



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

“COMPONENTES PRINCIPALES COMUNES: UN MÉTODO PARA LA COMPARACIÓN DE MATRICES DE VARIANZA Y COVARIANZA”.

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

M A T E M Á T I C O

P R E S E N T A :

LEONARDO KUBA DORFSMAN

DIRECTORA DE TESIS:

M. en C. LETICIA EUGENIA GRACIA MEDRANO VALDELAMAR



FACULTAD DE CIENCIAS UNAM

2006



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

ACT. MAURICIO AGUILAR GONZÁLEZ
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito: "Componentes Principales Comunes: un método para la comparación de matrices de varianza y covarianza".

realizado por Leonardo Kuba Dorfsman

con número de cuenta 402058229 , quien cubrió los créditos de la carrera de: Matemáticas.

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis

Propietario M.en C. Leticia Eugenia Gracia Medrano Valdelamar.

Leticia Gracia

Propietario Dra. Silvia Ruiz Velasco Acosta.

A BUA

Propietario M. en C. Jéssica Hernández Rojano.

Jessica Hernandez R

Suplente Mat. Margarita Elvira Chávez Cano.

M Elvira

Suplente Dr. Javier Páez Cárdenas.

Javier Paez

Consejo Departamental de Matemáticas.



M. en C. Alejandro Bravo Mojica

FACULTAD DE CIENCIAS
CONSEJO DEPARTAMENTAL
DE
MATEMÁTICAS

**Componentes Principales Comunes:
un método para la comparación de matrices
de varianza covarianza**

Leonardo Kuba Dorfsman

*A mis padres, por su apoyo;
y a mis maestros, por su enseñanza,
y a la perseverancia, sin la cual, lo demás hubiera sido inútil.*

Agradecimientos

Cualquier trabajo, por menor que parezca, requiere no sólo de mucho esfuerzo propio; sino también de apoyo directo e indirecto de otras personas. Este texto no es la excepción, por lo que lo consideraría incompleto sin una pequeña sección agradecimiento a las personas que, de una manera u otra, han sido indispensables para lograr ésto, que para mí significa la titulación, y para otros, un texto de apoyo más en Componentes Principales Comunes.

Antes que nada, debo agradecer a mis padres. Ellos me han apoyado económicamente durante toda mi vida, además de haberme criado y educado de la mejor manera que han sabido.

Creo que también debo agradecer a todos los que me apoyaron emocionalmente en tiempos de confusión. En este rubro han sido de gran importancia mi tía Gneshe, el Dr. Alejandro Tarrago (entre otros de sus colegas como lo son la Dra. Klein y el Dr. Litman), así como varios amigos entre los que puedo mencionar a Paulina, María José, Pablo y Karin Matarasso, quien además trabajó intensamente en la corrección de estilo de este texto.

Debo considerar también el apoyo de otras personas, que tal vez no fueron las que mejor han sabido apoyarme, pero siempre han las mejores intenciones con respecto a mí. Entre ellas se encuentra mi abuela Estela, entre muchos otros familiares.

Sin embargo, este trabajo no sólo se hizo sólo con apoyo económico y emocional, requirió también de una fuerte preparación intelectual. Por ello, debo agradecer a varios maestros que me formaron, cada uno y refinando el trabajo que el anterior había hecho. En este rubro debo empezar por agradecer a la Maestra Guadalupe Romero Zárate: ella me enseñó en la secundaria que las matemáticas eran más que una serie de procedimientos misteriosos que era conveniente conocer para obtener un número “útil” que, entre otras cosas, servían para pasar el examen.

Esa visión fue reforzada en la Universidad Hebrea de Jerusalén, donde Yael y Albert, maestros de los cuales ya no recuerdo sus apellidos, me enseñaron que el verdadero aprendizaje matemático, como un procedimiento analítico que conducía a resultados útiles o abstractos, requiere paciencia y “amor al arte”.

Sin embargo la U.N.A.M. es el lugar en el que se ha consolidado mi formación matemática. En ella, he tenido grandes maestros. De ellos, por brevedad, menciono sólo a dos: el Dr. Javier Páez Cárdenas, que me enseñó a llevar la formalidad como una parte medular en el proceso analítico matemático y a la Dra. Ana Meda Guardiola, que me enseñó a utilizar esas herramientas analíticas para resolver problemas reales.

Considerando ahora, no a los que me han formado en mi carrera y en la vida, sino, a los que me ayudaron directamente en la realización de este trabajo; agradezco a la M. en C. Leticia Eugenia Gracia Valdelamar. Ella fué mi directora de tesis orientándome y apoyándome en la realización del trabajo. El Biólogo Armando López Velázquez colaboró con sus valiosos datos, que permitieron dar un ejemplo práctico de a técnica que se explica. Conté también con sinodales que revisaron con esmero y cuidado mi trabajo y lo mejoraron con sus comentarios: la Dra. Silvia Ruiz, que, con su conocimiento, esclareció algunos puntos sobre los Componentes Principales Comunes; el Dr. Javier Páez, que colaboró con la claridad didáctica en algunas demostraciones; y la Mat. Margarita Chávez Cano y la M. en C. Jéscica Hernández Rojano, quienes ayudaron mucho en mejoras en cuanto la redacción y presentación del trabajo.

Índice general

Índice de figuras	v
Introducción	vi
CAPÍTULO 1. Componentes Principales	1
1.1. Introducción	1
1.2. Definición en vectores aleatorios	2
1.3. Estimación de los Componentes Principales	5
CAPÍTULO 2. Componentes Principales Comunes	9
2.1. Introducción	9
2.2. Igualdad de Varianzas	12
2.3. Componentes Principales Comunes	14
2.4. Matrices de varianza proporcionales	17
2.5. Componentes Principales Comunes Parciales	21
2.6. El proceso de comparación de las matrices de varianza	26
CAPÍTULO 3. Análisis de datos con Componentes Principales Comunes	30
3.1. Poblaciones, muestras y datos	32
3.2. Análisis descriptivo de los datos	32
3.3. Análisis de componentes principales comunes	49
3.4. Conclusiones del ejemplo	82
CAPÍTULO 4. Conclusiones	84
Apéndice A. Generalización de la Esperanza, Covarianza y Varianza de variables a vectores aleatorios	86
Apéndice B. Teorema de la Descomposición Espectral	87
Apéndice C. Distribución Normal Multivariada	89
Apéndice D. Matrices Ortogonales	91
Apéndice E. Derivadas parciales simbólicas con respecto a vectores	92
Apéndice F. Potencias no Enteras de Matrices	93

Índice de figuras

1. Largo, caliz y pedicelo de la flor	33
2. Antera, pistilo y hercogamia de la flor	34
3. Gráfica matriz de las muestras 2003	35
4. Gráfica matriz de Taxco madre	37
5. Gráfica matriz de Taxco padre	38
6. Gráfica matriz Tula madre	39
7. Tula padre	40
8. Histogramas de algunas variables de Taxco	41
9. Histogramas de algunas variables de Ticumán	42
10. Diagrama de dispersión ancho vs. cáliz en Ticumán	43
11. Gráfica matriz Teotihuacán	45
12. Gráfica matriz Ticumán tempranas	46
13. Gráfica matriz Ticumán tardías	47
14. Histogramas Teotihuacán	48
15. Histogramas Ticumán tempranas	48
16. Histogramas Ticumán tardías	49
17. Diagrama de dispersión de los componentes 3 y 5 para Taxco M según CPC de Tula M y P y Taxco M con 7 variables	60
18. Diagrama de dispersión de los componentes 3 y 5 en Tula M para el modelo CPC de Tula M y Tula P	63

Introducción

Uno de los objetos más utilizados en la probabilidad y estadística multivariada es la matriz de varianzas y covarianzas (a la que en este texto, por simplificación, llamaremos matriz de varianzas). En dicha matriz se miden las varianzas de cada una de las variables incluídas en un modelo multivariado, así como la covarianza entre ellas. Además, a partir de ella se puede calcular la matriz de correlación. Las varianzas miden la dispersión de los datos, mientras que las covarianzas miden la relación entre dos variables (ver [10]). Las correlaciones miden la dependencia lineal entre las variables del modelo.

Como se puede ver en lo anterior, la matriz de varianzas contiene gran cantidad de información, y, por eso, no es extraño que tanto la probabilidad como la estadística multivariada le hayan prestado tanta atención. En particular, uno de los problemas importantes para la estadística multivariada es comparar matrices de varianzas de diferentes poblaciones. Por lo general, esas comparaciones se hacen con pruebas de igualdad cuyo objetivo es saber si varias matrices de varianzas de poblaciones diferentes son iguales o no.

Sin embargo, cuando se tienen datos que son vectores p -variados, su matriz de varianzas tiene $\frac{p(p+1)}{2}$ grados de libertad. De lo anterior se sigue que, en los modelos en que se utilizan las matrices de varianzas, existe una gran diferencia de grados de libertad entre el modelo que considera a las matrices de varianzas como iguales y el modelo que considera a las matrices de varianzas como distintas. Por eso, es conveniente hacer escalas de lo parecidas o diferentes que son las matrices de varianzas y covarianzas de varias poblaciones, con peldaños intermedios entre la igualdad y la desigualdad, para poder modelar matrices de varianzas que no son iguales, sin agregar tantos grados de libertad como si se consideraran completamente diferentes.

Se han hecho varios modelos al respecto. Algunos, por ejemplo, se basan en la comparación de la matriz de correlación como el que se sugiere en [6]. En este trabajo se trata un modelo hecho por Bernhard Flury, presentado en [2], que compara las matrices de varianzas según sus Componentes Principales.

Los Componentes Principales son un método para descomponer una matriz de varianzas según las direcciones en que se encuentra la variación de los datos. La idea es escribir los datos en una base ortonormal, ordenada, de manera que la mayor variación de los datos esté en la primera coordenada y la menor en la última; estando todas las variaciones intermedias ordenadas de mayor a menor en las demás coordenadas. En ciertos casos,

esta metodología puede ser utilizada para obtener gráficas en dos o tres dimensiones que representan razonablemente bien espacios de dimensiones mayores.

En el primer capítulo se explicará lo que son los Componentes Principales en general. En el segundo capítulo se introducirá una escala de similitud entre varias matrices de varianza a partir de los Componentes Principales, y se presentará la teoría para poder comparar varias matrices de varianza a partir de este método. En el tercer capítulo se aplicará el método de Componentes Principales Comunes en la comparación de varianzas de poblaciones de flores de la especie *Datura Stramonium*. En el último capítulo se presentarán las conclusiones en cuanto a este método de comparación de matrices de varianza.

Se espera que el lector tenga conocimientos básicos en probabilidad y estadística, así como en álgebra lineal. Sin embargo, la mayoría de los resultados importantes de estas áreas que se utilizarán en este trabajo se encuentran en los apéndices.

CAPÍTULO 1

Componentes Principales

1.1. Introducción

Supóngase que se tiene una muestra aleatoria x_1, \dots, x_n , donde x_i es un vector de p entradas $\forall i \in \{1, \dots, n\}$ y se quiere extraer información de dicha muestra.

En el formato dado, la información es muy difícil de analizar, pues si $p > 3$ no se puede graficar toda la información en gráficas cartesianas. Sin éstas últimas resulta muy complicado inferir patrones en los datos para proseguir el análisis de manera acertada.

Se han propuesto varias alternativas para tener una idea gráfica de los datos: éstas incluyen métodos tales como graficar las variables de dos en dos, las caras de Chernoff, las curvas de Andrews y otros que son comúnmente mencionados en casi cualquier libro de estadística multivariada. Sin embargo, todos estos métodos son algo limitantes por el hecho de que no permiten percibir fácilmente las relaciones entre los datos: en el mejor de los casos, ayudan a inferir grupos de datos pero no patrones en ellos. Además, estos métodos tienen la desventaja de que su resultado está influenciado por el orden de aparición que se escoja de las variables.

Otra alternativa para tener una idea gráfica de una muestra multivariada de dimensión mayor a tres es la que se presenta en este capítulo. Consiste en hacer un cambio en la base en la que se expresan los datos para encontrar una base en la que las primeras coordenadas sean las que contienen la mayor variación de los datos, y las últimas representen variaciones muy pequeñas en los datos. Así, cuando la variabilidad en las últimas coordenadas es mucho menor que en las primeras, se sospecha que la dimensión real de los elementos de la población es menor a p y se reduce la información. En especial, si la dimensión real es dos o tres, al reducir la información se cuenta con métodos muy eficientes para graficar la información en el nuevo formato.

Adicionalmente, el cambio de base que se le aplica a los datos tiene la propiedad de que genera una base en la que las coordenadas en las que está expresada la información no están correlacionadas. Así, cuando se quieren realizar regresiones en términos de variables altamente correlacionadas, se puede aplicar la transformación para obtener la misma información en términos de variables no correlacionadas y hacer la regresión en términos de estas nuevas variables para evitar problemas de multicolinealidad.

Además, como se hace más adelante en este texto, al tener muestras diferentes y aplicar la transformación mencionada a cada una, se pueden comparar los resultados

obtenidos para analizar más fácilmente las matrices de varianza y, en consecuencia, las diferencias entre las poblaciones que se desean estudiar.

1.2. Definición en vectores aleatorios

Sea X un vector aleatorio de p entradas con segundo momento finito, centrado (i.e. $E[X] = 0$) y sea $\Psi = \text{Var}[X]$ (ver Apéndice A). Se quiere simplificar la comprensión de dicho vector aleatorio a través de una transformación lineal que convierta a dicho vector en otro vector aleatorio U , tal que las diferentes coordenadas de U no estén correlacionadas, y que si $U = (u_1, \dots, u_p)$ entonces:

$$(1.1) \quad \text{Var}(u_1) \geq \text{Var}(u_2) \geq \dots \geq \text{Var}(u_p)$$

El primer paso para resolver el problema será buscar el vector de constantes $a \in \mathbb{R}^p$ que maximice $\text{Var}(a'X)$. Sin embargo, salvo en el caso en que $\text{Var}(X) = 0$, esto no es posible, pues aumentando indefinidamente cualquier coordenada del vector a que corresponda a una entrada de X con varianza distinta de cero, se puede aumentar $\text{Var}(a'X)$ indefinidamente. Por lo anterior se aumenta la restricción:

$$(1.2) \quad a'a = 1$$

Lo anterior, como se verá más adelante, garantizará que la transformación buscada sea ortonormal y, en consecuencia, preserve distancias.

Como Ψ es una matriz de varianza, se sabe que es simétrica. Partiendo de esta hipótesis, es posible aplicar el Teorema de la Descomposición Espectral (ver Apéndice B) para encontrar una matriz $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$, y una matriz $\beta = (\beta_1, \dots, \beta_p)$ ortogonal que cumplan:

$$(1.3) \quad \Psi = \beta\Lambda\beta'$$

Es posible escribir estas matrices de tal manera que la matriz Λ sea tal que:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$$

Sea $a \in \mathbb{R}^p$ arbitrario, tal que $a'a = 1$. Como $\{\beta_1, \dots, \beta_p\}$ es base de \mathbb{R}^p se puede expresar a en términos de esa base con la siguiente notación:

$$a = \alpha_1\beta_1 + \dots + \alpha_p\beta_p$$

Definiendo el vector α como $(\alpha_1, \dots, \alpha_p)$ se puede ver que:

$$(1.4) \quad 1 = \alpha'\alpha$$

Pues:

$$\begin{aligned}
 1 &= a'a \\
 &= \left\langle \sum_{i=1}^p \alpha_i \beta_i, \sum_{j=1}^p \alpha_j \beta_j \right\rangle \\
 &= \sum_{i=1}^p \sum_{j=1}^p \alpha_i \alpha_j \beta_i \beta_j \\
 &= \sum_{i=1}^p \alpha_i^2 \\
 &= \alpha' \alpha
 \end{aligned}$$

A partir de lo anterior es posible acotar $Var(a'X)$ de la siguiente manera:

$$\begin{aligned}
 Var(a'X) &= a' \Psi a \\
 &= \alpha' \beta' \Psi \beta \alpha \\
 &= \alpha' \Lambda \alpha \\
 (1.5) \quad &= \sum_{j=1}^p \lambda_j \alpha_j^2 \\
 &\leq \lambda_1 \sum_{j=1}^p \alpha_j^2
 \end{aligned}$$

$$\begin{aligned}
 Var(a'X) &\leq \lambda_1 \\
 (1.6)
 \end{aligned}$$

y, como $\beta_1 = 1 \beta_1 + 0 \beta_2 + \dots + 0 \beta_p$, de (1.5) se tiene que:

$$Var(\beta_1' X) = \lambda_1$$

Por lo tanto β_1 es un valor que maximiza $Var(a'X)$ sujeta a las restricciones impuestas. Con ese fundamento, definimos el primer componente principal U_1 como:

$$(1.7) \quad U_1 = \beta_1' X$$

Ahora, aprovechando que ya se tiene definido el primer componente principal, se pueden definir los demás Componentes Principales de manera recursiva como se hace a continuación:

Supóngase que se tienen ya definidos $h - 1$ Componentes Principales. Se define el h -ésimo componente principal como la transformación lineal $U_h = a'X$ con $a \in \mathbb{R}^p$ que maximiza $Var(U_h)$ dadas las siguientes restricciones:

1. $a'a = 1$
2. $Cov(U_h, U_i) = 0 \quad \forall i \in \{1, \dots, h-1\}$

A partir de lo anterior, se mostrará por inducción sobre h , (para $h \leq p$), que $U_k = \beta_k' X$ y $Var(U_k) = \lambda_k$ donde β_k son las columnas de la matriz β y $\lambda_k = \Lambda_{kk}$; Λ y β las matrices encontradas por el Teorema de la Descomposición Espectral.

Como la base de esta demostración por inducción (caso $h = 1$) ya se hizo arriba, se proseguirá con el paso inductivo:

Hipótesis de Inducción:

$$U_k = \beta_k X \text{ y } Var(U_k) = \lambda_k \quad \forall k \in \{1, 2, \dots, h-1\}$$

Por demostrar:

$$U_k = \beta_k X \text{ y } Var(U_k) = \lambda_k \quad \forall k \in \{1, 2, \dots, h\}$$

Sea a el vector correspondiente a la transformación lineal $U_h = a'X$. Usando la hipótesis de inducción se reformula la restricción (2) de la siguiente manera:

$$\begin{aligned} Cov(U_h, U_i) &= Cov(a'X, \beta_i X) \\ &= a' \Psi \beta_i \\ &= \lambda_i a' \beta_i \end{aligned}$$

En consecuencia, como $\lambda_i > 0 \forall i \in \{1, \dots, h-1\}$ (pues son varianzas) se tiene que:

$$Cov(U_h, U_i) = 0 \iff a' \beta_i = 0$$

Así, la restricción (2) se convierte en:

$$(1.8) \quad a' \beta_i = 0 \quad \forall i \in \{1, \dots, h-1\}$$

Dado que $\{\beta_1, \dots, \beta_p\}$ es base se puede escribir:

$$a = \alpha_1 \beta_1 + \dots + \alpha_p \beta_p$$

Y utilizando (1.8) se obtiene para $i \in \{1, \dots, h-1\}$:

$$0 = a' \beta_i = \sum_{j=1}^p \alpha_j \beta_j' \beta_i = \alpha_i$$

Por lo que:

$$a = \alpha_h \beta_h + \dots + \alpha_p \beta_p$$

Finalmente:

$$\begin{aligned} Var(a'X) &= \sum_{j=h}^p \lambda_j \alpha_j^2 && \text{Ver (1.5)} \\ &\leq \lambda_h \sum_{j=h}^p \alpha_j^2 \\ Var(a'X) &\leq \lambda_h \end{aligned}$$

Y es fácil ver que:

$$Var(\beta_h X) = \lambda_h$$

Por lo que $U_h = \beta_h' X$

□

Nota: Del Teorema de la Descomposición Espectral se sigue directamente que cuando las λ 's son distintas dos a dos, los Componentes Principales son únicos salvo por multiplicaciones por -1 . Para hacerlos únicos, por convención, se asigna β_k de manera que su primer coeficiente diferente a cero sea positivo.

En el caso en que $\lambda_{q+1} = \lambda_{q+2} = \dots = \lambda_{q+r}$, del Teorema de la Descomposición Espectral se sigue que $E_{\lambda_{q+1}}$ (el subespacio correspondiente al valor propio λ_{q+1} con la transformación Ψ) es único. Sin embargo, cualquier base ortonormal de este espacio puede ser usada para generar los componentes principales U_{q+1}, \dots, U_{q+r} . En este caso se dice que la variación en el espacio $E_{\lambda_{q+1}}$ es esférica.

En seguida se definirán un par de conceptos sobre la varianza de un vector aleatorio X y se demostrará que son invariantes bajo la transformación en Componentes Principales.

DEFINICIÓN 1.1. Sea X un vector aleatorio y sea $\Psi = \text{Var}(X)$; se definen la varianza total y la varianza generalizada, respectivamente, como:

$$\sigma_{total}^2 = \text{tr}(\Psi) \quad \sigma_{gen}^2 = \det(\Psi)$$

TEOREMA 1.1. La varianza total y la varianza generalizada son invariantes bajo la transformación en Componentes Principales.

Demostración:

$$\sigma_{gen}^2 = \det(\Psi) = \det(\beta\Lambda\beta') = \det(\beta) \det(\Lambda) \det(\beta') = \det(\Lambda)$$

$$\sigma_{tot}^2 = \text{tr}(\Psi) = \text{tr}(\beta\Lambda\beta') = \text{tr}(\beta'\beta\Lambda) = \text{tr}(\Lambda)$$

□

Finalmente, se puede agregar que la varianza total es una medida de la variación total de un vector aleatorio. Por lo que, teniendo los componentes principales, al reducir la dimensión del vector aleatorio U , de p a m coordenadas, es posible usar como medida de representatividad de la reducción, al cociente de la varianza total del vector original y la del vector reducido. En fórmulas: si $U = (U_1, \dots, U_p)$ es el vector de Componentes Principales y se reduce a $U' = (U_1, \dots, U_m)$ la proporción de la varianza total de U que aporta U' es:

$$(1.9) \quad \frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i}$$

1.3. Estimación de los Componentes Principales

En la sección anterior se definió el concepto de Componentes Principales para vectores aleatorios, sin embargo, ésta es una técnica estadística, y, como bien se sabe, en estadística no se conoce de antemano la distribución de los datos, y, mucho menos, su matriz exacta de varianza; lo único que se tiene de entrada son datos que se consideran generados por una distribución. Así, todo lo anterior sería de poca utilidad, a menos que se puedan utilizar datos, organizados a manera de muestras, para estimar los componentes principales. En esta sección se mostrará, *grosso modo*, la metodología para ese fin.

Lo primero que se debe hacer para estimar los Componentes Principales a partir de una muestra, dado que se desconoce Ψ (la matriz de varianza del vector aleatorio), es estimarla. Para lo anterior se propone el siguiente estimador:

$$(1.10) \quad \hat{\Psi} = S = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})'$$

Donde x_1, \dots, x_N es la muestra aleatoria y:

$$(1.11) \quad \bar{x} = \frac{1}{N} \sum_{j=1}^N x_j \quad \text{es el estimador de la media de la distribución.}$$

Es importante que se note que, dado que x_1, \dots, x_N son vectores p -variados \bar{x} también es un vector p -variado.

Se sabe que dichos estimadores son insesgados.

Además, cuando la distribución de la muestra aleatoria es normal multivariada, se tiene que:

$$\bar{x} \quad \text{y} \quad S^* = \frac{N-1}{N} S$$

son los estimadores de máxima verosimilitud de Ψ . Es posible mostrar que si $N \geq p$, S es definida positiva con p raíces características distintas.

Así, se le aplica a S el Teorema de la Descomposición Espectral para escribirla como:

$$(1.12) \quad S = BLB'$$

Donde $B = (b_1, \dots, b_p)$ es ortogonal y L es diagonal.

Suponiendo nuevamente que la distribución de la población es normal multivariada y que los valores propios de S son distintos dos a dos, se puede demostrar que B y L son los estimadores de máxima verosimilitud de β y Λ respectivamente, como se hace a continuación:

Dado que la distribución de la población es normal multivariada, la función de verosimilitud para la varianza, Ψ , tomando como parámetro para μ a su estimador de máxima verosimilitud (ver Apéndice C) es:

$$L(\Psi) = (2\pi)^{-\frac{np}{2}} (\det \Psi)^{-\frac{n}{2}} \exp \left\{ -\frac{n}{2} \text{tr}(\Psi^{-1}S) \right\}$$

Aplicándole el logaritmo a la función de verosimilitud, se obtiene:

$$\log L(\Psi) = -\frac{np}{2} \log(2\pi) - \frac{n}{2} [\log(\det \Psi) + \text{tr}(\Psi^{-1}S)]$$

Maximizar la función anterior es equivalente a minimizar:

$$f(\Psi) = \log(\det(\Psi)) + \text{tr}(\Psi^{-1}S)$$

Eso es equivalente a minimizar $f(\Psi) - f(S)$, pues éstas sólo difieren por una constante.

$$(1.13) \quad f(\Psi) - f(S) = \text{tr}(\Psi^{-1}S) - \log(\det(\Psi^{-1}S)) - \text{tr}(I_p)$$

$$(1.14) \quad = \text{tr}(\Psi^{-1}S) - \log(\det(\Psi^{-1}S)) - p$$

Sustituyendo $\Psi = \beta\Lambda\beta'$ y $S = BLB'$, se obtiene que:

$$(1.15) \quad f(\Psi) - f(S) = \text{tr}(\beta\Lambda^{-1}\beta'BLB') - \log(\det(\beta\Lambda^{-1}\beta'BLB')) - p$$

$$(1.16) \quad = \text{tr}(B'\beta\Lambda^{-1}\beta'BL) - \log(\det(B'\beta\Lambda^{-1}\beta'BL)) - p$$

Denotando $A = B'\beta$ y $L^{(\frac{1}{2})}$ como en el Apéndice F, se tiene que A es ortogonal (pues es producto de ortogonales) y lo anterior se convierte en:

$$(1.17) \quad f(\Psi) - f(S) = \text{tr}(A\Lambda^{-1}A'L) - \log(\det(A\Lambda^{-1}A'L)) - p$$

$$(1.18) \quad = \text{tr}(L^{\frac{1}{2}}A\Lambda^{-1}A'L^{\frac{1}{2}}) - \log(\det(L^{\frac{1}{2}}A\Lambda^{-1}A'L^{\frac{1}{2}})) - p$$

Si $\gamma_1, \dots, \gamma_p$ son los valores propios de $L^{\frac{1}{2}}A\Lambda^{-1}A'L^{\frac{1}{2}}$:

$$(1.19) \quad f(\Psi) - f(S) = \sum_{j=1}^p \gamma_j - \log \prod_{j=1}^p \gamma_j - p$$

$$(1.20) \quad = \sum_{j=1}^p (\gamma_j - \log \gamma_j - 1)$$

Ya que $\log(x) \leq x - 1$ y la igualdad se cumple sólo para $x = 1$, se tiene que el mínimo de la ecuación anterior se encuentra en $\gamma_1 = \dots = \gamma_p = 1$.

Eso es equivalente a que $L^{\frac{1}{2}}A\Lambda^{-1}A'L^{\frac{1}{2}} = I_p$. Despejando:

$$L^{\frac{1}{2}}A\Lambda^{-1}A'L^{\frac{1}{2}} = I_p$$

$$A\Lambda^{-1}A' = L^{-1}$$

$$\Lambda^{-1} = A'L^{-1}A$$

$$(1.21) \quad \Lambda = A'LA$$

Como A es ortogonal, la ecuación 1.21 debe de ser la descomposición espectral de L . Por hipótesis, los valores de la diagonal de L son distintos dos a dos, pues son los valores propios de S . De ello se concluye que $A = I_p$, salvo posiblemente por permutaciones de columnas y multiplicaciones por -1 de éstas.

Pero $A = B'\beta$, por lo que $A = I_p$ es equivalente a $B = \beta$. Y, sustituyendo $A = I_p$ en 1.21, se tiene que $L = \Lambda$. De allí que los estimadores de máxima verosimilitud de β y Λ sean B y L , respectivamente.

Lo anterior se denota como:

$$(1.22) \quad \hat{\beta} = B \quad \text{y} \quad \hat{\Lambda} = \frac{N-1}{N}L$$

Finalmente, es indispensable hacer notar que la transformación en Componentes Principales no es invariante bajo transformaciones de escala de una de las p entradas. Por ello, es diferente utilizar la matriz de varianza que la de correlación para aplicar la transformación en Componentes Principales. En general, se utiliza la matriz de varianza cuando se quieren analizar los datos en la escala dada y la de correlación cuando se quieren analizar en un contexto en el cual, cada una de las entradas tiene la misma importancia o cuando las escalas en que se miden las variables son muy distintas.

CAPÍTULO 2

Componentes Principales Comunes

2.1. Introducción

En el capítulo anterior se presentó una técnica para analizar más fácilmente a la matriz de varianza de una muestra aleatoria con p observaciones en n individuos. En este capítulo se presentarán modelos para comparar matrices de varianza de varias poblaciones en las que se consideran relevantes p variables. Las comparaciones a tratar se basan en el análisis de los Componentes Principales de las matrices de varianza.

Esta técnica sirve para formalizar que dos matrices de varianza tienen los mismos componentes principales. Es conveniente realizar este análisis al comienzo de un estudio estadístico para saber si las poblaciones varían en los mismos ejes.

Ejemplos de aplicaciones de esta técnica pueden ser: comparar matrices de varianza de machos y hembras de una misma especie; o de billetes falsos y originales.

Sean $\Psi_1, \Psi_2, \dots, \Psi_k$ matrices de varianza de diferentes poblaciones. Se define la siguiente escala de similitud entre ellas:

1. **Igualdad:** Las matrices de varianza son iguales, es decir:

$$\Psi_1 = \Psi_2 = \dots = \Psi_k$$

2. **Proporcionalidad:** Las matrices de varianza son proporcionales, es decir:

$$\Psi_i = \rho_i \Psi_1 \quad \forall i \in \{2, \dots, k\}$$

3. **Componentes Principales Comunes:** Las transformaciones a componentes principales de todas las poblaciones son las mismas, es decir:

$$\Psi_i = \beta \Lambda_i \beta' \quad \forall i \in \{1, \dots, k\}$$

Donde β es una matriz ortogonal y $\Lambda_i = \text{diag}(\lambda_{i1}, \dots, \lambda_{ik})$, pero no necesariamente se cumple que $\lambda_{i1} \geq \lambda_{i2} \geq \dots \geq \lambda_{ik}$

4. **Componentes Principales Comunes Parciales:** q de las p transformaciones en Componentes Principales son las mismas en todas las poblaciones, es decir:

$$\Psi_i = \beta^{(i)} \Lambda_i \beta^{(i)'}$$

Donde $\beta^{(i)} = (\beta_1, \dots, \beta_q, \beta_{q+1}^{(i)}, \dots, \beta_p^{(i)})$

5. Ψ_1, \dots, Ψ_k no están relacionadas con ninguna de las condiciones anteriores al considerarlas todas al mismo tiempo.

Es fácil notar que esta escala es incluyente es decir $1 \Rightarrow 2 \Rightarrow \dots \Rightarrow 5$. Además, es importante hacer notar que aun cuando no se cumple algún nivel de similitud entre el conjunto de las matrices $\{\Psi_1, \Psi_2, \dots, \Psi_k\}$ se puede cumplir ese nivel de similitud para alguno de sus subconjuntos. Por ello, es necesario hacer varias pruebas para tener claro qué relación existe entre las diferentes matrices.

Nótese que, del Teorema de Descomposición de dos Matrices Definidas Semipositivas, (ver [2, Apéndice A]), se deduce que: si Ψ_1, Ψ_2 son matrices de varianza (y en consecuencia son semidefinidas positivas) existe una matriz Γ de $p \times p$ que cumple con:

$$(2.1) \quad \Gamma' \Psi_1 \Gamma = I_p$$

$$(2.2) \quad \Gamma' \Psi_2 \Gamma = \Delta = \text{diag}(\delta_1, \dots, \delta_p) \quad \delta_j > 0 \quad \forall j \in \{1, \dots, p\}$$

De donde:

$$(2.3) \quad \Psi_1^{-1} \Psi_2 \Gamma = \Gamma \Delta$$

Así, con la transformación $U = \Gamma X$ se pueden obtener simultáneamente coordenadas no correlacionadas de dos muestras. Sin embargo, dado que Γ no es ortogonal, estas transformaciones no se parecen a las de Componentes Principales Comunes.

Considérense las siguientes equivalencias de los Componentes Principales Comunes:

TEOREMA 2.1. *Sean Ψ_1 y Ψ_2 matrices simétricas de orden $p \times p$ donde Ψ_1 no es singular. Entonces, las siguientes condiciones son equivalentes:*

1. *Los vectores característicos de Ψ_1 y Ψ_2 son iguales (o se pueden escoger así si tienen raíces múltiples).*
2. *Los vectores característicos de $\Psi_1^{-1} \Psi_2$ son ortogonales (o se pueden escoger así en caso de raíces múltiples).*
3. *$\Psi_1^{-1} \Psi_2$ es simétrica.*
4. *$\Psi_1 \Psi_2 = \Psi_2 \Psi_1$*

DEMOSTRACIÓN. Se demostrará $(1) \Rightarrow (3) \Rightarrow (2) \Rightarrow (4) \Rightarrow (1)$

▪ $(1) \Rightarrow (3)$

Dado (1), por el Teorema de la Descomposición Espectral:

$$\Psi_1 = \beta \Lambda_1 \beta'$$

$$\Psi_2 = \beta \Lambda_2 \beta'$$

En ese caso:

$$\begin{aligned} \Psi_1^{-1} \Psi_2 &= (\beta \Lambda_1 \beta')^{-1} (\beta \Lambda_2 \beta') \\ &= (\beta \Lambda_1^{-1} \beta') (\beta \Lambda_2 \beta') \\ &= \beta \Lambda_1^{-1} \Lambda_2 \beta' \end{aligned}$$

Y, al transponer, como $\Lambda_1^{-1} \Lambda_2$ es simétrica se tiene que:

$$(\beta \Lambda_1^{-1} \Lambda_2 \beta')' = \beta \Lambda_1^{-1} \Lambda_2 \beta'$$

▪ $(3) \Rightarrow (2)$

Es inmediato del Teorema de la Descomposición Espectral.

- (2) \Rightarrow (4)

Dado (2), se pueden ordenar los vectores característicos de $\Psi_1^{-1}\Psi_2$ como las columnas de una matriz Γ para escribir:

$$\Psi_1^{-1}\Psi_2 = \Gamma\Delta\Gamma'$$

Con Δ diagonal.

Entonces:

$$\Psi_2\Psi_1 = \Psi_1(\Psi_1^{-1}\Psi_2)\Psi_1 = \Psi_1\Gamma\Delta\Gamma'\Psi_1$$

Y como:

$$(\Psi_1\Gamma\Delta\Gamma'\Psi_1)' = \Psi_1\Gamma\Delta\Gamma'\Psi_1$$

Se tiene que:

$$\Psi_2\Psi_1 = (\Psi_2\Psi_1)'$$

de donde se concluye:

$$\Psi_2\Psi_1 = (\Psi_2\Psi_1)' = \Psi_1'\Psi_2' = \Psi_1\Psi_2$$

- (4) \Rightarrow (1)

Por el Teorema de la Descomposición Espectral existe una matriz $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ tal que:

$$\Psi_1 = \beta\Lambda\beta'$$

Sea $T = \beta\Psi_2\beta'$. Así $\Psi_2 = \beta T\beta'$.

Sustituyendo lo anterior en la condición $\Psi_1\Psi_2 = \Psi_2\Psi_1$ se tiene que:

$$\beta\Lambda T\beta' = \beta T\Lambda\beta'$$

De lo anterior se concluye que $\Lambda T = T\Lambda$. Denotando $T = (t_{ij})$, se puede escribir:

$$\lambda_i t_{ij} = \lambda_j t_{ij} \quad \text{para } i, j = 1, \dots, p$$

Lo anterior es equivalente a:

$$(2.4) \quad (\lambda_i - \lambda_j)t_{ij} = 0$$

Así, se tiene que:

$$(2.5) \quad (t_{ij}) = 0 \quad \forall i, j \text{ tales que } \lambda_i \neq \lambda_j$$

Si los valores propios de Ψ_1 son distintos dos a dos, se tiene que $(t_{ij}) = 0 \forall i \neq j$. Lo anterior quiere decir que T es diagonal y, en consecuencia, las columnas de β son también vectores propios de Ψ_2 .

Si existen valores propios repetidos de Ψ_1 , se reordena β de manera que sus primeras r columnas sean los vectores propios correspondientes al valor propio repetido λ^* .

Se escribe $\beta = (\beta_{(1)}, \beta_{(2)})$, donde $\beta_{(1)}$ son las primeras r columnas de β y $\beta_{(2)}$ las columnas restantes.

Con la misma metodología, se parte T como:

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}$$

Donde T_{11} es de $r \times r$ y T_{22} es de $(p - r) \times (p - r)$.

Se sabe que T es simétrica, en consecuencia T_{11} es simétrica. Por el Teorema de la Descomposición Espectral, existe una matriz ortogonal H de orden $r \times r$, tal que $H'T_{11}H$ es diagonal.

Sea $\beta^* = (\beta_{(1)}H, \beta_{(2)})$. Entonces β^* es ortogonal y:

$$\beta^{*'}\Psi_1\beta^* = \begin{bmatrix} \lambda^*I_r & 0 \\ 0 & \beta'_{(2)}\Psi_1\beta_{(2)} \end{bmatrix}$$

que es diagonal.

Además:

$$\beta^{*'}\Psi_2\beta^* = \begin{bmatrix} H'T_{11}H & H'T_{12} \\ T_{21}H & T_{22} \end{bmatrix}$$

Como $\Lambda T = T\Lambda$ y Λ tiene exactamente r entradas con el valor propio λ_r^* en su diagonal, se puede concluir, de 2.5, que $T_{12} = 0$ y $T_{21} = 0$.

Si λ^* es el único valor repetido en Λ , entonces, por el argumento seguido a partir de la ecuación (2.4), aplicado a T_{22} , las columnas de β^* son vectores propios de Ψ_1 y Ψ_2 . Si no, se repite el mismo proceso hecho con β con todos los valores repetidos de Λ para encontrar los vectores propios comunes a Ψ_1 y Ψ_2 .

□

En las siguientes secciones se tratarán los métodos para estudiar los diferentes niveles de similitud entre matrices de varianzas. Es importante hacer notar que el orden que se sigue no es de mayor a menor similitud según la escala o viceversa. En realidad, el orden de los temas expuestos se escogió por la didáctica matemática, ya que, a partir de secciones ya estudiadas se facilita la comprensión de las siguientes.

2.2. Igualdad de Varianzas

Sean $\Psi_1, \Psi_2, \dots, \Psi_k$ matrices definidas positivas de k poblaciones diferentes. Se supone que los datos en cada población se distribuyen como normales multivariados y se desea probar la hipótesis de que las k matrices de varianzas son iguales, formalmente:

$$(2.6) \quad H_- : \quad \Psi_1 = \Psi_2 = \dots = \Psi_k$$

Como los datos se distribuyen normales multivariados, la función de verosimilitud de Ψ_1, \dots, Ψ_k (matrices de varianza de las poblaciones), dadas las matrices de varianza estimadas S_i , en su forma insesgada, y que las distribuciones están centradas (ver Apéndice C), es:

$$(2.7) \quad L(\Psi_1, \dots, \Psi_k) = C \times \prod_{i=1}^k \exp \left\{ \operatorname{tr} \left(-\frac{n_i}{2} \Psi_i^{-1} S_i \right) \right\} (\det \Psi_i)^{-n_i/2}$$

Donde C es una constante que no depende de Ψ_1, \dots, Ψ_k

Maximizar la función de verosimilitud es equivalente a minimizar:

$$(2.8) \quad g(\Psi_1, \dots, \Psi_k) = -2 \log L + 2 \log C$$

$$(2.9) \quad = \sum_{i=1}^k n_i [\log(\det \Psi_i) + \operatorname{tr}(\Psi_i^{-1} S_i)]$$

Así, maximizar la función de verosimilitud bajo H_0 , denotando $\Psi = \Psi_1 = \dots = \Psi_k$, es equivalente a minimizar:

$$(2.10) \quad g(\Psi, \dots, \Psi) = \sum_{i=1}^k n_i [\log(\det \Psi) + \operatorname{tr}(\Psi^{-1} S_i)]$$

$$(2.11) \quad = n [\log(\det \Psi) + \operatorname{tr}(\Psi^{-1} S)]$$

donde $n = n_1 + n_2 + \dots + n_k$ y $S = \frac{\sum_{i=1}^k n_i S_i}{n}$.

Para minimizar 2.11 se debe considerar la función:

$$f(\Psi) = \log(\det \Psi) + \operatorname{tr}(\Psi^{-1} S)$$

que se desea minimizar en el espacio de las matrices definidas positivas.

Minimizar la función anterior es equivalente a minimizar $f(\Psi) - f(S)$, pues éstas sólo difieren por una constante. Supóngase que $\lambda_1, \dots, \lambda_k$ son los valores propios de $\Psi^{-1} S$, entonces:

$$\begin{aligned} f(\Psi) - f(S) &= \log(\det \Psi) + \operatorname{tr}(\Psi^{-1} S) - \log(\det S) - \operatorname{tr}(I_p) \\ &= -\log[\det(\Psi^{-1} S)] + \operatorname{tr}(\Psi^{-1} S) - p \\ &= -\log \left(\prod_{i=1}^k \lambda_i \right) + \sum_{i=1}^k \lambda_i - p \\ &= \sum_{i=1}^p (-\log \lambda_i + \lambda_i - 1) \end{aligned}$$

Como $\log(x) \leq x - 1$ para todo $x > 0$ y la igualdad sólo ocurre cuando $x = 1$; el mínimo de $f(\Psi) - f(S)$ se da cuando $\lambda_i = 1$ para todo $i \in \{1, \dots, k\}$. Lo anterior es equivalente a que $\Psi = S$. Así, el estimador de máxima verosimilitud de Ψ es S , que en adelante se denominará como la mejor matriz común de varianza para las poblaciones $1, 2, \dots, k$.

El estadístico dado por el logaritmo del cociente de verosimilitudes para probar

$$H_{=} \text{ vs. } H_a: \Psi_1, \dots, \Psi_k \text{ no tienen ninguna semejanza}$$

es:

$$(2.12) \quad T_{=} = -2 \log \frac{L(S, \dots, S)}{L(S_1, \dots, S_k)}$$

$$(2.13) \quad = n \log(\det S) - \sum_{i=1}^k n_i \log(\det S_i)$$

Los grados de libertad al estimar k matrices de varianza, una para cada población, son $kp(p-1)/2$, mientras que los grados de libertad al estimar una sólo matriz de varianza, común a las k poblaciones, son $p(p-1)/2$, por lo que la diferencia en los grados de libertad entre k matrices diferentes y una sólo son $(k-1)p(p-1)/2$. Así, según la Teoría de Pruebas de Cocientes de Verosimilitudes, (ver [9, págs. 418,419]), $T_{=}$ se distribuye bajo $H_{=}$ como una ji-cuadrada con $(k-1)p(p-1)/2$ grados de libertad.

2.3. Componentes Principales Comunes

Sean Ψ_1, \dots, Ψ_k matrices definidas positivas de varianza de k poblaciones diferentes. Se define la hipótesis de Componentes Principales Comunes como:

$$(2.14) \quad H_{CPC} : \quad \Psi_i = \beta \Lambda_i \beta' \quad \forall i \in \{1, \dots, k\}$$

Donde $\Lambda_i = \text{diag}(\lambda_{i1}, \dots, \lambda_{ip})$

Suponiendo que las muestras provienen de distribuciones normales multivariadas y que son de tamaño n_i , si S_i es la matriz estimada de varianza de la muestra i , se tiene que la función de verosimilitud de Ψ_1, \dots, Ψ_k dados S_1, \dots, S_k , es:

$$(2.15) \quad L(\Psi_1, \dots, \Psi_k) = C \times \prod_{i=1}^k \exp \left\{ \text{tr} \left(-\frac{n_i}{2} \Psi_i^{-1} S_i \right) \right\} (\det \Psi_i)^{-\frac{n_i}{2}}$$

Donde C no depende de Ψ_1, \dots, Ψ_k

Definiendo nuevamente:

$$(2.16) \quad g(\Psi_1, \dots, \Psi_k) = -2 \log L + 2 \log C$$

se tiene que, maximizar L es equivalente a minimizar g , siendo esto último más sencillo, pues:

$$(2.17) \quad g(\Psi_1, \dots, \Psi_k) = \sum_{i=1}^k n_i [\log(\det \Psi_i) + \text{tr}(\Psi_i^{-1} S_i)]$$

Suponiendo cierta H_{CPC} y escribiendo $\beta = (\beta_1, \dots, \beta_p)$, se tiene que:

$$(2.18) \quad \det(\Psi_i) = \det(\beta \Lambda_i \beta') = \det(\Lambda_i) = \prod_{j=1}^p \lambda_{ij}$$

$$(2.19) \quad \begin{aligned} \text{tr}(\Psi_i^{-1} S_i) &= \text{tr}((\beta \Lambda_i \beta')^{-1} S_i) = \text{tr}((\beta \Lambda_i^{-1} \beta') S_i) = \text{tr}(\Lambda_i^{-1} \beta' S_i \beta) \\ &= \sum_{j=1}^p \frac{[\beta' S_i \beta]_{jj}}{\lambda_{ij}} = \sum_{j=1}^p \frac{\beta'_j S_i \beta_j}{\lambda_{ij}} \end{aligned}$$

De donde:

$$(2.20) \quad g(\beta, \Lambda_1, \dots, \Lambda_k) = \sum_{i=1}^k n_i \sum_{j=1}^p \left(\log \lambda_{ij} + \frac{\beta'_j S_i \beta_j}{\lambda_{ij}} \right)$$

Así, para minimizar g sujeta a las restricciones de ortonormalidad de β se utilizan los multiplicadores de Lagrange. Para ello se define la función:

$$(2.21) \quad g^*(\beta, \Lambda_1, \dots, \Lambda_k) = g(\beta, \Lambda_1, \dots, \Lambda_k) - \sum_{j=1}^p \delta_j (\beta'_j \beta_j - 1) - 2 \sum_{h < k} \delta_{hj} \beta'_h \beta_j$$

Derivando:

$$(2.22) \quad \frac{\partial g^*}{\partial \lambda_{im}} = n_i \left[\frac{1}{\lambda_{im}} - \frac{\beta'_j S_i \beta_m}{\lambda_{im}^2} \right] = n_i \left[\frac{\lambda_{im} - \beta'_m S_i \beta_m}{\lambda_{im}^2} \right]$$

Igualando a cero, se obtiene que:

$$(2.23) \quad \lambda_{im} = \beta'_m S_i \beta_m \quad \forall i \in \{1, \dots, k\} \quad m \in \{1, \dots, p\}$$

Calculando la derivada parcial con respecto a β_m (ver Apéndice E) e igualando a cero se obtiene que:

$$(2.24) \quad \sum_{i=1}^k \frac{n_i S_i \beta_m}{\lambda_{im}} - \sum_{h=1, h \neq m}^p \delta_{mh} \beta_h - \delta_m \beta_m = 0 \quad \forall m \in \{1, \dots, p\}$$

Donde para $m > h$ se define $\delta_{mh} = \delta_{hm}$.

Así, multiplicando por β'_j la ecuación 2.24, para $j \neq m$, se obtiene:

$$(2.25) \quad \sum_{i=1}^k \frac{n_i \beta'_j S_i \beta_m}{\lambda_{im}} = \delta_{mj}$$

Análogamente:

$$(2.26) \quad \sum_{i=1}^k \frac{n_i \beta'_m S_i \beta_j}{\lambda_{ij}} = \delta_{jm}$$

Como S_i es simétrica, se tiene que $\beta'_j S_i \beta_m = \beta'_m S_i \beta_j$. Substituyendo lo anterior en las ecuaciones 2.25 y 2.26 y despejando, se obtiene la ecuación base de los Componentes Principales Comunes:

$$(2.27) \quad \beta'_m \left(\sum_{i=1}^k n_i \frac{\lambda_{im} - \lambda_{ij}}{\lambda_{im} \lambda_{ij}} S_i \right) \beta_j = 0 \quad \forall m \neq j; m, j \in \{1, \dots, p\}$$

La ecuación anterior se debe resolver sujeta a las restricciones:

$$(2.28) \quad \beta'_m \beta_j = \begin{cases} 0, & m \neq j \\ 1, & m = j \end{cases}$$

Para resolver lo anterior existe un algoritmo propuesto por Flury y Gatuchi conocido como el algoritmo FG (ver [2]).

Dicho algoritmo provee un máximo local de 2.20, sin embargo, no se tiene la certeza de que ese máximo local sea un máximo global de 2.20.

Suponiendo que los estimadores máximo-verosímiles son únicos, se denotarán por $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$ y $\hat{\Lambda}_i = \text{diag}(\hat{\lambda}_{i1}, \dots, \hat{\lambda}_{ip})$. Así, los estimadores máximo verosímiles de Ψ_i bajo H_{CPC} son:

$$(2.29) \quad \hat{\Psi}_i = \hat{\beta}' \hat{\Lambda}_i \hat{\beta}' \quad \text{para } i = 1, \dots, k$$

Además de las ecuaciones 2.23 y 2.19 se tiene que:

$$(2.30) \quad \text{tr}(\Psi_i^{-1} S_i) = p$$

Por lo que el máximo valor de la ecuación de verosimilitud restringida a H_{CPC} es:

$$(2.31) \quad L(\hat{\Psi}_1, \dots, \hat{\Psi}_k) = C \times \prod_{i=1}^k \exp \left\{ -\frac{pn_i}{2} \right\} (\det(\hat{\Psi}_i))^{-\frac{n_i}{2}}$$

Mientras que el máximo global de la ecuación de verosimilitud es:

$$(2.32) \quad L(S_1, \dots, S_k) = C \times \prod_{i=1}^k \exp \left\{ -\frac{pn_i}{2} \right\} (\det(S_i))^{-\frac{n_i}{2}}$$

Así, el estadístico del cociente de logverosimilitud para probar H_{CPC} es:

$$(2.33) \quad T_{CPC} = -2 \log \frac{L(\hat{\Psi}_1, \dots, \hat{\Psi}_k)}{L(S_1, \dots, S_k)} = \sum_{i=1}^k n_i \log \left(\frac{\det \hat{\Psi}_i}{\det S_i} \right)$$

Y, de acuerdo con la Teoría de Pruebas de Cocientes de Verosimilitudes (ver [9, págs. 418,419]), T_{CPC} se distribuye asintóticamente como una ji-cuadrada con $(k-1)p(p-1)/2$ grados de libertad.

Usando los estimadores máximo verosímiles para $\hat{\beta}$, podemos definir los componentes principales comunes muestrales como:

$$(2.34) \quad U = \hat{\beta}' X$$

Entonces, la matriz de varianza de U en el grupo i es:

$$(2.35) \quad F_i = \hat{\beta}' S_i \hat{\beta}$$

Y se cumple:

$$(2.36) \quad \hat{\Lambda}_i = \text{diag}(F_i)$$

De lo anterior se puede reescribir el estadístico del cociente de verosimilitudes como:

$$(2.37) \quad T_{CPC} = \sum_{i=1}^k n_i \log \left(\frac{\det(\text{diag } F_i)}{\det F_i} \right)$$

En la ecuación anterior es claro que ese estadístico es una medida de la diagonalizabilidad simultánea de las matrices S_i .

La matriz de correlación de la U se puede calcular como:

$$(2.38) \quad R_i = \hat{\Lambda}_i^{(-\frac{1}{2})} F_i \hat{\Lambda}_i^{(-\frac{1}{2})} \quad \text{para } i = 1, \dots, k$$

(Ver Apéndice F para la definición de $\hat{\Lambda}_i^{(-\frac{1}{2})}$)

Bajo H_{CPC} , esta matriz tiende a ser I_p , por lo que calcularla puede ayudar a comprender, cuando no se cumple H_{CPC} , en qué variables y/o en qué muestras se encuentra el problema.

2.4. Matrices de varianza proporcionales

Sean Ψ_1, \dots, Ψ_k matrices de varianza definidas positivas. Se desea probar la hipótesis de que las k matrices de varianza son proporcionales, formalmente:

$$(2.39) \quad H_{PROP} : \quad \Psi_i = \rho_i \Psi_1 \quad \forall i \in \{2, 3, \dots, k\}$$

donde ρ_i son constantes positivas.

Utilizando el Teorema de la Descomposición Espectral:

$$(2.40) \quad \Psi_i = \beta \Lambda_i \beta'$$

Donde $\Lambda_i = \text{diag}(\lambda_{i1}, \dots, \lambda_{ip})$.

En ese caso, suponiendo cierta H_{PROP} , se tiene que:

$$(2.41) \quad \Lambda_i = \rho_i \Lambda_1 \quad \forall i \in \{2, \dots, k\}$$

O equivalentemente:

$$(2.42) \quad \lambda_{ij} = \rho_i \lambda_j$$

Es claro, que la matriz a la que se le asigna la constante 1, Ψ_1 , es elegida arbitrariamente; pero esto no afecta al modelo.

La manera como se planteó el problema en las ecuaciones 2.41 y 2.42 permite aprovechar los resultados de la sección de Componentes Principales Comunes para derivar las pruebas de hipótesis y los estimadores de máxima verosimilitud.

Nuevamente se supone una distribución normal multivariada, por lo que se conserva la misma función de verosimilitud.

Para maximizarla, se aprovecharán los resultados de la sección anterior. Sustituyendo 2.42 en la ecuación 2.20 se obtiene que: encontrar los estimadores de máxima verosimilitud, es equivalente a minimizar:

$$(2.43) \quad g(\beta, \Lambda, \rho_1, \dots, \rho_k) = \sum_{i=1}^k n_i \sum_{j=1}^p \left[\log(\rho_i \lambda_j) + \frac{\beta'_j S_i \beta_j}{\rho_i \lambda_j} \right]$$

con las restricciones $\beta' \beta = I_p$ y $\rho_1 = 1$.

Utilizando la ecuación 2.22 de la sección anterior y la Regla de la Cadena se calculan las derivadas parciales con respecto a las ρ_i :

$$(2.44) \quad \begin{aligned} \frac{\partial g}{\partial \rho_i} &= \sum_{j=1}^p \frac{\partial g}{\partial \lambda_{ij}} \cdot \frac{\partial \lambda_{ij}}{\partial \rho_i} \\ &= n_i \sum_{j=1}^p \lambda_j \left(\frac{1}{\lambda_{ij}} - \frac{\beta'_j S_i \beta_j}{\lambda_{ij}^2} \right) \quad \forall i \in \{2, \dots, k\} \end{aligned}$$

Igualando a cero y sustituyendo $\lambda_{ij} = \rho_i \lambda_j$, se obtiene:

$$(2.45) \quad \rho_i = \frac{1}{p} \sum_{j=1}^p \frac{\beta'_j S_i \beta_j}{\lambda_j} \quad \text{para } i = 2, \dots, k$$

Análogamente:

$$(2.46) \quad \begin{aligned} \frac{\partial g}{\partial \lambda_j} &= \sum_{i=1}^k \frac{\partial g}{\partial \lambda_{ij}} \cdot \frac{\partial \lambda_{ij}}{\partial \lambda_j} \\ &= \sum_{i=1}^k n_i \rho_i \left(\frac{1}{\lambda_{ij}} - \frac{\beta'_j S_i \beta_j}{\lambda_{ij}^2} \right) \end{aligned}$$

Igualando a cero, sustituyendo $\lambda_{ij} = \rho_i \lambda_j$ y denotando $n = n_1 + \dots + n_k$, se obtiene:

$$(2.47) \quad \lambda_j = \frac{1}{n} \sum_{i=1}^k \frac{n_i}{\rho_i} \beta'_j S_i \beta_j \quad j = 1, \dots, p$$

La ecuación 2.27, que es la que toma en cuenta que:

$$\frac{\partial g}{\partial \beta_j} = 0$$

se convierte en:

$$(2.48) \quad \beta'_m \left(\sum_{i=1}^k n_i \frac{\rho_i \lambda_m - \rho_i \lambda_j}{\rho_i^2 \lambda_m \lambda_j} S_i \right) \beta_j = 0 \quad m \neq j$$

Lo anterior se simplifica, quedando como:

$$(2.49) \quad \left(\frac{1}{\lambda_j} - \frac{1}{\lambda_m} \right) \beta'_m \left(\sum_{i=1}^k \frac{n_i}{\rho_i} S_i \right) \beta_j = 0 \quad m \neq j$$

Cuando las raíces λ_j son distintas se pueden resolver las ecuaciones 2.45, 2.47 y 2.49 bajo las restricciones $\beta' \beta = I_p$ y $\rho_1 = 1$, a través de un algoritmo iterativo ([2, sección 9]).

Denotando a los estimadores de máxima verosimilitud por: $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$, $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_p)$ y $\hat{\rho}_i$, respectivamente, los estimadores de máxima verosimilitud restringidos a H_{PROP} son:

$$(2.50) \quad \hat{\Psi}_i = \hat{\rho}_i \hat{\beta} \hat{\Lambda} \hat{\beta}' \quad i = 1, \dots, k$$

Así, el estadístico para la prueba de hipótesis:

$$H_{PROP} \text{ vs. } H_a : \Psi_i \text{ no tienen ninguna relación}$$

es:

$$(2.51) \quad T_{PROP} = -2 \log \frac{L(\hat{\Psi}_1, \dots, \hat{\Psi}_k)}{L(S_1, \dots, S_k)}$$

$$(2.52) \quad = \sum_{i=1}^k n_i \log \frac{\det \hat{\Psi}_i}{\det S_i}$$

$$(2.53) \quad = n \left\{ \sum_{j=1}^p \log \hat{\lambda}_j + \sum_{i=1}^k r_i [p \log \hat{\rho}_i - \log(\det S_i)] \right\}$$

Lo anterior se despejó de acuerdo a 2.40 y 2.42; y se definió:

$$r_i = \frac{n_i}{n}$$

Según la Teoría General de Pruebas de Cocientes de Verosimilitudes, la distribución asintótica de T_{PROP} es ji-cuadrada con $(k-1)(p^2 + p - 2)/2$ grados de libertad, siempre y cuando las r_i no sean cercanas a cero ó a uno.

Cuando se sospecha que las raíces λ_j de Ψ no son distintas, suponiendo que hay esfericidad en r componentes (en notación matemática $H_s : \lambda_{p-r+1} = \dots = \lambda_p$) se puede aplicar el siguiente modelo de estimación de los parámetros y prueba de hipótesis:

Sea λ^* la raíz de multiplicidad r , de 2.43, se tiene que la ecuación de verosimilitud restringida a H_s y H_{PROP} es:

$$(2.54) \quad g(\beta, \Lambda, \rho_1, \dots, \rho_k) = \sum_{i=1}^k \left[\sum_{j=1}^{p-r} \left((\log \rho_i \lambda_j) + \frac{\beta'_j S_i \beta_j}{\rho_i \lambda_j} \right) + \sum_{j=p-r+1}^p \left(\log(\rho_i \lambda^*) + \frac{\beta'_j S_i \beta_j}{\rho_i \lambda^*} \right) \right]$$

Siguiendo el mismo proceso de los multiplicadores de Lagrange, se tiene que, para encontrar los estimadores de máxima verosimilitud, es necesario resolver el siguiente sistema de ecuaciones:

$$(2.55) \quad \rho_i = \frac{1}{p} \left[\sum_{j=1}^{p-r} \frac{\beta'_j S_i \beta_j}{\lambda_j} + \frac{1}{\lambda^*} \sum_{j=p-r+1}^p \beta'_j S_i \beta_j \right] \quad i = 2, \dots, k$$

$$(2.56) \quad \lambda_j = \frac{1}{n} \sum_{i=1}^k \frac{n_i}{\rho_i} \beta'_j S_i \beta_j \quad j = 1, \dots, p-r$$

$$(2.57) \quad \lambda^* = \frac{1}{n} \sum_{i=1}^k \frac{n_i}{\rho_i} \left(\frac{1}{r} \sum_{j=p-r+1}^p \beta'_j S_i \beta_j \right) \text{ donde } \rho_1 = 1$$

$$(2.58) \quad \left(\frac{1}{\lambda_j} - \frac{1}{\lambda_m} \right) \beta'_m \left(\sum_{i=1}^k \frac{n_i}{\rho_i} S_i \right) \beta_j = 0 \quad 1 \leq j < m \leq p-r$$

$$(2.59) \quad \left(\frac{1}{\lambda_j} - \frac{1}{\lambda^*} \right) \beta'_m \left(\sum_{i=1}^k \frac{n_i}{\rho_i} S_i \right) \beta_j = 0 \quad 1 \leq j \leq p-r < m \leq p$$

$$(2.60) \quad \beta' \beta = I_p$$

En las ecuaciones anteriores, la solución para $\beta_{p-r+1}, \dots, \beta_p$ no es única. Sin embargo, si se denota una solución de ese sistema de ecuaciones como $\bar{\beta}, \bar{\lambda}_j, \bar{\lambda}^*, \bar{\rho}_i$ para las diferentes incógnitas respectivamente, el estimador de máxima verosimilitud de Ψ_i es:

$$(2.61) \quad \bar{\Psi}_i = \bar{\rho}_i \bar{\beta} \bar{\Lambda} \bar{\beta}' \quad \forall i \in \{1, \dots, k\}$$

Donde $\Lambda = \text{diag}(\bar{\lambda}_1, \dots, \bar{\lambda}_{p-r}, \bar{\lambda}^*, \dots, \bar{\lambda}^*)$.

Para la prueba de hipótesis:

H_s vs. H_a : Hay p raíces distintas de Ψ (pero se cumple H_{PROP})

se define el siguiente estadístico de logverosimilitud:

$$(2.62) \quad T_S = -2 \log \left[\frac{L(\bar{\Psi}_1, \dots, \bar{\Psi}_k)}{L(\hat{\Psi}_1, \dots, \hat{\Psi}_k)} \right]$$

$$(2.63) \quad = \sum_{i=1}^k n_i \log \left[\frac{\det \bar{\Psi}_i}{\det \hat{\Psi}_i} \right]$$

$$(2.64) \quad = \sum_{i=1}^k n_i \left[p \log \frac{\bar{\rho}_i}{\hat{\rho}_i} + \log \frac{(\bar{\lambda}^*)^r \prod_{j=1}^{p-r} \bar{\lambda}_j}{\prod_{j=1}^p \hat{\lambda}_j} \right]$$

Se considera, por lo general, que el estadístico anterior se distribuye como una ji-cuadrada con $r(r+1)/2 - 1$ grados de libertad, pero no se ha encontrado alguna prueba formal de ello.

2.5. Componentes Principales Comunes Parciales

Cuando se infiere que no todos los componentes de las matrices son iguales, pero que algunos de ellos podrían serlo, se puede utilizar el siguiente modelo parcial de Componentes Principales Comunes: su hipótesis es que, de las p componentes principales de cada matriz, q coinciden y los otros $p - q$ son específicos de cada matriz.

Formalizando:

Sean Ψ_1, \dots, Ψ_k matrices de varianza definidas positivas, se define la hipótesis de que existen q componentes principales comunes a ellas como:

$$(2.65) \quad H_{CPC}(q) : \quad \Psi_i = \beta^{(i)} \Lambda_i \beta^{(i)'} \quad \forall i \in \{1, \dots, k\}$$

donde:

$$(2.66) \quad \Lambda_i = \text{diag}(\lambda_{i1} \dots, \lambda_{ip})$$

$$(2.67) \quad \beta^{(i)} = (\beta_c, \beta_s^{(i)})$$

con β_c , una matriz fija de $p \times q$; $\beta_s^{(i)}$ matrices de dimensión $p \times (p - q)$; tales que las $\beta^{(i)}$ sean matrices de $p \times p$ ortogonales.

De la ortogonalidad se concluye que $H_{CPC}(p - 1)$ implica H_{CPC} . Por lo anterior, el rango relevante de q es $1 \leq q \leq p - 2$. Ello implica que este modelo es interesante cuando la dimensión de cada observación es, al menos, tres.

Se separan las β 's por columnas de la siguiente manera:

$$(2.68) \quad \beta_c = (\beta_1, \dots, \beta_q)$$

$$(2.69) \quad \beta_s^{(i)} = (\beta_{q+1}^{(i)}, \dots, \beta_p^{(i)})$$

Suponiendo nuevamente una distribución normal multivariada, se hace algo similar a lo que se hizo con la ecuación 2.15. Maximizar la ecuación de verosimilitud restringida a $H_{CPC}(q)$ es equivalente a minimizar:

$$(2.70) \quad g(\beta_c, \beta_s^{(1)}, \dots, \beta_s^{(k)}, \Lambda_1, \dots, \Lambda_k) = \sum_{i=1}^k n_i \left[\sum_{j=1}^p \log \lambda_{ij} + \sum_{j=1}^q \frac{\beta_j' S_i \beta_j}{\lambda_{ij}} + \sum_{j=q+1}^p \frac{\beta_j^{(i)'} S_i \beta_j^{(i)}}{\lambda_{ij}} \right]$$

La función anterior se debe minimizar sujeta a las condiciones de ortonormalidad de las $\beta^{(i)}$ que se pueden escribir como:

$$(2.71) \quad \beta_h' \beta_j = \begin{cases} 0 & \text{si } h \neq j \\ 1 & \text{si } h = j \end{cases} \quad 1 \leq h, j \leq q$$

$$(2.72) \quad \beta_h^{(i)'} \beta_j^{(i)} = \begin{cases} 0 & \text{si } h \neq j \\ 1 & \text{si } h = j \end{cases} \quad q \leq h, j \leq p; \quad i = 1, \dots, k$$

$$(2.73) \quad \beta_h' \beta_j^{(i)} = 0 \quad 1 \leq h \leq q < j \leq p \quad i = 1, \dots, k$$

Utilizando los multiplicadores de Lagrange, se llega a que, para resolver el problema, se debe encontrar el mínimo de la siguiente función:

$$(2.74) \quad G = g - \sum_{h=1}^q \gamma_h (\beta'_h \beta_h - 1) - 2 \sum_{1 \leq h < j \leq q} \gamma_{hj} \beta'_h \beta_j \\ - \sum_{i=1}^k \left[\sum_{h=q+1}^p \gamma_h^{(i)} (\beta_h^{(i)'} \beta_h^{(i)} - 1) + 2 \sum_{q < h < j \leq p} \gamma_{hj}^{(i)} \beta_h^{(i)'} \beta_j^{(i)} + \sum_{1 \leq h \leq q < j \leq p} \delta_{hj}^{(i)} \beta'_h \beta_j^{(i)} \right]$$

Calculando las parciales de las λ_{ij} e igualándolas a cero, se obtiene que:

$$(2.75) \quad \lambda_{ij} = \begin{cases} \beta'_j S_i \beta_j & \text{para } j = 1, \dots, q \\ \beta_j^{(i)'} S_i \beta_j^{(i)} & \text{para } j = q+1, \dots, p \end{cases} \quad i = 1, \dots, k$$

Por lo que se obtiene un resultado análogo al de 2.30:

$$\text{tr}(\Psi^{-1} S_i) = p \quad \forall i \in \{1, \dots, k\}$$

La ecuación $\frac{\partial G}{\partial \beta_j^{(i)}} = 0$ es equivalente, para cada $\beta_j^{(i)}$, a:

$$(2.76) \quad \frac{n_i S_i \beta_j^{(i)}}{\lambda_{ij}} - \gamma_j^{(i)} \beta_j^{(i)} - \sum_{h=q+1, h \neq j}^p \gamma_{hj}^{(i)} \beta_h^{(i)} - \sum_{h=1}^q \delta_{hj}^{(i)} \beta_h = 0$$

donde se define $\gamma_{hj}^{(i)} = \gamma_{jh}^{(i)}$ para $j < h$

Al multiplicar la ecuación anterior por la izquierda por β'_m con ($1 \leq m \leq q$) se obtiene:

$$(2.77) \quad \delta_{mh}^{(i)} = \frac{n_i \beta'_m S_i \beta_j^{(i)}}{\lambda_{ij}}$$

Si ahora se multiplica la misma ecuación por $\beta_h^{(i)'}$ por la izquierda, con $h \neq j$, se concluye:

$$(2.78) \quad \gamma_{hj}^{(i)} = \frac{n_i \beta_h^{(i)'} S_i \beta_j^{(i)}}{\lambda_{ij}}$$

Análogamente:

$$(2.79) \quad \gamma_{jh}^{(i)} = \frac{n_i \beta_j^{(i)'} S_i \beta_h^{(i)}}{\lambda_{ih}}$$

Y como S_i es simétrica, se tiene que:

$$(2.80) \quad \beta_h^{(i)'} S_i \beta_j^{(i)} = \beta_j^{(i)'} S_i \beta_h^{(i)}$$

De las tres ecuaciones anteriores se tiene que:

$$(2.81) \quad \beta_h^{(i)'} S_i \beta_j^{(i)} = 0 \quad \text{cuando } \lambda_{ij} \neq \lambda_{ih}$$

Si se calcula la parcial de G con respecto a las β_m y se iguala a cero, para $m \in \{1, \dots, q\}$ se obtiene:

$$(2.82) \quad \sum_{i=1}^k \frac{n_i S_i \beta_m}{\lambda_{im}} - \gamma_m \beta_m - \sum_{h=1, h \neq m}^q \gamma_{mh} \beta_h - \sum_{i=1}^k \sum_{h=q+1}^p \delta_{mh}^{(i)} \beta_h^{(i)} = 0$$

Donde nuevamente se define $\gamma_{hj} = \gamma_{jh}$ para $j < h$. Multiplicando por β'_h con $h \neq m$ se obtiene:

$$(2.83) \quad \gamma_{mh} = \sum_{i=1}^k \frac{n_i \beta'_h S_i \beta_m}{\lambda_{im}}$$

Análogamente:

$$(2.84) \quad \gamma_{hm} = \sum_{i=1}^k \frac{n_i \beta'_m S_i \beta_h}{\lambda_{ih}}$$

De las dos ecuaciones anteriores y el hecho de que: $\beta'_m S_i \beta_h = \beta'_h S_i \beta_m$ se tiene que:

$$(2.85) \quad \beta'_m \left(\sum_{i=1}^k \frac{\lambda_{im} - \lambda_{ih}}{\lambda_{im} \lambda_{ih}} n_i S_i \right) \beta_h = 0 \quad \text{para } 1 \leq m, h \leq q; \quad m \neq h$$

Multiplicando 2.82 por $\beta_j^{(r)}$ y sustituyendo $\delta_{mj}^{(r)}$ usando 2.77 se obtiene:

$$(2.86) \quad \left(\frac{1}{\lambda_{rj}} - \frac{1}{\lambda_{rm}} \right) n_r \beta'_m S_r \beta_j^{(r)} = \beta_j^{(r)'} \left[\sum_{i=1, i \neq r}^k \left(\frac{n_i S_i \beta_m}{\lambda_{im}} - \sum_{h=q+1}^p \delta_{mh}^{(i)} \beta_h^{(i)} \right) \right]$$

para $r = 1, \dots, k \quad 1 \leq m \leq q < j \leq p$

Del proceso anterior se concluye que, para resolver el modelo de Componentes Principales Comunes Parciales, se debe encontrar la solución de 2.75, 2.77, 2.81, 2.85 y 2.86, bajo las restricciones de ortonormalidad de las $\beta^{(i)}$ mencionadas arriba. Esto requiere de algoritmos mucho más complicados que los que se necesitan para resolver el modelo general de los Componentes Principales Comunes. Algunos de ellos se explican en [2].

Otra opción es aproximar la solución al modelo parcial de los Componentes Principales Comunes calculando la solución del modelo general de Componentes Principales Comunes y aprovechando el siguiente lema (demostrado en [2, sección 6.6]).

LEMA 2.1. Sean S_1, \dots, S_k matrices simétricas definidas positivas de orden $p \times p$ y sean a_1, \dots, a_q vectores propios comunes a S_1, \dots, S_k , entonces la matriz β de máxima verosimilitud restringida a H_{CPC} , dadas las S_i , contiene a a_1, \dots, a_q como columnas (o puede ser elegida de manera que las contenga si no es única).

Cuando las matrices de varianza, Ψ_1, \dots, Ψ_k , comparten q vectores propios, no necesariamente sus matrices estimadas, S_1, \dots, S_k , también los comparten. Pero si las muestras son grandes, por las Leyes de los Grandes Números, se esperaría que S_1, \dots, S_q tengan

q vectores propios muy parecidos. En consecuencia, dado que las ecuaciones de máxima verosimilitud restringidas a H_{CPC} son continuas, los estimados de β van a tener aproximaciones de los q vectores propios que comparten las matrices de varianza.

Así, se pueden encontrar aproximaciones de las $\beta^{(i)}$ de la siguiente manera:

1. Calcular $B = (B_1 : B_2)$, donde B es el estimador de máxima verosimilitud de β bajo H_{CPC} , y B está ordenado de tal manera que B_1 son las q columnas que se creen vectores propios comunes a S_1, \dots, S_k .
2. Encontrar matrices Q_i tales que $Q_i' B_2' S_i B_2 Q_i$ sea diagonal y definir $B_2^{(i)} := B_2 Q_i$ para $i = 1, \dots, k$.
3. Definir los aproximados de los estimadores de máxima verosimilitud como:

$$(2.87) \quad \bar{\beta}^{(i)} := (B_1; B_2^{(i)})$$

$$(2.88) \quad \bar{\Lambda}_i := \text{diag}(\bar{\beta}^{(i)'} S_i \bar{\beta}^{(i)})$$

$$(2.89) \quad \bar{\lambda}_{ij} := (\bar{\Lambda}_i)_{jj}$$

$$(2.90) \quad \bar{\Psi}_i := \bar{\beta}^{(i)} \bar{\Lambda}_i \bar{\beta}^{(i)} \quad i = 1, \dots, k$$

Denotando $\hat{\Psi}_1, \dots, \hat{\Psi}_k$ a los estimadores de máxima verosimilitud de las varianzas bajo $H_{CPC}(q)$, es posible construir el estadístico que se muestra en la ecuación 2.91 para la prueba de hipótesis:

$H_{CPC}(q)$ vs. H_a : las matrices de varianza no están relacionadas

como:

$$(2.91) \quad T_{CPC}(q) = -2 \log \frac{L(\hat{\Psi}_1, \dots, \hat{\Psi}_k)}{L(S_1, \dots, S_k)} = \sum_{i=1}^k n_i \log \frac{\det \hat{\Psi}_i}{\det S_i}$$

La distribución asintótica del estadístico anterior es ji-cuadrada con $(k-1)q(2p-q-1)/2$ grados de libertad.

Si no se cuenta con los estimadores de máxima verosimilitud exactos, sino tan sólo con sus aproximados, se utiliza el estadístico de prueba aproximado:

$$(2.92) \quad \bar{T}_{CPC}(q) = \sum_{i=1}^k n_i \log \frac{\det \bar{\Psi}_i}{\det S_i}$$

Este estadístico siempre es mayor al exacto, porque no utiliza los estimadores exactos de máxima verosimilitud, por lo que estrictamente, se puede utilizar para decidir que no se tiene evidencia suficiente para rechazar $H_{CPC}(q)$, pero no para rechazarla.

2.5.1. Análisis de Subespacios Comunes. Hasta ahora el enfoque que se ha seguido para comparar las matrices de varianza es utilizar su descomposición espectral para ver si comparten vectores propios. Esto es equivalente a buscar que componentes principales de varias matrices de varianza diferentes generen subespacios de dimensión uno iguales. Ahora se generalizará esta visión: no sólo se buscará que los componentes

principales de varias matrices de varianza compartan subespacios de dimensión uno, sino que compartan subespacios de dimensión q . Esto es: si Ψ_1, \dots, Ψ_k son matrices de varianza, se buscará en cada matriz un grupo de q vectores propios, tales que los subespacios generados por estos grupos sean iguales.

Para formalizar la idea anterior, se define la hipótesis de un subespacio común de dimensión q de la siguiente manera:

DEFINICIÓN 2.1. Sean Ψ_1, \dots, Ψ_k matrices de varianza, y sean sus descomposiciones espectrales $\Psi_i = \beta^{(i)} \Lambda_i \beta^{(i) \prime}$ para cada $i \in \{1, \dots, k\}$. Pudiendo reordenar las columnas de las $\beta^{(i)}$ como mejor convenga, se escriben $\beta^{(i)} = (\beta_{s1}^{(i)}, \beta_{s2}^{(i)})$, donde $\beta_{s1}^{(i)}$ es una matriz de orden $p \times q$ y $\beta_{s2}^{(i)}$ es de orden $p \times (q - p)$ se define la hipótesis de un subespacio común de dimensión q como:

$$(2.93) \quad H_{SC}(q) : \quad \beta_{s1}^{(i) \prime} \beta_{s2}^{(m)} = 0 \quad \forall i, m \in \{1, 2, \dots, k\}$$

La definición anterior es equivalente a decir que existen para $i \in \{1, \dots, k\}$, matrices ortogonales $H_1^{(i)}$ y $H_2^{(i)}$ de tamaños $q \times q$ y $(p - q) \times (p - q)$ respectivamente, tales que:

$$(2.94) \quad \beta_{s1}^{(i)} = \beta_{s1}^{(1)} H_1^{(i)} \text{ y } \beta_{s2}^{(i)} = \beta_{s2}^{(1)} H_2^{(i)} \quad \text{para todo } i \in \{1, \dots, k\}$$

Es importante hacer notar que, el hecho de que esta hipótesis sea una generalización de los componentes principales comunes, hace que $H_{SC}(1)$ implique $H_{CPC}(1)$ y que, por la ortogonalidad de los componentes principales, $H_{SC}(p - 1)$ implique $H_{CPC}(1)$. Por ello, el modelo es relevante sólo cuando $p \geq 4$, pues cuando $p \leq 3$ el modelo siempre coincide con algún modelo parcial de Componentes Principales Comunes.

Para encontrar subespacios comunes es posible asumir nuevamente que los datos se distribuyen normal multivariado y aplicar el método de la máxima verosimilitud. Sin embargo, las ecuaciones resultantes son aún más complicadas de resolver que las que se necesitan resolver para el caso $H_{CPC}(q)$; por lo que se prefiere mostrar otro enfoque, que para muestras grandes, da una aproximación del resultado bastante buena.

LEMA 2.2. Sean S_1, \dots, S_k matrices definidas positivas y simétricas de orden $p \times p$, con q vectores propios en cada una que generan un mismo subespacio \mathcal{A}_q , entonces, la matriz $\hat{\beta}$, estimador de máxima verosimilitud bajo H_{CPC} de los vectores propios en la descomposición espectral, tiene q columnas que generan el subespacio \mathcal{A}_q .

(Demostrado en [2, sección 6.6]).

Evidentemente, las matrices estimadas de varianza S_1, \dots, S_k no cumplen exactamente las hipótesis del lema, pero las cumplen aproximadamente para muestras grandes.

Lo anterior implica que se puede seguir el procedimiento a continuación para resolver $H_{SC}(q)$:

1. Calcular $B = (B_1 : B_2)$, donde B es el estimador de máxima verosimilitud de β bajo H_{CPC} , y B está ordenado de tal manera que B_1 son las q columnas que se cree que generan subespacios comunes de q vectores propios de S_1, \dots, S_k .
2. Encontrar matrices P_i y Q_i tales que $P_i' B_1' S_i B_1 P_i$ y $Q_i' B_2' S_i B_2 Q_i$ sean diagonales.
3. Definir $\bar{\beta}_{s_1}^{(i)} := B_1 P_i$ y $\bar{\beta}_{s_2}^{(i)} := B_2 Q_i$ para $i = 1, \dots, k$.
4. Definir los aproximados de los estimadores de máxima verosimilitud como:

$$(2.95) \quad \bar{\beta}^{(i)} := (\bar{\beta}_{s_1}^{(i)}; \bar{\beta}_{s_2}^{(i)})$$

$$(2.96) \quad \bar{\Lambda}_i := \text{diag}(\bar{\beta}^{(i)'} S_i \bar{\beta}^{(i)})$$

$$(2.97) \quad \bar{\lambda}_{ij} := (\bar{\Lambda}_i)_{jj}$$

$$(2.98) \quad \bar{\Psi}_i := \bar{\beta}^{(i)} \bar{\Lambda}_i \bar{\beta}^{(i)} \quad i = 1, \dots, k$$

Se define ahora el estadístico para la prueba de hipótesis:

$$H_{SC}(q) \text{ vs. } H_a : \Psi_1, \dots, \Psi_k \text{ son arbitrarias}$$

como:

$$(2.99) \quad T_{SC}(q) = -2 \log \frac{L(\hat{\Psi}_1, \dots, \hat{\Psi}_k)}{L(S_1, \dots, S_k)} = \sum_{i=1}^k n_i \log \frac{\det \hat{\Psi}_i}{\det S_i}$$

que es el estadístico de logverosimilitud que se ha estado usando.

Este estadístico se distribuye como una ji-cuadrada con $(k-1)q(p-q)$ grados de libertad, y se puede aproximar con los aproximados de los estadísticos de máxima verosimilitud para obtener:

$$(2.100) \quad \bar{T}_{SC}(q) = \sum_{i=1}^k n_i \log \frac{\det \bar{\Psi}_i}{\det S_i}$$

El estadístico aproximado siempre es mayor al exacto y por eso estrictamente nos sirve para no rechazar H_{SC} pero no para rechazarla.

2.6. El proceso de comparación de las matrices de varianza

Por lo general, en el análisis multivariado, cuando se comparan dos matrices de varianza, si éstas no son iguales se consideran simplemente como diferentes. Dado que la diferencia en el número de parámetros estimados entre un modelo y otro es muy grande, $(k-1)p(p-1)/2$, lo anterior no es una práctica muy conveniente y, por eso, se han propuesto los diferentes niveles de similitud entre varias matrices de varianza. A continuación se explicarán algunos métodos útiles en la comparación de varias matrices de varianza.

Antes que nada, es conveniente recordar la escala de similitud entre matrices de varianza, junto con el número de parámetros a estimar de cada uno de sus niveles. Para ello incluimos la TABLA 1.

TABLA 1. Niveles de similitud entre matrices de varianza

Nivel	Modelo	Número de Parámetros
1	Igualdad	$\frac{p(p-1)}{2} + p$
2	Proporcionalidad	$\frac{p(p-1)}{2} + p + k - 1$
3	CPC	$\frac{p(p-1)}{2} + kp$
4	CPC(q)	$\frac{p(p-1)}{2} + kp + \frac{1}{2}(k-1)(p-q)(p-q-1)$
4*	SC(q)	$\frac{p(p-1)}{2} + kp + \frac{1}{2}(k-1)(p-q)(p-q-1) + \frac{1}{2}(k-1)q(q-1)$
5	Matrices arbitrarias	$k \left[\frac{1}{2}p(p-1) + p \right]$

A continuación se descompondrá el estadístico de logverosimilitud de igualdad de varianzas, que se puede ver en la ecuación 2.12, de acuerdo con la parte de desigualdad que le corresponde a cada peldaño de la escala propuesta. Se puede definir el estadístico del cociente de logverosimilitud para la prueba:

H_0 : Las matrices tienen el nivel de similitud A

vs.

H_a : Las matrices cumplen con el nivel de similitud B

donde B es un modelo menor en la escala que A, como:

$$(2.101) \quad T(A|B) = \sum_{i=1}^k n_i \log \frac{\det \hat{\Psi}_i}{\det \hat{\Psi}_i}$$

$$(2.102) \quad = \sum_{i=1}^k n_i \sum_{j=1}^p (\log \hat{\lambda}_{ij} - \log \hat{\lambda}_{ij})$$

Donde $\hat{\Psi}_i$ y $\hat{\lambda}_{ij}$ son los estimadores del modelo A, mientras que $\hat{\Psi}_i$ y $\hat{\lambda}_{ij}$ son los estimadores del modelo B.

Se puede inferir de la Teoría de Pruebas de Cocientes de Verosimilitudes que el estadístico anterior se distribuye como una ji-cuadrada con los grados de libertad correspondientes a la diferencia de grados de libertad entre el modelo A y el B.

Así, el estadístico para:

H_0 : Las matrices de varianza son iguales

vs.

H_a : Las matrices de varianza no están relacionadas

que es:

$$(2.103) \quad T_{TOTAL} = \sum_{i=1}^k n_i \log \frac{\det S}{\det S_i}$$

se puede descomponer de la siguiente manera:

$$(2.104) \quad T_{TOTAL} = T(\text{desigualdad de constantes de proporcionalidad} | \text{proporcionalidad}) \\ + T(\text{no proporcionalidad} | \text{CPC}) \\ + T(\text{desigualdad de los últimos } p - q \text{ componentes} | \text{CPC}(q)) \\ + T(\text{desigualdad de los últimos } q \text{ componentes})$$

Es importante hacer notar que, aunque en el estadístico anterior sólo se calculó un modelo de $CPC(q)$, en realidad se puede descomponer ese estadístico en $p-2$ estadísticos de la forma:

$$T(\text{desigualdad del componente } q \mid \text{CPC}(q-1))$$

para ser más específicos en la descomposición del estadístico total. Estos estadísticos son especialmente útiles cuando son extremadamente significativos, pues en los demás casos no justifican contundentemente la elección un modelo en lugar de otro. Es conveniente indicar también que se deben dividir los estadísticos entre los grados de libertad de estos para comparar más adecuadamente sus magnitudes relativas.

Finalmente conviene resaltar que las diferentes partes del estadístico total no necesariamente son independientes.

2.6.1. Selección del modelo según el Criterio de Información Akaike. Otro método alternativo para elegir un modelo adecuado para un conjunto de k matrices de varianza, es utilizar el Criterio de Información de Akaike. La idea subyacente a este criterio es establecer una medida de bondad de modelo en base al siguiente razonamiento:

El máximo de la función de logverosimilitud es un estimador sesgado de la verosimilitud esperada promedio, donde el sesgo es asintóticamente el número de parámetros. Así, según Akaike, al tener que elegir entre varios modelos se debe escoger el que maximice la función que resulte de restar el número de parámetros del modelo al máximo de la función de logverosimilitud restringida al modelo. Por razones prácticas, Akaike recomendaba minimizar la función mencionada multiplicada por -2. En pocas palabras, el criterio tradicional de Akaike era elegir el modelo que minimiza la función:

$$(2.105) \quad AIC = -2 \times (\text{máximo de la función de logverosimilitud restringido al modelo}) \\ + 2 \times (\text{número de parámetros estimados del modelo})$$

Cabe remarcar que el criterio de Akaike está fundamentado filosóficamente en que estadísticamente no hay un modelo verdadero y lo que realmente se busca es el modelo que mejor se ajuste a los datos.

Para aplicar las ideas de Akaike al problema de diferenciar matrices de varianza se modifica el criterio de Akaike de la siguiente manera:

Supóngase que se tienen r modelos diferentes entre los cuales hay que elegir el “más adecuado”. Sea m_i el número de parámetros estimados del i -ésimo modelo y L_i el máximo de la función de verosimilitud que alcanza ese modelo. Los modelos a probar cumplen con una escala de la forma $m_1 < m_2 < \dots < m_r$ y $L_1 \leq L_2 \leq \dots \leq L_r$. El criterio tradicional de Akaike recomienda escoger el modelo i tal que:

$$-2 \log L_i + 2m_i$$

sea mínimo.

El criterio modificado es seleccionar el modelo que minimice:

$$(2.106) \quad AIC(i) = -2 \log \frac{L_i}{L_r} + 2(m_i - m_1)$$

Así, $AIC(r)$ sólo depende de los parámetros de más que estima el modelo r en comparación con el modelo 1; y $AIC(1)$ sólo depende del logaritmo del cociente de log-verosimilitudes, mientras que $AIC(i)$ siempre es un valor no negativo que, teóricamente, pudiera llegar a ser cero únicamente en el modelo 1. Además, es muy fácil calcular el valor del criterio a partir del valor del estadístico del logaritmo del cociente de verosimilitudes, tan sólo conociendo los parámetros de los diferentes modelos.

Como los Componentes Principales Comunes no son invariantes bajo transformaciones a escala, la jerarquía establecida para las matrices de varianza puede no tener sentido si se analizan variables en diferentes escalas simultáneamente. Lo anterior se debe a que, cambiar la escala de medida de algunas variables, cambiaría el valor de los componentes principales, así como de los estadísticos correspondientes a los modelos de Componentes Principales Comunes, de Componentes Principales Comunes parciales y el modelo de Subespacios Comunes. Por otro lado, se puede afirmar que los modelos de igualdad y proporcionalidad no dependen de la escala, por lo que se pueden usar con confianza independientemente de las unidades en las que estén medidas las variables.

CAPÍTULO 3

Análisis de datos con Componentes Principales Comunes

En los capítulos anteriores se explicaron los detalles teóricos de los Componentes Principales y los Componentes Principales Comunes. En este capítulo se presentará un ejemplo de la utilización del método de Componentes Principales Comunes en un problema de biología evolutiva con el fin de que sea más claro cómo se usan adecuadamente esos conocimientos teóricos, así como sus posibles aplicaciones.

Una de las aplicaciones más frecuentes de los Componentes Principales Comunes es en biología evolutiva. Los datos del ejemplo que aquí se presenta forman parte del proyecto de investigación que, para obtener el grado de doctor, realiza, dentro del Instituto de Ecología de la U.N.A.M., el Biólogo Armando López Velázquez. Fue él quien realizó los experimentos y las tomas de datos para obtener las muestras; y por eso los derechos de propiedad intelectual sobre éstas le corresponden. Sin embargo, ha permitido cortésmente que se utilicen para este ejemplo. Armando López, el investigador, tiene el objetivo de comparar poblaciones de distintas localidades de la flor *Datura stramonium*.

Esta flor es mejor conocida en México como el “toloache” y es famosa en el saber popular por sus supuestas propiedades para facilitar el enamoramiento de un ser deseado. Le recordamos al lector que las semillas de esta planta son altamente venenosas, por lo que le sugerimos no intentar “enamorar” a nadie con “toloache”.

La *Datura stramonium* es una hierba de ciclo anual que se reproduce exclusivamente por semillas. Su ciclo de vida comienza en la temporada de lluvias (mayo-junio); su periodo de floración, durante el cual se reproduce, ocurre entre junio y noviembre; y su muerte se da en el momento que maduran sus frutos, alrededor de diciembre.

Sus flores son hemafroditas. Por eso se pueden polinizar de dos maneras: ya sea que en una misma flor la parte masculina polinice a la parte femenina (autopolinización) o que algún insecto o colibrí tome el polen de una flor y lo lleve al aparato femenino de otra flor para polinizarla (entrecruzamiento).

Los insectos y aves principales que visitan esta flor y que la pueden polinizar son abejas (*Apis mellifera*), abejas carpinteras (*Xilocopa* sp), colibríes y varias especies de esfíngidos. Éstos últimos se acercan a las flores para obtener de ellas una recompensa que puede ser su néctar o su polen. Además, existen características de la flor que atraen a los polinizadores como son su tamaño, su aroma y, en ocasiones, combinaciones de estos.

De acuerdo con los biólogos, la reproducción por entrecruzamiento es más ventajosa en términos evolutivos, pues casi cualquier código genético tiene algunos defectos (genes deletéreos) y, cuando un código genético defectuoso se mezcla con un código genético no

defectuoso, los caracteres defectuosos no se expresan en la flor. Eso ayuda a que ésta pueda sobrevivir hasta su reproducción y perpetúe la especie.

Como consecuencia de esto último las flores deben competir por los polinizadores, tratando de ofrecer la mayor cantidad de néctar y las características exteriores más atractivas para los polinizadores.

De acuerdo a la teoría biológica, según el ambiente, las plantas se van adaptando para ser lo más atractivas posible para los polinizadores específicos de dicho ambiente. Sin embargo, genéticamente tienen una estructura que limita sus posibilidades de adaptación. Estas limitaciones genéticas se expresan en la manera que los distintos atributos de la flor están correlacionados entre sí, ya sea de manera positiva o negativa. Así, en ocasiones, el mejorar un atributo atractivo para polinizadores va ligado a la mejora o empeora de otro atributo atractivo. En especial, cuando la mejora de un atributo implica la empeora de otro atributo, la flor enfrenta disyuntivas en cuanto a qué atributo le conviene más “mejorar”.

El objetivo del investigador, al comparar las flores de las poblaciones de *Datura stramonium*, es ver cómo ha influido su estructura genética en sus cambios. Justamente, esas comparaciones se hacen a través de las matrices de varianza.

Para entender bien las matrices de varianza que se presentarán, es necesario saber cómo se han extraído los datos del experimento. Los biólogos utilizan dos tipos de matriz de varianza en estudios evolutivos: las fenotípicas, (que se denotan con P por su inicial en inglés); y las genotípicas (denotadas por la letra G). Las matrices G tratan de medir la influencia genética en una población, mientras que las matrices P miden la manera en que se expresan los genes en la planta, tomando en cuenta no sólo los factores genéticos, sino también los ambientales y la interacción entre ambos.

En cuanto a la manera de medir las matrices G y P, se puede decir que, al calcular una matriz de varianza de diferentes características de un ser vivo, se considera que se tiene una matriz P; mientras que para obtener los datos para calcular una matriz G es necesario realizar un diseño experimental en el que se genere una muestra controlando quién es el padre y la madre de cada elemento, así como el medio ambiente, para poder descartar la información ambiental. El cálculo de estas últimas matrices requiere de diseños experimentales más elaborados y de una cantidad de elementos mucho más grande para poder lograr un tamaño suficiente de muestra para estimar bien la matriz de varianza.

En este ejemplo, por la manera en que se obtuvieron las muestras, todas las matrices son fenotípicas. En fases posteriores de la investigación biológica se realizarán pruebas con las matrices genotípicas, pero, por cuestiones de tiempo, no están incluidas en este ejemplo.

Sabemos que, en los pocos estudios que se han realizado para comparar las matrices G de diferentes animales, se ha encontrado que éstas difieren significativamente cuando pertenecen a especies distintas pero no difieren mucho cuando son de la misma especie.

Este tipo de estudios en plantas han sido aún menos abundantes por lo que no se sabe bien qué conclusiones se pueden esperar.

3.1. Poblaciones, muestras y datos

Para encontrar información de lo anterior se obtuvieron las siguientes muestras:

En el año 2003 se extrajeron muestras de tres poblaciones de *Datura stramonium* en lugares diferentes: Taxco Viejo, Guerrero; Tula, Hidalgo; y Santo Domingo, Morelos. En cada una de estas muestras silvestres se midieron tres atributos: largo de la flor (largo), ancho de la flor (ancho) y hercogamia.

En el año 2004 se midieron dos muestras cultivadas en invernadero. Una obtenida a partir de semillas de Taxco Viejo, a la que se denominó Taxco, y otra obtenida con semillas de Tula. Se le midieron, a cada flor, diez atributos, entre los que se encuentran: los tres medidos el año anterior, además de largo de cáliz (cáliz), largo de pedicelo (pedicelo), antera, pistilo, cámara de néctar (cámara), volumen de azúcar en el néctar (néctar) y porcentaje de azúcar en el néctar (sacarosa).

Más específicamente, las muestras de Tula son de un lugar llamado San Andrés Tul-tepec, sin embargo, el término común que usaba el ecólogo para denominarlas era Tula, por lo que, en este trabajo, se utiliza el mismo nombre para referirlas.

También en el 2004, se midieron estos diez atributos a dos muestras silvestres de la misma flor: una tomada en Ticumán, Morelos y otra en Teotihuacán, Estado de México.

Es importante mencionar que los datos del 2003 no se pueden comparar con los del 2004, ni se pueden comparar los datos de las muestras silvestres del 2004 con los de las muestras de invernadero del 2004, pues, aunque en todos hay algunos atributos iguales, la metodología de medición no permaneció constante.

3.2. Análisis descriptivo de los datos

3.2.1. Los datos. De cada una de las muestras se recibieron dos bases de datos: una con todos los datos medidos y otra que no contenía algunos datos que el biólogo investigador había excluido por considerarlos sospechosos. Las razones por las que un dato podía haber sido considerado sospechoso fueron: que no tuviera sentido, como por ejemplo, que una flor tuviera largo igual a cero, o que el dato mostrara una magnitud muy grande que, a los ojos del investigador, pareciera inverosímil. Se decidió confiar en la intuición del investigador, pues es él quien conoce bien las plantas de estudio. Y, efectivamente, al hacer pruebas de normalidad, los datos sin las observaciones sospechosas dieron mejores resultados.

3.2.2. Las variables. Las muestras del 2003 tenían mediciones de tres variables inherentes a la flor: largo, ancho y hercogamia. Las muestras del 2004, tanto de invernadero como silvestres, tenían más variables: largo, ancho, cáliz, hercogamia, pedicelo, néctar, sacarosa, largo de la antera, largo del pistilo y largo de la cámara de néctar.

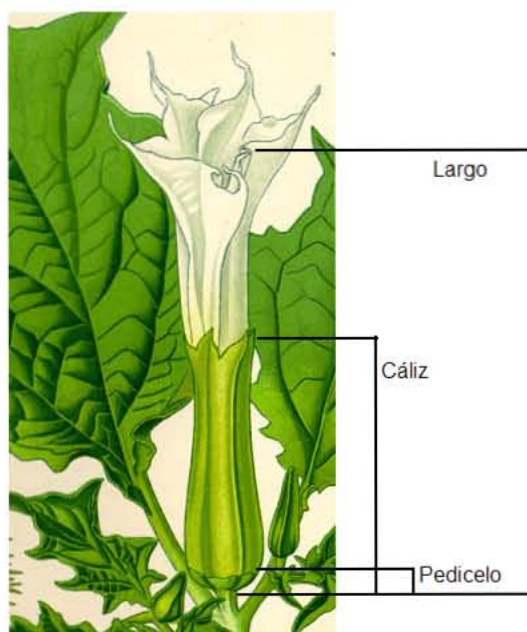


FIGURA 1. Largo, cáliz y pedicelo de la flor

Las variables largo, ancho, cáliz, hercogamia, pedicelo, largo de antera, largo de pistilo y largo de la cámara de néctar fueron medidas en milímetros. El néctar fue medido en microlitros y la sacarosa en unidades porcentuales Brix.

La manera en que fue medido el largo, cáliz y pedicelo de la flor se puede ver en la FIGURA 1. Lo que significa la antera, el pistilo y la hercogamia de una flor se puede ver en la FIGURA 2. El ancho es el “diámetro” de la flor cuando está completamente abierta, de manera que sus pétalos son perpendiculares a su tallo. La cámara es una parte de la flor que se encuentra abajo de los pétalos y está cubierta por el cáliz, por lo que no es visible en su superficie.

La hercogamia, como se puede ver en la FIGURA 2, es la distancia entre la parte masculina de la flor (que es la que contiene el polen y se llama antera) y la parte femenina (llamada pistilo). Esta medida es positiva si el pistilo es más largo que la antera y negativa en el caso contrario. El hecho de que esta medida sea negativa implica que la parte masculina de la flor es más alta que la femenina y, en consecuencia, es fácil que se autopolinice. Muchas de las flores tienen su antera y pistilo a la misma altura, por lo que su hercogamia es 0.

El néctar es la sustancia dulce acuosa en las flores de la cual se alimentan los insectos y las aves.

La sacarosa es un tipo de azúcar que contiene el néctar.

Se decidió excluir del análisis de componentes principales comunes las variables néctar, sacarosa y hercogamia; las primeras dos por estar medidas en unidades diferentes a

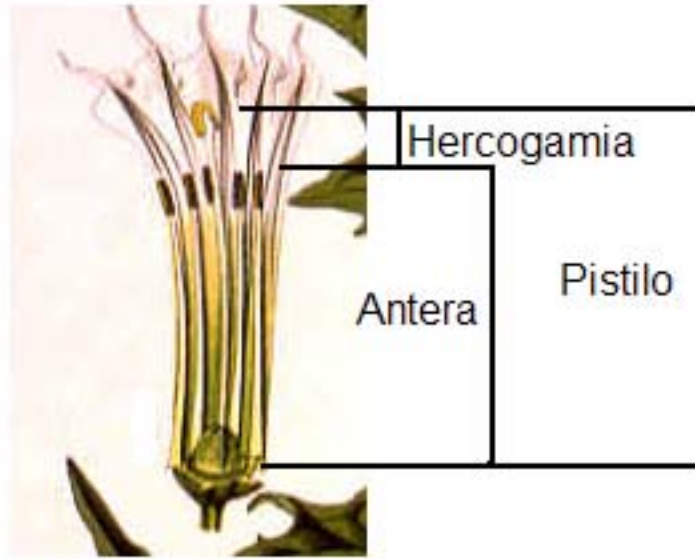


FIGURA 2. Antera, pistilo y hercogamia de la flor

todas las demás y la tercera por tener, en todas las muestras, un gran porcentaje de sus elementos con medida cero, lo cual hace evidente que la muestra, en esa característica, no sea ni remotamente normal.

Es importante hacer notar que el utilizar los Componentes Principales Comunes sólo con variables medidas en la misma escala es seguir un enfoque muy purista. En la práctica se utilizan los Componentes Principales Comunes muy frecuentemente con variables medidas en varias escalas. De hecho, al desarrollar el modelo matemático, en ningún momento se supuso que las variables están medidas en la misma escala. Lo que ocurre es que los Componentes Principales, y en consecuencia los Componentes Principales Comunes, son un método dependiente de la escala. Al modificarla, se modifican tanto los Componentes Principales estimados, como el valor de los estadísticos de prueba para los modelos de Componentes Principales Comunes y Componentes Principales Comunes Parciales. Por ello, cuando no hay una escala estándar, estos estadísticos pueden perder sentido. Teniendo la precaución de evitar el problema anterior, se considera que una escala estándar hace sentido cuando todas las variables están medidas en la misma escala. Ello restringe el uso del modelo al caso en que todas las variables son dimensionalmente congruentes. En muchas ocasiones se utiliza esta técnica con variables medidas en diferentes escalas, eso no es necesariamente un error, pero dado que este es un ejemplo ilustrativo de los Componentes Principales Comunes preferimos tratar de utilizar el modelo en su forma más pura y con la mayor cautela posible.

3.2.3. Las muestras del 2003. Las bases de datos originales de este año contenían seis muestras: Acolman, Estado de México; Quilamula, Morelos; San Cristobal, Chiapas; Santo Domingo, Morelos; Taxco Viejo, Guerrero; y Tula, Hidalgo.

Al considerar las bases sin los datos sospechosos se eliminaron tres muestras: Acolman, Quilamula y Santo Domingo, pues éstas contenían muy pocas observaciones: 40, 60 y 35 respectivamente. Esas bases, en lugar de contener datos respectivos a la medición de cada flor, contenían el promedio de los atributos de la planta. Sin embargo, no todas las plantas tienen el mismo número de flores. Por lo anterior, esa muestra contenía en cada renglón información respectiva a un promedio de un número diferente de individuos. Además, se sabía que los métodos de medición que se utilizaron para esas bases fueron modificados por ser considerados inadecuados por otros ecólogos que ayudaban al investigador en su trabajo. Por todo lo anterior, así como porque sólo contenían dos variables utilizables, que eran largo y ancho, se decidió no utilizar las muestras del 2003 en el análisis de componentes principales comunes.

De cualquier manera se incluye una gráfica matriz de las bases de datos sin los datos sospechosos en la FIGURA 3. En esta gráfica se puede observar que las flores más grandes, así como las que tienen una mayor hercogamia, son las de Taxco Viejo.

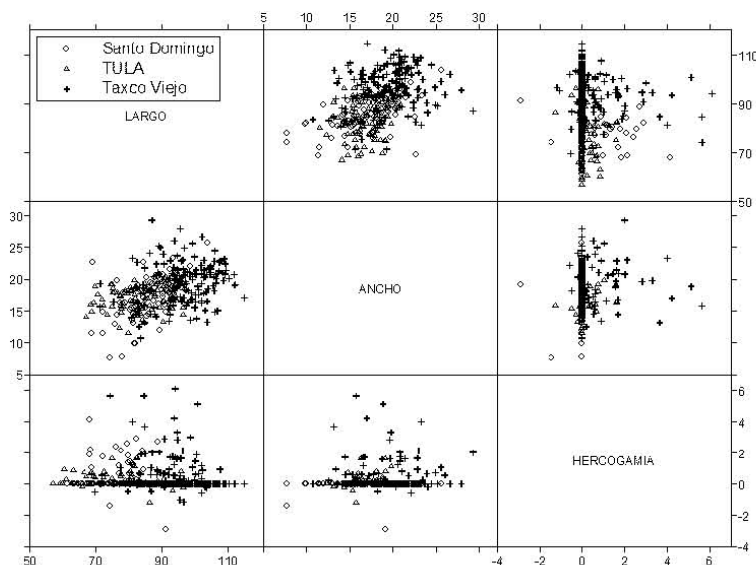


FIGURA 3. Gráfica matriz de las muestras 2003

3.2.4. Muestras de invernadero en el 2004. En el 2004 se cultivaron en invernadero dos muestras: una de semillas de Tula y otra de semillas de Taxco. Para realizar experimentos posteriores, las semillas de cada lugar se dividieron en dos grupos: uno se plantó en macetas más grandes que el otro. Al grupo que se sembró en macetas más grandes se le llamó el *grupo padre* y al de las macetas más pequeñas se le llamó el *grupo madre*; a la variable que denominaba la diferencia entre un grupo y el otro se le llamó sexo. Aunque, para el observador común, no había ninguna diferencia entre las plantas del

grupo padre y las plantas del *grupo madre*, según el ecólogo, el *grupo padre* tenía hojas más anchas y un tamaño mayor. Por ello, no es posible garantizar que los grupos *padre* y *madre* puedan ser considerados como que pertenezcan a una sólo población.

Para saber si la percepción del ecólogo era correcta, se realizaron, con las muestras de cada lugar, análisis discriminantes del sexo usando las demás variables como discriminadoras. En estos se encontró que, en la muestra de Tula se podía diferenciar razonablemente bien al *grupo padre* del *grupo madre*; sin embargo, en Taxco, los análisis discriminantes dejaron bastante que desear.

Para tener control sobre la posible variación ambiental entre las plantas padre y las plantas madre, se consideró que, en el 2004, se tenían cuatro muestras de invernadero: Taxco madre, Taxco padre, Tula madre y Tula padre.

Los resultados para la prueba Kolmogorov-Smirnov de bondad de ajuste para una normal con sus parámetros de máxima verosimilitud se muestran en la TABLA 1. Éstos fueron calculados en S-plus, un programa preciso en esta prueba sobre todo cuando los valores p se encuentran entre 0 y 0.1, por lo que los valores mayores a 0.1 pueden no ser muy precisos.

TABLA 1. Valores p de las pruebas de normalidad para las poblaciones de invernadero en 2004

Variable	Taxco Madre	TaxcoPadre	Tula Madre	Tula Padre
Largo	0.974	0.856	0.0536	0.913
Ancho	0.066	0.521	0.141	0.39
Cáliz	0.739	0.567	0.13	0.119
Hercogamia	0	0	0	0
Pedículo	0.317	0.001	0.323	0.249
Néctar	0.358	0.0004	0.249	0.227
Sacarosa	0	0	0.0345	0.0602
Antera	0.0541	0.673	0.147	0.505
Pistilo	1	0.578	0.0343	0.282
Cámara de néctar	0.225	0.0223	0.164	0.689

El objetivo de hacer pruebas de normalidad es que el método de Componentes Principales Comunes supone que los datos provienen de una distribución normal multivariada. La normalidad en todas las variables no implica la normalidad multivariada, pero, para que una muestra sea normal multivariada, sí es necesario que todas sus variables se distribuyan de manera normal (ver Apéndice C).

La técnica de Componentes Principales Comunes es exploratoria, por lo que no es necesario que la normalidad multivariada se cumpla al cien por ciento para aplicarla, pero sí es conveniente saber qué tanto se acercan los datos a este supuesto. En esta misma dirección, se revisan las gráficas de dispersión de las variables que se utilizan en los análisis de Componentes Principales Comunes. Si la distribución de los datos fuera

Normal Multivariada, en estas gráficas se verían aproximaciones a elipses en cada cuadro. Observando las FIGURAS 4, 5, 6 y 7 se puede notar que ese no es siempre el caso, aunque ocurre en muchas ocasiones.

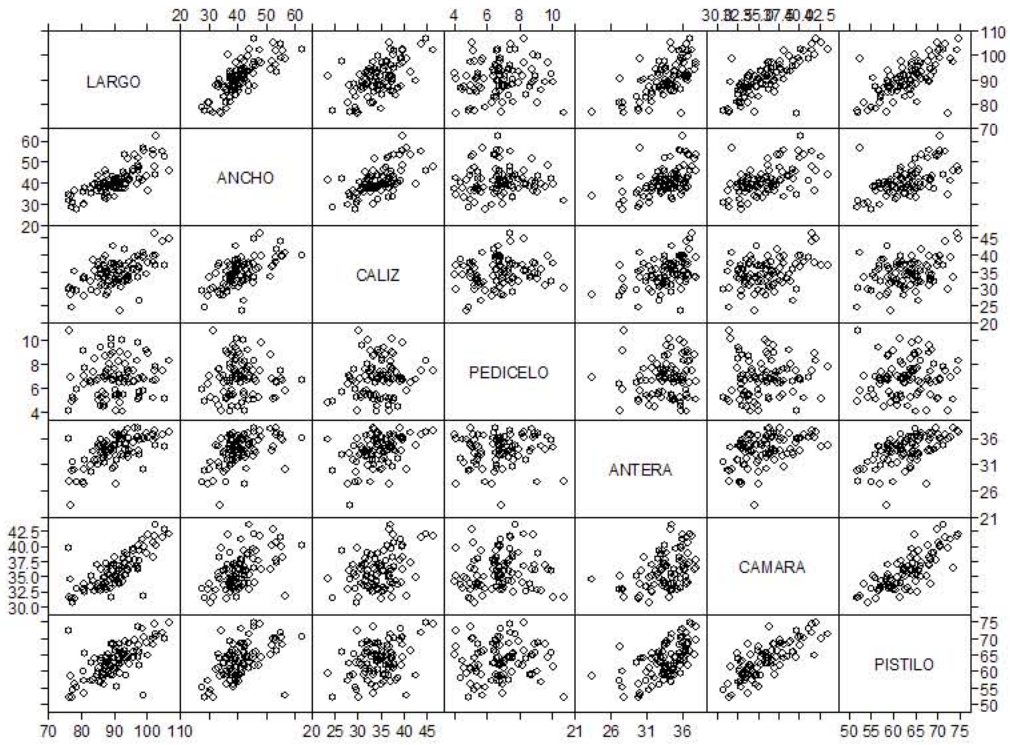


FIGURA 4. Gráfica matriz de Taxco madre

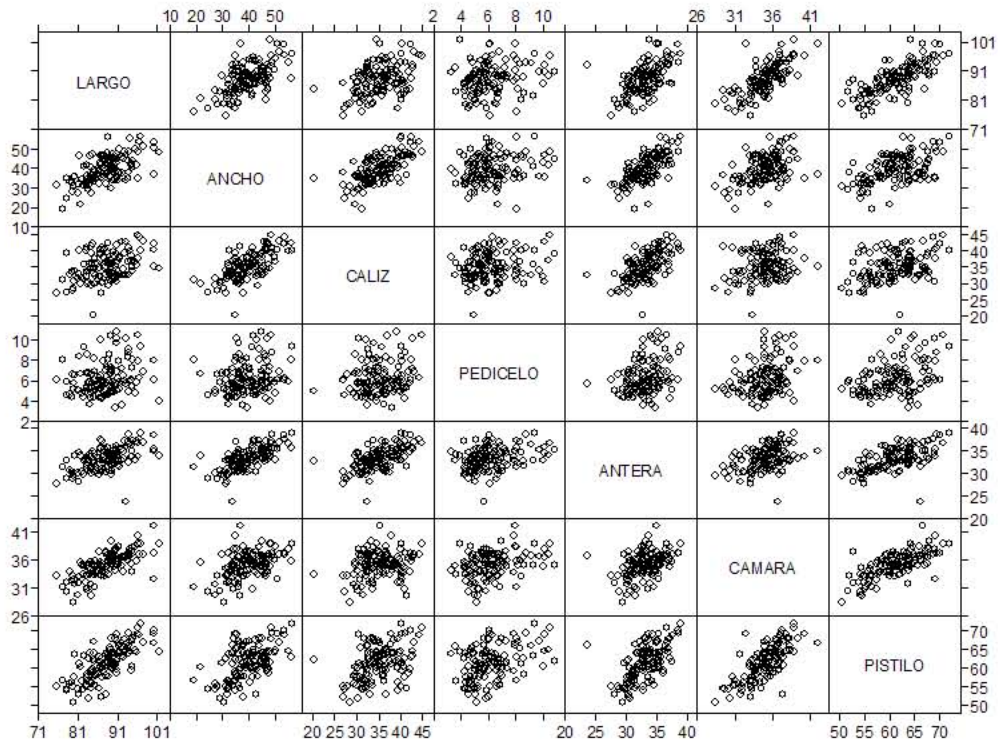


FIGURA 5. Gráfica matriz de Taxco padre

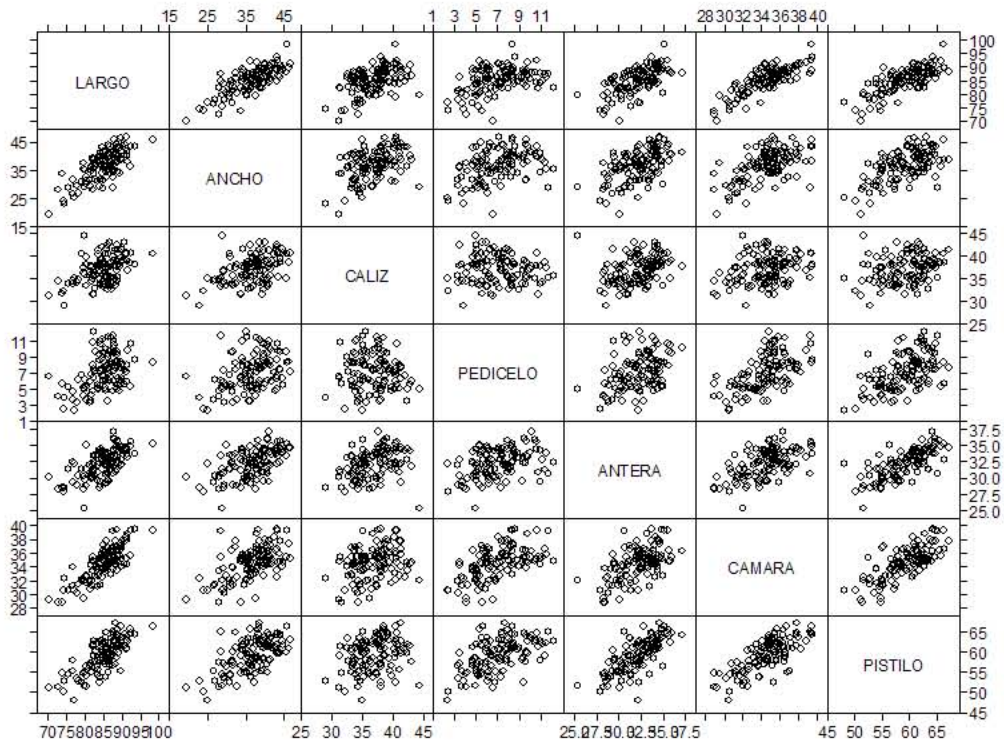


FIGURA 6. Gráfica matriz Tula madre

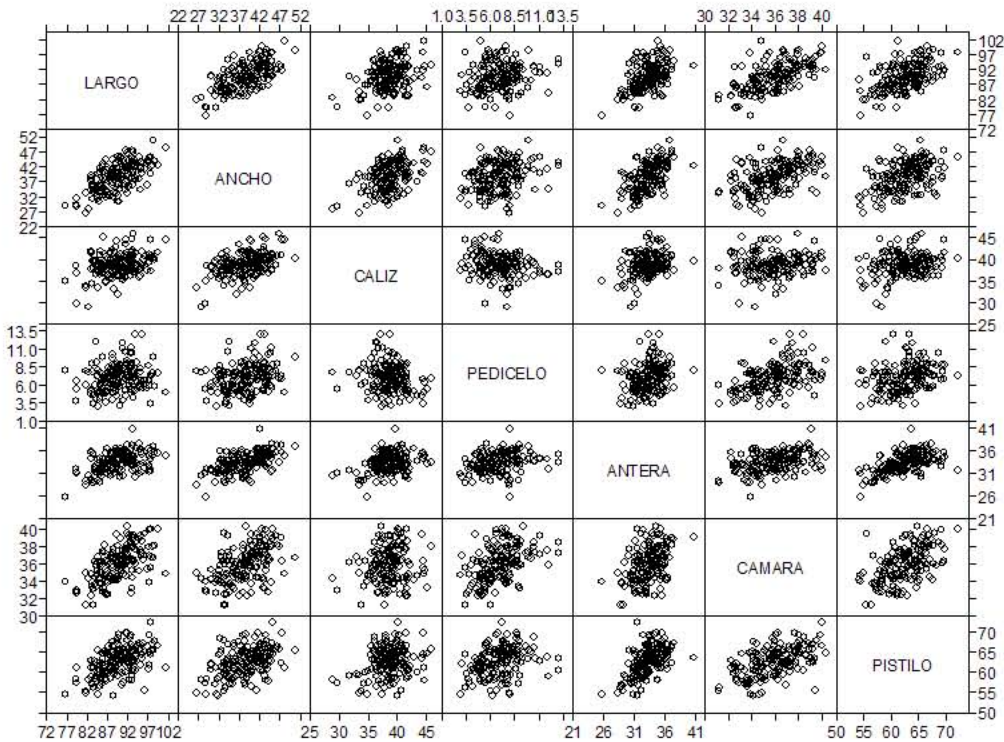


FIGURA 7. Tula padre

Para dar una idea de los problemas que tienen estas variables individualmente con la normalidad, se presentan algunos histogramas de las variables que se utilizan en el análisis de componentes principales comunes y no pasaron la prueba de bondad de ajuste Kolmogorov-Smirnov con un nivel de significancia de 0.90. La FIGURA 8 muestra que en los datos del ancho de la población de Taxco madre hay todo un grupo de observaciones que forman una pequeña campana normal a la derecha de la campana principal. La misma FIGURA muestra que, por unas observaciones muy a la izquierda, la distribución de la antera en Taxco madre es sesgada hacia la derecha. Asimismo, dicha FIGURA muestra que en el pedicelo de Taxco padre hay muchas observaciones muy a la derecha de lo que se consideraría una distribución normal usual; y que la distribución de la cámara de néctar en las flores de Taxco padre es muy pronunciada para ser una normal, pero es casi simétrica.

Finalmente, se pueden aprovechar las gráficas matrices de los atributos para notar que el pedicelo no tiene una relación lineal fuerte con ninguna de las otras variables, más bien, es un parámetro neutro.

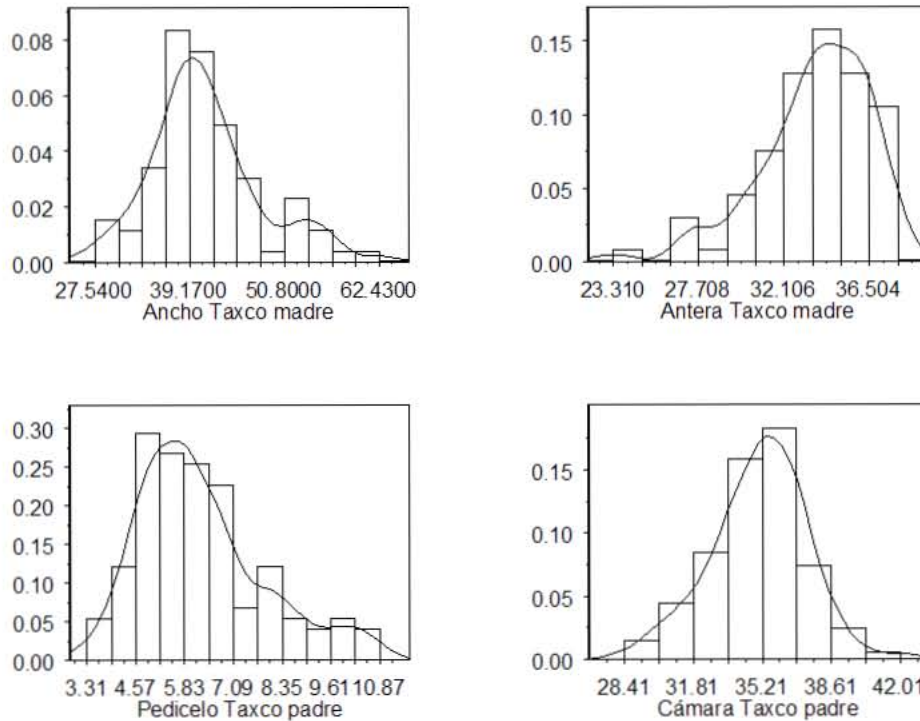


FIGURA 8. Histogramas de algunas variables de Taxco

3.2.5. Muestras silvestres en el 2004. En el 2004 también se sembraron y observaron dos muestras más, ubicadas en el sitio de las semillas de la muestra. Una de ellas en Teotihuacán y la otra en Ticumán. Al hacer las pruebas Kolmogorov-Smirnov de bondad de ajuste para una distribución normal se obtuvieron los valores p que se muestran en la TABLA 2.

Es claro, a partir de esos resultados, que la normalidad funcionó mucho mejor en Teotihuacán que en Ticumán, por lo que, para encontrar la causa de la no normalidad, se observaron los histogramas de algunas de las variables de Ticumán. En los histogramas del ancho y cáliz de la flor, mostrados en la FIGURA 9 se observaron indicios de dos modas.

TABLA 2. Valores p para las pruebas de normalidad para las poblaciones silvestres de 2004

Variable	Teotihuacán	Ticumán
Largo	0.392	0.0253
Ancho	0.9	0.0176
Cáliz	0.279	0
Hercogamia	0	0
Pedicelo	0.001	0.0091
Néctar	0.165	0
Sacarosa	0	0.0376
Antera	0.0381	0.0005
Pistilo	0.0337	0.161
Cámara de néctar	0.0588	0.007

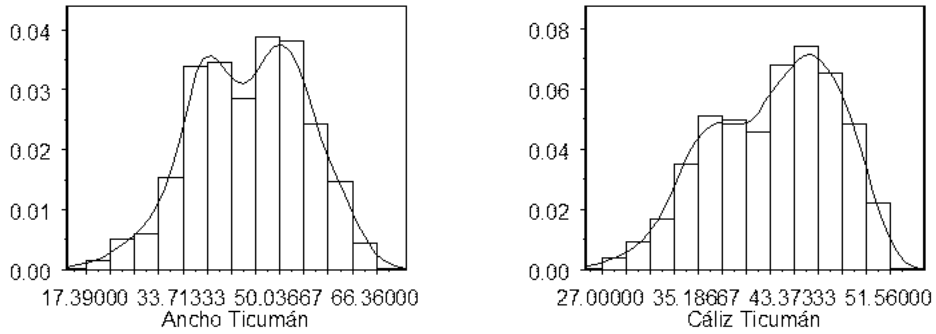


FIGURA 9. Histogramas de algunas variables de Ticumán

Por esa razón, se realizaron los diagramas de dispersión según varias variables. Hubo dos visitas a Ticumán para recabar datos: la primera, del 4 al 7 de septiembre de 2004, y la segunda, el 24 y 25 del mismo mes. Al dividir el diagrama de dispersión de ancho contra caliz, según la fecha en que se tomó la medición se obtiene la FIGURA 10. En ella se ve claramente que la distribución de los datos depende de la fecha en que fueron tomados. Al comentar eso con el investigador, él explicó que las plantas van creciendo. En ocasiones florecen y, después de florecer, sus ramas se bifurcan y crecen aún más para volver a florecer, pero esta vez con flores más pequeñas. Por eso, en la gráfica se puede observar que las flores de la segunda visita, que serían de bifurcaciones más altas, son más pequeñas que las de la primera visita.

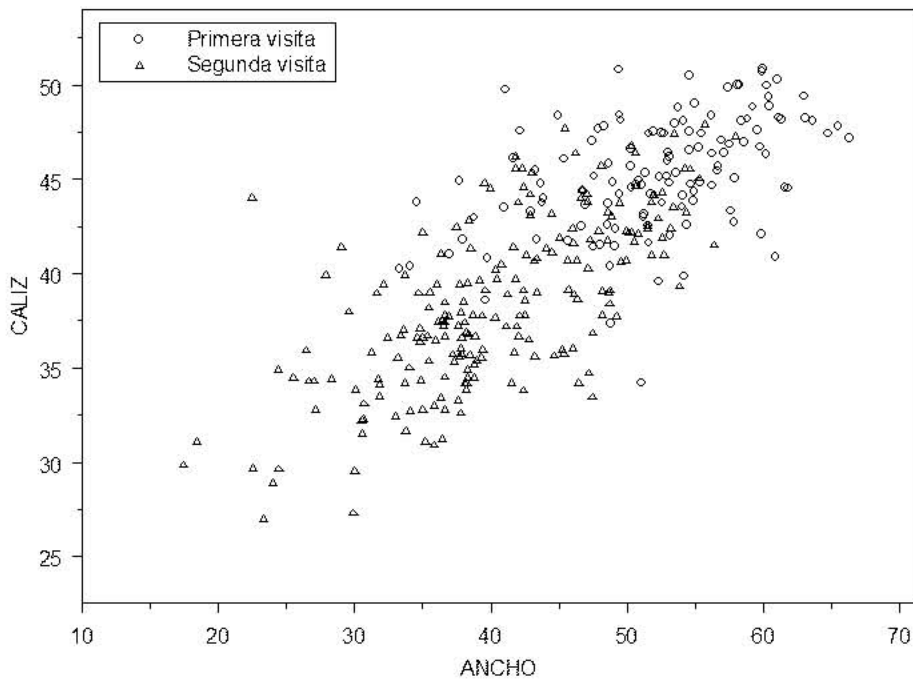


FIGURA 10. Diagrama de dispersión ancho vs. cáliz en Ticumán

Debido a lo anterior, se decidió separar la muestra de Ticumán en dos: las observaciones tempranas (de la primera visita) y las observaciones tardías (de la segunda visita) y realizar los análisis por separado, tomando en cuenta, de esa manera, la variación por las fechas. Los valores p de la prueba Kolmogorov-Smirnov de bondad de ajuste para una distribución normal con los parámetros estimados de máxima verosimilitud se muestran en la TABLA 3.

TABLA 3. Valores p para las pruebas de normalidad para las Poblaciones de Ticumán según la fecha en la que fueron tomadas

Variable	Tempranas	Tardías
Largo	0.6	0.0093
Ancho	0.319	0.105
Cáliz	0.131	0.173
Hercogamia	0	0
Pedicelo	0.011	0.0729
Néctar	0.0004	0
Sacarosa	0.0132	0.882
Antera	0.0238	0.0022
Pistilo	0.0089	0.404
Cámara de néctar	0.0378	0.0827

Se incluyen también en las FIGURAS 11, 12, y 13, las gráficas de dispersión de los atributos utilizados de dos en dos, de las muestras de Teotihuacán, Ticumán tempranas y Ticumán tardías, respectivamente.

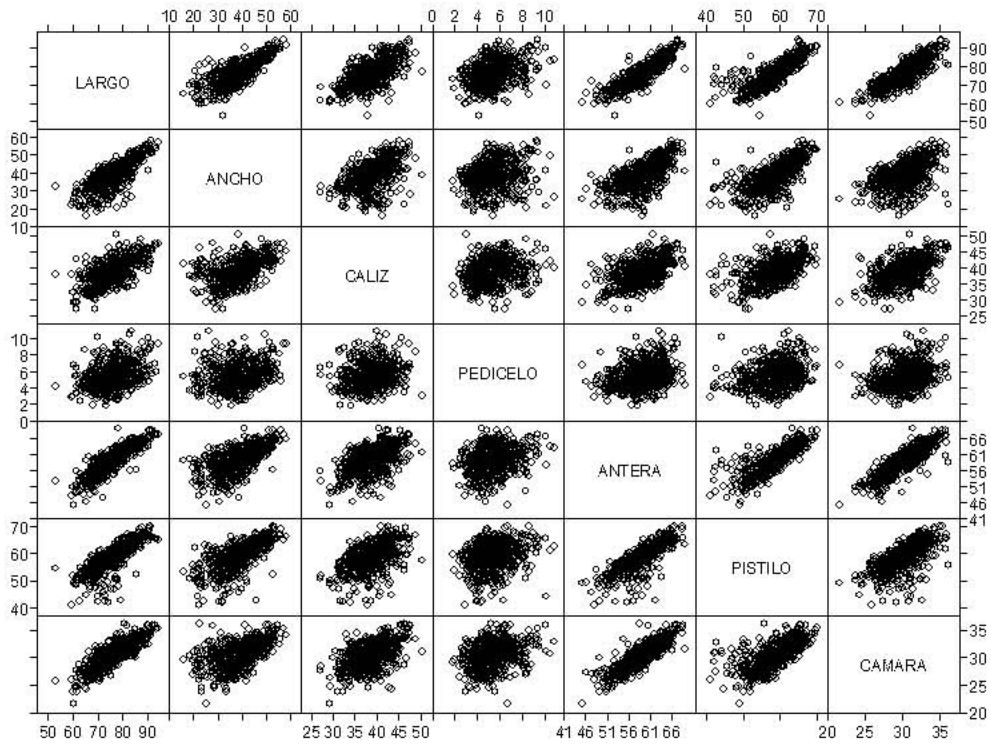


FIGURA 11. Gráfica matriz Teotihuacán

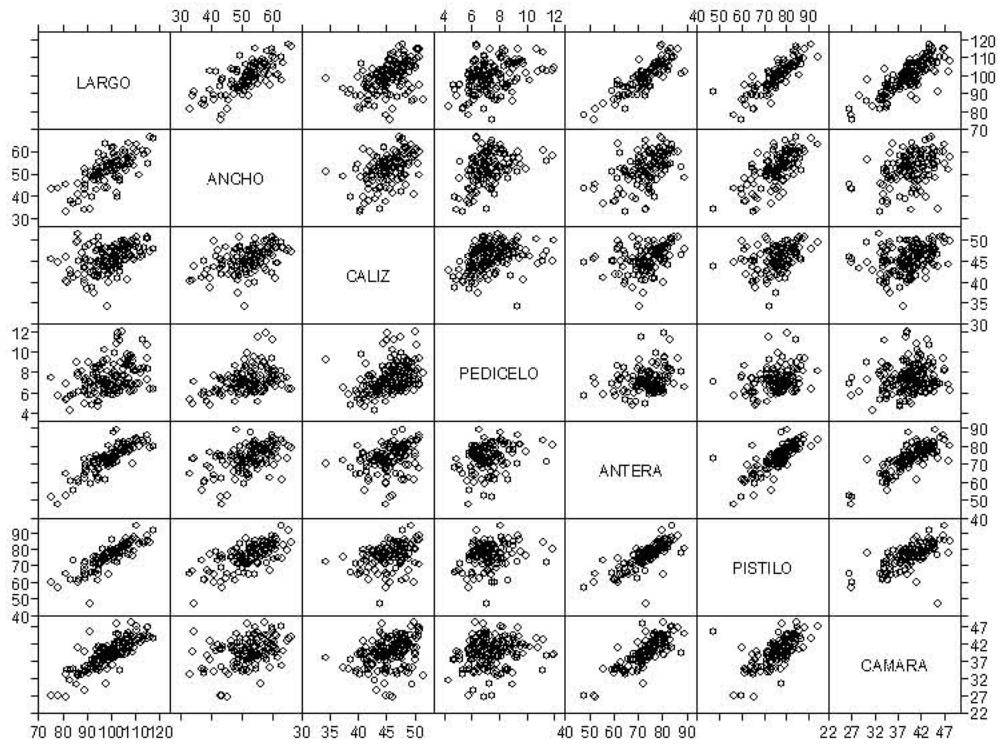


FIGURA 12. Gráfica matriz Ticumán tempranas

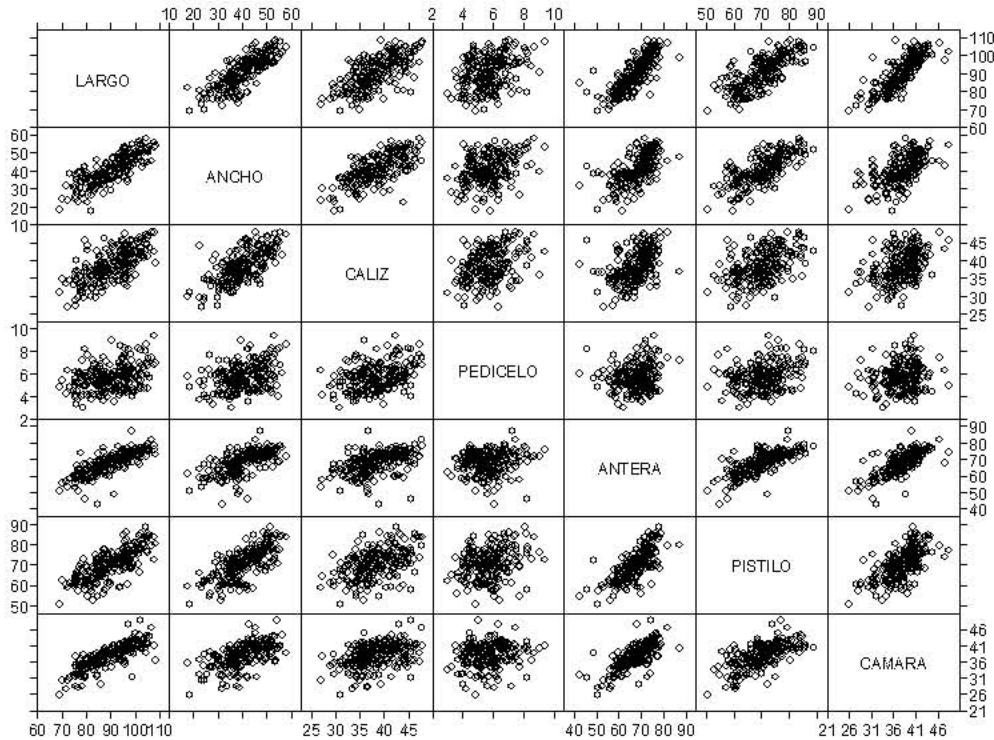


FIGURA 13. Gráfica matriz Ticumán tardías

Finalmente, se incluyen los histogramas de las variables que no pasaron la prueba Kolmogorov-Smirnov de bondad de ajuste para una distribución normal con los parámetros estimados de máxima verosimilitud, en las FIGURAS 14, 15 y 16. Se puede observar que, en las tres muestras, los histogramas del pedicelo están sesgados a la izquierda; los de cámara y antera están sesgados a la derecha. El pistilo tiene un sesgo a la derecha en Teotihuacán y Ticumán tempranas; y el largo en Ticumán tardías tiene un pico a la derecha y es una gráfica muy achatada.

En cuanto a las formas de los histogramas que parecen sospechosas a primera vista, y que podrían llevar a pensar que los datos fueron medidos erróneamente, el biólogo explicó que se podrían deber a los tipos de selección. Esto no es otra cosa que las medidas más favorables para que una planta pueda sobrevivir y reproducirse. Estas medidas no siempre son cercanas a la media ni continuas, por lo que los histogramas que se observan son plausibles.

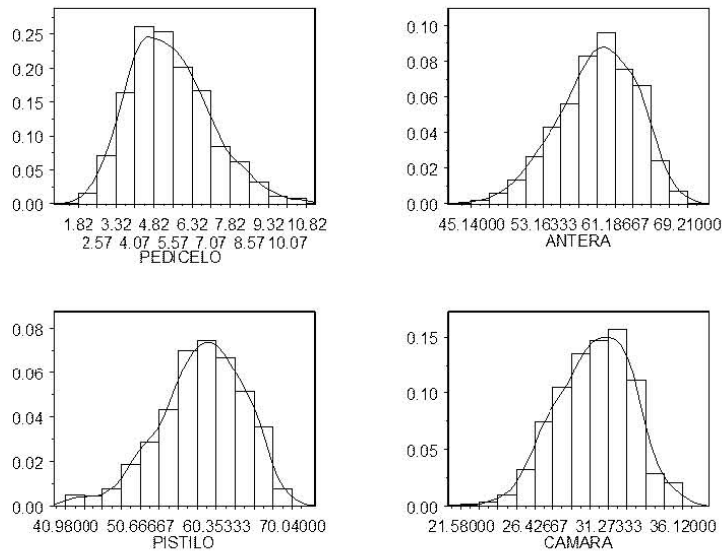


FIGURA 14. Histogramas Teotihuacán

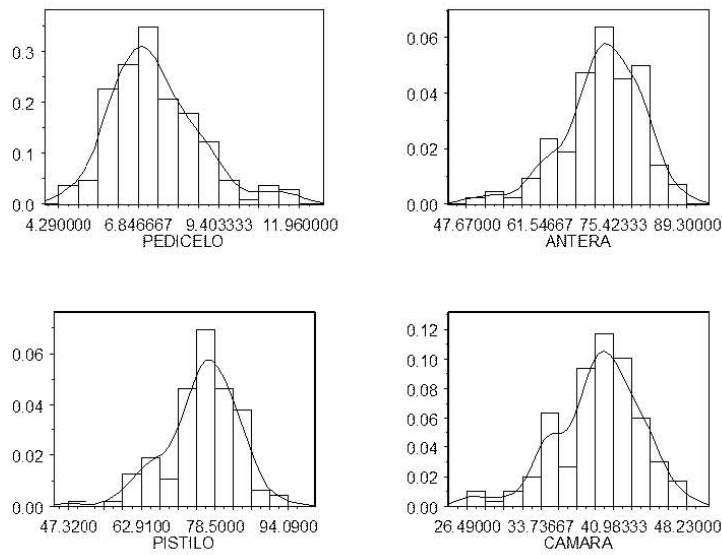


FIGURA 15. Histogramas Ticumán tempranas

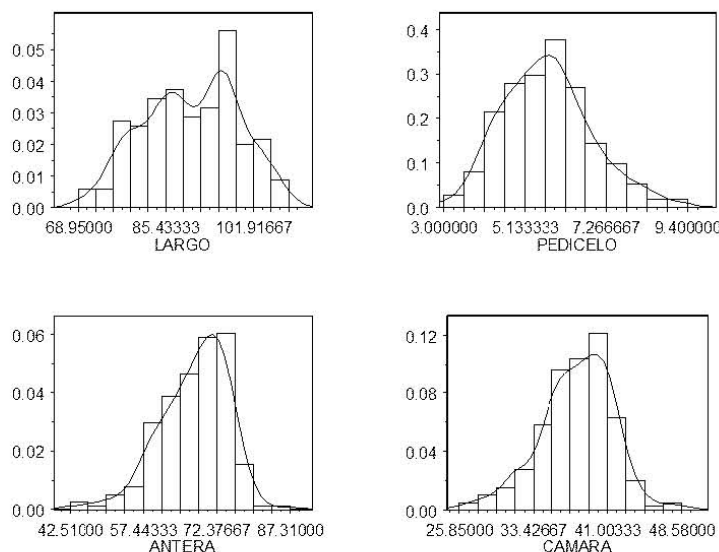


FIGURA 16. Histogramas Ticumán tardías

3.3. Análisis de componentes principales comunes

3.3.1. Muestras de Invernadero 2004. Como ya se había mencionado, se tienen cuatro muestras de invernadero en el 2004: Tula madre, Tula padre, Taxco madre y Taxco padre. A partir de este momento se denota con la letra M a las muestras madre y con la letra P a las muestras padre.

Antes que nada, se presentan las matrices de varianza de cada una de las muestras, así como sus componentes principales, tomando en cuenta las siete variables que se utilizan en los estudios de componentes principales comunes con mayor número de variables.

Tula M:

Matriz de varianza(calculada con 114 datos):

	Largo	Ancho	Cáliz	Pedicelo	Antera	Cámara	Pistilo
Largo	22.35	19.10	6.77	4.45	6.41	9.02	14.39
Ancho	19.10	29.33	7.63	4.71	6.46	7.29	13.26
Cáliz	6.77	7.63	9.19	-0.43	2.24	2.15	3.51
Pedicelo	4.45	4.71	-0.43	4.87	1.95	2.95	4.85
Antera	6.41	6.46	2.24	1.95	4.60	2.89	6.26
Cámara	9.01	7.29	2.15	2.95	2.89	5.55	7.19
Pistilo	14.39	13.26	3.51	4.85	6.26	7.19	16.31

Componentes Principales:

Valores propios:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
64.22	10.82	7.33	4.21	2.67	1.82	1.13

Vectores propios:

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
Largo	0.54	0.13	0.31	0.66	-0.15	-0.13	0.31
Ancho	0.60	-0.55	-0.54	-0.01	-0.06	0.06	-0.09
Cáliz	0.19	-0.46	0.71	-0.32	0.37	0.09	0.05
Pedicelo	0.14	0.30	-0.30	-0.02	0.83	-0.1	0.33
Antera	0.19	0.14	0.07	-0.33	-0.10	-0.88	-0.18
Cámara	0.23	0.25	0.09	0.21	0.27	0.16	-0.85
Pistilo	0.42	0.53	0.07	-0.55	-0.26	0.39	0.14

Tula P:*Matriz de varianza (calculada con 156 datos):*

	Largo	Ancho	Cáliz	Pedicelo	Antera	Cámara	Pistilo
Largo	21.14	14.28	4.27	1.76	4.82	5.58	8.21
Ancho	14.28	22.60	5.73	2.55	6.08	4.66	7.62
Cáliz	4.27	5.73	7.12	-0.98	1.82	0.79	2.27
Pedicelo	1.76	2.55	-0.98	3.69	1.1	1.83	2.03
Antera	4.83	6.08	1.82	1.1	4.67	1.8	4.36
Cámara	5.58	4.66	0.8	1.83	1.8	4.05	3.7
Pistilo	8.21	7.62	2.27	2.03	4.36	3.7	12.47

Componentes Principales:

Valores propios:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
46.21	9.25	7.61	6.24	1.46	2.71	2.27

Vectores propios:

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
Largo	0.6	0.16	-0.76	-0.06	-0.09	-0.05	0.16
Ancho	0.62	-0.54	0.35	0.34	-0.18	0.21	-0.05
Cáliz	0.19	-0.37	0.1	-0.74	0.49	-0.01	0.14
Pedicelo	0.09	0.2	0.14	0.46	0.59	-0.08	0.60
Antera	0.22	0.07	0.24	-0.03	-0.11	-0.93	-0.11
Cámara	0.2	0.20	-0.05	0.17	0.56	0.07	-0.76
Pistilo	0.35	0.68	0.46	-0.3	-0.19	0.28	0.07

Taxco M:*Matriz de varianza (calculada con 91 datos):*

	Largo	Ancho	Cáliz	Pedicelo	Antera	Cámara	Pistilo
Largo	49.27	35.89	16.60	1.07	12.26	15.77	25.37
Ancho	35.89	44.41	16.77	0.34	9.27	10.19	18.51
Cáliz	16.60	16.77	17.12	0.94	4.76	4.49	8.02
Pedicelo	1.07	0.34	0.94	2.25	0.4	0.34	0.5
Antera	12.26	9.27	4.76	0.4	7.77	3.63	8.72
Cámara	15.77	10.19	4.49	0.34	3.63	8.58	12.1
Pistilo	25.37	18.51	8.02	0.5	8.72	12.1	27.69

Componentes Principales:

Valores propios:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
122.21	18.74	9.25	8.48	4.58	2.14	1.67

Vectores propios:

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
Largo	0.63	0.12	-0.09	-0.71	-0.02	-0.04	-0.26
Ancho	0.56	-0.58	0.50	0.30	-0.01	0.06	0.07
Cáliz	0.26	-0.34	-0.84	0.28	-0.12	-0.08	0.01
Pedicelo	0.01	0.01	-0.11	-0.02	0.02	0.99	-0.04
Antera	0.17	0.12	-0.11	0.04	0.91	-0.02	0.35
Cámara	0.21	0.29	0	-0.07	-0.4	0.04	0.84
Pistilo	0.38	0.66	0.04	0.56	-0.06	0	-0.31

Taxco P:

Matriz de varianza (calculada con 119 datos):

	Largo	Ancho	Cáliz	Pedicelo	Antera	Cámara	Pistilo
Largo	28.08	23.36	9.12	1.81	7.63	8.83	18.27
Ancho	23.36	50.48	20.45	2.55	12.3	7.41	17.59
Cáliz	9.12	20.45	18.48	1.32	6.59	2.44	7.72
Pedicelo	1.81	2.55	1.32	2.66	1.30	1.22	3
Antera	7.63	12.30	6.59	1.3	6.44	2.42	7.1
Cámara	8.83	7.41	2.44	1.22	2.42	5.44	7.86
Pistilo	18.27	17.59	7.72	3	7.1	7.86	21.83

Componentes Principales

Valores propios:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
89.94	20.79	8.74	6.29	2.66	2.21	1.66

Vectores propios:

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
Largo	0.46	0.51	-0.21	-0.66	-0.11	0.01	0.19
Ancho	0.69	-0.5	-0.46	0.24	0.1	-0.02	0.04
Cáliz	0.32	-0.4	0.76	-0.36	0.19	0.02	0
Pedicelo	0.05	0.05	0.12	0.23	-0.12	0.78	0.55
Antera	0.20	-0.03	0.17	0.10	-0.88	0.07	-0.38
Cámara	0.16	0.25	-0.02	0.05	0.37	0.51	-0.71
Pistilo	0.38	0.52	0.35	0.57	0.13	-0.34	0.12

Primero que nada se observan cuidadosamente los componentes principales de cada matriz. En todas las matrices el primer componente principal tiene todas sus variables con constantes positivas, por lo que podemos considerar que representa el tamaño de la flor. El segundo componente principal tiene todas las constantes positivas, excepto el ancho y el cáliz. En el caso de Taxco P también tiene una constante negativa en la antera, pero esta es de una magnitud muy pequeña (-0.03). En general, a simple vista, se ve que los primeros dos Componentes Principales de las matrices se parecen en muchas constantes (por ejemplo: el primer componente principal de Tula M, Tula P y Taxco M tiene valores para el largo de alrededor de 0.60), aunque no siempre ocurre eso (por ejemplo: en las muestras de Tula M y Tula P el pedicelo tiene valores de 0.3 y 0.2 respectivamente, que son relativamente cercanos; sin embargo, la misma variable en Taxco M y Taxco P tiene valores de 0.01 y 0.05 respectivamente, estos son parecidos entre sí, pero no se parecen a los de las muestras de Tula). En los demás componentes principales no se alcanzan a distinguir claramente patrones de similitud. Sin embargo, hay que considerar que los primeros dos componentes principales, de acuerdo a la ecuación 1.9, representan el 81 % de la variación de los datos de Tula M, el 73 % de la variación en Tula P, el 84 % de la variación en Taxco M y el 84 % de la variación en Taxco P. Ello hace que sea mucho más significativo el parecido entre estos componentes principales que el parecido entre los demás componentes principales de las matrices.

Ahora comienza el análisis de Componentes Principales Comunes. Para todos los análisis de Componentes Principales Comunes se utilizó el programa de Patrick Phillips de Componentes Principales Comunes de la Universidad de Oregon que se puede ver en la referencia [8].

La primera comparación se hace con las cuatro matrices de varianza muestrales, tomando en cuenta siete variables: largo, ancho, cáliz, pedicelo, largo antera, largo de cámara y largo pistilo. Los resultados se encuentran en las TABLAS 4 y 5.

TABLA 4. Criterio de Información de Akaike para la comparación Tula M, Tula P, Taxco M y Taxco P con 7 variables

Modelo		T	GL	Valor p	T/GL	AIC
Mayor (H_0 :)	Menor (H_a :)					
Igualdad	Proporcionalidad	61.356	3	0.0000	20.452	273.412
Proporcionalidad	CPC	54.526	18	0.0000	3.029	218.056
CPC	CPC(5)	0.385	3	0.9432	0.128	199.531
CPC(5)	CPC(4)	30.945	6	0.0000	5.157	205.145
CPC(4)	CPC(3)	17.252	9	0.0449	1.917	186.200
CPC(3)	CPC(2)	19.113	12	0.0858	1.593	186.948
CPC(2)	CPC(1)	50.413	15	0.0000	3.361	191.835
CPC(1)	No relacionadas	39.422	18	0.0025	2.190	171.422
No relacionadas						168.000*

TABLA 5. Valores p para la comparación Tula M, Tula P, Taxco M y Taxco P con 7 variables

Modelo		T	GL	valor p
Mayor (H_0 :)	Menor (H_a :)			
Igualdad	No relacionadas	273.412	84	0.0000
Proporcionalidad	No relacionadas	212.056	81	0.0000
CPC	No relacionadas	157.531	63	0.0000
CPC(5)	No relacionadas	157.145	60	0.0000
CPC(4)	No relacionadas	126.200	54	0.0000
CPC(3)	No relacionadas	108.948	45	0.0000
CPC(2)	No relacionadas	89.835	33	0.0000
CPC(1)	No relacionadas	39.422	18	0.0025

La TABLA 4 muestra que el modelo que tiene el menor valor para en el Criterio de Información de Akaike (AIC, en la forma en que fue modificado en la sección 2.6.1 del capítulo 2), marcado con un asterisco, es el que considera a las matrices de varianza como no relacionadas. Esa información es confirmada por la de la TABLA 5, en la que se ve que son muy bajos los valores p de las pruebas del tipo:

H_0 : El modelo mayor se cumple vs. H_a : Las matrices son no relacionadas

Al hacer diferentes pruebas quitando una de esas matrices y comparando las otras tres, se observó que, al comparar únicamente Tula M, Tula P y Taxco M se obtiene un modelo

que comparte algunos componentes principales. Los resultados de esa comparación se muestran en las TABLAS 6 y 7.

TABLA 6. Criterio de Información de Akaike para la comparación Tula M, Tula P y Taxco M con 7 variables

Modelo		T	GL	Valor p	T/GL	AIC
Mayor (H_0 :)	Menor (H_a :)					
Igualdad	Proporcionalidad	57.985	2	0.0000	28.992	193.975
Proporcionalidad	CPC	30.706	12	0.0022	2.559	139.991
CPC	CPC(5)	0.954	2	0.6205	0.477	133.285
CPC(5)	CPC(4)	12.341	4	0.0150	3.085	136.330
CPC(4)	CPC(3)	52.577	6	0.0000	8.763	131.989
CPC(3)	CPC(2)	21.537	8	0.0058	2.692	91.412
CPC(2)	CPC(1)	9.514	10	0.4841	0.951	85.875*
CPC(1)	No relacionadas	8.362	12	0.7563	0.697	96.362
No relacionadas						112.000

TABLA 7. Valores p para la comparación Tula M, Tula P y Taxco M con 7 variables

Modelo		T	GL	valor p
Mayor (H_0 :)	Menor (H_a :)			
Igualdad	No relacionadas	193.975	56	0.0000
Proporcionalidad	No relacionadas	135.991	54	0.0000
CPC	No relacionadas	105.285	42	0.0000
CPC(5)	No relacionadas	104.330	40	0.0000
CPC(4)	No relacionadas	91.989	36	0.0000
CPC(3)	No relacionadas	39.412	30	0.1168
CPC(2)	No relacionadas	17.875	22	0.7134
CPC(1)	No relacionadas	8.362	12	0.7563

En la TABLA 6 se observa que el modelo con el menor índice de Akaike es en el que las tres matrices de varianza comparten dos componentes principales.

Existe otro criterio de decisión que se denomina en inglés *jump-up approach* y consiste en escoger el modelo tal que, tanto ese modelo como todos los anteriores, tienen un valor p aceptable para la prueba:

H_0 : se cumple el modelo vs. H_a las matrices no están relacionadas.

Según ese criterio, considerando la TABLA 7, a un nivel de significancia del 95 %, es decir, aceptando con cualquier valor p mayor a 0.05, se escogería el modelo CPC(3). Sin embargo, los dos primeros modelos tienen valores p muy altos (del orden de 0.7) en

las pruebas mencionadas comparados con el valor p del modelo CPC(3), que tiene un valor p de 0.1168. Por esa razón, se puede inferir que el modelo CPC(3) tiene un valor p aceptable por la “resaca” de lo bien aceptados que fueron los otros dos modelos. Esa información es confirmada por la TABLA 6. En ella vemos que las pruebas:

$$H_0 : \text{CPC}(1) \text{ vs. } H_a : \text{No relacionadas}$$

y

$$H_0 : \text{CPC}(2) \text{ vs. } H_a : \text{CPC}(1)$$

tienen con valores p muy altos, (0.7563 y 0.4841 respectivamente), en comparación con el valor p para la prueba:

$$H_0 : \text{CPC}(3) \text{ vs. } H_a : \text{CPC}(2)$$

de 0.0058.

Se presentan a continuación, las aproximaciones de los componentes principales por el Método de la Máxima Verosimilitud para el modelo CPC(2):

Tula M:

Valores propios:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
4.22	1.95	1.16	2.64	64.14	10.78	7.34

Vectores propios:

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
Largo	0.69	0.1	-0.25	-0.14	0.57	0.09	0.31
Ancho	-0.16	-0.09	0.07	-0.07	0.6	-0.55	-0.54
Cáliz	-0.36	-0.13	-0.09	0.36	0.18	-0.43	0.71
Pedicelo	0.03	-0.02	-0.32	0.83	0.14	0.3	-0.3
Antera	-0.22	0.9	0.28	0.08	0.21	0.11	0.06
Cámara	0.12	-0.33	0.83	0.21	0.23	0.26	0.1
Pistilo	-0.56	-0.21	-0.22	-0.32	0.4	0.58	0.07

Tula P:

Valores propios:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
7.44	2.36	1.48	2.67	46.14	9.24	6.43

Vectores propios:

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
Largo	0.69	0.1	-0.17	-0.07	0.57	0.18	0.34
Ancho	-0.15	-0.09	0.03	-0.24	0.63	-0.56	-0.45
Cáliz	-0.36	-0.13	-0.15	0.48	0.2	-0.38	0.65
Pedicelo	0.03	-0.02	-0.6	0.6	0.09	0.19	-0.48
Antera	-0.22	0.9	0.24	0.19	0.19	0.1	-0.03
Cámara	0.12	-0.33	0.72	0.5	0.2	0.2	-0.15
Pistilo	-0.56	-0.21	-0.09	-0.26	0.38	0.65	0.09

Taxco M:

Valores propios:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
8.5	4.75	1.69	2.15	112.1	18.69	9.2

Vectores propios:

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
Largo	0.69	0.1	-0.27	-0.06	0.61	0.17	0.2
Ancho	-0.15	-0.09	0.08	0.06	0.56	-0.59	-0.54
Cáliz	-0.36	-0.13	0.02	-0.08	0.27	-0.37	0.8
Pedicelo	0.03	-0.02	-0.03	0.99	0.01	0.01	0.11
Antera	-0.22	0.9	0.29	0.02	0.2	0.12	0.03
Cámara	0.12	-0.33	0.86	0.01	0.2	0.29	0.04
Pistilo	-0.56	-0.21	-0.3	0.01	0.4	0.62	-0.12

De acuerdo a la definición 1.1 y a la fórmula 1.9, es posible calcular la proporción de varianza que cubre cada vector común en cada modelo. Para que se tenga un ejemplo práctico se muestra la realización de ese cálculo para Tula M:

$$\frac{4.22}{4.22 + 1.95 + 1.16 + 2.64 + 64.14 + 10.78 + 7.34} = 0.05 = 5\%$$

Según el cálculo anterior la proporción de la varianza total cubierta por el primer componente principal común es 5%. Con un cálculo similar se encontró que la proporción de la varianza total cubierta por el segundo componente principal común (también en Tula M) es de 2%. La proporción de la varianza total cubierta por los dos vectores comunes es igual a la suma de la proporción de cada vector, que resulta ser, en este caso, el 7%. Haciendo las mismas cuentas para las otras dos muestras, se concluye que en Tula P, la proporción de la varianza total cubierta por los vectores comunes es de 13%, y en Taxco M del 8%.

Para más seguridad en el modelo, se comparan, también, los componentes principales, según esta aproximación por máxima verosimilitud, con los componentes principales de cada matriz por separado (mostrados al principio de esta sección). Aunque el primer componente principal común puede ser el sexto en Tula M y el cuarto en Taxco M; y que el segundo componente principal común puede ser el sexto componente principal de las matrices de Tula M y Tula P por separado, no es claro, que los componentes comunes, según las aproximaciones por máxima verosimilitud, correspondan a componentes principales comunes “verdaderos” según los cálculos en cada matriz por separado. Por ello, no es muy claro que haya dos componentes principales comunes, pero sí, que casi el total de la variación está en el espacio representado por los componentes principales no comunes, que es un espacio de dimensión cinco. Con base en lo anterior, se puede sugerir un modelo de subespacio común de dimensión dos, aunque el programa que se utiliza no permita realizar esta prueba de hipótesis.

Un problema que se puede presentar en los modelos basados en componentes principales comunes es que los *scores* de los componentes principales comunes no sean ortogonales. Para ejemplificar cómo detectar ese problema, en uno de esos modelos, se muestra el modelo completo de componentes principales comunes, que ya se vio que no se ajusta bien a los datos:

Los componentes principales comunes, según el método de máxima verosimilitud son:

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
Largo	0.59	0.17	0.25	0.7	-0.09	0.09	-0.23
Ancho	0.6	-0.6	-0.48	-0.16	-0.1	-0.13	0.04
Cáliz	0.22	-0.35	0.77	-0.35	0.33	-0.02	-0.04
Pedicelo	0.08	0.14	-0.33	-0.02	0.83	0.2	-0.36
Antera	0.2	0.13	-0.01	-0.23	-0.13	0.88	0.31
Cámara	0.21	0.25	-0.01	0.11	0.36	-0.29	0.82
Pistilo	0.39	0.62	-0.01	-0.55	-0.19	-0.27	-0.23

Y sus valores propios son:

Para Tula M:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
63.75	10.62	7.6	4.24	2.92	1.9	1.2

Para Tula P:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
46.04	9.2	6.23	7.46	2.82	2.3	1.7

Para Taxco M:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
111.25	18.36	8.42	8.51	3.75	4.93	1.86

Las matrices de correlación entre los componentes principales comunes estimados según el modelo son:

Para Tula M:

$$\begin{bmatrix} 1 & 0.06 & -0.15 & -0.12 & 0.22 & -0.07 & -0.01 \\ 0.06 & 1 & -0.07 & 0.03 & 0.18 & 0.13 & -0.03 \\ -0.15 & -0.07 & 1 & 0.03 & -0.05 & 0.04 & 0.15 \\ -0.12 & 0.03 & 0.03 & 1 & 0.01 & -0.1 & 0.09 \\ 0.22 & 0.18 & -0.05 & 0.01 & 1 & 0.03 & -0.03 \\ -0.07 & 0.13 & 0.04 & -0.1 & 0.03 & 1 & 0.06 \\ -0.01 & -0.03 & 0.15 & 0.09 & -0.03 & 0.06 & 1 \end{bmatrix}$$

Para Tula P:

$$\begin{bmatrix} 1 & -0.05 & -0.1 & 0.06 & -0.02 & 0.13 & -0.02 \\ -0.05 & 1 & -0.02 & -0.01 & 0.02 & -0.07 & -0.16 \\ -0.1 & -0.02 & 1 & 0.06 & -0.19 & -0.08 & -0.11 \\ 0.06 & -0.01 & 0.06 & 1 & 0.01 & 0.02 & -0.06 \\ -0.02 & 0.02 & -0.19 & 0.01 & 1 & -0.02 & 0.18 \\ 0.13 & -0.07 & -0.08 & 0.02 & -0.02 & 1 & 0.11 \\ -0.02 & -0.16 & -0.11 & -0.06 & 0.18 & 0.11 & 1 \end{bmatrix}$$

Para Taxco M:

$$\begin{bmatrix} 1 & 0 & 0.27 & 0.11 & -0.2 & -0.11 & 0.04 \\ 0 & 1 & 0.06 & -0.05 & -0.19 & -0.09 & 0.22 \\ 0.27 & 0.06 & 1 & 0 & 0.4 & 0.12 & -0.07 \\ 0.11 & -0.05 & 0 & 1 & -0.03 & 0.1 & 0.01 \\ -0.2 & -0.19 & 0.4 & -0.03 & 1 & 0.03 & -0.17 \\ -0.11 & -0.09 & 0.12 & 0.1 & 0.03 & 1 & -0.09 \\ 0.04 & 0.22 & -0.07 & 0.01 & -0.17 & -0.09 & 1 \end{bmatrix}$$

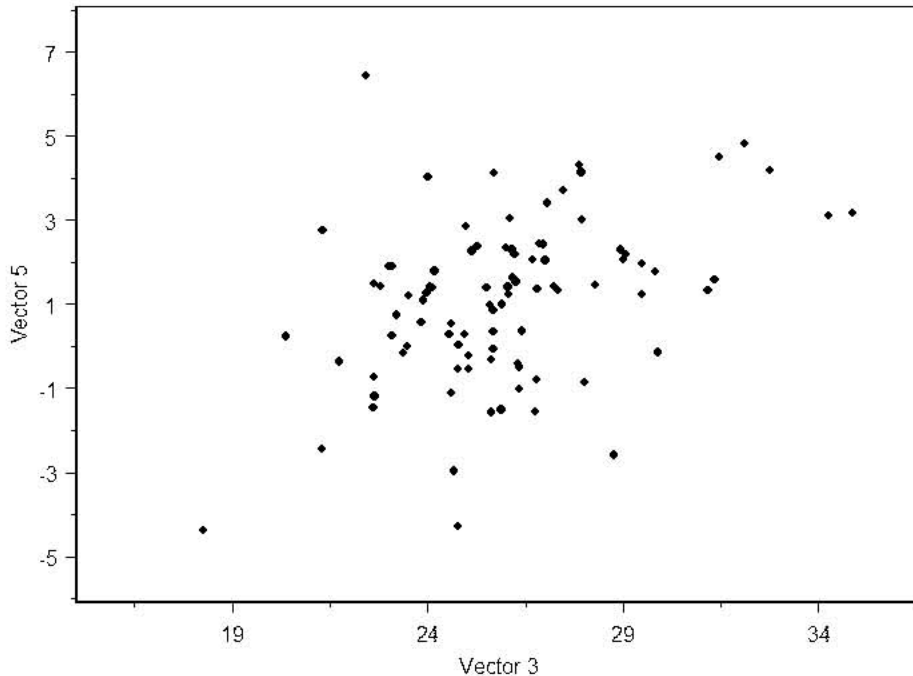


FIGURA 17. Diagrama de dispersión de los componentes 3 y 5 para Taxco M según CPC de Tula M y P y Taxco M con 7 variables

Las matrices anteriores muestran que las mayores correlaciones se encuentran en la matriz que corresponde a Taxco M. La mayor es entre el tercer y el quinto componente principal, y es del orden de 0.4. Si se observa el diagrama de dispersión de los valores calculados para dichos componentes principales, en la FIGURA 17, se observa que hay una correlación positiva entre estos componentes.

Debido a que las matrices de correlación se ajustaron mejor a Tula M y Tula P, parece conveniente realizar una comparación entre las matrices de Tula M y Tula P. Los resultados de esa comparación se muestran en las TABLAS 8 y 9. Según el Criterio de Información de Akaike, el mejor modelo es el de Componentes Principales Comunes. Y, efectivamente, al observar la TABLA 9 se comprueba que el modelo CPC se acepta con un valor p de 0.1076.

TABLA 8. Criterio de Información de Akaike para la comparación Tula M, Tula P con 7 variables

Modelo		T	GL	Valor p	T/GL	AIC
Mayor (H_0 :)	Menor (H_a :)					
Igualdad	Proporcionalidad	0.396	1	0.5292	0.396	46.845
Proporcionalidad	CPC	17.175	6	0.0087	2.862	48.449
CPC	CPC(5)	0.142	1	0.7064	0.142	43.274 *
CPC(5)	CPC(4)	3.397	2	0.1830	1.698	45.132
CPC(4)	CPC(3)	3.170	3	0.3662	1.057	45.736
CPC(3)	CPC(2)	10.062	4	0.0394	2.516	48.566
CPC(2)	CPC(1)	3.436	5	0.6331	0.687	46.504
CPC(1)	No Relacionadas	9.068	6	0.1698	1.511	53.068
No Relacionadas	—					56.000

TABLA 9. Valores p para la comparación Tula M, Tula P con 7 variables

Modelo		T	GL	valor p
Mayor (H_0 :)	Menor (H_a :)			
Igualdad	No relacionadas	46.845	28	0.0142
Proporcionalidad	No relacionadas	46.449	27	0.0114
CPC	No relacionadas	29.274	21	0.1076
CPC(5)	No relacionadas	29.132	20	0.0852
CPC(4)	No relacionadas	25.736	18	0.1060
CPC(3)	No relacionadas	22.566	15	0.0938
CPC(2)	No relacionadas	12.504	11	0.3270
CPC(1)	No relacionadas	9.068	6	0.1698

Para probar qué tan bueno es ese modelo, se calculan las matrices de correlación en cada muestra de los Componentes Principales Comunes estimados por el método de máxima verosimilitud, obteniendo los siguientes resultados:

Tula Madre:

$$(3.1) \quad \begin{bmatrix} 1 & -0.06 & -0.02 & -0.07 & 0.11 & -0.11 & 0.02 \\ -0.06 & 1 & 0.03 & -0.01 & -0.07 & -0.1 & -0.03 \\ -0.02 & 0.03 & 1 & 0.03 & 0.14 & 0.08 & 0.11 \\ -0.07 & -0.01 & 0.03 & 1 & 0.02 & -0.07 & 0.12 \\ 0.11 & -0.07 & 0.14 & 0.02 & 1 & 0 & -0.11 \\ -0.11 & -0.1 & 0.08 & -0.07 & 0 & 1 & -0.02 \\ 0.02 & -0.03 & 0.11 & 0.12 & -0.11 & -0.02 & 1 \end{bmatrix}$$

Tula Padre:

$$(3.2) \quad \begin{bmatrix} 1 & 0.05 & 0.02 & 0.09 & -0.1 & 0.11 & -0.02 \\ 0.05 & 1 & -0.02 & 0.02 & 0.06 & 0.09 & 0.03 \\ 0.02 & -0.02 & 1 & 0.06 & -0.12 & -0.07 & -0.1 \\ 0.09 & 0.02 & 0.06 & 1 & -0.01 & 0.03 & -0.06 \\ -0.1 & 0.06 & -0.12 & -0.01 & 1 & 0 & 0.11 \\ 0.11 & 0.09 & -0.08 & 0.03 & 0 & 1 & 0.02 \\ -0.02 & 0.03 & -0.1 & -0.06 & 0.11 & 0.02 & 1 \end{bmatrix}$$

La máxima correlación entre dos componentes principales estimados se da entre el tercer y el quinto componente de Tula Madre y que es del orden de 0.14. Para un problema así, se consideran altas las correlaciones arriba de 0.25, por lo que ninguna de estas correlaciones implica un problema para el modelo.

A continuación se reproducen los estimados de máxima verosimilitud dado el modelo CPC:

Valores Propios:

Tula Madre:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
64.09	10.82	7.23	4.19	2.77	1.91	1.22

Tula Padre:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
46.05	9.22	6.29	7.54	2.8	2.35	1.52

Vectores propios de la matriz CPC:

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
Largo	0.56	-0.16	0.29	0.7	-0.12	0.06	-0.22
Ancho	0.62	0.56	-0.5	-0.15	-0.13	-0.11	0.06
Cáliz	0.2	0.42	0.7	-0.3	0.43	-0.05	-0.1
Pedicelo	0.11	-0.26	-0.4	-0.02	0.72	0.07	-0.48
Antera	0.2	-0.11	0.03	-0.25	-0.07	0.93	0.13
Cámara	0.22	-0.24	-0.02	0.12	0.44	-0.13	0.82
Pistilo	0.4	-0.59	0.13	-0.55	-0.26	-0.31	-0.11

Es claro que los primeros dos componentes principales comunes, según el modelo de máxima verosimilitud, corresponden respectivamente a los dos primeros componentes principales de cada matriz calculados por separado (mostrados al principio de la sección 3.3.1). En el caso de Tula M se infiere, aunque ya de manera menos clara, que los últimos

cinco componentes principales comunes, en el modelo de máxima verosimilitud, se parecen a los respectivos componentes calculados directamente de la matriz. Además, los valores propios correspondientes son similares. Lo mismo sucede en Tula P, excepto en los casos del tercer y cuarto componentes, que no es muy claro que coincidan. Aún comparando el tercer componente del modelo común con el cuarto del modelo por separado; y el cuarto del modelo común con el tercero del modelo por separado; no es claro que esos vectores propios sean los mismos. Los valores propios del modelo común son muy similares a los de los modelos por separado, aunque no en el mismo orden de la relación de los vectores propios, ello hace dudar si la comparación de los vectores propios en Tula P está bien ordenada y qué tanto se ajusta el modelo.

En el diagrama de dispersión del tercer y quinto componentes en Tula M (FIGURA 18), se observa que la correlación es baja, pues no es clara a simple vista.

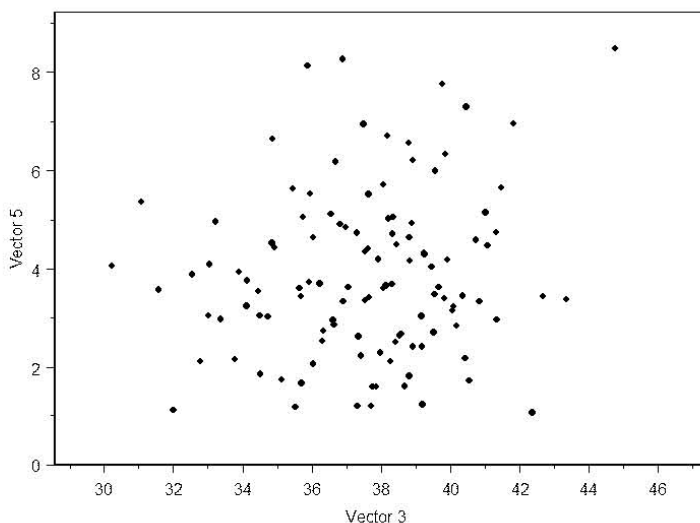


FIGURA 18. Diagrama de dispersión de los componentes 3 y 5 en Tula M para el modelo CPC de Tula M y Tula P

Ahora se repite la comparación de las matrices de varianza de Tula M y Tula P, pero sólo con tres variables: largo, ancho y cáliz. Los resultados de dicho análisis se encuentran en las TABLAS 10 y 11.

Como se puede ver en la TABLA 10, el modelo con el menor valor para el criterio de Akaike es el de igualdad. La TABLA 11 confirma esa información, pues con un nivel de significancia del 90 % se aceptan todos los modelos hasta la igualdad. En particular, el modelo de igualdad se acepta con un valor p de 0.3412, que es bastante alto.

TABLA 10. Criterio de Información de Akaike para la comparación Tula M, Tula P con 3 variables

Modelo		T	GL	Valor p	T/GL	AIC
Mayor (H_0 :)	Menor (H_a :)					
Igualdad	Proporcionalidad	0.822	1	0.3645	0.822	6.785*
Proporcionalidad	CPC	4.772	2	0.0920	2.386	7.963
CPC	CPC(1)	0.118	1	0.7315	0.118	7.191
CPC(1)	No relacionadas	1.073	2	0.5848	0.536	9.073
No relacionadas						12.000

TABLA 11. Valores p para la comparación Tula M, Tula P con 3 variables

Modelo		T	GL	valor p
Mayor (H_0 :)	Menor (H_a :)			
Igualdad	No Relacionadas	6.785	6	0.3412
Proporcionalidad	No Relacionadas	5.963	5	0.3098
CPC	No Relacionadas	1.191	3	0.7553
CPC(1)	No Relacionadas	1.073	2	0.5848

Suponiendo que las matrices son iguales, la mejor matriz estimada para las dos poblaciones juntas es la siguiente:

	Largo	Ancho	Cáliz
Largo	21.65	16.31	5.33
Ancho	16.31	25.44	6.54
Cáliz	5.33	6.54	8

Los componentes principales de la matriz anterior son los siguientes:

Valores propios:

Vector 1	Vector 2	Vector 3
7,22	5,82	42,1

Vectores propios:

	Vector 1	Vector 2	Vector 3
Largo	0.76	-0.04	-0.64
Ancho	-0.59	0.35	-0.73
Cáliz	-0.26	-0.94	-0.24

Las matrices de correlación para esos componentes principales en cada una de las muestras son:

En Tula M:

$$\begin{bmatrix} 1 & -0.04 & 0.07 \\ -0.04 & 1 & 0.02 \\ 0.07 & 0.02 & 1 \end{bmatrix}$$

En Tula P:

$$\begin{bmatrix} 1 & 0.03 & -0.05 \\ 0.03 & 1 & -0.02 \\ -0.05 & -0.02 & 1 \end{bmatrix}$$

La mayor correlación, en valor absoluto, se da entre el primer y el tercer vector de Tula M y es del orden de 0.07, que es muy bajo, por lo que, al menos en sus componentes principales, el modelo se ajusta bien.

Finalmente, se presentan las matrices de varianza de cada una de las muestras por separado para comparalas con la mejor matriz común de varianza y ver qué tanto difieren:

Tula M:

	Largo	Ancho	Cáliz
Largo	22.35	19.1	6.77
Ancho	19.1	29.34	7.63
Cáliz	6.77	7.63	9.19

Tula P:

	Largo	Ancho	Cáliz
Largo	21.14	14.28	4.28
Ancho	14.28	22.61	5.74
Cáliz	4.28	5.74	7.13

Al comparar esas matrices entre sí, así como con la matriz estimada de máxima verosimilitud se confirma que es bastante plausible que las matrices sean iguales. Esto, a pesar de que la covarianza de largo con ancho, y la varianza del ancho, difieren mucho entre una matriz y otra.

En los siguientes análisis, se comparan las matrices de las dos poblaciones de Taxco para saber si se comportan entre sí de la misma manera que las matrices de Tula. Los resultados de aplicar la prueba entre Taxco M y Taxco P, tomando en cuenta siete variables, se encuentran en las TABLAS 12 y 13.

TABLA 12. Criterio de Información de Akaike para la comparación Taxco M, Taxco P, con 7 variables

Modelo		T	GL	Valor p	$T/$ GL	AIC
Mayor (H_0 :)	Menor (H_a :)					
Igualdad	Proporcionalidad	4.341	1	0.0372	4.341	60.040
Proporcionalidad	CPC	9.964	6	0.1262	1.661	57.698
CPC	CPC(5)	3.959	1	0.0466	3.959	59.734
CPC(5)	CPC(4)	16.959	2	0.0002	8.480	57.775
CPC(4)	CPC(3)	11.058	3	0.0114	3.686	44.816
CPC(3)	CPC(2)	10.093	4	0.0389	2.523	39.757
CPC(2)	CPC(1)	1.572	5	0.9046	0.314	37.664*
CPC(1)	No relacionadas	2.092	6	0.9111	0.349	46.092
No relacionadas						56.000

TABLA 13. Valores p para la comparación Taxco M, Taxco P, con 7 variables

Modelo		T	GL	valor p
Mayor (H_0 :)	Menor (H_a :)			
Igualdad	No relacionadas	60.040	28	0.0004
Proporcionalidad	No relacionadas	55.698	27	0.0009
CPC	No relacionadas	45.734	21	0.0014
CPC(5)	No relacionadas	41.775	20	0.0030
CPC(4)	No relacionadas	24.816	18	0.1301
CPC(3)	No relacionadas	13.757	15	0.5440
CPC(2)	No relacionadas	3.664	11	0.9787
CPC(1)	No relacionadas	2.092	6	0.9111

La TABLA 12 muestra que el modelo con el menor criterio de Akaike es CPC(2). La TABLA 13 apoya la información anterior, ya que en el modelo CPC(2) se tiene un valor p de 0.9787. Si para escoger modelo se usara el criterio *jump-up*, a un nivel de significancia del 90 %, se aceptaría el modelo CPC(4), de acuerdo a la TABLA 13. Sin embargo, en la TABLA 12 se puede observar que las pruebas de hipótesis:

$$H_0 : \text{CPC}(1) \text{ vs. } H_a : \text{No relacionadas y } H_0 : \text{CPC}(2) \text{ vs. } H_a : \text{CPC}(1)$$

se aceptan con valores p muy altos, sin embargo las pruebas:

$$H_0 : \text{CPC}(3) \text{ vs. } H_a : \text{CPC}(2) \text{ y } H_0 : \text{CPC}(4) \text{ vs. } H_a : \text{CPC}(3)$$

tienen valores p extremadamente bajos. Esto da una idea de que la aceptación de los modelos CPC(3) y CPC(4) se debe a la “resaca” de la excelente aceptación de CPC(2). Así mismo, si se observa la descomposición de la T total de 60.040, los pasos CPC(4) a CPC(3); y CPC(3) a CPC(2) representan valores de 11.058 y 10.093 respectivamente,

que son muy grandes para los grados de libertad que representan. Por todo lo anterior, el modelo elegido es CPC(2); sin embargo, como ya se reprodujo un ejemplo de un modelo parcial de componentes principales comunes, no se reproduce este modelo en su totalidad, sólo se reproducen sus valores propios y los componentes principales comunes estimados:
Valores Propios:

Taxco M:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
9.31	4.71	2.14	1.68	112.2	18.56	8.47

Taxco P:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
8.8	2.68	2.23	1.67	90.69	20.92	6.41

Componentes principales comunes estimados:

	Vector 1	Vector 2
Largo	-0.12	-0.11
Ancho	-0.46	0.09
Cáliz	0.83	0.18
Pedicelo	0.1	-0.03
Antera	0.15	-0.88
Cámara	-0.02	0.39
Pistilo	0.22	0.12

Los componentes principales comunes corresponden a los componentes principales 3 y 5, respectivamente, en ambas muestras por separado, sin embargo, sus valores propios estimados según máxima verosimilitud no son los de mayor varianza. En realidad representan sólo un 9% de la varianza total en cada muestra.

Ahora se presenta la comparación de las mismas dos muestras, pero considerando únicamente tres variables: largo, ancho y cáliz. Los resultados de esa prueba se encuentran en las TABLAS 14 y 15.

TABLA 14. Criterio de Información de Akaike para la comparación Taxco M y Taxco P con 3 variables

Modelo		T	GL	Valor p	T/GL	AIC
Mayor (H_0 :)	Menor (H_a :)					
Igualdad	Proporcionalidad	0.085	1	0.7705	0.085	19.116
Proporcionalidad	CPC	3.243	2	0.1976	1.622	21.030
CPC	CPC(1)	14.374	1	0.0001	14.374	21.787
CPC(1)	No relacionadas	1.414	2	0.4932	0.707	9.414*
No relacionadas						12.000

TABLA 15. Valores p para la comparación Taxco M, Taxco P con 3 variables

Modelo		T	GL	valor p
Mayor (H_0 :)	Menor (H_a :)			
Igualdad	No relacionadas	19.116	6	0.0040
Proporcionalidad	No relacionadas	19.030	5	0.0019
CPC	No relacionadas	15.787	3	0.0013
CPC(1)	No relacionadas	1.414	2	0.4932

En la TABLA 14 se observa que el modelo con el menor criterio de Akaike es el que tiene un componente principal común a las dos poblaciones. La TABLA 15 confirma que ese modelo se acepta con un valor p de 0.4932, y que, el siguiente modelo no se acepta por tener un valor p muy bajo. No se reproduce el modelo aceptado, pero se menciona que el componente principal común a las dos poblaciones es el de menor importancia representando el 8% de la varianza total en cada población.

Vale la pena destacar que con siete variables se compartieron dos componentes principales, y, al reducir a tres variables, sólo hubo un componente principal común. Es decir, reducir el número de variables a considerar no implica que se vayan a compartir más, o al menos el mismo número de componentes principales. Para entender esto hay que recordar que, cuando se tienen siete variables, se tienen siete componentes principales, si dos son comunes a ambas matrices, dos séptimos de los componentes son comunes. Si se tienen tres variables y un componente principal común, entonces un tercio de los componentes principales son comunes. También es importante recalcar que los componentes principales que se comparten en el modelo con siete variables, en este caso en particular, no son los de mayor importancia, en realidad, tienen una importancia pequeña, por lo que, al reducir el modelo a tres variables, puede que esos dos componentes principales comunes no representen casi nada de la nueva matriz de varianza.

De toda la información anterior, se puede concluir que las matrices de las poblaciones de Tula se parecen mucho más entre sí que las matrices de Taxco. Además, al comparar grupos de tres matrices, se obtuvieron los mejores resultados excluyendo a Taxco P. Lo

anterior, hace suponer que hubo algo que generó un cambio en la muestra de Taxco P. Esto pudo haberse debido a una interacción genético-ambiental que hizo que el cambio de macetas pequeñas a grandes haya influido en Taxco, pero no en Tula; o a que, por alguna razón, la muestra de Taxco haya sido expuesta a circunstancias diferentes en el invernadero, posibilidad descartada por el ecólogo. Sin embargo, tal vez se debió simplemente a que la muestra obtenida es atípica y, en consecuencia, su matriz de varianzas no es representativa. De cualquier manera, para descartar esa variación, que no se comprende muy bien, se utilizarán únicamente las muestras madre para comparar las poblaciones de Taxco y Tula.

Primero se realiza una prueba con siete variables. Los resultados se muestran en las TABLAS 16 y 17.

TABLA 16. Criterio de Información de Akaike para la comparación Taxco M y Tula M con 7 variables

Modelo		T	GL	Valor p	T/GL	AIC
Mayor (H_0 :)	Menor (H_a :)					
Igualdad	Proporcionalidad	39.485	1	0.0000	39.485	118.521
Proporcionalidad	CPC	16.715	6	0.0104	2.786	81.036
CPC	CPC(5)	21.010	1	0.0000	21.010	76.321
CPC(5)	CPC(4)	7.614	2	0.0222	3.807	57.311
CPC(4)	CPC(3)	21.582	3	0.0001	7.194	53.697
CPC(3)	CPC(2)	4.147	4	0.3865	1.037	38.115*
CPC(2)	CPC(1)	2.852	5	0.7228	0.570	41.969
CPC(1)	No relacionadas	5.117	6	0.5290	0.853	49.117
No relacionadas						56.000

TABLA 17. Valores p para la comparación Taxco M, Tula M con 7 variables

Modelo		T	GL	valor p
Mayor (H_0 :)	Menor (H_a :)			
Igualdad	No relacionadas	118.521	28	0.0000
Proporcionalidad	No relacionadas	79.036	27	0.0000
CPC	No relacionadas	62.321	21	0.0000
CPC(5)	No relacionadas	41.311	20	0.0034
CPC(4)	No relacionadas	33.697	18	0.0137
CPC(3)	No relacionadas	12.115	15	0.6703
CPC(2)	No relacionadas	7.969	11	0.7161
CPC(1)	No relacionadas	5.117	6	0.5290

En las TABLAS 16 y 17 se puede ver que, según los dos criterios de selección de modelos ya mencionados, el mejor modelo es CPC(3). Se presentan, del modelo CPC(3) estimado por máxima verosimilitud, todos los valores propios y los vectores propios comunes:

Tula M:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
1,84	4,24	1,19	64,2	10,8	7,32	2,64

Taxco M:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
4,92	8,5	1,73	112,09	18,59	9,12	2,12

Vectores propios comunes:

	Vector 1	Vector 2	Vector 3
Largo	0,15	0,66	-0,27
Ancho	-0,07	-0,09	0,07
Cáliz	-0,11	-0,41	-0,02
Pedicelo	0,04	0	-0,25
Antera	0,87	-0,27	0,34
Cámara	-0,38	0,15	0,82
Pistilo	-0,25	-0,54	-0,26

El primer componente principal común corresponde al quinto de las muestras por separado. El segundo y el tercero se asemejan ligeramente al cuarto y al séptimo de las muestras por separado, respectivamente, aunque, no puede decirse con seguridad que corresponden a ellos. Los tres componentes principales comunes, en conjunto, representan el 8% de la varianza total en Tula M, y el 10% en Taxco M, por lo que no son muy significativos.

Se repite la comparación de las mismas muestras, pero ahora sólo con cinco variables: largo, ancho, cáliz, antera y pistilo. Los resultados se encuentran en las TABLAS 18 y 19.

Según el Criterio de Akaike, de acuerdo a la información de la TABLA 18, el mejor modelo es el de proporcionalidad. En la misma TABLA se puede ver que el modelo de igualdad tiene un valor en el criterio de Akaike mucho más alto. La diferencia con respecto al valor de Akaike del modelo CPC es más moderada, pero significativa. Siguiendo ahora el *jump-up approach*, se observa en la TABLA 19 que se escogería el mismo modelo que con el Criterio de Akaike. Según la misma tabla, la igualdad de las matrices se descarta en forma tajante, por lo que éste es un claro ejemplo de proporcionalidad.

A continuación se muestra el modelo de proporcionalidad según máxima verosimilitud:

Taxco M:

Matriz de varianza estimada:

	Largo	Ancho	Cáliz	Antera	Pistilo
Largo	44.83	35.55	14.32	12.03	26.05
Ancho	35.55	49.86	15.28	10.76	21.86
Cáliz	14.32	15.28	17.04	4.42	7.17
Antera	12.03	10.76	4.42	8.18	10.31
Pistilo	26.05	21.85	7.17	10.31	29.06

Constante de proporcionalidad: 1

Tula M:

Matriz de varianza estimada:

	Largo	Ancho	Cáliz	Antera	Pistilo
Largo	24.26	19.24	7.75	6.51	14.1
Ancho	19.24	26.99	8.27	5.82	11.83
Cáliz	7.75	8.27	9.22	2.39	3.88
Antera	6.51	5.82	2.39	4.43	5.58
Pistilo	14.1	11.83	3.88	5.58	15.73

Constante de proporcionalidad: 0.54

TABLA 18. Criterio de Información de Akaike para la comparación Taxco M y Tula M con 5 variables

Modelo		T	GL	Valor p	T/GL	AIC
Mayor (H_0 :)	Menor (H_a :)					
Igualdad	Proporcionalidad	46.217	1	0.0000	46.217	60.954
Proporcionalidad	CPC	3.887	4	0.4215	0.972	16.737*
CPC	CPC(3)	2.376	1	0.1232	2.376	20.849
CPC(3)	CPC(2)	0.305	2	0.8587	0.152	20.474
CPC(2)	CPC(1)	0.773	3	0.8559	0.258	24.169
CPC(1)	No relacionadas	7.396	4	0.1164	1.849	29.396
No relacionadas						30.000

TABLA 19. Valores p para la comparación Taxco M, Tula M con 5 variables

Modelo		T	GL	valor p
Mayor (H_0 :)	Menor (H_a :)			
Igualdad	No relacionadas	60.954	15	0.0000
Proporcionalidad	No relacionadas	14.737	14	0.3964
CPC	No relacionadas	10.849	10	0.3694
CPC(3)	No relacionadas	8.474	9	0.4872
CPC(2)	No relacionadas	8.169	7	0.3179
CPC(1)	No relacionadas	7.396	4	0.1164

Como punto de referencia para saber qué tan adecuado es el modelo de proporcionalidad, se muestran las matrices de varianza estimadas para cada población por separado:

Taxco M:

	Largo	Ancho	Cáliz	Antera	Pistilo
Largo	49.27	35.89	16.6	12.26	25.37
Ancho	35.89	44.41	16.77	9.27	18.51
Cáliz	16.6	16.77	17.12	4.76	8.01
Antera	12.26	9.27	4.76	7.78	8.72
Pistilo	25.37	18.51	8.01	8.72	27.69

Tula M:

	Largo	Ancho	Cáliz	Antera	Pistilo
Largo	22.35	19.1	6.77	6.42	14.39
Ancho	19.1	29.34	7.63	6.46	13.27
Cáliz	6.77	7.63	9.19	2.24	3.51
Antera	6.42	6.46	2.24	4.6	6.27
Pistilo	14.39	13.27	3.51	6.27	16.32

Al comparar las matrices estimadas por máxima verosimilitud suponiendo proporcionalidad, con las matrices de varianza estimadas por separado, se concluye que el modelo de proporcionalidad se ajusta bastante bien.

3.3.2. Muestras Silvestres de 2004. Se cuenta con tres muestras que corresponden a: Ticumán tempranas, Ticumán tardías y Teotihuacán. A continuación se muestran sus matrices de varianza y componentes principales calculados para cada matriz individualmente:

Ticumán Tempranas:

Matriz de varianza (calculada con 100 datos):

	Largo	Ancho	Cáliz	Pedicelo	Antera	Pistilo	Cámara
Largo	70.81	42.26	12.38	4.16	56.65	56.41	29.27
Ancho	42.26	50.12	10.76	3.78	29.38	39.31	11.2
Cáliz	12.38	10.76	9.21	1.51	8.76	9.6	4.75
Pedicelo	4.16	3.78	1.51	1.67	3.33	2.91	1.25
Antera	56.65	29.38	8.76	3.33	63.74	48.89	28.27
Pistilo	56.41	39.31	9.6	2.91	48.89	63.52	21.94
Cámara	29.27	11.2	4.75	1.25	28.27	21.94	18.9

Componentes Principales:

Valores propios:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
215.28	30.71	12.98	8.45	6.03	3.31	1.22

Vectores propios:

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
Largo	0.55	-0.06	0.28	0.55	0.33	-0.45	0.07
Ancho	0.37	0.77	0.29	-0.32	0.14	0.25	0.03
Cáliz	0.10	0.13	0.28	0.24	-0.91	-0.08	0.11
Pedicelo	0.03	0.04	0.09	-0.02	-0.08	-0.13	-0.98
Antera	0.49	-0.50	0.14	-0.67	-0.12	-0.14	0.05
Pistilo	0.50	0.12	-0.83	0.11	-0.17	0.09	-0.06
Cámara	0.23	-0.34	0.22	0.29	0.03	0.83	-0.11

Ticumán Tardías:*Matriz de varianza (calculada con 208 datos):*

	Largo	Ancho	Cáliz	Pedicelo	Antera	Pistilo	Cámara
Largo	78.54	56.79	26.3	3.55	47.38	48.41	26.58
Ancho	56.79	63.49	24.27	3.39	34.44	42.34	17.33
Cáliz	26.3	24.27	19.35	1.81	13.97	15.36	6.86
Pedicelo	3.55	3.39	1.81	1.36	1.05	2.36	0.66
Antera	47.38	34.44	13.97	1.05	46.87	37.57	18.88
Pistilo	48.41	42.34	15.36	2.36	37.57	53.95	17.67
Cámara	26.58	17.33	6.86	0.66	18.88	17.67	13.87

Componentes Principales:

Valores propios:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
214.93	24.03	17.06	9.17	7.80	3.40	1.04

Vectores propios:

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
Largo	0.58	-0.06	0.52	-0.47	0.12	0.39	0.08
Ancho	0.49	-0.60	-0.31	0.37	0.40	-0.07	0.01
Cáliz	0.21	-0.40	0.17	0.08	-0.84	-0.24	0.03
Pedicelo	0.03	-0.06	-0.01	-0.07	-0.03	0.10	-0.99
Antera	0.40	0.56	0.20	0.68	-0.10	0.10	-0.06
Pistilo	0.44	0.34	-0.71	-0.36	-0.23	-0.01	0.03
Cámara	0.20	0.18	0.24	-0.19	0.24	-0.87	-0.09

Teotihuacán:*Matriz de varianza (calculada con 489 datos):*

	Largo	Ancho	Cáliz	Pedicelo	Antera	Pistilo	Cámara
Largo	48.84	43.23	17.88	4.26	26.21	29.78	14.35
Ancho	43.23	63.68	15.89	2.58	21.71	28.98	10.06
Cáliz	17.88	15.89	15.44	1.36	9.42	11.45	5.23
Pedicelo	4.26	2.58	1.36	2.45	1.94	2.48	1.17
Antera	26.21	21.71	9.42	1.94	18.55	18.61	9.01
Pistilo	29.78	28.98	11.45	2.48	18.61	28.52	8.43
Cámara	14.35	10.06	5.23	1.17	9.01	8.43	5.89

Componentes Principales:

Valores propios:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
142.08	19.4	9.1	7.14	3.08	1.73	0.84

Vectores propios:

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
Largo	0.56	-0.30	-0.13	-0.47	0.44	0.34	0.22
Ancho	0.60	0.78	-0.01	0.01	-0.12	-0.10	-0.07
Cáliz	0.22	-0.21	-0.82	0.46	-0.10	-0.07	0
Pedicelo	0.04	-0.07	-0.01	-0.08	0.51	-0.85	-0.06
Antera	0.31	-0.35	0.19	-0.18	-0.68	-0.37	0.35
Pistilo	0.39	-0.28	0.52	0.67	0.18	0.09	-0.12
Cámara	0.16	-0.23	-0.03	-0.27	-0.20	-0.01	-0.90

Primero se observan los componentes principales comunes por separado para ver qué se puede inferir. Lo primero que se observa es que el primer componente principal en las tres matrices tiene todos sus elementos positivos, por lo que es posible relacionarlo con el tamaño de la flor. Posteriormente, atendiendo a los detalles, se ve que hay cierto parecido entre el primer componente principal de las tres muestras, así como entre los últimos componentes principales de las muestras de Ticumán tempranas y Ticumán tardías.

Una vez observados los resultados de calcular los componentes principales de las matrices de varianzas estimadas, se procede a aplicar la teoría de los componentes principales comunes, en especial las pruebas de hipótesis y el criterio de Akaike.

Lo primero que hay que saber es la relación que existe entre las matrices de varianzas de las dos muestras de Ticumán. Para ello, se corre un análisis de componentes principales comunes considerando las matrices de las muestras Ticumán tempranas y Ticumán tardías. Los resultados de dicho análisis se encuentran en las TABLAS 20 y 21.

TABLA 20. Criterio de Información de Akaike para la comparación Ticumán Tempranas y Ticumán Tardías con 7 variables

Modelo		T	GL	Valor p	T/GL	AIC
Mayor (H_0 :)	Menor (H_a :)					
Igualdad	Proporcionalidad	0.113	1	0.7366	0.113	72.699
Proporcionalidad	CPC	10.251	6	0.1145	1.709	74.586
CPC	CPC(5)	4.451	1	0.0349	4.451	76.335
CPC(5)	CPC(4)	18.601	2	0.0001	9.301	73.883
CPC(4)	CPC(3)	10.591	3	0.0142	3.530	59.282
CPC(3)	CPC(2)	11.462	4	0.0218	2.865	54.691
CPC(2)	CPC(1)	7.714	5	0.1727	1.543	51.229*
CPC(1)	No relacionadas	9.515	6	0.1466	1.586	53.515
No relacionadas						56.000

TABLA 21. Valores p para la comparación Ticumán Tempranas y Ticumán Tardías con 7 variables

Modelo		T	GL	valor p
Mayor (H_0 :)	Menor (H_a :)			
Igualdad	No relacionadas	72.699	28	0.0000
Proporcionalidad	No relacionadas	72.586	27	0.0000
CPC	No relacionadas	62.335	21	0.0000
CPC(5)	No relacionadas	57.883	20	0.0000
CPC(4)	No relacionadas	39.282	18	0.0026
CPC(3)	No relacionadas	28.691	15	0.0176
CPC(2)	No relacionadas	17.229	11	0.1013
CPC(1)	No relacionadas	9.515	6	0.1466

Según cualquiera de los dos criterios, el modelo elegido es CPC(2). Los valores propios calculados considerando ese modelo son:

Tempranas:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
6,1	8,9	3,5	1,3	215,21	30,59	12,38

Tardías:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7
7,95	9,13	3,41	1,05	214,91	24	16,99

Los vectores propios comunes, correspondientes a los dos primeros valores propios de cada población, en el orden mostrado arriba, son:

	Vector 1	Vector 2
Largo	-0,26	-0,44
Ancho	-0,27	0,43
Cáliz	0,84	-0,12
Pedicelo	0,03	-0,04
Antera	0,2	0,66
Pistilo	0,2	-0,37
Cámara	-0,26	-0,16

Estos vectores propios son de importancia pequeña, conteniendo un 5 % de la varianza total en las observaciones tempranas y un 6 % en las tardías.

Aunque estos vectores propios podrían parecer similares a los componentes principales quinto y cuarto, respectivamente, de cada matriz por separado, la similitud no es muy clara. Lo anterior implica que las poblaciones no se parecen casi nada. Eso es extremadamente inesperado, considerando que las muestras fueron tomadas de las mismas plantas, con tan sólo un mes de diferencia.

Se puede tratar de explicar lo anterior especulando que las flores en las bifurcaciones altas de la planta no son iguales a las flores en las partes bajas de la planta, ó que estas matrices tienen un componente ambiental mucho más fuerte que el genético y representan una variación ambiental en las tres semanas entre las dos tomas de muestras.

Cabe agregar como posible hipótesis de la falta de parecido, que el modelo de Componentes Principales Comunes no fue un buen modelo para analizar los factores en juego. Esto puede ocurrir porque el supuesto de que los factores que intervienen en el desarrollo de las flores son ortogonales no es muy realista, y, como se puede ver en [4], el considerarlos ortogonales afecta mucho las conclusiones del análisis. Modelos alternativos para este tipo de comparaciones se pueden encontrar en [12].

De cualquier manera se realizó también la comparación de dichas poblaciones con cinco y tres variables.

En el caso de cinco variables se consideraron largo, ancho, cáliz, antera y pistilo.

TABLA 22. Criterio de Información de Akaike para la comparación Ticumán Tempranas y Ticumán Tardías con 5 variables

Modelo		T	GL	Valor p	T/GL	AIC
Mayor (H_0 :)	Menor (H_a :)					
Igualdad	Proporcionalidad	0.383	1	0.5360	0.383	37.452
Proporcionalidad	CPC	4.623	4	0.3282	1.156	39.069
CPC	CPC(3)	4.969	1	0.0258	4.969	42.446
CPC(3)	CPC(2)	18.781	2	0.0001	9.391	39.476
CPC(2)	CPC(1)	5.306	3	0.1507	1.769	24.695*
CPC(1)	No relacionadas	3.389	4	0.4950	0.847	25.389
No relacionadas						30.000

TABLA 23. Valores p para la comparación Ticumán Tempranas, Ticumán Tardías con 5 variables

Modelo		T	GL	valor p
Mayor (H_0 :)	Menor (H_a :)			
Igualdad	No relacionadas	37.452	15	0.0011
Proporcionalidad	No relacionadas	37.069	14	0.0007
CPC	No relacionadas	32.446	10	0.0003
CPC(3)	No relacionadas	27.476	9	0.0012
CPC(2)	No relacionadas	8.695	7	0.2753
CPC(1)	No relacionadas	3.389	4	0.4950

Como se ve en las TABLAS 22 y 23, según cualquiera de los dos criterios, el mejor modelo es CPC(2). Los valores propios correspondientes a ese modelo y sus vectores propios comunes son:

Valores propios:

Tempranas:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5
8.7	6.46	202.31	27.01	11.44

Tardías:

Vector 1	Vector 2	Vector 3	Vector 4	Vector 5
8.91	7.55	206.3	23.3	16.14

Vectores propios compartidos:

	Vector 1	Vector 2
Largo	-0.49	-0.41
Ancho	0.5	-0.18
Cáliz	-0.18	0.86
Antera	0.59	0.19
Pistilo	-0.37	0.18

Se constata que hay muy poco parecido entre las matrices de varianza, pues sólo hay dos componentes comunes y no son de gran importancia.

Finalmente, se hace la comparación de las muestras con las tres variables que mejor cumplieron la normalidad: largo, ancho y cáliz.

TABLA 24. Criterio de Información de Akaike para la comparación Ticumán Tempranas y Ticumán Tardías con 3 variables

Modelo		T	GL	Valor p	T/GL	AIC
Mayor (H_0 :)	Menor (H_a :)					
Igualdad	Proporcionalidad	0.436	1	0.5091	0.436	19.489
Proporcionalidad	CPC	4.316	2	0.1155	2.158	21.053
CPC	CPC(1)	12.902	1	0.0003	12.902	20.737
CPC(1)	No relacionadas	1.836	2	0.3994	0.918	9.836*
No relacionadas						12.000

TABLA 25. Valores p para la comparación Ticumán Tempranas y Ticumán Tardías con 3 variables

Modelo		T	GL	valor p
Mayor (H_0 :)	Menor (H_a :)			
Igualdad	No relacionadas	19.489	6	0.0034
Proporcionalidad	No relacionadas	19.053	5	0.0019
CPC	No relacionadas	14.737	3	0.0021
CPC(1)	No relacionadas	1.836	2	0.3994

Las TABLAS 24 y 25 muestran que con cualquiera de los dos criterios el modelo que se elige es CPC(1). Los valores propios y el vector propio común en dicho modelo son:

Valores propios:

Tempranas:

Vector 1	Vector 2	Vector 3
17.1	105.59	6.71

Tardías:

Vector 1	Vector 2	Vector 3
13.84	138.98	8.56

Vector propio común:

	Vector 1
Largo	-0.66
Ancho	0.74
Cáliz	0.068

Estas matrices no se parecen mucho, pues su componente principal común no es el de mayor varianza, representando 13% de la varianza en las observaciones tempranas y 9% en las tardías.

Finalmente, se compara la población de Teotihuacán con las dos poblaciones de Ticumán tomando en cuenta estas tres variables, que son las que mejor se ajustan a la distribución normal.

TABLA 26. Criterio de Información de Akaike para la comparación Ticumán Tempranas y Teotihuacán con 3 variables

Modelo		T	GL	Valor p	T/GL	AIC
Mayor (H_0 :)	Menor (H_a :)					
Igualdad	Proporcionalidad	2.521	1	0.1123	2.521	51.929
Proporcionalidad	CPC	3.564	2	0.1683	1.782	51.407
CPC	CPC(1)	25.917	1	0.0000	25.917	51.843
CPC(1)	No relacionadas	19.926	2	0.0000	9.963	27.926
No relacionadas						12.000*

TABLA 27. Valores p para la comparación Ticumán Tempranas y Teotihuacán con 3 variables

Modelo		T	GL	valor p
Mayor (H_0 :)	Menor (H_a :)			
Igualdad	No relacionadas	51.929	6	0.0000
Proporcionalidad	No relacionadas	49.407	5	0.0000
CPC	No relacionadas	45.843	3	0.0000
CPC(1)	No relacionadas	19.926	2	0.0000

En las TABLAS 26 y 27 se encuentran los resultados de las comparaciones de las matrices de varianza de Ticumán tempranas y Teotihuacán. La conclusión, según cualquiera de los dos criterios, es que no están relacionadas, y, por lo tanto, no se prosigue el análisis con más variables.

TABLA 28. Criterio de Información de Akaike para la comparación Ticumán Tardías y Teotihuacán con 3 variables

Modelo		T	GL	Valor p	T/GL	AIC
Mayor (H_0 :)	Menor (H_a :)					
Igualdad	Proporcionalidad	6.480	1	0.0109	6.480	49.228
Proporcionalidad	CPC	8.858	2	0.0119	4.429	44.748
CPC	CPC(1)	13.109	1	0.0003	13.109	39.891
CPC(1)	No relacionadas	20.782	2	0.0000	10.391	28.782
No relacionadas						12.000*

TABLA 29. Valores p para la comparación Ticumán Tardías, Teotihuacán con 3 variables

Modelo		T	GL	valor p
Mayor (H_0 :)	Menor (H_a :)			
Igualdad	No relacionadas	49.228	6	0.0000
Proporcionalidad	No relacionadas	42.748	5	0.0000
CPC	No relacionadas	33.891	3	0.0000
CPC(1)	No relacionadas	20.782	2	0.0000

En las TABLAS 28 y 29 se encuentran los resultados del análisis de componentes principales comunes para Ticumán tardías y Teotihuacán. No hay ninguna similitud entre estas matrices; y dado que estas variables son las que mejor cumplen con los supuestos del modelo no vale la pena proseguir el análisis con más variables.

3.4. Conclusiones del ejemplo

Trabajando con las muestras de invernadero se encontró que las matrices de varianza de Tula M y Tula P son muy similares, obteniendo el modelo de componentes principales comunes con siete variables y la igualdad con largo, ancho y cáliz. Las poblaciones Taxco M y Taxco P no se asemejaron prácticamente en nada; pues, en el análisis, tomando en cuenta siete variables, sólo se obtuvieron dos componentes principales comunes que representaban el 9% de la varianza total. Al reducir las variables a largo, ancho y cáliz, se obtuvo un sólo componente principal común que representaba el 8% de la varianza total.

Dada la diferencia abismal entre las muestras madres y padres de Taxco se decidió utilizar sólo las muestras madres para hacer la comparación entre Taxco y Tula. Al hacer la prueba con siete variables se observó poca similitud, tan sólo obteniendo tres componentes comunes a ambas poblaciones y siendo la proporción de la varianza total cubierta por éstos bastante baja. Sin embargo, al quitar dos variables (pedicelo y cámara), se obtuvo un resultado muy diferente: se aceptó la proporcionalidad entre las matrices de varianza con un valor p de 0.3964. Es posible que esos resultados tan disonantes se deban a que las variables cámara y pedicelo son variables muy neutras, en el sentido de que están poco correlacionadas con las demás, como se puede ver en las gráficas matriz en las FIGURAS 4 y 6. Esto hace que interfieran en los análisis de componentes principales comunes.

A diferencia de lo que ocurrió con las poblaciones de invernadero, no hubo casi rasgos de similitud entre las matrices de varianza de las poblaciones silvestres. Lo más que se encontró fue un modelo de dos componentes principales comunes, que implicaban, en conjunto, el 5% o 6% de la varianza total según la muestra. Por ello, es posible inferir que, en estas matrices (que son fenotípicas), se observó más el efecto ambiental que el genético: por ello, en poblaciones con ambientes similares se detectaron mucho más similitudes que en poblaciones con ambientes totalmente diferentes. Sin embargo, como se realizaron relativamente pocas pruebas, no se tiene una prueba absoluta de que éste haya sido el caso.

No está de más recalcar que todas las pruebas de hipótesis que se utilizaron suponen que los datos se distribuyen de manera normal multivariada. Como se vio en la sección 3.2 de este capítulo, no todas las variables pasaron las pruebas de normalidad, requisito indispensable para la normalidad multivariada. Una opción sería aplicar transformaciones para normalizar los datos de la mejor manera posible. Sin embargo, esta opción presenta un problema: para aplicar la técnica efectivamente es necesario que en todas las poblaciones a comparar, la variable sea transformada de la misma manera, pues de otra manera se analizarían matrices de varianza de variables distintas y el análisis carecería de sentido real. Como se puede observar en la sección mencionada, no siempre los datos de poblaciones diferentes correspondientes a la misma variable muestran la misma distribución. La estadística cuenta con métodos sencillos y comunes para poder encontrar transformaciones para normalizar datos de una variable en una población. Sin embargo, el

encontrar una misma transformación que, para varias poblaciones, mejore la normalidad de los datos, en cada población, no es un problema nada trivial.

Es importante, por último, hacer hincapié en que la distribución de los estadísticos de los cocientes de verosimilitud es asintótica; es decir, los estadísticos se distribuyen de la manera descrita cuando las muestras se aproximan a un tamaño infinito. Lo anterior implica que se deben usar únicamente con muestras “grandes”, donde “grandes” es un término muy relativo. Lo anterior es importante tenerlo en cuenta como una debilidad del modelo de Componentes Principales Comunes que pudo interferir en el análisis realizado.

CAPÍTULO 4

Conclusiones

Al comparar matrices de varianza, se debe ocupar métodos de comparación con grados intermedios entre la igualdad y la desigualdad, ya que entre estas dos situaciones hay una gran diferencia de grados de libertad al momento de estimar.

Uno de ellos es comparar estas matrices según la cantidad de componentes principales iguales entre ellas, así como por la igualdad o desigualdad de sus valores propios. En general, los componentes principales son una manera de descomponer las matrices de varianza según las direcciones en las que varía el objeto de estudio. En muchas ocasiones, al observarlos, se puede inferir qué parte de la realidad refleja un componente principal, como en muchos casos en biología se puede inferir que el componente principal más importante es el tamaño.

Sin embargo, se debe tener en cuenta que, en realidad, la descomposición en componentes principales es más un modelo teórico matemático que una manera de detectar los factores reales que influyen en el desarrollo de los individuos de una población. Lo anterior es porque el supuesto de que los factores influyentes en una población son ortogonales es poco realista. En ocasiones, para medir los factores reales que influyen en los individuos de una población, se debe hacer otros análisis estadísticos como el análisis de factores o una técnica aún no desarrollada del todo llamada análisis confirmatorio de factores. Para más información sobre estos temas, consultar [4].

Como todo el Análisis de Componentes Principales Comunes se basa en las matrices de varianza, es necesario asegurarse de que éstas estén bien calculadas, y para ello, es conveniente contar con métodos para la estimación robusta de matrices de varianza cuyo objetivo es impedir que observaciones extremas afecten la estimación de esta matriz. Se han realizado algunos estudios sobre ello y se propone como bibliografía para ese tema [5], [3], [11] y [7].

La Descomposición en Componentes Principales se puede usar, tanto para matrices de varianza, como para matrices de correlación. Utilizar las matrices de correlación es especialmente útil cuando no todas las variables están medidas en las mismas unidades, más aún cuando el orden de magnitud de sus varianzas difiere mucho según la variable. Sin embargo, El Método de Componentes Principales Comunes, tal como se presentó en este texto, está desarrollado especialmente para matrices de varianza, pues se basa en maximizar sus funciones de verosimilitud. Lo anterior restringe su uso más concienzudo a matrices en las que todas las variables estén medidas en las mismas unidades, de manera que, el concepto de Componente Principal tenga sentido. En algunos casos es

posible aventurarse a aplicarlo a matrices de varianza de variables en diferentes unidades, siempre y cuando, los órdenes de las varianzas de cada variable sean similares. Incluso, se puede pensar que es razonable utilizar las aproximaciones de los componentes principales comunes con matrices de correlación, dado que se puede demostrar, a partir de la ecuación 2.23, que maximizar la función 2.15 es equivalente a minimizar en β la función:

$$\Phi(\beta) = \prod_{i=1}^k \left[\frac{\det \text{diag}(\beta' S_i \beta)}{\det(\beta' S_i \beta)} \right]^{n_i}$$

Sin embargo, en ese caso, los estimadores no son los de máxima verosimilitud ni sus distribuciones asintóticas son las mencionadas. Existe la posibilidad de cambiar la escala de las variables, dividiendo a cada variable entre su desviación estándar. Así, la matriz de varianza con las variables en esta nueva escala es la de correlación en la escala original; y, dado que es una matriz de varianza, se puede aplicar la técnica de los Componentes Principales Comunes por máxima verosimilitud sin problemas.

Otra opción para abordar los Componentes Principales Comunes en matrices de correlación, es calcularlos no en base a una función de máxima verosimilitud, sino a partir de un método de mínimos cuadrados que se muestra en [1, Capítulo 4]. No obstante, dado que este método no es por máxima verosimilitud carece de pruebas de hipótesis.

Los Componentes Principales Comunes también sirven para hacer modelos estadísticos que no suponen que las matrices de varianza sean idénticas, pero tampoco agregan tantos grados de libertad como si se consideraran completamente diferentes. Un ejemplo de este tipo de aplicaciones tiene lugar en el análisis discriminante, donde se requiere modelar distintas estructuras de covarianza.

Al comparar varias poblaciones por componentes principales comunes, es necesario escoger las variables que van a intervenir en el análisis. No hay una regla que obligue a aumentar o disminuir variables, por lo que se debe hacer varias pruebas. Primero, se recomienda escoger las variables relevantes. Después, encontrar las que mejor se ajustan al modelo; posteriormente comenzar a hacer pruebas con ellas, y si es necesario, con subconjuntos de ellas. En caso de que se quisieran incluir variables que no se ajustan definitivamente al modelo, se recomienda, primero, hacer un análisis con el subconjunto de todas las variables a considerar que se ajustan al modelo y después otro con todas las que se desean incluir para comparar resultados. Si las conclusiones son las mismas no hay ningún problema, si no se debe identificar, con la pericia del investigador, si lo que falla es que las variables no se ajustan al modelo o si realmente el uso de esas variables cambia por completo el panorama de la investigación. Finalmente, se recomienda reducir las variables a números razonables, pues en la literatura revisada, en la que se usan los Componentes Principales Comunes de la manera más concienzuda, no hay ejemplos con más de cinco o seis variables.

APÉNDICE A

Generalización de la Esperanza, Covarianza y Varianza de variables a vectores aleatorios

Sea X un vector aleatorio de p entradas. Se define la esperanza de X , $E[X]$, como un vector aleatorio de p entradas tal que $E[X]_i = E[X_i]$, donde i es el índice que denota cada entrada del vector.

En ese caso, si A , B y C son matrices de constantes tales que, por sus dimensiones, $AXB + C$ está bien definido se cumple que:

$$(A.1) \quad E[AXB + C] = AE[X]B + C$$

Lo anterior es muy fácil de ver utilizando la linealidad del operador Esperanza.

Ahora considérese otro vector aleatorio Y de q entradas, donde no necesariamente $q = p$; se define la matriz de covarianza entre los vectores X y Y , $Cov(X, Y)$, como una matriz de tamaño $p \times q$ donde $Cov(X, Y)_{ij} = Cov(X_i, Y_j)$.

Si A es una matriz de $r \times p$ y B es una matriz de $s \times q$, donde $r, s \in \mathbb{N}$, $r, s \geq 1$; utilizando la bilinealidad de la covarianza se deduce que:

$$(A.2) \quad Cov(AX, BY) = ACov(X, Y)B'$$

Se define la matriz de varianza del vector aleatorio X , $Var[X]$, como $Var[X] = Cov(X, X)$. Entonces, del último resultado mencionado, se deduce inmediatamente, por la definición, que $Var[AX] = AVar[X]A'$.

APÉNDICE B

Teorema de la Descomposición Espectral

TEOREMA B.1. (*Descomposición Espectral*) Sea A una matriz real simétrica de dimensión $p \times p$. Entonces, existe una matriz ortogonal B y una matriz diagonal Λ tales que: $B'AB = \Lambda$.

Si todas las raíces características de A son distintas, la matriz B está únicamente determinada salvo por multiplicaciones por -1 y permutación de columnas.

Las columnas de B son los vectores propios normalizados de A y los elementos de la diagonal de Λ son los valores propios de A .

Para probar el teorema anterior se necesitan los siguientes lemas:

LEMA B.1. Los valores propios de las matrices simétricas con entradas reales son reales.

DEMOSTRACIÓN. Sea A una matriz simétrica con entradas reales y sea $c(\lambda)$ su polinomio característico. Como los coeficientes de $c(\lambda)$ son reales; si $c(\lambda) = 0$, entonces $c(\bar{\lambda}) = 0$ y si x es un vector propio correspondiente al valor propio λ , entonces \bar{x} es un vector propio correspondiente al valor propio $\bar{\lambda}$.

Sea λ un valor propio de A y x un vector propio que le corresponda.

Si se descomponen $\lambda = \alpha + i\beta$, donde α y β son reales; y $x = u + iv$ donde u y v son vectores de reales, se puede hacer lo siguiente:

$$(B.1) \quad \bar{x}'Ax = \lambda\bar{x}'x$$

$$(B.2) \quad x'A\bar{x} = \bar{\lambda}x'\bar{x}$$

Se puede, además, ver que:

$$(B.3) \quad \bar{x}'Ax = (\bar{x}'Ax)' \quad \text{pues se está transponiendo un número}$$

$$(B.4) \quad = x'A'\bar{x}$$

$$(B.5) \quad = x'A\bar{x} \quad \text{pues } A \text{ es simétrica}$$

Sustituyendo B.1 y B.2 en la igualdad B.5; y tomando en cuenta que $\bar{x}'x = x'\bar{x} = u'u + v'v > 0$, se tiene que:

$$(\alpha + i\beta)(u'u + v'v) = (\alpha - i\beta)(u'u + v'v)$$

De donde se sigue inmediatamente que:

$$i\beta(u'u + v'v) = 0$$

Por lo que $\beta = 0$ y, en consecuencia, λ es real. \square

LEMA B.2. *Los vectores propios asociados a dos valores propios diferentes de una matriz simétrica con entradas reales son ortogonales.*

DEMOSTRACIÓN. Sean λ_1 y λ_2 dos valores propios de A , tales que $\lambda_1 \neq \lambda_2$ y sean x_1 y x_2 los vectores propios asociados a λ_1 y λ_2 respectivamente.

En ese caso:

$$x_2'Ax_1 = \lambda_1x_2'x_1$$

y

$$x_1'Ax_2 = \lambda_2x_1'x_2$$

Ya que A es simétrica, se tiene que:

$$x_1'Ax_2 = x_2'Ax_1$$

Por lo que:

$$\begin{aligned}\lambda_1x_2'x_1 &= \lambda_2x_1'x_2 \\ (\lambda_1 - \lambda_2)x_1'x_2 &= 0\end{aligned}$$

De donde se deduce que $x_1'x_2 = 0$ que es el resultado buscado. \square

De esos dos lemas se sigue el teorema:

DEMOSTRACIÓN. Del primer lema se puede seguir fácilmente que existen p vectores propios de A . Denótese b_1, \dots, b_p a estos vectores propios en su forma normalizada. Sean $\lambda_1, \dots, \lambda_p$ sus respectivos valores propios. Del segundo lema se tiene que b_1, \dots, b_p son ortogonales. Definimos $B = (b_1, \dots, b_p)$.

El hecho de que b_1, \dots, b_p sean ortonormales, implica que B es ortogonal. Observando a B como un cambio de base se tiene que:

$$B'AB = \Lambda$$

Falta únicamente probar la unicidad de B . Supóngase que existe otra matriz ortogonal B^* tal que $B^{*'}AB^* = \Lambda^* = \text{diag}(\lambda_1^*, \dots, \lambda_p^*)$, con B^* ortogonal. Entonces: $AB^* = B^*\Lambda^*$. En ese caso si $B^* = (b_1^*, \dots, b_p^*)$, se tiene que:

$$Ab_j^* = \lambda_j^*b_j^* \quad j = 1, \dots, p$$

De allí se concluye que λ_j^* son los valores propios de A y b_j^* sus respectivos vectores propios; y de la unicidad de éstos, cuando los valores propios son distintos dos a dos, se concluye la unicidad de B y Λ , salvo por permutación de columnas y multiplicaciones por -1 en las columnas de B . \square

APÉNDICE C

Distribución Normal Multivariada

DEFINICIÓN C.1. Sea Y un vector aleatorio de p entradas. Se dice que Y sigue una distribución Normal Multivariada si y sólo si para todo vector de constantes con p entradas a , $a'Y$ sigue una distribución normal univariada. En el caso de que $\text{Var}[Y]$ sea una matriz definida positiva (es decir, que la forma cuadrática asociada a la matriz $\text{Var}[Y]$ sea definida positiva), si denotamos $\text{Var}[Y] = \Psi$ y $E[Y] = \mu$, se denota lo anterior como:

$$Y \sim N_p(\mu, \Psi)$$

Cuando $\text{Var}[Y]$ no es definida positiva se tiene una Distribución Normal Multivariada singular, sin embargo, por lo general, se trabajan con distribuciones con $\text{Var}[Y]$ definidas positivas. En ese caso, es posible probar, aunque no se hace en este texto, por no ser una prueba trivial, que si $E[Y] = \mu$ y $\text{Var}[Y] = \Psi$, entonces, la función de densidad de Y es:

$$(C.1) \quad f(y) = (2\pi)^{-\frac{p}{2}} (\det \Psi)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (y - \mu)' \Psi^{-1} (y - \mu) \right\} \\ -\infty < y_j < \infty \quad \text{para } j = 1, 2, \dots, d$$

A partir de lo anterior se puede deducir que la función de verosimilitud para μ y Ψ dada una muestra aleatoria x_1, \dots, x_n es la siguiente:

$$\begin{aligned} L(\mu, \Psi) &= (2\pi)^{-\frac{np}{2}} (\det \Psi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \mu)' \Psi^{-1} (x_i - \mu) \right\} \\ &= (2\pi)^{-\frac{np}{2}} (\det \Psi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \text{tr}[(x_i - \mu)' \Psi^{-1} (x_i - \mu)] \right\} \\ &= (2\pi)^{-\frac{np}{2}} (\det \Psi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{tr}[\Psi^{-1} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)'] \right\} \end{aligned}$$

Calculando ahora:

$$\begin{aligned} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)' &= \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)(x_i - \bar{x} + \bar{x} - \mu)' \\ &= \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' + n(\bar{x} - \mu)(\bar{x} - \mu)' \\ &= nS + n(\bar{x} - \mu)(\bar{x} - \mu)' \end{aligned}$$

Donde:

$$S = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

Suponiendo que Ψ es definida positiva, se tiene que Ψ^{-1} es definida positiva. Por lo tanto:

$$\text{tr}[\Psi^{-1}(\bar{x} - \mu)(\bar{x} - \mu)'] = (\bar{x} - \mu)' \Psi^{-1} (\bar{x} - \mu) \geq 0$$

Y, por lo tanto, para cualquier Ψ definida positiva, dada S , $L(\mu, \Psi)$ se maximiza con $\hat{\mu} = \bar{x}$. Así:

$$L(\hat{\mu}, \Psi) = (2\pi)^{-\frac{np}{2}} (\det \Psi)^{-\frac{n}{2}} \exp \left\{ -\frac{n}{2} \text{tr}(\Psi^{-1}S) \right\}$$

Aplicando el logaritmo a la función de verosimilitud, se tiene que:

$$\log[L(\hat{\mu}, \Psi)] = -\frac{np}{2} \log(2\pi) - \frac{n}{2} [\log(\det \Psi) + \text{tr}(\Psi^{-1}S)]$$

Se sabe que S es definida positiva con probabilidad 1 (ver [11, Apéndice 5.13]). Entonces, se puede encontrar el máximo de $\log L(\hat{\mu}, \Psi)$ con respecto a Ψ considerando que eso es equivalente a encontrar el mínimo de:

$$f(\Psi) = \log(\det \Psi) + \text{tr}(\Psi^{-1}S)$$

Eso se logra con el siguiente método:

Minimizar la función anterior es equivalente a minimizar $f(\Psi) - f(S)$, pues éstas sólo difieren por una constante. Supóngase que $\lambda_1, \dots, \lambda_p$ son los valores propios de $\Psi^{-1}S$, entonces:

$$\begin{aligned} f(\Psi) - f(S) &= \log(\det \Psi) + \text{tr}(\Psi^{-1}S) - \log(\det S) - \text{tr}(I_p) \\ &= -\log[\det(\Psi^{-1}S)] + \text{tr}(\Psi^{-1}S) - p \\ &= -\log\left(\prod_{i=1}^p \lambda_i\right) + \sum_{i=1}^p \lambda_i - p \\ &= \sum_{i=1}^p (-\log \lambda_i + \lambda_i - 1) \end{aligned}$$

Como $\log(x) \leq x - 1$ para todo $x > 0$ y la igualdad sólo ocurre cuando $x = 1$, se tiene que el mínimo de $f(\Psi) - f(S)$ se da cuando $\lambda_i = 1$ para todo $i \in \{1, \dots, k\}$, que es equivalente a que $\Psi = S$. Así, el estimador de máxima verosimilitud de Ψ , que se denotará $\hat{\Psi}$, es S .

Como una cuestión de notación, cabe hacer hincapié en que, en este texto, cuando se escribe $L(\Psi)$, se hace referencia de manera abreviada, a $L(\Psi, \hat{\mu})$.

Para finalizar, supóngase que se tienen k muestras de distribuciones normales multivariadas independientes de tamaños n_1, \dots, n_k , y se desea encontrar la función de verosimilitud conjunta de sus matrices de varianzas. En ese caso, considerando a las Ψ 's como independientes:

$$\begin{aligned} L(\Psi_1, \dots, \Psi_k) &= \prod_{i=1}^k L(\Psi_i) \\ (C.2) \quad &= (2\pi)^{-\frac{p \sum_{i=1}^k n_i}{2}} \prod_{i=1}^k \exp \left\{ \text{tr} \left(-\frac{n_i}{2} \Psi_i^{-1} S_i \right) \right\} (\det \Psi_i)^{-n_i/2} \end{aligned}$$

APÉNDICE D

Matrices Ortogonales

DEFINICIÓN D.1. *Sea A una matriz real de $n \times n$ y A' su transpuesta. Se dice que A es una matriz ortogonal si $AA' = A'A = I$.*

Nótese que si se denotan A_1, \dots, A_n a los renglones de dicha matriz, el hecho de que A sea ortogonal es equivalente a que A_1, \dots, A_n formen una base ortonormal de \mathbb{R}^n , pues:

$$A_i \bullet A_j = \sum_{k=1}^n A_{ik}A_{jk} = \sum_{k=1}^n A_{ik}(A')_{kj} = (AA')_{ij} = I_{ij}$$

Por lo que:

$$A_i \bullet A_j = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

Análogamente A es una matriz ortogonal si y sólo si sus columnas forman una base ortonormal de \mathbb{R}^n , que es un resultado muy utilizado en este texto.

APÉNDICE E

Derivadas parciales simbólicas con respecto a vectores

DEFINICIÓN E.1. Sea $f : \mathbb{R}^m \rightarrow \mathbb{R}$ una función y x un vector en \mathbb{R}^m , denótese $\frac{\partial f}{\partial x_j}$ a la derivada parcial de f con respecto a x_j , en el caso de que ésta exista.

Se define la derivada parcial de f con respecto a x como:

$$(E.1) \quad \frac{\partial f}{\partial x} := \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_m} \end{bmatrix}$$

En el caso de que $f(x) = a'x$, donde a es un vector constante con m entradas, se tiene que:

$$f(x) = \sum_{i=1}^m a_i x_i$$

Por lo que:

$$\frac{\partial f}{\partial x_j} = a_j$$

Y entonces:

$$\frac{\partial f}{\partial x} = a$$

En el caso de que $f(x) = x'Ax$, donde A es una matriz simétrica de $m \times m$, se tiene que:

$$f(x) = \sum_{j=1}^m \sum_{i=1}^m x_i x_j A_{ij}$$

Por lo que:

$$\begin{aligned} \frac{\partial f}{\partial x_k} &= \sum_{j=1, j \neq k}^m x_j A_{kj} + \sum_{i=1, i \neq k}^m x_i A_{ik} + 2A_{kk}x_k \\ &= \sum_{j=1, j \neq k}^m x_j A_{kj} + \sum_{i=1, i \neq k}^m x_i A_{ki} + 2A_{kk}x_k \\ &= 2 \sum_{j=1, j \neq k}^m x_j A_{kj} + 2A_{kk}x_k \\ &= 2 \sum_{j=1}^m A_{kj}x_j \\ &= (2Ax)_k \end{aligned}$$

Y entonces:

$$\frac{\partial f}{\partial x} = 2Ax$$

APÉNDICE F

Potencias no Enteras de Matrices

DEFINICIÓN F.1. Sea L una matriz de orden $p \times p$, diagonal, con entradas reales; y sea r un número real; se define la matriz $L^{(r)}$, (L elevada a la potencia r) como:

$$(F.1) \quad (L^{(r)})_{ii} := (L)_{ii}^r \quad \forall i \in \{1, \dots, p\}$$

$$(F.2) \quad (L^{(r)})_{ij} := 0 \quad \forall i, j \in \{1, \dots, p\}; \quad i \neq j$$

Es claro que, dada la definición anterior, se cumple:

$$(F.3) \quad \underbrace{L \dots L}_{n \text{ veces}} = L^{(n)} \quad \forall n \in \mathbb{N}$$

Además, cuando L es invertible también se cumple que:

$$(F.4) \quad L^{(-1)}L = I_p$$

Con lo anterior, es fácil ver que cuando r es entero, la definición anterior coincide con la definición común de elevar una matriz a potencias enteras.

DEFINICIÓN F.2. Sea A una matriz simétrica de orden $p \times p$, B una matriz ortogonal y L una matriz diagonal (ambas de orden $p \times p$) tales que $A = B'LB$. Si r es un número real, se define $A^{(r)}$ (la matriz A elevada a la potencia r) como:

$$(F.5) \quad A^r := B' L^{(r)} B$$

A continuación se demuestra que, para r entero, la definición anterior coincide con la definición común de elevar una matriz a potencias enteras. Lo anterior se resume a demostrar dos propiedades:

$$1. \quad A^{(n)} = \underbrace{A \dots A}_{n \text{ veces}} \quad \forall n \in \mathbb{N}:$$

Se demuestra por inducción sobre n :

Base (caso $n = 1$):

$$A^{(1)} = B' L^{(1)} B = B' L B = A$$

Paso inductivo:

Hipótesis de inducción:

$$A^{(n)} = \underbrace{A \dots A}_{n \text{ veces}}$$

Por demostrar:

$$A^{(n+1)} = \underbrace{A \dots A}_{n+1 \text{ veces}}$$

Demostración:

$$\begin{aligned} \underbrace{A \dots A}_{n+1 \text{ veces}} &= \underbrace{A \dots A}_n A \\ &= B' L^{(n)} B A \\ &= B' L^{(n)} B B' L B \\ &= B' L^{(n+1)} B \\ &= A^{(n+1)} \end{aligned}$$

2. Cuando A es invertible $A^{(-1)}A = I_p$:

$$\begin{aligned} A^{(-1)}A &= (B'LB)^{(-1)}(B'LB) \\ &= (B'L^{(-1)}B)(B'LB) \\ &= B'L^{(-1)}LB \\ &= B'B \\ &= I_p \end{aligned}$$

Bibliografía

- [1] Beaghen, Michael Jr., *Canonical Variate Analysis and Related Methods with Longitudinal Data*, Tesis de Doctorado, Virginia Tech, Blacksburg, Virginia, 1997. (Se puede encontrar en línea en <http://scholar.lib.vt.edu/theses/available/etd-11997-212717>).
- [2] Flury, Bernhard, *Common Principal Components and Related Multivariate Models*, John Wiley & Sons Inc., 1988.
- [3] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A., *Robust Statistics*, Wiley, New York. 1986.
- [4] Houle, D., Mezey, J. y Galpern, P., *Interpretation of the Results of Common Principal Components Analyses*, International Journal of Organic Evolution. **56** (2002), no. 3, 433-440.
- [5] Huber, P.J., *Robust Statistical Procedures*, Society for Industrial and Applied Mathematics, Philadelphia, 1977.
- [6] Looney, S. W., *A comparison of estimators of a common correlation coefficient*, Communications in Statistics B-Computation and Simulation. **15** (1986), 531-543.
- [7] Peña, Daniel y Prieto, Francisco J., *Multivariate Outlier detection and Robust Covariance Matrix Estimation*, tech. report, American Statistical Association and the American Society for Quality TECHNOMETRICS. **43**(2001).
- [8] Phillips, Patrick, *CPC - Common Principal Component Analysis Program.*, en línea, <http://darkwing.uoregon.edu/pphil/programs/cpc/cpc.htm> . 2000.
- [9] Rao, C.R., *Linear Statistical Inference and Its Applications*, segunda edición, Wiley, New York, 1973.
- [10] Ross, Sheldon, *A First Course in Probability*, sexta edición, Prentice Hall, 2002.
- [11] Seber, G.A.F., *Multivariate Observations*, Wiley, New York, 1984.
- [12] Steppan, S., Phillips, P. C. y Houle, D, *Comparative Quantitative Genetics: Evolution of the G Matrix* TRENDS in Ecology & Evolution. **17** (2002), no. 7, 320-327.