de regresión se realiza una regresión simple de la variable Exp con cada una de las variables independientes (SupTotal, PromProd, Monto, PromProd10, Monto10, Pob y Marg). Como resultado de estos modelos (Tabla 3.29) se tiene que todas las variables independientes son significativas, excepto Marg, lo cual sugiere que se realice un análisis de regresión múltiple con todas las variables independientes. El modelo con el R^2 y el R^2-ajustado máximos es el que tiene a la variable monto del apoyo a productores (Monto, véase Figura 3.36).

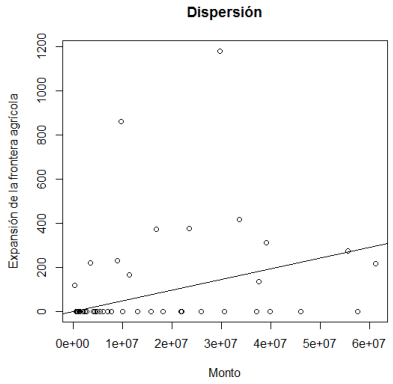

Figura 3.36: Diagrama de dispersión de la expansión de la frontera agrícola (Exp) frente el monto del apoyo (Monto), con la recta ajustada por el modelo de regresión, en el tercer periodo.

Tabla 3.29. *Resumen de los modelo en los cuales Exp es la variable dependiente con cada una de las variables independientes como predictor, en el tercer periodo, donde: Estimación es el coeficiente estimado para el predictor en ese modelo; CS es el código significativo: 0 '***'0.001 '**'0.01 '*'0.05 '.'0.1 ' '1.*

Predictor (es)	*Estimación*	*Error estándar*	*Valor t*	$Pr(>\|t\|)$	*CS*	R^2	R^2 *ajustado*
SupTotal	*0.0008*	*0.0003*	*2.558*	*0.0141*	*	*0.1321*	*0.1119*
PromProd	*0.05425*	*0.01792*	*3.027*	*0.00416*	**	*0.1757*	*0.1565*
Monto	$4.84*10^{-6}$	$1.47*10^{-6}$	*3.29*	*0.00201*	**	*0.2011*	*0.1825*
PromProd10	*3.66*	*1.28*	*2.86*	*0.00651*	**	*0.1598*	*0.1403*
Monto10	$4.15*10^{-5}$	$1.493*10^{-5}$	*2.78*	*0.00803*	**	*0.1524*	*0.1326*
Pob	*0.00115*	*0.000569*	*2.019*	*0.0497*	*	*0.0866*	*0.0654*
Intercepto	*122.48*	*44.49*	*2.753*	*0.00869*	**	*0.0048*	*-0.019*
Marg	*-20.27*	*45.02*	*-.45*	*0.65494*			

La regresión paso a paso de los datos estandarizados proporciona el modelo que tiene como predictor al promedio de productores con más de 10 hectáreas (PromProd10), además de contener al intercepto (Tabla 3.30), sin embargo la variable no es significativa, por lo que el R^2 y el R^2-ajustado son muy pequeños, por lo tanto no se encuentra ningún modelo aceptable por este método.

Tabla 3.30. *Resumen del modelo elegido por regresión paso a paso, en el tercer periodo, donde: Estimación es el coeficiente estimado para el predictor en ese modelo; CS es el código significativo: 0 '***'0.001 '**'0.01 '*'0.05 '.'0.1 ' '1.*

Predictor (es)	*Estimación*	*Error estándar*	*Valor t*	$Pr(>\|t\|)$	*CS*	R^2	R^2 *ajustado*
PromProd10	*0.2443*	*0.1479*	*1.652*	*0.106*		*0.0599*	*0.0378*
Intercepto	*-1.68*$*10^{-16}$	*1.479*$*10^{-1}$	*0.000*	*1.00*		*0.0597*	*0.037*
PromProd10	*2.443*$*10^{-1}$	*1.496*$*10^{-1}$	*1.633*	*0.11*			

El análisis de regresión mediante componentes principales indica que ninguna de estas componentes principales es significativa y que el intercepto tampoco es significativo para explicar la variable Exp. El R^2 y el R^2-ajustado son muy bajos y tienen variables no significativas, por lo se concluye que las componentes principales no explican la variable Exp en este periodo (Tabla 3.31).

Tabla 3.31. *Resumen del modelo en el que Exp es la variable dependiente con las tres primeras componentes principales como variables independientes, en el tercer periodo, donde: CPi es la i-ésima componente principal, $i = 1, 2, 3$, Estimación es el coeficiente estimado para el predictor en ese modelo; CS es el código significativo: 0 '***'0.001 '**'0.01 '*'0.05 '.'0.1 ' '1.*

Predictor (es)	*Estimación*	*Error estándar*	*Valor t*	$Pr(>\|t\|)$	*CS*	R^2	R^2 *ajustado*
Intercepto	*-1.77*$*10^{-16}$	*1.51*$*10^{-1}$	*0.000*	*1.000*		*0.0707*	*0.0010*
CP1	*-1.24*$*10^{-1}$	*8.1*$*10^{-2}$	*-1.528*	*0.134*			
CP2	*1.09*$*10^{-3}$	*1.19*$*10^{-1}$	*0.009*	*0.993*			
CP3	*1.21*$*10^{-1}$	*1.43*$*10^{-1}$	*0.842*	*0.405*			

Resultados del tercer periodo

A diferencia de los periodos anteriores, al considerar a todas las variables no se encontraron modelos aceptables, sólo se logra encontrar un modelo de regresión lineal

simple que explica a la expansión de la frontera agrícola mediante el monto del apoyo (Monto) cuyo R^2 y R^2-ajustado son menores que 0.21. Además, el análisis de componentes principales arroja componentes con significados útiles similares a los periodos anteriores.

3.10. Resultados

El análisis exploratorio indica que Centro posee condiciones peculiares, por lo cual no se consideró en los análisis subsiguientes. Los análisis de componentes principales arrojan componentes similares, con interpretaciones similares y puntuaciones similares para los diferentes municipios considerados, como se observa en las Figuras 3.37, 3.38 y 3.39, en las cuales se comparan las puntuaciones de las tres primeras componentes principales de los diferentes periodos; además, estas componentes explican alrededor del 91 % de la variación total de las variables independientes consideradas, por lo que se puede decir que el comportamiento de las variables independientes se ha mantenido al transcurrir de los años, a pesar de que:

- En promedio, el número de productores anuales ha disminuido en cada periodo de $499781/7 = 71397$ (primer periodo) a $334077/5 = 66815.4$ (segundo periodo) y luego a $255649/4 = 63912.25$ (tercer periodo).

- En monto del apoyo a los productores anualmente ha aumentado en los periodos de $761499010/7 = 108785572.86$ (primer periodo) a $885028129/5 = 177005625.8$ (segundo periodo) y luego a $748597345/4 = 187149336.2$ (tercer periodo).

- En promedio, el número de productores anuales con más de 10 hectáreas ha aumentado y luego disminuido, es decir, del valor $4283/7 = 611.9$ (primer periodo) aumentó a $3370/5 = 674$ (segundo periodo) y luego disminuyó a $2410/4 = 602.5$ (tercer periodo).

- En monto del apoyo a los productores con más de 10 hectáreas anualmente ha aumentado y luego disminuyó en los periodos, de $61680245/7 = 8811463.59$ (primer

periodo) a $68719415/5 = 13743882.99$ (segundo periodo) y luego a $50267916/4 = 12566979.06$ (tercer periodo).

- La población ha aumentado en los periodos considerados de 1614451 (primer periodo) a 1713421 (segundo periodo) y luego a 1900455 (tercer periodo). Esta información se encuentra en la Tabla 3.32.

Es decir, pese a que en promedio, el número de productores beneficiados por PROCAMPO ha disminuido, el monto del apoyo ha aumentado y la forma de distribución de PROCAMPO ha variado, además del crecimiento poblacional, se ha mantenido gran parte de la variación de estas variables a lo largo de los años.

Tabla 3.32. *Resultados por periodo, donde Productores es el total de productores apoyados en el periodo correspondiente y Monto es su respectivo apoyo, Productores 10 es el total de productores con más de 10 hectáreas apoyados en el periodo correspondiente y Monto 10 es su respectivo apoyo, y Expansión es el total de expansión de la frontera agrícola en el periodo correspondiente.*

Periodo	*Productores*	*Monto*	*Productores 10*	*Monto 10*	*Población*	*Expansión*
1993-2002	*499781*	*761499010*	*4283*	*61680245*	*1614451*	*19759*
2003-2007	*334077*	*885028129*	*3370*	*68719415*	*1713421*	*31861*
2008-2011	*255649*	*748597345*	*2410*	*50267916*	*1900455*	*4874*

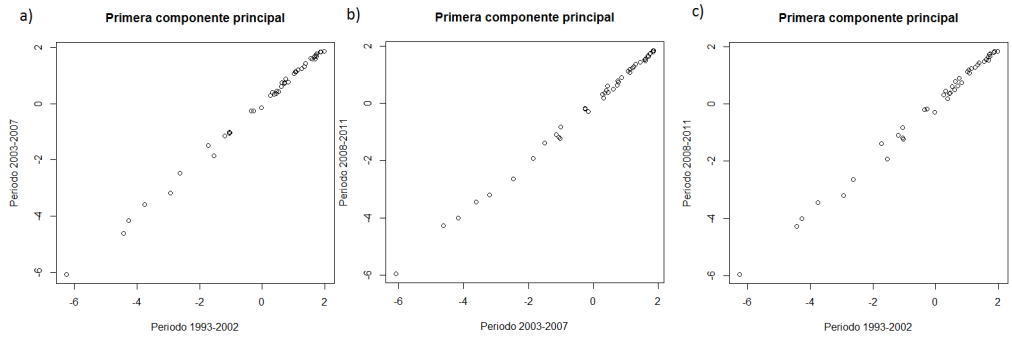

Figura 3.37: Gráficas comparativas de la primera componente principal: a) 2003-2007 frente 1993-2002, b) 2008-2011 frente 2003-2007 y c) 2008-2011 frente 1993-2002.

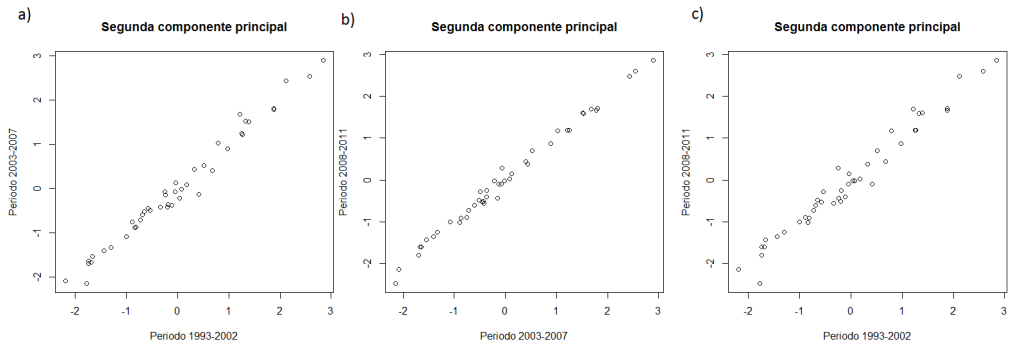

Figura 3.38: Gráficas comparativas de la segunda componente principal, se graficaron las puntuaciones negativas del tercer periodo debido a que los coeficientes de esa componente tienen signos opuestos a los coeficientes de las componentes de los otros periodos: a) 2003-2007 frente 1993-2002, b) 2008-2011 frente 2003-2007 y c) 2008-2011 frente 1993-2002.

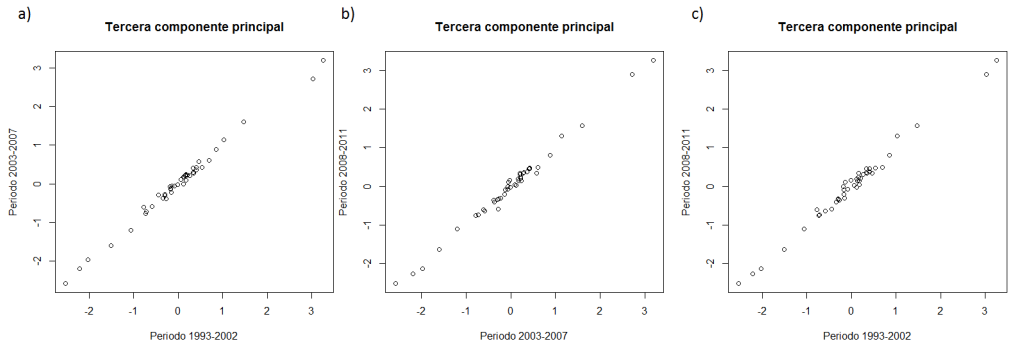

Figura 3.39: Gráficas comparativas de la tercera componente principal: a) 2003-2007 frente 1993-2002, b) 2008-2011 frente 2003-2007 y c) 2008-2011 frente 1993-2002.

El municipio Huimanguillo tuvo una expansión de frontera agrícola sobresaliente en el primer periodo, por lo cual no se considero en el análisis de regresión correspondiente. En los tres periodos existe una relación lineal entre la expansión de la frontera agrícola y cada una de las variables independientes consideradas, con diferentes grados de significancia, excepto el índice de marginación, que no tiene relación lineal significante en el tercer periodo.

Para los dos primeros periodos, la regresión paso a paso y la regresión mediante el análisis de componentes principales proporcionan modelos con problemas en los supuestos necesarios para poder considerar al modelo de regresión como aceptable, los

cuales no explican completamente la expansión de la frontera agrícola, a diferencia del tercer periodo, en el cual no se encontró ningún modelo aceptable. Dado que las variables independientes han mantenido gran porcentaje de su variabilidad al transcurrir de los años, lo que ha cambiado es la expansión de la frontera agrícola, lo cual es cierto, ya que el promedio anual del primer periodo: $19759/9 = 2195$ aumentó a $31861/5 = 6372$ (promedio anual del segundo periodo) y luego disminuyó a $4874/4 = 1219$ (promedio anual del tercer periodo, Figura 3.40).

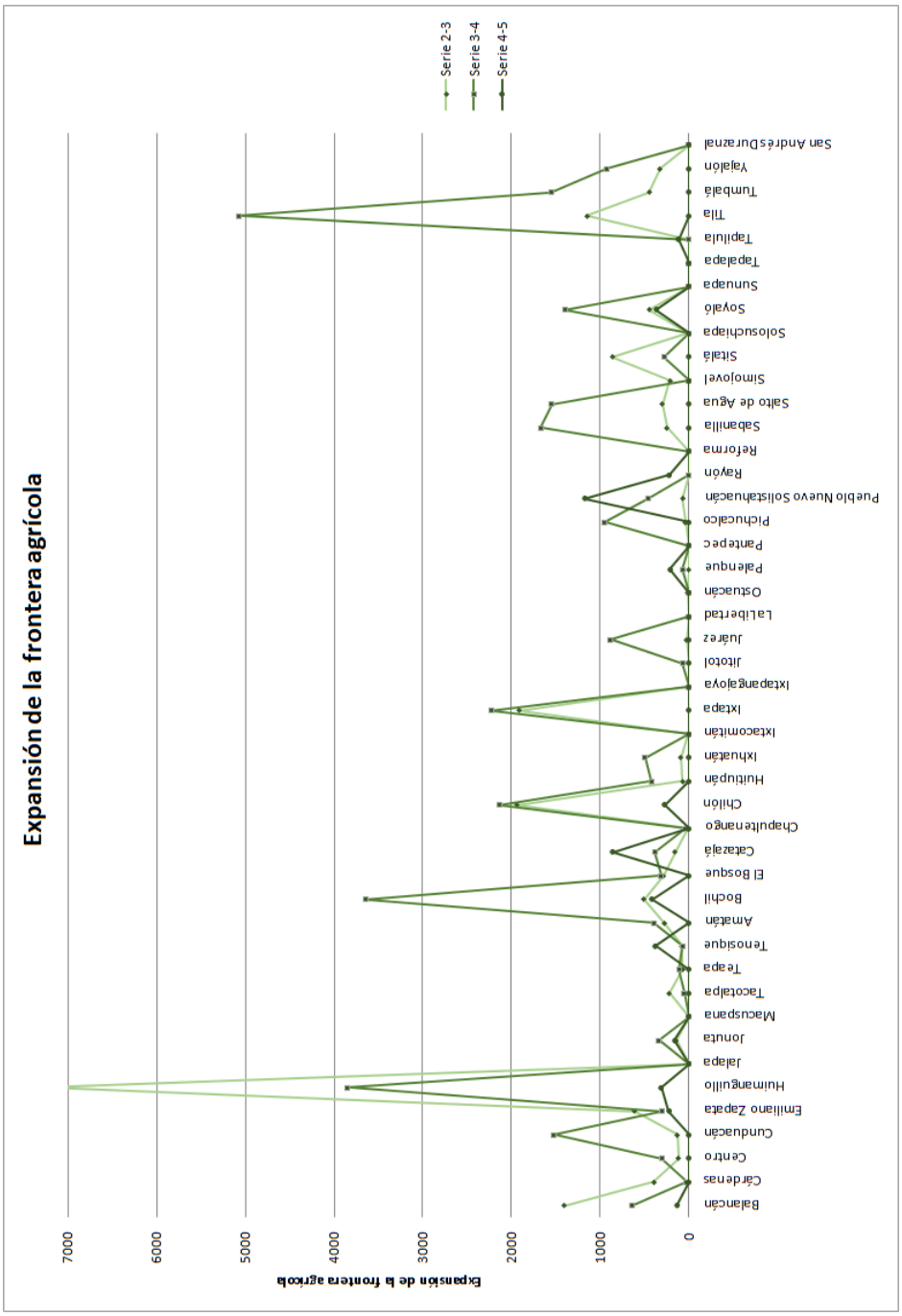

Figura 3.40: Gráfica comparativa de la expansión de la frontera agrícola en los periodos 1993-2002 (Serie 2-3), 2003-2007 (Serie 3-4), 2008-2011 (Serie 4-5). Elaboración propia con base en [17].

Conclusiones

El análisis de regresión es muy extenso debido a las aplicaciones que tiene; en este trabajo se estudió un caso especial: el análisis de regresión lineal múltiple. Se presentan diversos resultados útiles, tales como: la construcción del modelo; los métodos de selección de modelos; las estimaciones, propiedades, pruebas de hipótesis, predicciones y comprobación de supuestos, que constituyen uno de los primeros pasos a efectuar cuando se desea hallar una relación de dependencia entre variables.

Por otro lado, también se presentan los principios del análisis de componentes principales (ACP), su construcción y las diferentes propiedades útiles que poseen, entre las cuales se cuenta la reducción de la dimensión, es decir, analizar un número menor de variables sin pérdida de mucha información, lo cual resulta muy útil, ventajoso y complementario para diferentes análisis. También se estudian pruebas de hipótesis y algunas de sus aplicaciones, por ejemplo, el Análisis de regresión lineal múltiple, en el que las variables originales son sustituidas por las componentes principales cuando éstas se hallan altamente correlacionadas; evitando con ello una de las principales causas de inestabilidad en la regresión, que es la multicolinealidad. En este trabajo se logró el desarrollo de las pruebas de los teoremas referidos a las pruebas de hipótesis en el análisis de regresión, así como en el análisis de componentes principales.

Como una aplicación de los métodos antes mencionados, se analizó la expansión de la frontera agrícola, que junto con la pecuaria son la principal causa de deforestación en México. El análisis se efectuó con datos de la región transfronteriza entre Tabasco y Chiapas, donde los objetos de estudio son los municipios. Al analizar si existe una relación lineal entre la expansión de la frontera agrícola y la superficie total, el apoyo PROCAMPO, la población total y el índice de marginación.Se obtuvieron modelos de los tres periodos que proporcionan una posible explicación al fenómeno: las variables del subsidio PROCAMPO tiene más influencia en la expansión de la frontera agrícola que el resto de las variables consideradas; sin embargo, estos modelos no son del todo satisfactorios debido a su bajo coeficiente R^2. Con el análisis de componentes principales

se obtuvieron importantes descripciones del fenómeno en estudio, tales como identificar puntos atípicos, se logró además una mejor comprensión de las variables independientes que se consideran para explicar el comportamiento de la expansión de la frontera agrícola.

Se obtuvieron modelos de regresión erráticos en cuanto a explicar la expansión de la frontera agrícola en términos de características de los municipios como “superficie total”, “promedio de productores”, “monto del apoyo”, “promedio de productores con más de 10 hectáreas”, “monto del apoyo a productores con más de 10 hectáreas” y “marginación”, lo cual lleva a plantear la aplicación de otros métodos estadísticos o aplicar regresión a grupos de municipios homogéneos.

Una opción para continuar con el análisis (en nuevas tesis o investigaciones) es formar una agrupación en la cual se estudien municipios con características similares (por ejemplo: por extensión territorial) con el fin de obtener mejores resultados, ya que debido a la gran diferencia existente entre algunas características, aplicar un criterio para toda la zona no es lo más conveniente.

Otro tema de interés es la expansión de la frontera pecuaria, pues se tiene que ésta aumentó más que la agrícola en los periodos considerados (Figura 3.41) por lo que sería un importante tema de estudio posterior.

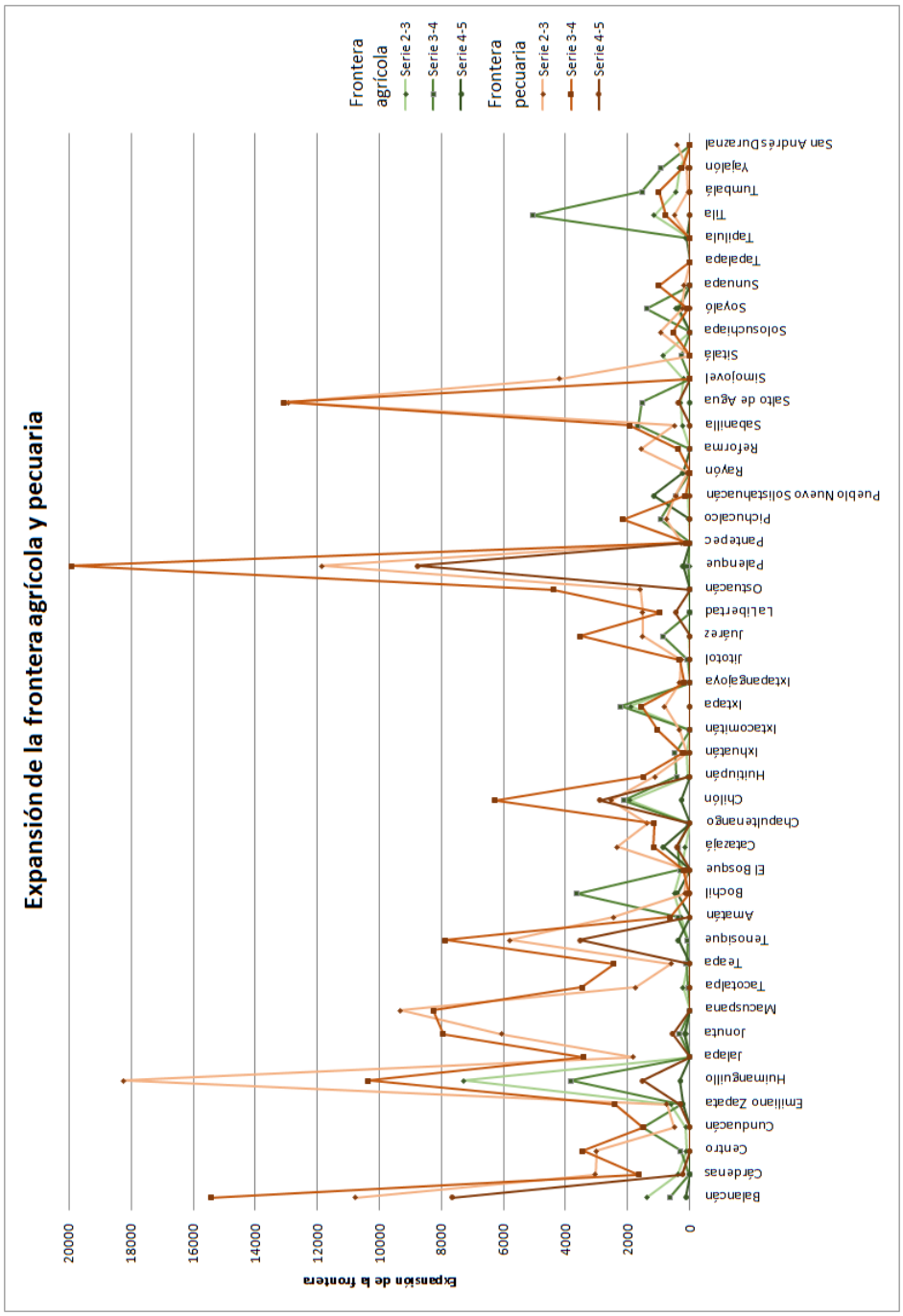

Figura 3.41: Gráfica comparativa de la expansión de la frontera agrícola y la pecuaria en los periodos 1993-2002 (Serie 2-3), 2003-2007 (Serie 3-4), 2008-2011 (Serie 4-5). Elaboración propia con base en [17].

Apéndice A

Datos utilizados

El objetivo de este apéndice es describir el procedimiento que se llevó a cabo para obtener los datos usados en el análisis del Capítulo 3. Las bases de datos constan de información de los municipios de la región Tabasco-Chiapas (Tabla A.1), de los cuales se investigó la superficie total, el total de la población, el índice de marginación, y la expansión de la frontera agrícola, además se obtuvo información del Programa de Apoyos Directos al Campo (PROCAMPO). La Tabla A.1 enlista los municipios de la región Tabasco-Chiapas, además de la entidad y microrregión a la que pertenecen, la columna con encabezado Mun muestra una abreviatura propuesta del nombre del municipio, la cual se usa a lo largo del análisis.

Tabla A.1. *Municipios de la región Tabasco-Chiapas.*

Entidad	***Municipio***	***Mun***
Tabasco	*Balancán*	*Bal*
Tabasco	*Cárdenas*	*Car*
Tabasco	*Centro*	*Cen*
Tabasco	*Cunduacán*	*Cun*
Tabasco	*Emiliano Zapata*	*EmZ*
Tabasco	*Huimanguillo*	*Huim*
Tabasco	*Jalapa*	*Jal*
Tabasco	*Jonuta*	*Jon*

Tabla A.1 Continuación.

Entidad	Municipio	Mun
Tabasco	Macuspana	Mac
Tabasco	Tacotalpa	Tac
Tabasco	Teapa	Tea
Tabasco	Tenosique	Ten
Chiapas	Amatán	Ama
Chiapas	Bochil	Boc
Chiapas	El Bosque	ElB
Chiapas	Catazajá	Cat
Chiapas	Chapultenango	Cha
Chiapas	Chilón	Chi
Chiapas	Huitiupán	Huit
Chiapas	Ixhuatán	Ixh
Chiapas	Ixtacomitán	Ixta
Chiapas	Ixtapa	Ixtapa
Chiapas	Ixtapangajoya	Ixtapan
Chiapas	Jitotol	Jit
Chiapas	Juárez	Jua
Chiapas	La Libertad	LaL
Chiapas	Ostuacán	Ost
Chiapas	Palenque	Pal
Chiapas	Pantepec	Pan
Chiapas	Pichucalco	Pic
Chiapas	Pueblo Nuevo Solistahuacán	PNS
Chiapas	Rayón	Ray
Chiapas	Reforma	Ref
Chiapas	Sabanilla	Sab
Chiapas	Salto de Agua	SdA

Tabla A.1 Continuación.

Entidad	Municipio	Mun
Chiapas	Simojovel	Sim
Chiapas	Sitalá	Sit
Chiapas	Solosuchiapa	Sol
Chiapas	Soyaló	Soy
Chiapas	Sunuapa	Sun
Chiapas	Tapalapa	Tapa
Chiapas	Tapilula	Tapi
Chiapas	Tila	Til
Chiapas	Tumbalá	Tum
Chiapas	Yajalón	Yaj
Chiapas	San Andrés Duraznal	SAD

La superficie total (SupTotal) de los municipios hace referencia la extensión de cada municipio expresado en hectáreas.

El total de la población (Pob) de cada municipio se consultó en la página web del INEGI (www.inegi.org.mx) y es el conjunto de personas que residen en el lugar en el momento de la entrevista, ya sean nacionales o extranjeros. Se incluye a los mexicanos que cumplen funciones diplomáticas fuera del país y a los familiares que vivan con ellos; así como a los que cruzan diariamente la frontera para trabajar en otro país, y también a la población sin vivienda. No se incluye a los extranjeros que cumplen con un cargo o misión diplomática en el país ni a sus familiares. Esta información se obtuvo para los años 1995, 2000, 2005 y 2010.

El índice de marginación (Marg) de cada municipio se consultó en la página web del SNIM (www.snim.rami.gob.mx) y se obtuvo para los años 2000, 2005 y 2010.

La expansión de la frontera agrícola (Exp) la proporcionó CentroGeo [17], y consta de la expansión de la frontera agrícola (en hectáreas) que hubo en los periodos comprendidos entre la serie I y la serie II; la serie II y la serie III (1993-2002); la serie III y la serie IV (2002-2007), y la serie IV y la serie V (2007-2011) de las Cartas de uso de suelo y vegetación del INEGI, por municipio.

Para obtener información acerca del apoyo PROCAMPO se consultó la página web de SAGARPA (www.sagarpa.gob.mx), que contiene las listas de beneficiarios de PROCAMPO de los dos periodos, primavera-verano y otoño-invierno, desde 1994 a 2013. En este estudio sólo se consideró el periodo de primavera-verano, por lo que se descargaron las listas de beneficiarios de los estados de Tabasco y Chiapas de todos los años disponibles, es decir, de 1995 a 2013 excepto 1998.

Las listas de beneficiarios de PROCAMPO contiene la siguiente información para cada uno de los productores beneficiados:

- Nombre del estado
- Nombre del DDR
- Nombre del CADER
- Nombre del municipio
- Nombre del Ejido
- Nombre Productor
- Superficie apoyada
- Importe apoyado
- Ciclo
- Nombre del cultivo
- Régimen Hídrico
- Programa.

Mediante la creación de tablas dinámicas en Excel, para cada año del cual se tiene información se realizaron los siguientes cálculos: 1) se contaron los nombres de los productores por municipio, 2) se sumó el importe apoyado por municipio, 3) se contaron los nombres de los productores con más de 10 hectáreas de superficie apoyada por municipio y 4) se sumó el importe apoyado a productores con más de 10 hectáreas de superficie apoyada por municipio.

Dado que la expansión de la frontera agrícola está dada por periodos, la información de PROCAMPO también se calculó para los periodos 1995-2002, 2003-2007 y 2008-2011, es decir, usando la información de las tablas dinámicas para cada uno de estos periodos, se promediaron las cuentas de los nombres de los productores (PromProd), se sumó el importe del apoyo (Monto), se promediaron las cuentas de los nombres de los productores con más de 10 hectáreas de superficie apoyada (PromProd10), se sumó el importe apoyado a productores con más de 10 hectáreas de superficie apoyada (Monto10).

Por último, se construyeron tres bases de datos, todas con los encabezados Mun, SupTotal, PromProd, Monto, PromProd10, Monto10, Pob, Marg, Exp. En la primera base

de datos (Tabla A.2) se colocó la información de PromProd, Monto, PromProd10, Monto10 del periodo 1995-2002, de Pob y Marg del año 2000 y de Exp entre la serie II a la serie III (1993-2002). En la segunda base de datos (Tabla A.3) se colocó la información de PromProd, Monto, PromProd10, Monto10 del periodo 2003-2007, de Pob y Marg del año 2005 y de Exp entre la serie III a la serie IV (2002-2007). En la tercera base de datos (Tabla A.4) se colocó la información de PromProd, Monto, PromProd10, Monto10 del periodo 2008-2011, de Pob y Marg del año 2010 y de Exp entre la serie IV a la serie V (2007-2011).

También se realizaron tablas en las que se muestra el orden de los municipios (excepto el municipio Centro) respecto al tamaño cada variable, en orden decreciente (Tabla A.5, Tabla A.6 y Tabla A.7).

Tabla A.2. *Información de 1993 a 2002.*

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
Bal	357695.523	2664	45235728.82	109.4285714	13966610.96	54265	-0.12047	1401
Car	204693.744	2554.714286	29799981.19	47.14285714	3742651.02	217261	-0.5062	396.2
Cen	171876.253	1586.285714	10967475.87	0	0	520308	-1.47823	118
Cun	59558.70164	895.142857	6351978.97	1	92776	104360	-0.17398	128
EmZ	59323.14699	495.714286	9847285.81	17.42857143	5425908.16	26951	-0.93732	614
Huim	371412.648	4166.857143	43252947.79	33.14285714	2626316.61	158573	0.23528	7278
Jal	59158.57934	465.428571	2775860.85	0	0	32840	-0.43712	0
Jon	163542.366	1402.285714	11839910.38	9	2019854	27807	0.14918	141
Mac	242777.721	1413.714286	9414864.74	0.28571429	20424	133985	-0.40745	0
Tac	73505.43055	1747.428571	13170372.48	2.42857143	187052.5	41296	-0.03262	220.9
Tea	42178.20264	223.857143	1511809.82	0	0	45834	-0.54712	70.19
Ten	188225.615	2123.571429	26751738.14	30.57142857	2685883.8	55712	-0.37821	75.81
Ama	31704.83595	1888.142857	19139422.06	5.28571429	361150.5	18778	1.75796	279.9
Boc	36648.42095	2129	34020274.15	94.57142857	7294317.66	22722	0.57721	509.8
ElB	15875.81056	2238.857143	20441539.21	0	0	14993	0.89123	288.2
Cat	63256.10576	721.714286	8779444.23	12.57142857	1964389	15709	0.43747	158.2
Cha	17889.85717	373.571429	1838903.1	0	0	6965	1.19021	5.35396
Chi	168101.109	5133.857143	56250268.27	33.14285714	2827493	77686	2.04647	1936

Tabla A.2 Continuación.

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
Huit	33849.85621	2808.571429	29453315.7	0	0	20041	1.79575	71.63
Ixh	9472.790976	552.714286	4284589.2	0	0	8877	0.93346	91.27
Ixt	12758.018	160.285714	794991.8	0	0	9143	0.57135	10.44
Ixtapa	27849.01527	2726.142857	39903380.2	46.85714286	4430877.38	18533	0.59721	1908
Ixtapan	10683.70966	99.285714	822260.24	0.85714286	62363	4707	1.34714	0.53949
Jit	23600.0088	1611.428571	14462287.76	12	1131028	13076	0.93599	0.1134
Jua	74030.25455	458.571429	7596463.32	10	832523	19956	0.40491	25.72
LaL	45531.35488	227.428571	2699287.88	3.14285714	371336	5288	0.45358	0
Ost	60062.21375	778	10499651.31	10.57142857	775728.5	17026	1.23439	15.98
Pal	288601.793	4777.571429	65330983.23	50.42857143	3876863	85464	0.50717	3.71694
Pan	10666.29203	630	5129675.77	0	0	8566	1.75371	0
Pic	59258.98016	201.857143	1566003.1	0.28571429	18722	29357	0.40957	41.2
PNS	24633.64857	2868	27240450.83	12.28571429	966499	24405	1.25533	72.28
Ray	6863.619674	369	4064681.66	12.42857143	900536	6870	1.08637	0
Ref	43268.42604	321.142857	5937324.66	12.57142857	972658	34809	-0.67515	1.44801
Sab	24920.02167	2803.428571	20726103.36	4.71428571	287007	21156	1.41512	250.2
SdA	122958.966	3359.571429	42479708.57	15.71428571	1038108.06	49300	1.74947	297.2
Sim	31495.02891	5252.142857	56717113.34	0	0	31615	1.32428	212.9

Tabla A.2 Continuación.

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
Sit	10568.08259	441.571429	4535904.1	9.28571429	1222190	7987	2.65901	865.3
Sol	15720.5103	301.571429	2694601.65	0	0	7784	0.86077	9.2327
Soy	9577.397064	984.571429	14817109.6	4.14285714	283048	7767	0.55361	450.8
Sun	7908.367581	75.285714	894734	0.42857143	21520	1936	1.73762	0
Tapa	6527.218485	87.428571	683846.5	0	0	3639	1.01277	0
Tapi	4195.751134	61.571429	316566.5	0	0	10349	0.23488	0
Til	80190.72387	5264	33909461.14	2.85714286	510120	58153	1.35945	1152
Tum	40227.9647	2702.285714	17920059.68	2.57142857	429142	26866	1.6726	449.8
Yaj	20878.319	836	5596124.93	4.71428571	335149	26044	0.71975	326.9
SAD	3727.2496					3423	1.27767	0

Tabla A.3. *Información de 2003 a 2007.*

Mun	*SupTotal*	*PromProd*	*Monto*	*PromProd10*	*Monto10*	*Pob*	*Marg*	*Exp*
Bal	*357695.523*	*2236.6*	*48311881.65*	*124.8*	*15239924.03*	*53038*	*-0.2485*	*644.9*
Car	*204693.744*	*2258.6*	*32147706.01*	*51.6*	*4467990.57*	*219563*	*-0.5441*	*25.16*
Cen	*171876.253*	*1357.4*	*12715504.33*	*0.8*	*45888*	*558524*	*-1.4908*	*306.5*
Cun	*59558.70164*	*776*	*7040760.65*	*1*	*104038*	*112036*	*-0.3793*	*1520*
EmZ	*59323.14699*	*441.2*	*10961503.05*	*18*	*6223577.85*	*26576*	*-0.9321*	*307.2*
Huim	*371412.648*	*3706.2*	*47265717.38*	*38.8*	*3032743.9*	*163462*	*0.0375*	*3844*

Tabla A.3 Continuación.

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
Jal	59158.57934	420.4	3175353.25	0	0	33596	-0.5656	0
Jon	163542.366	1291	13558935.22	10.8	2342037.52	28403	0.0358	335.6
Mac	242777.721	1158	10722316.57	1	56748	142954	-0.4839	0
Tac	73505.43055	1667.4	15121815.66	0.8	109214	42833	-0.1421	52.35
Tea	42178.20264	177.6	1530957.39	0	0	49262	-0.5677	113.6
Ten	188225.615	1885.2	28613890	32.6	2474756.04	55601	-0.5512	71.69
Ama	31704.83595	1700	21440108.58	5.4	389609.5	19637	1.7543	397.7
Boc	36648.42095	1974.2	38665151.94	94.6	8020610.25	26446	0.4781	3644
ElB	15875.81056	2179.2	24419988.8	0	0	14932	1.0742	313.8
Cat	63256.10576	792.6	11859188.1	15	2582242	15876	0.502	385.3
Cha	17889.85717	343.6	2206228.95	0	0	7124	0.8236	36.44
Chi	168101.109	4778.6	66834863.7	40.8	3409188.25	95907	2.1526	2140
Huit	33849.85621	2707.6	34824674.5	0	0	20087	1.6624	423.9
Ixh	9472.790976	489.6	4813431.75	0.8	52620	8734	0.8729	493.1
Ixt	12758.018	145.8	1015400.35	0	0	9696	0.6378	0
Ixtapa	27849.01527	2680	51726082.93	57.2	5461040.5	21705	0.5014	2231
Ixtapan	10683.70966	84.6	907040.35	1	69752.75	4911	1.2272	0
Jit	23600.0088	1618.4	17695682.4	11	1186536	15005	0.9543	70.16

Tabla A.3 Continuación.

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
Jua	74030.25455	463.4	8570377.2	7	680976	20173	0.2797	881.8
LaL	45531.35488	205.4	2946598	3	374365	5286	0.497	0
Ost	60062.21375	736.8	11927947.1	13.6	975663	16392	1.1358	0
Pal	288601.793	4362.8	75368795.18	64.8	4744358.71	97991	0.6044	74.09
Pan	10666.29203	644.6	6691977.5	0	0	9785	1.3733	0
Pic	59258.98016	184	1743589.3	1.6	89766	29583	0.2571	949.9
PNS	24633.64857	2748.2	33358028.7	15.2	1205323	27832	1.1851	459.7
Ray	6863.619674	355.6	4489520.55	13	970595	7965	0.7222	0
Ref	43268.42604	319.8	6498956.62	11.8	894042.32	34896	-0.7757	0
Sab	24920.02167	2834.6	25644571.03	4.6	284915.25	23675	1.5195	1671
SdA	122958.966	3103.6	48064810.99	16	1060721	53547	1.6932	1547
Sim	31495.02891	4925	65665786.05	1	60930	32451	1.3528	0
Sit	10568.08259	411.2	5102395.9	8.2	1199784.25	10246	3.3551	280.9
Sol	15720.5103	275.8	3118856.15	0	0	7900	0.9149	0
Soy	9577.397064	1023.8	18331587.08	1	94580	8852	0.5537	1400
Sun	7908.367581	70.8	1025907	1	63841.5	2088	1.0543	0
Tapa	6527.218485	81.2	811538.75	0	0	3928	0.7478	0
Tapi	4195.751134	62.6	421861.25	0	0	9934	0.2796	3.33668

Tabla A.3 Continuación.

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
Til	80190.72387	5250.4	43119141.75	0	0	63172	1.4233	5069
Tum	40227.9647	2533	21185746.42	3	485438.5	28884	1.8238	1552
Yaj	20878.319	710.4	6081457.1	4	311486.25	31457	0.9219	922.7
SAD	3727.2496					3145	1.4981	0

Tabla A.4. *Información de 2008 a 2011.*

Mun	*SupTotal*	*PromProd*	*Monto*	*PromProd10*	*Monto10*	*Pob*	*Marg*	*Exp*
Bal	*357695.523*	*1979.75*	*37581250.64*	*121.75*	*11620373.65*	*56739*	*-0.0207*	*136.1*
Car	*204693.744*	*2068*	*25906855.21*	*45*	*3306242.77*	*248481*	*-0.5708*	*0*
Cen	*171876.253*	*1268.75*	*10400609.8*	*0.25*	*11556*	*640359*	*-1.4824*	*0*
Cun	*59558.70164*	*731*	*6022228.05*	*0.75*	*63558*	*126416*	*-0.3651*	*0*
EmZ	*59323.14699*	*420.25*	*8949194.03*	*17*	*4858821.23*	*29518*	*-0.878*	*230*
Huim	*371412.648*	*3438.25*	*39139364.12*	*33*	*2110968.71*	*179285*	*-0.0159*	*312.2*
Jal	*59158.57934*	*394.5*	*2729456*	*0*	*0*	*36391*	*-0.5662*	*0*
Jon	*163542.366*	*1210.75*	*11416597.15*	*7.75*	*1754953.2*	*29511*	*-0.0428*	*164.7*
Mac	*242777.721*	*956.5*	*7756812.1*	*0*	*0*	*153132*	*-0.3792*	*0*
Tac	*73505.43055*	*1601.5*	*12985111.71*	*0*	*0*	*46302*	*0.0084*	*0*
Tea	*42178.20264*	*158.5*	*1250489.6*	*0*	*0*	*53555*	*-0.5076*	*0*
Ten	*188225.615*	*1739.5*	*23431781.15*	*30*	*1885147.03*	*58960*	*-0.3582*	*376*

Tabla A.4 Continuación.

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
Ama	31704.83595	1633.75	18214599.81	6	351036	21275	1.5584	0
Boc	36648.42095	1932.75	33648560.66	90.25	6793310.77	30642	0.5395	417
ElB	15875.81056	2152	21767915.2	0	0	18559	1.0482	0
Cat	63256.10576	794.25	9644271.2	9.75	1389127.5	17140	0.3623	859.5
Cha	17889.85717	333.5	2033197.8	0	0	7332	0.9249	0
Chi	168101.109	4622.25	55638764.25	35.75	2161425	111554	2.0887	273.8
Huit	33849.85621	2639.75	30574506.97	0	0	22536	1.3131	0
Ixh	9472.790976	480.5	4395996.4	1	60600	10239	0.7028	0
Ixt	12758.018	140	883010	0	0	10176	0.4097	0
Ixtapa	27849.01527	2691	46058246.25	55.25	3778768.2	24517	0.6228	0
Ixtapan	10683.70966	81.5	779467.2	1	56817	5478	1.0495	0
Jit	23600.0088	1610.25	15802140.8	11.25	990866	18683	0.7521	0
Jua	74030.25455	435.75	7010505.7	6.5	498352.5	21084	0.1949	0
LaL	45531.35488	193.25	2416304	2.75	269640	4974	0.3721	0
Ost	60062.21375	694.75	9969271.2	13	757416.5	17067	0.9511	0
Pal	288601.793	4068.75	61195858.33	50.5	3017204.96	110918	0.6196	216.8
Pan	10666.29203	640.5	6053844.9	0	0	10870	1.217	0
Pic	59258.98016	177.25	1437357.6	1.25	58585	29813	-0.0293	0

Tabla A.4 Continuación.

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
PNS	24633.64857	2712.25	29627205.95	14.5	942256.5	31075	0.9289	1178
Ray	6863.619674	342	3572985.25	9.25	561413.25	9002	0.5298	218.5
Ref	43268.42604	311.5	5476584.56	11.5	726072.36	40711	-0.728	0
Sab	24920.02167	2783.75	21973013.79	0.75	39001.5	25187	1.3443	0
SdA	122958.966	2933.5	39808679.78	11.25	619975.85	57253	1.3947	0
Sim	31495.02891	4780.5	57518948.4	1	55660	40297	1.0827	0
Sit	10568.08259	401.75	4186434.75	7.25	859236.75	12269	3.1224	0
Sol	15720.5103	268.5	2715235.6	0	0	8065	0.4173	0
Soy	9577.397064	1023.5	16877522.74	0.25	19260	9740	0.3808	373
Sun	7908.367581	67	872956	1	52002	2235	0.9466	0
Tapa	6527.218485	76.25	687240	0	0	4121	0.7511	0
Tapi	4195.751134	62.25	377805	0	0	12170	-0.0196	118.5
Til	80190.72387	5060.25	37133667.72	0	0	71432	1.2749	0
Tum	40227.9647	2421.75	18144385.08	3	396038	31723	1.7534	0
Yaj	20878.319	647.25	4931722.1	3.25	213786	34028	1.2249	0
SAD	3727.2496					4545	1.3723	0

Tabla A.5. *El número indica la posición que le corresponde a cada municipio (excepto Centro) al estar ordenados de forma decreciente en cada variable del primer periodo.*

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
Bal	2	12	4	1	1	9	36	4
Car	5	13	10	4	6	1	41	11
Cun	15	23	27	28	28	4	37	21
EmZ	16	29	23	9	3	18	44	7
Huim	1	5	5	7	9	2	32	1
Jal	18	30	34	38	38	14	40	44
Jon	8	21	21	19	10	17	34	20
Mac	4	20	24	31	31	3	39	43
Tac	12	18	20	27	27	12	35	17
Tea	21	38	39	41	41	11	42	26
Ten	6	16	13	8	8	8	38	23
Ama	25	17	16	20	23	26	4	15
Boc	23	15	8	2	2	22	24	8
ElB	33	14	15	35	35	30	20	14
Cat	13	26	25	11	11	29	29	19
Cha	32	33	37	40	40	39	15	32
Chi	7	3	3	6	7	6	2	2
Huit	24	8	11	34	34	24	3	25
Ixh	40	28	32	37	37	34	19	22
Ixt	35	40	42	42	42	33	25	30
Ixtapa	27	10	7	5	4	27	23	3
Ixtapan	36	41	41	29	29	42	11	35
Jit	30	19	19	15	13	31	18	36
Jua	11	31	26	17	18	25	31	28
LaL	19	37	35	24	22	41	28	41
Ost	14	25	22	16	19	28	14	29

Tabla A.5 *Continuación.*

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
Pal	3	4	1	3	5	5	27	33
Pan	37	27	30	36	36	35	5	37
Pic	17	39	38	32	32	16	30	27
PNS	29	7	12	14	16	21	13	24
Ray	42	34	33	13	17	40	16	39
Ref	20	35	28	12	15	13	43	34
Sab	28	9	14	21	25	23	9	16
SdA	9	6	6	10	14	10	6	13
Sim	26	2	2	33	33	15	12	18
Sit	38	32	31	18	12	36	1	6
Sol	34	36	36	39	39	37	21	31
Soy	39	22	18	23	26	38	26	9
Sun	41	43	40	30	30	44	7	38
Tapa	43	42	43	43	43	43	17	40
Tapi	44	44	44	44	44	32	33	42
Til	10	1	9	25	20	7	10	5
Tum	22	11	17	26	21	19	8	10
Yaj	31	24	29	22	24	20	22	12

Tabla A.6. *El número indica la posición que le corresponde a cada municipio (excepto Centro) al estar ordenados de forma decreciente en cada variable del segundo periodo.*

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
Bal	2	13	5	1	1	10	36	14
Car	5	12	12	5	6	1	39	30
Cun	15	24	27	29	26	4	37	9
EmZ	16	30	24	9	3	21	44	22
Huim	1	5	7	7	8	2	33	2
Jal	18	31	34	38	38	14	41	43

Tabla A.6 *Continuación.*

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
Jon	8	20	21	17	11	19	34	20
Mac	4	21	25	28	32	3	38	42
Tac	12	18	20	32	25	12	35	28
Tea	21	39	39	41	41	11	42	24
Ten	6	16	13	8	10	8	40	26
Ama	25	17	16	20	21	27	4	18
Boc	23	15	9	2	2	22	29	3
ElB	33	14	15	36	36	31	14	21
Cat	13	23	23	12	9	29	26	19
Cha	32	34	37	40	40	40	20	29
Chi	7	3	2	6	7	6	2	5
Huit	24	9	10	35	35	26	6	17
Ixh	40	28	32	33	33	37	19	15
Ixt	35	40	41	42	42	35	23	40
Ixtapa	27	10	4	4	4	24	27	4
Ixtapan	36	41	42	31	29	42	11	34
Jit	30	19	19	16	14	30	16	27
Jua	11	29	26	19	19	25	30	13
LaL	19	37	36	24	22	41	28	41
Ost	14	25	22	13	16	28	13	35
Pal	3	4	1	3	5	5	24	25
Pan	37	27	28	37	37	34	9	32
Pic	17	38	38	25	28	17	32	11
PNS	29	8	11	11	12	20	12	16
Ray	42	33	33	14	17	38	22	39
Ref	20	35	29	15	18	13	43	44
Sab	28	7	14	21	24	23	7	6

Tabla A.6 *Continuación.*

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
SdA	9	6	6	10	15	9	5	8
Sim	26	2	3	26	31	15	10	33
Sit	38	32	31	18	13	32	1	23
Sol	34	36	35	39	39	39	18	37
Soy	39	22	18	27	27	36	25	10
Sun	41	43	40	30	30	44	15	36
Tapa	43	42	43	43	43	43	21	38
Tapi	44	44	44	44	44	33	31	31
Til	10	1	8	34	34	7	8	1
Tum	22	11	17	23	20	18	3	7
Yaj	31	26	30	22	23	16	17	12

Tabla A.7. *El número indica la posición que le corresponde a cada municipio (excepto Centro) al estar ordenados de forma decreciente en cada variable del tercer periodo.*

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
Bal	2	14	7	1	1	10	34	12
Car	5	13	12	5	5	1	42	43
Cun	15	24	28	30	24	4	38	39
EmZ	16	30	24	9	3	21	44	8
Huim	1	5	6	7	8	2	32	6
Jal	18	32	34	39	39	15	41	42
Jon	8	20	21	17	10	22	36	11
Mac	4	22	25	36	36	3	39	40
Tac	12	19	20	35	35	12	31	37
Tea	21	39	39	41	41	11	40	41
Ten	6	16	13	8	9	8	37	4
Ama	25	17	16	20	21	26	4	16
Boc	23	15	9	2	2	19	23	3

Tabla A.7 *Continuación.*

Mun	SupTotal	PromProd	Monto	PromProd10	Monto10	Pob	Marg	Exp
ElB	33	12	15	34	34	29	13	25
Cat	13	23	23	15	11	30	29	2
Cha	32	34	37	40	40	40	17	28
Chi	7	3	3	6	7	5	2	7
Huit	24	10	10	33	33	25	7	19
Ixh	40	28	31	26	25	35	20	31
Ixt	35	40	40	42	42	36	26	34
Ixtapa	27	9	4	3	4	24	21	32
Ixtapan	36	41	42	28	27	41	12	24
Jit	30	18	19	14	12	28	18	29
Jua	11	29	26	19	19	27	30	36
LaL	19	37	36	23	22	42	28	35
Ost	14	25	22	11	15	31	14	26
Pal	3	4	1	4	6	6	22	10
Pan	37	27	27	37	37	34	10	22
Pic	17	38	38	24	26	20	35	38
PNS	29	8	11	10	13	18	16	1
Ray	42	33	33	16	18	38	24	9
Ref	20	35	29	12	16	13	43	44
Sab	28	7	14	29	30	23	6	18
SdA	9	6	5	13	17	9	5	17
Sim	26	2	2	25	28	14	11	23
Sit	38	31	32	18	14	32	1	14
Sol	34	36	35	38	38	39	25	33
Soy	39	21	18	31	31	37	27	5
Sun	41	43	41	27	29	44	15	27
Tapa	43	42	43	43	43	43	19	30

Tabla A.7 *Continuación.*

Mun	***SupTotal***	***PromProd***	***Monto***	***PromProd10***	***Monto10***	***Pob***	***Marg***	***Exp***
Tapi	*44*	*44*	*44*	*44*	*44*	*33*	*33*	*13*
Til	*10*	*1*	*8*	*32*	*32*	*7*	*8*	*20*
Tum	*22*	*11*	*17*	*22*	*20*	*17*	*3*	*15*
Yaj	*31*	*26*	*30*	*21*	*23*	*16*	*9*	*21*

Apéndice B

Conceptos matemáticos

B.1. Derivadas de formas lineales y formas cuadráticas

Si $f(\underline{x})$ es una función del vector $\underline{x}$, puede ser deseable expresar las derivadas parciales como un vector, por lo que, se define $\partial f(\underline{x})/\partial \underline{x}$ a continuación:

Definición B.1 (Derivada de una función con respecto a un vector)**.** *Sea $f(\underline{x})$ una función de n variables reales independientes, x_1, x_2, ..., x_n. La derivada de $f(\underline{x})$ con respecto al vector $\underline{x} = (x_1, x_2, \cdots, x_n)'$ se denota por $\partial f(\underline{x})/\partial \underline{x}$ y está definida por:*

$$\partial f(\underline{x})/\partial \underline{x} = (\partial f(\underline{x})/\partial x_1, \partial f(\underline{x})/\partial x_2, \cdots, \partial f(\underline{x})/\partial x_n)'.$$

A continuación se enlistan algunos resultados:

1. Sea $f(\underline{x})$ una función lineal de n variables independientes definida por $f(\underline{x}) = \sum_{i=1}^{n} a_i x_i = \underline{a}'\underline{x} = \underline{x}'\underline{a}$, donde $\underline{a}' = (a_1, a_2, \cdots, a_n)$ y a_i es cualquier constante, $i = 1, 2, \ldots, n$. Entonces $\partial f(\underline{x})/\partial \underline{x} = \underline{a}$.

2. Sea $f(\underline{x})$ una forma cuadrática en las n variables reales independientes x_1, x_2, $\cdots$, x_n definida por $f(\underline{x}) = \underline{x}' A \underline{x}$, donde $A_{n\times n} = [a_{ij}]$ es una matriz simétrica de constantes. Entonces $\partial f(\underline{x})/\partial \underline{x} = 2A\underline{x}$.

B.2. Derivación matricial

Definición B.2 (Derivada de una función con respecto a una matriz)**.** *Sea $f(X)$ una función de la matriz $X_{n\times p} = [x_{ij}]$. La derivada de $f(X)$ con respecto a X se denota por $\partial f(X)/\partial X$ y está definida como la matriz:*

$$\partial f(X)/\partial X = (\partial f(X)/\partial x_{ij})\,.$$

Se tienen los siguientes resultados:

1.
$$\frac{\partial |X|}{\partial x_{ij}} = \begin{cases} X_{ij} \text{ si todos los elementos de } X_{n\times n} \text{ son distintos} \\ \left\{ \begin{array}{l} X_{ij},\ \ i = j \\ 2X_{ij},\ i \neq j \end{array} \right\} \text{ si } X \text{ es simétrica.} \end{cases} \tag{B.1}$$

Donde X_{ij} es el ij-ésimo cofactor de X.

2.
$$\frac{\partial trXY}{\partial X} = \begin{cases} Y' \text{ si todos los elementos de } X_{n\times p} \text{ son distintos} \\ Y + Y' - diag(Y) \text{ si } X_{n\times n} \text{ es simétrica.} \end{cases} \tag{B.2}$$

B.3. Álgebra matricial

Teorema B.3. *Si A es una matriz de $n \times n$ y $A^2 = mA$, entonces $tr(A) = m\ rang(A)$ [7].*

Teorema B.4 (Rouché-Frobenius)**.** *Dada una matriz $\mathcal{M} \in \mathbb{K}^{(m,n)}$ y un vector $\underline{b} \in \mathbb{K}^{(m,1)}$, el sistema lineal $\mathcal{M}x = \underline{b}$ es:*

1. *Compatible y determinado si y sólo si $rang(\mathcal{M}) = rang(\mathcal{M}|\underline{b}) = n$*
2. *Compatible e indeterminado si y sólo si $rang(\mathcal{M}) = rang(\mathcal{M}|\underline{b}) < n$*
3. *Incompatible si y sólo si $rang(\mathcal{M}) < rang(\mathcal{M}|\underline{b})$ [2].*

Teorema B.5. *Sea δ un determinante en $M_{n\times n}(F)$ y sea $A \in M_{n\times n}(F)$. Entonces las siguientes condiciones son equivalentes:*

1. $\delta(A) = 0$
2. A *no es invertible*
3. $rango(A) < n$

La demostración se puede consultar en [6], pág. 208.

Teorema B.6. *Sea $A_{n\times n}$ una matriz simétrica. Cada una de las condiciones (1) y (2) son necesarias y suficientes para que A sea una matriz semidefinida positiva. Cada condición (i), (ii) y (iii) es necesaria y suficiente para que A sea una matriz definida positiva.*

1. *Existe una matriz $B_{n\times n}$ de rango menor que n tal que $B'B = A$.*
2. *Las raíces características de A son no negativas y al menos una raíz es igual a cero.*

(i) Existe una matriz $B_{n\times n}$ de rango n tal que $B'B = A$

(ii) Las raíces características de A son todas positivas.

(iii) $a_{11} > 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \ldots, |A| > 0$ *[7].*

Teorema B.7. *[Teorema de la descomposición espectral o Teorema de la descomposición de Jordan] Cualquier matriz simétrica $A_{p\times p}$ puede ser escrita como:*

$$A = \Gamma\Lambda\Gamma' = \sum \lambda_i \gamma_{(i)}\gamma'_{(i)},$$

donde Λ es una matriz diagonal de valores propios de A y Γ es una matriz ortogonal cuyas columnas son vectores propios estandarizados.

La demostración se puede consultar en [11], pág. 469.

B.4. Inversa generalizada

Definición B.8 (Inversa generalizada)**.** *Sea $A_{m\times n}$ una matriz. Si existe una matriz denotada por A^- que satisface las cuatro condiciones de abajo, esta matriz es una inversa generalizada de A.*

1. AA^- *es simétrica;*

2. A^-A *es simétrica;*

3. $AA^-A = A$*;*

4. $A^-AA^- = A^-$*.*

Teorema B.9. *El rango de la inversa generalizada de* A *es igual al rango de* A*.*

La demostración se puede consultar en [7], pág. 27.

Teorema B.10. *Si* $A_{m\times n}$ *es una matriz de rango* m*, entonces* $A^- = A'(AA')^{-1}$ *y* $AA^- = I$*. Si el rango de* A *es* n*, entonces* $A^- = (A'A)^{-1}A'$ *y* $A^-A = I$*.*

Demostración. Si A es de rango m, entonces $rang(AA') = rang(A'A) = rang(A) = m$, luego $AA'_{m\times m}$ tiene inversa, así, $A^- = A'(AA')^{-1}$ está bien definida. Luego:

1. $AA^- = AA'(AA')^{-1} = I_{m\times n}$ es simétrica;

2. $A^-A = A'(AA')^{-1}A$ es simétrica ya que $(AA')^{-1}$ es simétrica;

3. $AA^-A = AA'(AA')^{-1}A = A$;

4. $A^-AA^- = A'(AA')^{-1}AA'(AA')^{-1} = A^-$,

por lo que $A^- = A'(AA')^{-1}$ es la inversa generalizada de A.

Si A es de rango n, entonces $rang(AA') = rang(A'A) = rang(A) = n$, luego $A'A_{n\times n}$ tiene inversa, así, $A^- = (A'A)^{-1}A'$ está bien definida. Luego:

1. $AA^- = A(A'A)^{-1}A'$ es simétrica ya que $(AA')^{-1}$ es simétrica;

2. $A^-A = (A'A)^{-1}A'A = I_{n\times n}$ es simétrica;

3. $AA^-A = A(A'A)^{-1}A'A = A$;

4. $A^-AA^- = (A'A)^{-1}A'A(A'A)^{-1}A' = A^-$,

por lo que $A^- = (A'A)^{-1}A'$ es la inversa generalizada de A. □

B.5. Inversa condicional

Definición B.11 (Inversa condicional)**.** *Sea A una matriz de tamaño $m \times n$. Una matriz A^c es definida como inversa condicional de A si y sólo si satisface $AA^cA = A$ [7].*

Teorema B.12. *Para cualquier matriz $X_{m\times n}$ de rango $r > 0$, se define K por:*

$$K = X(X'X)^cX'.$$

Algunos resultados relacionados con la matriz K son:

1. $K = K'$ *y* $K = K^2$*, es decir, K es una matriz simétrica idempotente*
2. $rang(K) = tr(K) = rang(X) = r$
3. $KX = X$, $X'K = X'$
4. $(X'X)^cX'$ *es una inversa condicional de X para cualquier inversa condicional de $X'X$*
5. $X(X'X)^c$ *es una inversa condicional de X' para cualquier inversa condicional de $X'X$ [7].*

Apéndice C

Estadística

C.1. La distribución multinormal

La más importante distribución de probabilidad multivariada es la distribución normal multivariada, que se define como sigue:

Definición C.1. *El vector aleatorio $\underline{X}$ se dice que tiene una distribución normal p-variada (o multinormal p-dimensional o normal multivariada) con media $\underline{\mu}$ y matriz de covarianza Σ si su función de densidad de probabilidad está dada por:*

$$f(\underline{x}) = |2\pi\Sigma|^{-1/2} exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})\right\},$$

donde $\Sigma > 0$ [11].

Teorema C.2. *Si $\underline{t}$ es asintóticamente normal con media $\underline{\mu}$ y matriz de covarianza V/n, y si $f = (f_1, \ldots, f_q)'$ son funciones de valor real que son diferenciables en $\underline{\mu}$, entonces $f(\underline{t})$ es asintóticamente normal con media $f(\underline{\mu})$ y matriz de covarianzas $D'VD/n$, donde $d_{ij} = \partial f_j/\partial t_i$ evaluada en $\underline{t} = \underline{\mu}$.*

La demostración se puede consultar en [11], pág. 52.

C.2. La distribución Wishart

En esta sección se considerarán a las funciones cuadráticas que son de la forma $\mathbb{X}'C\mathbb{X}$, donde C es una matriz simétrica. La más importante de estas funciones es la

matriz de covarianza muestral S obtenida cuando $C = n^{-1}H$, donde H es la matriz centrante. Las formas cuadráticas frecuentemente conducen a la distribución Wishart, que es una generalización matricial de la distribución univariada χ^2, y tiene propiedades similares [11].

Definición C.3. *Si $M_{p\times p}$ puede ser escrita como $M = \mathbb{X}'\mathbb{X}$, donde $\mathbb{X}_{m\times p}$ es una matriz de datos de una $N_p(\underline{0}, \Sigma)$, entonces se dice que M tiene una distribución Wishart con matriz de escala Σ y m grados de libertad, esto se denotará como $M \sim W_p(\Sigma, m)$. Cuando $\Sigma = I_{p\times p}$ se dice que la distribución está en forma estandarizada.*

Nótese que cuando $p = 1$, la distribución $W_1(\sigma^2, m)$ está dada por $\mathbb{X}'\mathbb{X}$, donde los elementos de $\mathbb{X}_{m\times 1}$ son variables $iiiN_1(0, \sigma^2)$, esto es, que la distribución $W_1(\sigma^2, m)$ es la misma que la distribución $\sigma^2\chi^2_m$ [11]. También se cumple el siguiente teorema:

Teorema C.4. *Si $\mathbb{X}_{n\times p}$ es una matriz de datos de una $N_p(\underline{0}, \Sigma)$, y si $C_{n\times n}$ es una matriz simétrica, entonces:*

i) *$\mathbb{X}'C\mathbb{X}$ tiene la misma distribución que la suma ponderada de matrices independientes $W_p(\Sigma, 1)$, donde los pesos son los valores propios de C;*

ii) *$\mathbb{X}'C\mathbb{X}$ tiene una distribución Wishart si y sólo si C es idempotente, en ese caso $\mathbb{X}'C\mathbb{X} \sim W_p(\Sigma, r)$, donde $r = tr(C) = rang(C)$;*

iii) *Si $S = n^{-1}\mathbb{X}'H\mathbb{X}$ es la matriz de covarianza muestral, entonces $nS \sim W_p(\Sigma, n-1)$.*

La demostración se puede consultar en [11], págs. 68-69.

Teorema C.5. *Si $M \sim W_p(\Sigma, m)$ y $m \geqslant p$, entonces $|M|$ es $|\Sigma|$ veces p variables aleatorias independiente χ^2 con grados de libertad m, $m-1$, ..., $m-p+1$.*

La demostración puede consultarse en [11], pág. 73.

Corolario C.6. *Si $M \sim W_p(\Sigma, m)$, $\Sigma > 0$ y $m > p$, entonces $M > 0$ con probabilidad uno.*

La demostración se puede consultar en [11], pág. 73.

C.3. Verosimilitud

C.3.1. La función de verosimilitud

Definición C.7. *Supongamos que* $X' = (\underline{x_1}, \ldots, \underline{x_n})$ *es una muestra aleatoria de una población con función de densidad de probabilidad* $f(\underline{x}, \underline{\theta})$*, donde* $\underline{\theta}$ *es un vector paramétrico. La función de verosimilitud de toda la muestra es:*

$$L(\underline{\theta}; X) = \prod_{i=1}^{n} f(\underline{x_i}; \underline{\theta}). \tag{C.1}$$

El logaritmo de la función de verosimilitud es:

$$l(\underline{\theta}; X) = logL(\underline{\theta}; X) = \sum_{i=1}^{n} logf(\underline{x_i}; \underline{\theta}). \tag{C.2}$$

Dada una muestra X, ambas, $L(\underline{\theta}; X)$ y $l(\underline{\theta}; X)$, son consideradas como funciones del parámetro $\underline{\theta}$. Para el caso especial $n = 1$, $L(\underline{\theta}; X) = f(X; \underline{\theta})$ y la diferencia entre la función de densidad de probabilidad y la función de verosimilitud es: $f(X; \underline{\theta})$ es interpretada como la función de densidad de probabilidad cuando se fija $\underline{\theta}$ y se varía X y es interpretada como la función de verosimilitud cuando se fija X y se varía $\underline{\theta}$. Como se sabe, la función de densidad de probabilidad juega un papel clave en la teoría de probabilidad, mientras la verosimilitud es central para la inferencia estadística [11].

Teorema C.8. *Supongamos que* $X' = (\underline{x_1}, \ldots, \underline{x_n})$ *es una muestra aleatoria de una* $N_p(\underline{\mu}, \Sigma)$*. Entonces (C.1) y (C.2) son respectivamente,*

$$L(\underline{\mu}, \Sigma; X) = |2\pi\Sigma|^{-n/2} exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\underline{x_i} - \underline{\mu})' \Sigma^{-1} (\underline{x_i} - \underline{\mu}) \right\} \tag{C.3}$$

y

$$\begin{aligned} l(\underline{\mu}, \Sigma; X) &= -\tfrac{n}{2} log|2\pi\Sigma| - \tfrac{1}{2} \textstyle\sum_{i=1}^{n} (\underline{x_i} - \underline{\mu})' \Sigma^{-1} (\underline{x_i} - \underline{\mu}) \\ &= -\tfrac{n}{2} log|2\pi\Sigma| - \tfrac{n}{2} tr(\Sigma^{-1} S) - \tfrac{n}{2} (\overline{\underline{x}} - \underline{\mu})' \Sigma^{-1} (\overline{\underline{x}} - \underline{\mu}). \end{aligned} \tag{C.4}$$

Demostración. Se hacen algunos cálculos

$$
\begin{aligned}
&(\underline{x_i}-\underline{\mu})'\Sigma^{-1}(\underline{x_i}-\underline{\mu})\\
&=(\underline{x_i}-\overline{\underline{x}}+\overline{\underline{x}}-\underline{\mu})'\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}}+\overline{\underline{x}}-\underline{\mu})\\
&=((\underline{x_i}-\overline{\underline{x}})'+(\overline{\underline{x}}-\underline{\mu})')\Sigma^{-1}((\underline{x_i}-\overline{\underline{x}})+(\overline{\underline{x}}-\underline{\mu}))\\
&=(\underline{x_i}-\overline{\underline{x}})'\Sigma^{-1}((\underline{x_i}-\overline{\underline{x}})+(\overline{\underline{x}}-\underline{\mu}))+(\overline{\underline{x}}-\underline{\mu})'\Sigma^{-1}((\underline{x_i}-\overline{\underline{x}})+(\overline{\underline{x}}-\underline{\mu}))\\
&=(\underline{x_i}-\overline{\underline{x}})'\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}})+(\underline{x_i}-\overline{\underline{x}})'\Sigma^{-1}(\overline{\underline{x}}-\underline{\mu})+(\overline{\underline{x}}-\underline{\mu})'\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}})\\
&\quad+(\overline{\underline{x}}-\underline{\mu})'\Sigma^{-1}(\overline{\underline{x}}-\underline{\mu}),
\end{aligned}
$$

además $(\underline{x_i}-\overline{\underline{x}})'\Sigma^{-1}(\overline{\underline{x}}-\underline{\mu})$ es un escalar, por lo que es igual a su transpuesto $(\overline{\underline{x}}-\underline{\mu})'\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}})$, así, se cumple la siguiente identidad:

$$
\begin{aligned}
(\underline{x_i}-\underline{\mu})'\Sigma^{-1}(\underline{x_i}-\underline{\mu}) &= (\underline{x_i}-\overline{\underline{x}})'\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}})+(\overline{\underline{x}}-\underline{\mu})'\Sigma^{-1}(\overline{\underline{x}}-\underline{\mu})\\
&\quad+2(\overline{\underline{x}}-\underline{\mu})'\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}}),
\end{aligned}
$$

la cual se suma sobre el indice $i=1,\ldots,n$, el último término del lado derecho desaparece, dando:

$$
\begin{aligned}
&\textstyle\sum_{i=1}^{n}(\underline{x_i}-\underline{\mu})'\Sigma^{-1}(\underline{x_i}-\underline{\mu})\\
&=\textstyle\sum_{i=1}^{n}\left((\underline{x_i}-\overline{\underline{x}})'\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}})+(\overline{\underline{x}}-\underline{\mu})'\Sigma^{-1}(\overline{\underline{x}}-\underline{\mu})+2(\overline{\underline{x}}-\underline{\mu})'\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}})\right)\\
&=\textstyle\sum_{i=1}^{n}(\underline{x_i}-\overline{\underline{x}})'\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}})+n(\overline{\underline{x}}-\underline{\mu})'\Sigma^{-1}(\overline{\underline{x}}-\underline{\mu})+2(\overline{\underline{x}}-\underline{\mu})'\Sigma^{-1}\\
&\quad\textstyle\sum_{i=1}^{n}(\underline{x_i}-\overline{\underline{x}})\\
&=\textstyle\sum_{i=1}^{n}(\underline{x_i}-\overline{\underline{x}})'\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}})+n(\overline{\underline{x}}-\underline{\mu})'\Sigma^{-1}(\overline{\underline{x}}-\underline{\mu}).
\end{aligned}
\tag{C.5}
$$

Ya que cada término $(\underline{x_i}-\overline{\underline{x}})'\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}})$ es un escalar, es igual a su traza. Por lo tanto:

$$
(\underline{x_i}-\overline{\underline{x}})'\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}})=tr((\underline{x_i}-\overline{\underline{x}})'\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}}))=tr(\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}})(\underline{x_i}-\overline{\underline{x}})'). \tag{C.6}
$$

Sumando (C.6) sobre el índice i y substituyendo en (C.5), se llega a:

$$
\begin{aligned}
&\textstyle\sum_{i=1}^{n}(\underline{x_i}-\underline{\mu})'\Sigma^{-1}(\underline{x_i}-\underline{\mu})\\
&=\textstyle\sum_{i=1}^{n}tr(\Sigma^{-1}(\underline{x_i}-\overline{\underline{x}})(\underline{x_i}-\overline{\underline{x}})')+n(\overline{\underline{x}}-\underline{\mu})'\Sigma^{-1}(\overline{\underline{x}}-\underline{\mu})\\
&=tr(\Sigma^{-1}\left\{\textstyle\sum_{i=1}^{n}(\underline{x_i}-\overline{\underline{x}})(\underline{x_i}-\overline{\underline{x}})'\right\})+n(\overline{\underline{x}}-\underline{\mu})'\Sigma^{-1}(\overline{\underline{x}}-\underline{\mu}).
\end{aligned}
\tag{C.7}
$$

Escribiendo a:

$$
\sum_{i=1}^{n}(\underline{x_i}-\overline{\underline{x}})(\underline{x_i}-\overline{\underline{x}})'=nS \tag{C.8}
$$

y usando (C.7) y (C.8) en (C.4), se obtiene:

$$\begin{aligned} l(\underline{\mu}, \Sigma; X) &= -\tfrac{n}{2} log|2\pi\Sigma| - \tfrac{1}{2}\left(tr(\Sigma^{-1}nS) + n(\underline{\overline{x}} - \underline{\mu})'\Sigma^{-1}(\underline{\overline{x}} - \underline{\mu})\right) \\ &= -\tfrac{n}{2} log|2\pi\Sigma| - \tfrac{n}{2}\left(tr(\Sigma^{-1}S) + (\underline{\overline{x}} - \underline{\mu})'\Sigma^{-1}(\underline{\overline{x}} - \underline{\mu})\right). \end{aligned}$$

□

C.3.2. Estimación de máxima verosimilitud

El estimador de máxima verosimilitud de un parámetro desconocido es el valor del parámetro que maximiza la verosimilitud de las observaciones dadas. En la mayoría de los casos, este máximo se puede hallar por diferenciación y, dado que $l(\underline{\theta}; X)$ es maximizada cuando $L(\underline{\theta}; X)$ es maximizada, la ecuación $(\partial l/\partial \underline{\theta}) = \underline{0}$ se resuelve para $\underline{\theta}$. El estimador de máxima verosimilitud $\widehat{\underline{\theta}}$ de $\underline{\theta}$ es el valor que proporciona un máximo global [11].

El caso normal multivariado sin restricciones

Teorema C.9. *Sea* $X' = (\underline{x_1}, \underline{x_2}, \ldots, \underline{x_n})$ *una muestra aleatoria de la distribución normal multivariada* $N_p(\underline{\mu}, \Sigma)$*, donde el espacio paramétrico es* $\overline{\underline{\Theta}} = \{(\underline{\mu}, \Sigma) | \underline{\mu} \in \mathbb{R}^p, \Sigma \in M_{p\times p}(\mathbb{R})$ *es definida positiva* $\}$*. Los estimadores de máxima verosimilitud de* $\underline{\mu}$ *y* Σ *son* $\widehat{\mu} = \underline{\overline{x}}$ *y* $\widehat{\Sigma} = S$*, si* $n > p + 1$*.*

Demostración. La distribución normal multivariada tiene la siguiente función logaritmo de verosimilitud con parámetros $\underline{\mu}$ y Σ, en la cual $(\underline{\overline{x}} - \underline{\mu})'\Sigma^{-1}(\underline{\overline{x}} - \underline{\mu})$ es un escalar por lo que es igual a su traza:

$$\begin{aligned} l(\underline{\mu}, \Sigma; X) &= -\tfrac{n}{2} log|2\pi\Sigma| - \tfrac{n}{2} tr(\Sigma^{-1}S) - \tfrac{n}{2}(\underline{\overline{x}} - \underline{\mu})'\Sigma^{-1}(\underline{\overline{x}} - \underline{\mu}) \\ &= -\tfrac{n}{2} log|2\pi\Sigma| - \tfrac{n}{2} tr(\Sigma^{-1}S) - \tfrac{n}{2} tr\left((\underline{\overline{x}} - \underline{\mu})'\Sigma^{-1}(\underline{\overline{x}} - \underline{\mu})\right) \\ &= -\tfrac{n}{2} log|2\pi\Sigma| - \tfrac{n}{2} tr(\Sigma^{-1}S) - \tfrac{n}{2} tr\left(\Sigma^{-1}(\underline{\overline{x}} - \underline{\mu})(\underline{\overline{x}} - \underline{\mu})'\right). \end{aligned}$$

Usando nuevos parámetros $\underline{\mu}$ y V con $V = \Sigma^{-1}$ se sigue que:

$$l(\underline{\mu}, V; X) = -\tfrac{n}{2} log|2\pi V^{-1}| - \tfrac{n}{2} tr(VS) - \tfrac{n}{2} tr\left(V(\underline{\overline{x}} - \underline{\mu})(\underline{\overline{x}} - \underline{\mu})'\right),$$

además,

$$\begin{aligned} -\tfrac{n}{2}log|2\pi V^{-1}| &= -\tfrac{n}{2}log\left((2\pi)^p|V^{-1}|\right) = -\tfrac{np}{2}log(2\pi) - \tfrac{n}{2}log|V^{-1}| \\ &= -\tfrac{np}{2}log(2\pi) - \tfrac{n}{2}log(|V|^{-1}) = -\tfrac{np}{2}log(2\pi) + \tfrac{n}{2}log|V|, \end{aligned}$$

por lo que:

$$l(\underline{\mu}, V; X) = -\frac{np}{2}log(2\pi) + \frac{n}{2}log|V| - \frac{n}{2}tr(VS) - \frac{n}{2}tr\left(V(\overline{\underline{x}} - \underline{\mu})(\overline{\underline{x}} - \underline{\mu})'\right). \qquad \text{(C.9)}$$

Se calcula $\partial l/\partial \underline{\mu}$ y $\partial l/\partial V$

$$\begin{aligned} \partial l/\partial \underline{\mu} &= \partial\left(-\frac{n}{2}tr\left(V(\overline{\underline{x}} - \underline{\mu})(\overline{\underline{x}} - \underline{\mu})'\right)\right)/\partial\underline{\mu} \\ &= -\frac{n}{2}\partial\ tr\left(V(\overline{\underline{x}}\,\overline{\underline{x}}' - \overline{\underline{x}}\underline{\mu}' - \underline{\mu}\overline{\underline{x}}' + \underline{\mu}\underline{\mu}')\right)/\partial\underline{\mu} \\ &= \frac{n}{2}\partial\ tr\left(V(\overline{\underline{x}}\underline{\mu}' + \underline{\mu}\overline{\underline{x}}' - \underline{\mu}\underline{\mu}')\right)/\partial\underline{\mu} \\ &= \frac{n}{2}\partial\left(\ tr(V\overline{\underline{x}}\underline{\mu}') + tr(V\underline{\mu}\overline{\underline{x}}') - tr(V\underline{\mu}\underline{\mu}')\right)/\partial\underline{\mu} \\ &= \frac{n}{2}\partial\left(\ tr(\underline{\mu}(V\overline{\underline{x}})') + tr(\underline{\mu}\overline{\underline{x}}'V) - tr(V\underline{\mu}\underline{\mu}')\right)/\partial\underline{\mu}, \end{aligned}$$

usando (B.2) se sigue que:

$$\partial l/\partial\underline{\mu} = \frac{n}{2}(V\overline{\underline{x}} + V'\overline{\underline{x}}) - \frac{n}{2}\partial tr(V\underline{\mu}\underline{\mu}')/\partial\underline{\mu}.$$

Por otro lado,

$$tr(V\underline{\mu}\underline{\mu}') = \sum_{j=1}^{p}\sum_{i=1}^{p} v_{ji}\mu_j\mu_i,$$

además $V = \Sigma^{-1}$ es simétrica, luego:

$$\begin{aligned} -\frac{n}{2}\partial\left(tr(V\underline{\mu}\underline{\mu}')\right)/\partial\mu_k &= v_{1k}\mu_1 + v_{2k}\mu_2 + \ldots + v_{k-1\ k}\mu_{k-1} + (v_{k1}\mu_1 + v_{k2}\mu_2 \\ &\quad + \ldots + v_{k\ k-1}\mu_{k-1} + 2v_{kk}\mu_k + v_{k\ k+1}\mu_{k+1} + \ldots + \\ &\quad v_{kp}\mu_p) + v_{k+1\ k}\mu_{k+1} + \ldots + v_{pk}\mu_p \\ &= 2(v_{1k}\mu_1 + v_{2k}\mu_2 + \ldots + v_{k-1\ k}\mu_{k-1} + v_{kk}\mu_k + \\ &\quad v_{k+1\ k}\mu_{k+1} + \ldots + v_{pk}\mu_p) \\ &= 2v'_{(k)}\underline{\mu},\ k = 1, 2, \cdots, p. \end{aligned}$$

Se sigue que:

$$\partial tr(V\underline{\mu}\underline{\mu}')/\partial\underline{\mu} = \begin{bmatrix} 2v'_{(1)}\underline{\mu} \\ 2v'_{(2)}\underline{\mu} \\ \vdots \\ 2v'_{(p)}\underline{\mu} \end{bmatrix} = 2\begin{bmatrix} v'_{(1)} \\ v'_{(2)} \\ \vdots \\ v'_{(p)} \end{bmatrix}\underline{\mu} = 2V'\underline{\mu} = 2V\underline{\mu}.$$

Por lo tanto,

$$\frac{\partial l}{\partial\underline{\mu}} = \frac{n}{2}(V\overline{\underline{x}} + V'\overline{\underline{x}}) - \frac{n}{2}\partial tr(V\underline{\mu}\underline{\mu}')/\partial\underline{\mu} = \frac{n}{2}(V\overline{\underline{x}} + V\overline{\underline{x}}) - \frac{n}{2}2V\underline{\mu} = nV(\overline{\underline{x}} - \underline{\mu}). \quad \text{(C.10)}$$

Para calcular $\partial l/\partial V$, se examina por separado cada término del lado derecho de (C.9). Dado que V es simétrica, usando (B.1) se sigue que:

$$\frac{\partial log|V|}{\partial v_{ij}} = \frac{1}{|V|}\frac{\partial|V|}{\partial v_{ij}} = \begin{cases} 2V_{ij}/|V|, & \text{si } i \neq j, \\ V_{ij}/|V|, & \text{si } i = j, \end{cases}$$

donde V_{ij} es el ij-ésimo cofactor de V. Ya que V es simétrica, la matriz con elementos $V_{ij}/|V|$ es igual a $V^{-1} = \Sigma$. Así,

$$\frac{\partial log|V|}{\partial V} = 2\Sigma - diag\Sigma. \quad \text{(C.11)}$$

De (B.2) se tiene que:

$$\frac{\partial trVS}{\partial V} = S + S' - diagS = 2S - diagS \quad \text{(C.12)}$$

y

$$\frac{\partial trV(\overline{\underline{x}} - \underline{\mu})(\overline{\underline{x}} - \underline{\mu})'}{\partial V} = 2(\overline{\underline{x}} - \underline{\mu})(\overline{\underline{x}} - \underline{\mu})' - diag((\overline{\underline{x}} - \underline{\mu})(\overline{\underline{x}} - \underline{\mu})'). \quad \text{(C.13)}$$

Combinando las ecuaciones (C.11), (C.12) y (C.13) se puede ver que:

$$\begin{aligned} \frac{\partial l}{\partial V} &= \tfrac{n}{2}(2\Sigma - diag\Sigma) - \tfrac{n}{2}(2S - diagS + 2(\overline{\underline{x}} - \underline{\mu})(\overline{\underline{x}} - \underline{\mu})' - \\ &\quad diag((\overline{\underline{x}} - \underline{\mu})(\overline{\underline{x}} - \underline{\mu})')) \\ &= \tfrac{n}{2}(2(\Sigma - S - (\overline{\underline{x}} - \underline{\mu})(\overline{\underline{x}} - \underline{\mu})') - diag(\Sigma - S - (\overline{\underline{x}} - \underline{\mu})(\overline{\underline{x}} - \underline{\mu})')) \\ &= \frac{n}{2}(2M - diagM), \end{aligned} \quad \text{(C.14)}$$

donde $M = \Sigma - S - (\underline{\overline{x}} - \underline{\mu})(\underline{\overline{x}} - \underline{\mu})'$.

Para encontrar los puntos críticos de $l(\underline{\mu}, \Sigma; X)$ se debe resolver:

$$\frac{\partial l}{\partial \underline{\mu}} = \underline{0} \quad \text{y} \quad \frac{\partial l}{\partial V} = 0.$$

De (C.10) se puede ver que $V(\underline{\overline{x}} - \underline{\mu}) = \underline{0}$, por lo tanto $\underline{\mu} = \underline{\overline{x}}$ y de (C.14) $\partial l/\partial V = \underline{0}$ implica que $2M - diagM = \underline{0}$, por lo que $M = 0$, es decir, $\Sigma = S + (\underline{\overline{x}} - \underline{\mu})(\underline{\overline{x}} - \underline{\mu})'$. Ya que $\underline{\mu} = \underline{\overline{x}}$, se obtiene $\Sigma = S$.

Dado que por hipótesis Σ es definida positiva, por iii) del Teorema C.4 se tiene que $nS \sim W_p(\Sigma, n-1)$, por el Corolario C.6 se sigue que $nS > 0$ con probabilidad uno, si $n > p+1$. Por lo tanto el punto crítico de $l(\underline{\mu}, \Sigma; X)$ es $(\underline{\overline{x}}, S)$ y

$$\begin{aligned} l(\underline{\overline{x}}, S; X) &= -\tfrac{n}{2} log|2\pi S| - \tfrac{n}{2} tr(S^{-1}S) - \tfrac{n}{2} tr\left(S^{-1}(\underline{\overline{x}} - \underline{\overline{x}})(\underline{\overline{x}} - \underline{\overline{x}})'\right) \\ &= -\tfrac{n}{2} log|2\pi S| - \tfrac{np}{2}. \end{aligned}$$

Por otro lado se tiene que:

$$\begin{aligned} l(\underline{\mu}, \Sigma; X) &= -\tfrac{n}{2} log|2\pi\Sigma| - \tfrac{n}{2} tr(\Sigma^{-1}S) - \tfrac{n}{2}(\underline{\overline{x}} - \underline{\mu})'\Sigma^{-1}(\underline{\overline{x}} - \underline{\mu}) \\ &\leqslant -\tfrac{n}{2} log|2\pi\Sigma| - \tfrac{n}{2} tr(\Sigma^{-1}S) = l_1, \end{aligned}$$

ya que $(\underline{\overline{x}} - \underline{\mu})'\Sigma^{-1}(\underline{\overline{x}} - \underline{\mu}) \geqslant 0$, dado que Σ es definida positiva, Σ^{-1} también es definida positiva (de $\Sigma^{-1}\Sigma = I$ y $(\Sigma^{-1})'\Sigma = (\Sigma'\Sigma^{-1})' = (\Sigma\Sigma^{-1})' = I' = I$ se sigue que $(\Sigma^{-1})' = \Sigma^{-1}$, además, dado $\underline{w} \in \mathbb{R}^p - \{\underline{0}\}$ se cumple que $(\underline{w}'\Sigma^{-1})\Sigma(\Sigma^{-1}\underline{w}) = \underline{w}'\Sigma^{-1}\underline{w} > 0$). Luego:

$$\begin{aligned} l_1 &= -\tfrac{n}{2} log((2\pi)^p|\Sigma|) - \tfrac{n}{2} tr(\Sigma^{-1}S) \\ &= -\tfrac{np}{2} log(2\pi) + \tfrac{n}{2} log(|\Sigma^{-1}|) - \tfrac{n}{2} tr(\Sigma^{-1}S) \\ &= -\tfrac{np}{2} log(2\pi) + \tfrac{n}{2} log(|\Sigma^{-1}S||S^{-1}|) - \tfrac{n}{2} tr(\Sigma^{-1}S) \\ &= -\tfrac{np}{2} log(2\pi) + \tfrac{n}{2} log(|S^{-1}|) + \tfrac{n}{2} log(|\Sigma^{-1}S|) - \tfrac{n}{2} tr(\Sigma^{-1}S) \\ &= -\tfrac{np}{2} log(2\pi) + \tfrac{n}{2} log(|S|^{-1}) + \tfrac{n}{2} log(|\Delta|) - \tfrac{n}{2} tr(\Delta) \\ &= -\tfrac{n}{2} log((2\pi)^p) - \tfrac{n}{2} log|S| + \tfrac{n}{2}(log(|\Delta|) - tr(\Delta)), \end{aligned}$$

donde $\Delta = \Sigma^{-1}S$. Además $|\Delta| = \prod_{i=1}^{p} a_i > 0$ ($|\Delta| = |\Sigma^{-1}S| = |\Sigma^{-1}||S| > 0$, ya que Σ^{-1} y S son definidas positivas y son simétricas, su determinante es el producto de sus valores propios que son positivos) y $tr(\Delta) = \sum_{i=1}^{p} a_i$, donde a_i, $i = 1, \cdots, p$ son los

valores propios de Δ.

Si $\underline{w} \in \mathbb{R}^p$ es un vector propio de Δ, se cumple que $\Sigma^{-1}S\underline{w} = \Delta\underline{w} = \lambda\underline{w}$, donde $\lambda \in \mathbb{R}$, además como Σ^{-1} es definida positiva, se tiene que para $S\underline{w} \in \mathbb{R}^p$ se cumple $(S\underline{w})'\Sigma^{-1}S\underline{w} > 0$, por otro lado, $(S\underline{w})'\Sigma^{-1}S\underline{w} = \underline{w}'S'\lambda\underline{w} = \lambda\underline{w}'S\underline{w}$, agregando el hecho de que S es definida positiva, se concluye que $\lambda > 0$, por lo tanto los valores propios de Δ son positivos. Luego:

$$\begin{aligned} l_1 &= -\tfrac{n}{2}log((2\pi)^p|S|) + \tfrac{n}{2}\left(log\left(\textstyle\prod_{i=1}^p a_i\right) - \sum_{i=1}^p a_i\right) \\ &= -\tfrac{n}{2}log|2\pi S| + \tfrac{n}{2}(\textstyle\sum_{i=1}^p log(a_i) - \sum_{i=1}^p a_i) \\ &= -\tfrac{n}{2}log|2\pi S| + \tfrac{n}{2}\textstyle\sum_{i=1}^p (log(a_i) - a_i) \end{aligned}$$

Considérese $f(a_i) = log(a_i) - a_i$, se tiene que $f'(a_i) = 1/a_i - 1$ y $f''(a_i) = -(a_i)^{-2}$, al igualar a cero $f'(a_i)$ se sigue que $a_i = 1$, además $f''(1) = -(1)^{-2} = -1$, por lo tanto $a_i = 1$ es un máximo global de $f(a_i)$ en $\mathbb{R}_+$, $i = 1, 2, \ldots, p$. Se sigue que:

$$\begin{aligned} l_1 &\leqslant -\tfrac{n}{2}log|2\pi S| + \tfrac{n}{2}\textstyle\sum_{i=1}^p (log(1) - 1) = -\tfrac{n}{2}log|2\pi S| + \tfrac{n}{2}\sum_{i=1}^p (-1) \\ &= -\tfrac{n}{2}log|2\pi S| - \tfrac{np}{2} = l(\overline{\underline{x}}, S; X), \end{aligned}$$

dado que $-\frac{n}{2}log|2\pi S| - \frac{np}{2}$ no depende de parámetros desconocidos, por lo tanto

$$l(\underline{\mu}, \Sigma; X) \leqslant l_1 \leqslant l(\overline{\underline{x}}, S; X),$$

así $(\overline{\underline{x}}, S)$ es el estimador de máxima verosimilitud de $(\underline{\mu}, \Sigma)$. □

A continuación se considera una importante propiedad de los estimadores de máxima verosimilitud que no se tiene por otro método de estimación. Frecuentemente el parámetro de interés no es $\underline{\theta}$ sino una función $h(\underline{\theta})$. Si $\widehat{\underline{\theta}}$ es el estimador de máxima verosimilitud de $\underline{\theta}$ y si $\underline{\lambda} = h(\underline{\theta})$ es una función uno a uno de $\underline{\theta}$, la función inversa $h^{-1}(\underline{\lambda}) = \underline{\theta}$ está bien definida y se puede escribir la función de verosimilitud como una función de $\underline{\lambda}$. Se tiene $L^*(\underline{\lambda}; x) = L(h^{-1}(\underline{\lambda}); x)$ de modo que:

$$sup_{\underline{\lambda}} L^*(\underline{\lambda}; x) = sup_{\underline{\lambda}} L(h^{-1}(\underline{\lambda}); x) = sup_{\theta} L(\theta; x).$$

Se sigue que el supremo de L^* se obtiene cuando $\underline{\lambda} = h(\widehat{\theta})$. Así $h(\widehat{\underline{\lambda}})$ es el estimador de máxima verosimilitud de $h(\theta)$ [13].

En muchas aplicaciones $\underline{\lambda} = h(\underline{\theta})$ no es uno a uno. Aún así es tentador tomar a $\widehat{\underline{\lambda}} = h(\widehat{\underline{\theta}})$ como el estimador de máxima verosimilitud de $\underline{\lambda}$. El siguiente teorema es una justificación.

Teorema C.10. *Sea $\{f_{\underline{\theta}} : \underline{\theta} \in \Theta\}$ una familia de funciones de densidad de probabilidad (funciones de masa de probabilidad), y sea $L(\underline{\theta})$ la función de verosimilitud. Supongamos que $\Theta \subseteq \mathbb{R}^k$, $k \geqslant 1$. Sea $h : \Theta \to \Lambda$ un mapeo de Θ sobre Λ, donde Λ está contenido en $\mathbb{R}^p$ $(1 \leqslant p \leqslant k)$. Si $\widehat{\underline{\theta}}$ es un estimador de máxima verosimilitud de $\underline{\theta}$, entonces $h(\widehat{\underline{\theta}})$ es un estimador de máxima verosimilitud de $h(\underline{\theta})$.*

La demostración se puede consultar en [13], pág. 418.

C.4. La prueba de razón de verosimilitud (PRV)

El método general de la PRV es maximizar la verosimilitud bajo la hipótesis nula H_0, y también maximizar la verosimilitud bajo la hipótesis alternativa H_1. Estos principales resultados están dados en las siguientes definiciones y teorema:

Definición C.11. *Si la distribución de la muestra aleatoria $X = (\underline{x_1}, \ldots, \underline{x_n})'$ depende de un vector parámetro $\underline{\theta}$, y si $H_0 : \underline{\theta} \in \overline{\underline{\Theta}}_0$ y $H_1 : \underline{\theta} \in \overline{\underline{\Theta}}_1$ son cualesquiera dos hipótesis, entonces el estadístico de razón de verosimilitud (RV) para contrastar H_0 contra H_1 se define como:*

$$\lambda(x) = L_0^*/L_1^*, \tag{C.15}$$

donde L_i^* es el valor máximo que la función de verosimilitud toma en la región $\overline{\underline{\Theta}}_i$, $i = 0, 1$ [11].

El estadístico $\lambda(x)$ y $r(x) = \frac{L_0^*}{L^*}$, donde L^* es el valor máximo que la función de verosimilitud toma en el espacio paramétrico $\overline{\underline{\Theta}}$, conducen al mismo criterio para rechazar H_0 [13].

Equivalentemente, se puede usar el estadístico $-2log\lambda = 2(l_1^* - l_0^*)$, donde $l_0^* = log(L_0^*)$ y $l_1^* = log(L_1^*)$ [11].

En general uno tiende a favorecer H_1 cuando el estadístico de razón de verosimilitud es bajo, y H_0 cuando es alto. Una prueba basada en el estadístico de razón de verosimilitud se puede definir como sigue:

Definición C.12. *La prueba de razón de verosimilitud (PRV) de tamaño α para contrastar H_0 contra H_1 tiene región de rechazo $R = \{x|\lambda(x) < c\}$, donde c es determinada de manera que $sup_{\underline{\theta}\in\overline{\underline{\Theta}}_0} P_{\underline{\theta}}(x \in R) = \alpha$ [11].*

La PRV tiene la siguiente propiedad asintótica muy importante.

Teorema C.13. *Bajo algunas condiciones de regularidad en $f_{\underline{\theta}}(X)$, la variable aleatoria $-2log\lambda(X)$ bajo H_0 es asintóticamente distribuida como una variable aleatoria χ^2 con grados de libertad igual a la diferencia entre el número de parámetros independientes en $\overline{\underline{\Theta}}$ y el número en $\overline{\underline{\Theta}}_0$ [13].*

Hipótesis de una muestra sobre Σ

Teorema C.14. *Sea $\underline{x_1}, \ldots, \underline{x_n}$ una muestra aleatoria de $N_p(\underline{\mu}, \Sigma)$. Si H_0 y H_1 son hipótesis que conducen a $\widehat{\Sigma}$ y S como los estimadores de máxima verosimilitud para Σ, y si $\overline{\underline{x}}$ es el estimador de máxima verosimilitud de $\underline{\mu}$ bajo ambas hipótesis, entonces la PRV para contrastar H_0 contra H_1 está dada por $-2log\lambda = np(a - logg - 1)$, donde a y g son la media aritmética y geométrica de los valores propios de $\widehat{\Sigma}^{-1}S$.*

La demostración se puede consultar en [11], págs. 133-134.

Apéndice D

Software R

R (www.r-project.org) es un *software* estadístico libre comúnmente usado. R permite llevar acabo análisis estadísticos de un modo interactivo, y una programación simple. La pantalla de inicio de R se muestra en la Figura D.1. Los comandos deben ser escritos en la línea de comandos $>$ y para ejecutarlos se debe oprimir *enter*. Para agregar un comentario se usa el signo # seguido del comentario. El signo de pregunta (?) junto al comando proporciona la ayuda de ese comando en particular, por ejemplo: ?*plot*, también se puede usar la función *help()*: *help(plot)*.

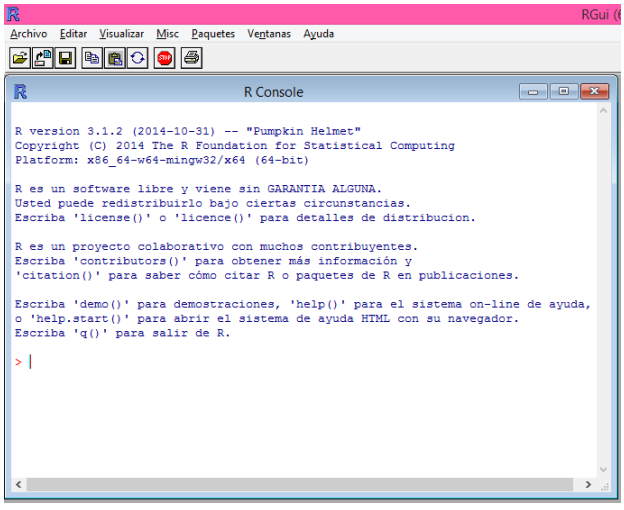

Figura D.1: Pantalla de inicio de R.

En algunos casos, especialmente si el número n de datos es grande, se dispone de los datos en un fichero. Supongamos que los datos están almacenados en un fichero

csv, donde las observaciones de cada variable están en una columna. Los datos pueden estar separados por espacios, comas, tabuladores, etc. y los decimales se pueden indicar por comas o puntos. Además, en ocasiones el fichero incluye una primera línea de cabecera en la que aparecen los nombres de las variables. Antes de leer los datos es necesario asegurarse que el directorio de trabajo es correcto; para elegir directorio se hace lo siguiente: archivo, cambiar directorio, y por último se elige la carpeta donde se encuentra el fichero.

La función para leer datos es *read.table()*, en su argumento se específica el nombre del fichero (*file*), si el fichero contiene cabecera (*header*) cuyas opciones son *TRUE* y *FALSE*, y la forma en que se separan los datos (*sep*), algunas posibilidades para éste son ",", "\t", los cuales indican que los datos se separan con coma y tabulador respectivamente. Un ejemplo de esta función es: *datos*$<-$*read.table(file="datos.csv", header=FALSE,sep=",")*.

Si el conjunto de datos multivariados llamado *datos* cumple que la i-ésima columna tiene el nombre *xi*, se puede acceder a esta columna con el comando *datos$xi*. La función *attach()* establece el conjunto de datos a utilizar, así, después de ejecutar el comando *attach(datos)*, se puede acceder a la i-ésima columna de *datos* directamente con su nombre, *xi*. La función $length()$ calcula el tamaño del objeto de su argumento: *length(xi)*. Si se desea calcular un vector cuyos elementos son los elementos al cuadrado de otro vector, por ejemplo *xi*, se ejecuta el siguiente comando *xi2* $<-$ *xi ^2*. En R se pueden unir *xi* y *xj* como columnas con el comando *cbind(xi,xj)* y como filas con el comando *rbind(xi,xj)*.

D.1. Análisis de regresión

En esta sección se enlistan las diferentes funciones básicas para llevar a cabo un análisis de regresión.

- La función *cov()* calcula la covarianza: *cov(xi,xj)* proporciona la covarianza de *xi* y *xj*. En R se trata de la covarianza muestral, que usa $n-1$ en el denominador.

Si se desea la covarianza que usa n, se usan dos comandos: *n=length(xi)* y *(n-1)cov(xi,xj)/n*.

- El coeficiente de correlación se calcula con la función *cor()*: *cor(xi,xj)*.
- Para obtener una representación gráfica de *xj* en función de *xi* se utiliza la función: *plot(xi,xj,xlab="Eje X",ylab="Eje Y")*.
- El comando *pairs()* con una matriz de datos en el argumento da como resultado la matriz de dispersiones de las diferentes combinaciones de las columnas: *pairs(datos)*.
- La función fundamental para la regresión lineal es *lm()* y utiliza como argumento principal una fórmula como las siguientes:
  - *xi* $\sim$ *xj* o *xi* $\sim$ *1+xj* ajusta *xi* respecto a *xj*.
  - *xi* $\sim$ *0+xj*, *xi* $\sim$ *-1+xj* o *xi* $\sim$ *xj-1* ajusta *xj* respecto a *xi* con una recta que pasa por el origen.
  - *log(xi)* $\sim$ *xj+xk* ajusta a la variable transformada *log(xi)* con *xj* y *xk* con el término intercepto implícito.
  - *xi* $\sim$ *xj + I(xj ^2)*, *xi* $\sim$ *I(xj) + I(xj ^2)* o *xi* $\sim$ *poly(xj,2)* la notación *I(xj ^2)* indica que se desea incluir en el modelo la potencia de grado 2 de *xj*.

  La salida de *lm()* es un objeto que contiene mucha información, el comando *ajuste=lm(fórmula)* se almacena este objeto con el nombre *ajuste*.
- El resumen del ajuste se obtiene con la función *summary()*: *summary (ajuste)*.
- Los valores ajustados son almacenados en el elemento *fitted* de la variable de salida de *lm()*: *ajuste$fitted* o *fitted(ajuste)*.
- Los residuos son almacenados en el elemento *resid* de la variable de salida de *lm()*: *ajuste$resid* o *resid(ajuste)*.
- Los coeficientes son almacenados en el elemento *coef* de la variable de salida de *lm()*: *ajuste$coef* o *coef(ajuste)*.

- El análisis de varianza se obtiene con la función *anova()*: *anova(ajuste)*.
- La función *confint()* proporciona los intervalos de confianza de los parámetros: *confint(ajuste,level=0.95)*.
- La representación gráfica de *xj* en función de *xi* con la recta ajustada se obtiene con las funciones:
  plot(xi,xj)
  abline(lm(xj $\sim$ *xi))*.
- Distintos gráficos de diagnóstico se obtienen con la función *plot()* aplicada a la salida de *lm()*: *plot(ajuste)* o *plot(ajuste,which=1)*, cambiando los valores del argumento *which* de 1 a 6 se obtienen los distintos gráficos, también se obtienen 4 gráficas de diagnóstico con las siguientes instrucciones:
  par(mfrow = c(2,2))
  plot(ajuste).
- La regresión paso a paso se realiza mediante la función *step()* aplicada a la salida de *ajuste*$<-$*lm(fórmula,data=datos)* e indicando la dirección del paso, la cual puede ser *backward* (hacia atrás), *forward* (hacia delante) o *both* (ambos): *step(ajuste,direction="both")*.

D.2. Análisis de componentes principales

Existen diferentes formas de realizar el análisis de componentes principales usando R, aquí se presenta sólo una para la cual no se necesita ninguna paquetería.

- Si el primer paso para llevar a cabo el análisis de componentes principales en un conjunto de datos multivariados llamado *datos* es estandarizar las variables bajo estudio, se utiliza la función *scale()* : *datosE* $<-$ *scale(datos)*, este comando estandariza *datos* y el resultado lo guarda en *datosE*.
- El ACP se puede llevar acabo usando la función *prcomp()*:
  acp $<-$ *prcomp(datos)*, si se lleva acabo el ACP en los datos, y

acp $<-$ *prcomp(datosE)*, si se lleva acabo el ACP en los datos estandarizados. Estos comandos realizan ACP y guardan su salida en *acp*.

- Se puede obtener un resumen de los resultados del análisis de componentes principales utilizando la función *summary()* de la salida de *prcomp()*: *summary(acp)*, esto proporciona la desviación estándar de cada componente, la proporción de la varianza explicada por cada componente y la proporción acumulada hasta ese componente.

- La desviación estándar de las componentes está almacenada en un elemento llamado *sdev* de la variable de salida fabricada por *prcomp*: *acp$sdev*.

- La variación de las componentes principales se puede obtener con: *(acp$sdev) ^2*.

- La varianza total explicada por las componentes es la suma de la varianza de las componentes, es decir: *sum((acp$sdev) ^2)*.

- Una gráfica de sedimentación se obtiene con la función *screeplot()*: *screeplot(acp)* o *screeplot(acp,type= "lines")*.

- Las cargas de las componentes principales está almacenada en un elemento de nombre *rotation* de la salida de *prcomp()*: *acp$rotation*. Éste contiene una matriz con las cargas de cada componente principal, donde la i-ésima columna de la matriz contiene las cargas de la i-ésima componente principal, es decir, el vector propio de la matriz de covarianza muestral o correlación muestral correspondiente al i-ésimo valor propio máximo.

- Las cargas de la i-ésima componente principal se pueden obtener con: *acp$rotation[,i]*.

- Las puntuaciones de las componentes principales son almacenadas en el elemento *x* de la variable de salida de *prcomp()*, éste contiene una matriz en la cual la i-ésima columna corresponde a las puntuaciones de la i-ésima componente principal, la cual se obtiene con: *acp$x[,i]*.

- El diagrama de dispersión de las dos primeras componentes principales se obtiene con: *plot(acp$x[,1],acp$x[,2])* [4].

Referencias

[1] Anderson, T. W. 1963. “Asymtotic theory for principal component analysis”. *Ann. Math. Statist.*, **34**, págs. 122-148.

[2] Arvesú Carballo, Jorge; Marcellán Español, Francisco; Sánchez Ruiz, Jorge. 2006. *Problemas Resueltos de Álgebra lineal* [pág. 162]. 1^a ed., España (Departamento de Matemáticas Universidad Carlos III de Madrid, Thomson.

[3] Bowerman, Bruce L.; O’Connell, Richard T.; Koehler,Anne B. 2007. *Pronósticos, series de tiempo y regresión. Un enfoque aplicado*. 4^a ed. Cengage Learning Editores S.A. de C.V.

[4] Coghlan, Avril. 2011. *A Little Book of R For Multivariate Analysis*. University College Cork, Cork, Ireland.

[5] Díaz Caraballo, José Neville. *Regresión Lineal con R*. Departamento de Matemáticas, Universidad de Puerto Rico. http://math.uprag.edu/Rsampleregression.pdf. Consultado en Junio de 2015.

[6] Friedberg, Stephen H.; Insel, Arnold J.; Spence, Lawrence E. 1982. *Álgebra lineal*. Illinois State University. 1^a ed., México, Publicaciones Cultural, S.A.

[7] Graybill, Franklin A. 1976. *Theory and application of the linear model*. Duxbury press.

[8] Johnson, Dallas E. 2000. *Métodos multivariados aplicados al análisis de datos*. 1^a ed., México, Thomson.

[9] Kessler, Mathieu. *Regresión lineal con R*. Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena. http://filemon.upct.es/∼mathieu/organizacion/practicas/practicasconR/practreglinealconR.pdf. Consultado en Junio de 2015.

[10] Kshirsagar, A. M. 1972. *Multivariate Analysis*. Marcell Dekker, New York.

[11] Mardia, K. V.; Kent, J. T.; Bibby, J. M. 1979. *Multivariate Analysis*. Academic press.

[12] Navidi, William. 2006. *Estadística para ingenieros y científicos*. 1^a ed., México, Mc Graw Hill.

[13] Rohatgi, Vijay K.; Ehsanes Saleh, A. K. Md. 2001. *An Introduction to Probability and Statistics*. 2^a ed., Canadá, Wiley-Interscience.

[14] Sánchez Rivera, José Antonio. 2012. *Análisis de Componentes Principales. Clasificación de Países según las carreras de atletismo*. Trabajo Fin de Máster, Universidad de Granada.

[15] *Tutorial 10: Regresión, Curso de Introducción a la Estadística*. http://www.postdatastatistics.com/. Consultado el 10 de Junio de 2015.

[16] Wackerly, Dennis D.; Mendenhall, William; Scheaffer, Richard L. 2008. *Estadística matemática con aplicaciones*. 7^a ed., Cengage Learning.

Referencias ambientales

[17] CentroGeo. 2015. *Programa de Ordenamiento Ecológico Regional para los Estados de Tabasco y Chiapas (Cuenca Grijalva-Usumacinta) capítulo Tabasco*. Villahermosa: Centro de Investigación en Geografía y Geomática “Ing. Jorge L. Tamayo” A.C., Comité de Planeación para el Desarrolo del Estado de Tabasco del Gobierno del Estado de Tabasco.

[18] Challenger, A.; Dirzo, R. 2009. “Factores de cambio y estado de la biodiversidad”. En Conabio, *Capital Natural de México*. (Vol. II: “Estado de conservación y tendencias de cambio”, págs. 37-73). México: Conabio.

[19] Conabio. 2012. *Desarrollo territorial sustentable: Programa Especial de Gestión en Zonas de Alta Biodiversdad*. México: Conabio.

[20] Galeana Pizaña, J. “Análisis territorial de los sistemas productivos desde una perspectiva de preservación del capital natural”. Obtenido de Plataforma Geoweb para la Red de Desarrollo en Sustentabilidad Alimentaria: http://asam.centrogeo.org.mx/index.php/resultado-1. Consultado el 10 de Junio de 2015.

[21] Grupo de Trabajo para el Marco Conceptual de la Evaluación de los Ecosistemas del Milenio. 2003. *Ecosistemas y bienestar humano: marco para la evaluación*. World Resources Institute.

[22] López, A. 2012. *Deforestación en México: un análisis preliminar*. México: Centro de Investigación y Docencia Económicas A.C.

[23] Sánchez Colón, S.; Flores Martínez, A.; Cruz-Leyva, I.; Velázquez, A. 2009. “Estado y transformación de los ecosistemas terrestres por causas humanas”. En Conabio, *Capital natural de México* (Vol. II: “Estado de conservación y tendencias de cambio”, págs. 75-129). México: Conabio.

[24] Semarnat. 2013. *Informe de la situación del medio ambiente en México. Compendio de estadísticas ambientales. Indicadores clave y desempeño ambiental*. Edición 2012. México: Semarnat.