



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS MATEMÁTICAS

FACULTAD DE CIENCIAS

ANÁLISIS ESTADÍSTICO DE LA CONSULTA INFANTIL Y JUVENIL 2000 DEL INSTITUTO FEDERAL ELECTORAL VÍA COMPONENTES PRINCIPALES

TESIS

**QUE PARA OBTENER EL GRADO ACADÉMICO DE
MAESTRO EN CIENCIAS**

PRESENTA

JORGE ISAAC MANUEL ORTEGA

DIRECTORA DE TESIS: DRA. SILVIA RUIZ-VELASCO ACOSTA

MEXICO, D.F.

ENERO, 2008



Universidad Nacional
Autónoma de México



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.

Agradecimientos

A mi esposa, Aurora

A mi madre, Lupita

A Silvia, Ignacio, Ramsés, Patricia, Leticia, Federico y Raul;
por la extraordinaria labor que como profesores hubieron de desempeñar
conmigo

A Victor, Lizbeth, Lucy, Isadora y Miguel;
compañeros de quienes también recibí aprendizajes valiosos para mi
formación

Introducción

La Consulta Infantil y Juvenil 2000 del Instituto Federal Electoral es un ejemplo de instrumento típico para recolección de información social y humana (experiencias, actitudes, intenciones, conductas, creencias, etc.), en especial cuando el tamaño de la muestra es superior al que un pequeño grupo de entrevistadores podría manejar. Este tipo de instrumentos suelen estar conformados por reactivos objetivos, es decir con un número limitado de respuestas posibles a elegir, que se consideran indicadores de la existencia de variables teóricas de interés para el estudio donde se utilice el instrumento. En el caso de la consulta se espera que los reactivos que la conforman sean representativos de las variables que mejor describan la experiencia en formación ciudadana de los adolescentes participantes, y tal vez, que estas variables se relacionen de tal manera que cumplan con supuestos teóricos previos o sugieran nuevas hipótesis en cuanto a cómo se forma un ciudadano.

Tomando en cuenta lo anterior, este trabajo pretende ejemplificar como, a través de herramientas de uso común para profesionales ajenos al área de la estadística matemática, es posible obtener gran cantidad de información útil para poner a prueba supuestos acerca de la construcción de instrumentos de investigación social, así como para sugerir relaciones entre las variables de la teoría sobre la cual se realiza la investigación.

Por ejemplo, el Análisis de Componentes Principales es un método popular en las ciencias sociales para reducir el número de variables de trabajo, y encontrar combinaciones lineales de las variables originales que aporten la mayor cantidad de varianza, con el fin de validar instrumentos de investigación a través de la exploración de grandes conjuntos de datos en búsqueda de posibles estructuras para “análisis factoriales confirmatorios”. La popularidad del método se debe principalmente a que es el *default* de muchos paquetes estadísticos usuales de las ciencias sociales dentro de la clasificación de “Análisis Factoriales” (a la cual no pertenece este método).

Sin embargo, a pesar de que muchos científicos sociales utilizan de manera más o menos frecuente el análisis de componentes principales, e incluso realizan rotaciones varimax en búsqueda de alternativas para una interpretación más sencilla de los datos de salida; en muchas ocasiones existe gran cantidad de información resumida en los Componentes Principales que es desaprove-

chada por la falta de experiencia en el manejo de estos objetos.

En la aplicación presentada al final del trabajo se ilustra como los Componentes Principales pueden ser rotados para satisfacer distintos criterios algebraicos que simplifican y clarifican la estructura de los datos; pueden ser retransformados y rotados virtualmente para realizar exploraciones visuales, a través de paquetes estadísticos comunes, en búsqueda de estructuras interesantes o inesperadas; o bien, pueden ser objeto de otros tipos de análisis estadísticos, como las pruebas de modelos de regresión, usando los componentes como materia prima de modelos adecuados para enriquecer las teorías sobre las cuáles se realiza la investigación social que requirió su extracción.

Así se hace evidente que, a pesar de las limitaciones teóricas que un científico social pudiera tener en la comprensión de modelos multivariados complejos, los Componentes Principales, apoyados de un conjunto de técnicas de relativa simplicidad y disponibles en la paquetería estadística de uso común, podrían aportarle gran cantidad de información valiosa para su labor de investigación.

Resumen

Este trabajo presenta el análisis estadístico de una consulta nacional para adolescentes, basándose en la técnica del Análisis de Componentes Principales. Para estos fines se realiza una revisión teórica de la técnica, de las características de los Componentes Principales y de otras relacionadas íntimamente con ella.

De manera inicial se define a qué se llama Componentes Principales y se describen los principios subyacentes a la técnica para su obtención; para luego realizar una revisión de sus propiedades matemáticas y estadísticas.

Dada cierta similitud entre el Análisis de Componentes Principales y el conjunto de procedimientos conocidos como Análisis Factorial, se procede entonces a realizar una revisión de éstos últimos presentando diferencias y similitudes entre ambos métodos, y enfocándose de manera particular en las técnicas de rotación, como procedimiento que puede ser aplicado tanto a Componentes Principales como a Factores.

La revisión continúa presentando un conjunto de criterios que pueden ser usados para elegir el número adecuado de Componentes Principales a conservar al realizar el análisis, para después mostrar como, tras su obtención, es posible realizar otros análisis en búsqueda de patrones informativos de los Componentes, ejemplificando esto con la técnica conocida como “*Projection Pursuit*”.

Para finalizar se presenta el análisis realizado sobre la consulta, ilustrando el uso del método en cuestión, la búsqueda de patrones informativos, la exploración gráfica de resultados y el apoyo en modelos de Regresión Multinomial Logística para contar con la información suficiente para tomar decisiones acerca de la calidad de la consulta como instrumento de medición social, así como de posibles significados de sus resultados.

Índice general

1. Análisis de Componentes Principales	7
1.1. Definición y derivación de componentes principales	7
2. Propiedades Matemáticas y Estadísticas de los Componentes Principales	11
2.1. Propiedades algebraicas óptimas de los componentes principales poblacionales	11
2.2. Componentes principales usando una matriz de correlación	14
2.3. Propiedades algebraicas óptimas de los componentes principales muestrales	17
3. Análisis de Componentes Principales y Análisis Factorial	21
3.1. Modelos para Análisis Factorial	21
3.2. Estimación de una solución inicial única	24
3.3. Rotación de factores	27
4. Elección de un Subconjunto de Componentes Principales	32
4.1. Porcentaje acumulado de varianza total	32
4.2. Tamaño de la varianza de los componentes principales	33
4.3. La gráfica de sedimentación y el diagrama log - valor propio .	35
4.4. Número de componentes con valores propios desiguales	36
4.5. Correlaciones parciales	37
5. Componentes Principales e Información	39
5.1. "Projection Pursuit"	39

5.2. Búsqueda de estructuras de interés	42
5.3. Modelos Lineales Generalizados	45
6. Aplicación	49
6.1. Descripción de la base de datos	49
6.2. Análisis y resultados	51
6.3. Conclusiones	70

Capítulo 1

Análisis de Componentes Principales

La idea central de realizar un análisis de componentes principales (ACP) es reducir la dimensionalidad de un conjunto de datos consistente en un número relativamente grande de variables correlacionadas, manteniendo en cuanto sea posible la variación presente en el conjunto original. Esto se logra transformando las variables originales en un nuevo conjunto de variables, los componentes principales (CPs), los cuales no están correlacionados, y se encuentran ordenados de tal manera que unos pocos de los primeros retienen la mayor parte de la variación presente en el conjunto original.

1.1. Definición y derivación de componentes principales

Supongamos que \mathbf{X} es un vector de p variables aleatorias, y que las varianzas de las p variables aleatorias y la estructura de covarianzas o correlaciones entre las p variables son de interés. A menos que p sea pequeña, o que la estructura sea muy simple, no sería de mucha ayuda simplemente observar las p varianzas y las $\frac{1}{2}p(p-1)$ correlaciones o covarianzas. Una alternativa es encontrar unas pocas ($\ll p$) variables derivadas tales que preserven la mayor parte de la información dada por estas varianzas y que no estén correlacionadas.

A pesar de que un ACP no ignora las covarianzas o correlaciones, se concentra en varianzas. El primer paso es buscar una función lineal $\alpha'_1 \mathbf{X}$ de elementos de \mathbf{X} que tenga varianza máxima, donde α_1 es un vector de p constantes $\alpha_{11}, \dots, \alpha_{1p}$ tales que:

$$\alpha'_1 \mathbf{X} = \alpha_{11}x_1 + \dots + \alpha_{1p}x_p = \sum_{j=1}^p \alpha_{1j}x_j$$

Ahora, se busca una función lineal $\alpha'_2 \mathbf{X}$ no correlacionada con $\alpha'_1 \mathbf{X}$ y con varianza máxima, y así sucesivamente hasta el k -ésimo paso: hallar una función $\alpha'_k \mathbf{X}$ de varianza máxima e independiente de $\alpha'_1 \mathbf{X}, \dots, \alpha'_{k-1} \mathbf{X}$. La k -ésima variable derivada es el k -ésimo CP. De esta manera pueden ser encontrados hasta p CPs, aunque se espera de manera general que la mayor parte de la varianza sea aportada por los m primeros CPs, con $m \ll p$.

Habiendo definido los CPs, se necesita ahora saber como hallarlos. Considérese, inicialmente, el caso donde el vector de variables aleatorias \mathbf{X} tiene una covarianza conocida Σ . Ésta es la matriz cuyo elemento (i, j) es la covarianza conocida entre el i -ésimo y el j -ésimo elemento de \mathbf{X} , cuando $i \neq j$, y la varianza del i -ésimo elemento cuando $i = j$. Resulta entonces que para $k = 1, \dots, p$, el k -ésimo CP está dado por $z_k = \alpha'_k \mathbf{X}$ donde α_k es un vector propio de Σ correspondiente a su k -ésimo valor propio más grande λ_k . Además, si α_k es elegido de longitud unitaria ($\alpha'_k \alpha_k = 1$), entonces $\mathbf{Var}(z_k) = \lambda_k$, donde $\mathbf{Var}(z_k)$ denota la varianza de z_k .

Para derivar la forma de los CPs, se considera primero $\alpha'_1 \mathbf{X}$; el vector α_1 maximiza $\mathbf{Var}(\alpha'_1 \mathbf{X}) = \alpha'_1 \Sigma \alpha_1$. Es claro que, de esta manera, el máximo no se alcanza para α_1 finita, así que una restricción de normalización debe ser impuesta. La restricción usada en esta derivación es $\alpha'_1 \alpha_1 = 1$, es decir, que la suma de los cuadrados de los elementos de α_1 sea igual a 1. Otras restricciones pueden ser más útiles en otras circunstancias, por ejemplo que $\max_j |\alpha_{1j}| = 1$ con $j = 1, \dots, p$, sin embargo, estas otras opciones suelen llevar a problemas más difíciles de optimización.

Para maximizar $\alpha'_1 \Sigma \alpha_1$ sujeto a $\alpha'_1 \alpha_1 = 1$, la manera estándar es usar la técnica de los multiplicadores de Lagrange. Maximizar $\alpha'_1 \Sigma \alpha_1 - \lambda_1 (\alpha'_1 \alpha_1 - 1)$, donde λ_1 es un multiplicador de Lagrange. La diferenciación con respecto a

α_1 , al ser igualada a cero, da

$$\Sigma\alpha_1 - \lambda_1\alpha_1 = 0$$

ó

$$(\Sigma - \lambda_1\mathbf{I}_p)\alpha_1 = 0,$$

donde \mathbf{I}_p es la matriz identidad de $p \times p$. Es decir, λ_1 es un valor propio de Σ y α_1 es el vector propio correspondiente. Para decidir cuál de los p valores propios da $\alpha_1'\mathbf{X}$ con varianza máxima, nótese que la cantidad que debe ser maximizada es

$$\alpha_1'\Sigma\alpha_1 = \alpha_1'\lambda_1\alpha_1 = \lambda_1\alpha_1'\alpha_1 = \lambda_1,$$

luego λ_1 debe ser tan grande como sea posible. Dado que Σ es positiva definida y simétrica, sus valores propios son reales positivos. Esto es, α_1 es el vector propio correspondiente al valor propio más grande de Σ , y $\mathbf{Var}(\alpha_1'\mathbf{X}) = \alpha_1'\Sigma\alpha_1 = \lambda_1$, el valor propio más grande.

En general, el k -ésimo CP de \mathbf{X} es $\alpha_k'\mathbf{X}$ y $\mathbf{Var}(\alpha_k'\mathbf{X}) = \lambda_k$, donde λ_k es el k -ésimo valor propio más grande de Σ , y α_k es el vector propio correspondiente. A manera de ilustración se prueba lo anterior para $k = 2$, la prueba para $k \geq 3$ es similar aunque su desarrollo es más complejo.

El segundo CP, $\alpha_2'\mathbf{X}$, maximiza $\alpha_2'\Sigma\alpha_2$ sujeto a no estar correlacionado con $\alpha_1'\mathbf{X}$, es decir, sujeto a $\mathbf{Cov}(\alpha_1'\mathbf{X}, \alpha_2'\mathbf{X}) = 0$, donde $\mathbf{Cov}(x, y)$ denota la covarianza entre las variables aleatorias x y y . Pero

$$\mathbf{Cov}(\alpha_1'\mathbf{X}, \alpha_2'\mathbf{X}) = \alpha_1'\Sigma\alpha_2 = \alpha_2'\Sigma\alpha_1 = \alpha_2'\lambda_1\alpha_1 = \lambda_1\alpha_2'\alpha_1 = \lambda_1\alpha_1'\alpha_2.$$

Esto es, cualquiera de las ecuaciones

$$\alpha_1'\Sigma\alpha_2 = 0, \alpha_2'\Sigma\alpha_1 = 0,$$

$$\alpha_1'\alpha_2 = 0, \alpha_2'\alpha_1 = 0,$$

podría ser utilizada para especificar que la correlación entre $\alpha_1'\mathbf{X}$ y $\alpha_2'\mathbf{X}$ es cero. Eligiendo arbitrariamente la última de ellas, y notando que de nuevo una restricción de normalización es necesaria, la cantidad a ser maximizada es

$$\alpha_2'\Sigma\alpha_2 - \lambda_2(\alpha_2'\alpha_2 - 1) - \phi\alpha_2'\alpha_1,$$

donde λ_2 , ϕ son multiplicadores de Lagrange. Diferenciando con respecto a α_2 se obtiene

$$\Sigma\alpha_2 - \lambda_2\alpha_2 - \phi\alpha_1 = 0,$$

y multiplicando el lado izquierdo de la ecuación por α_1' se obtiene

$$\alpha_1'\Sigma\alpha_2 - \lambda_2\alpha_1'\alpha_2 - \phi\alpha_1'\alpha_1 = 0,$$

lo cual, dado que los primeros dos términos son cero y que $\alpha_1'\alpha_1 = 1$, se reduce a $\phi = 0$. Así, $\Sigma\alpha_2 - \lambda_2\alpha_2 = 0$, o de manera equivalente, $(\Sigma - \lambda_2\mathbf{I}_p)\alpha_2 = 0$, por lo que λ_2 es una vez más un valor propio de Σ , y α_2 el vector propio correspondiente.

Una vez más, $\lambda_2 = \alpha_2'\Sigma\alpha_2$, así que λ_2 debe ser tan grande como sea posible. Ahora asumiremos que Σ no tiene valores propios repetidos, de esta manera, λ_2 es el segundo valor propio más grande de Σ , y α_2 es el vector propio correspondiente. Cuando existen valores propios repetidos para la matriz de correlación, o de varianza-covarianza, y por lo tanto existan varianzas iguales para distintos CPs, los CPs individuales correspondientes a un grupo de valores propios iguales no estarán definidos de manera única. Sin embargo este es un caso muy poco común en la práctica.

Continuando con el supuesto de que todos los valores propios son distintos entre sí, puede ser demostrado que para el tercer, cuarto, ..., p -ésimo CP, los vectores de coeficientes $\alpha_3, \dots, \alpha_p$ son los vectores propios de Σ correspondientes a $\lambda_3, \dots, \lambda_p$, del tercer valor propio más grande al más pequeño. Por otro lado, $\mathbf{Var}(\alpha_k'\mathbf{X}) = \lambda_k$ para $k = 1, \dots, p$. Finalmente, asumiendo que Σ es de rango completo, entonces existen exactamente p valores propios diferentes y diferentes de cero, y por tanto, p CPs.

Aun cuando la anterior definición y procedimientos fueron presentados usando la matriz de covarianzas poblacionales Σ , en el caso más realista, ésta es desconocida y se reemplaza por la matriz de covarianzas muestrales \mathbf{S} , de esta manera se obtienen CPs muestrales, los cuales serán estimadores de los CPs poblacionales, que bajo suposiciones distribucionales pueden usarse para hacer inferencias sobre la existencia de estos últimos y sobre la importancia de sus coeficientes.

Capítulo 2

Propiedades Matemáticas y Estadísticas de los Componentes Principales

A partir de este punto se considerará que ninguna de las varianzas de los CPs son iguales entre si o iguales a cero, a menos que se indique lo contrario.

2.1. Propiedades algebraicas óptimas de los componentes principales poblacionales

Sea \mathbf{Z} el vector cuyo k -ésimo componente es z_k , el k -ésimo CP, $k = 1, \dots, p$, según la derivación dada. Entonces

$$\mathbf{Z} = \mathbf{A}'\mathbf{X} \quad (2.1)$$

donde \mathbf{A} es la matriz ortogonal cuya k -ésima columna, α_k , es el k -ésimo vector propio de Σ . Esto es, los CPs están definidos por una transformación lineal ortonormal de \mathbf{X} . Además, se tiene directamente de la derivación presentada que

$$\Sigma\mathbf{A} = \mathbf{A}\Lambda \quad (2.2)$$

donde Λ es la matriz diagonal cuyo k -ésimo elemento diagonal es λ_k , el k -ésimo valor propio de Σ , y $\lambda_k = \mathbf{Var}(\alpha'_k \mathbf{X}) = \mathbf{Var}(z_k)$. Dos maneras alternativas de expresar (2.2) que son resultado de la ortogonalidad de \mathbf{A} , y que serán de utilidad más adelante son:

$$\mathbf{A}'\Sigma\mathbf{A} = \Lambda \quad (2.3)$$

y

$$\Sigma = \mathbf{A}\Lambda\mathbf{A}' \quad (2.4)$$

La transformación ortonormal de \mathbf{X} , (2.1), la cual define \mathbf{Z} , satisface un conjunto de propiedades algebraicas óptimas -que pueden ser consultadas en Jolliffe (2002)-, las primeras de las cuales se presentan a continuación:

P1 Para cualquier entero q , $1 \leq q \leq p$, considérese la transformación lineal

$$\mathbf{Y} = \mathbf{B}'\mathbf{X}, \quad (2.5)$$

donde \mathbf{Y} es un vector de q elementos y \mathbf{B}' es una matriz de $(q \times p)$. Sea $\Sigma_Y = \mathbf{B}'\Sigma\mathbf{B}$ la matriz de varianza-covarianza para \mathbf{Y} . Entonces la traza de Σ_Y , denotada $\text{tr}(\Sigma_Y)$, es maximizada tomando $\mathbf{B} = \mathbf{A}_q$, donde \mathbf{A}_q consiste de las primeras q columnas de \mathbf{A} .

P2 Considérese la transformación ortonormal $\mathbf{Y} = \mathbf{B}'\mathbf{X}$, con \mathbf{X} , \mathbf{B} , Σ_Y como ya han sido definidas. Entonces $\text{tr}(\Sigma_Y)$ se minimiza tomando $\mathbf{B} = \mathbf{A}_q^*$, donde \mathbf{A}_q^* consiste de las últimas q columnas de \mathbf{A} .

P3 Descomposición Espectral de Σ

$$\Sigma = \lambda_1\alpha_1\alpha_1' + \cdots + \lambda_p\alpha_p\alpha_p'. \quad (2.6)$$

P4 Considérese la transformación lineal $\mathbf{Y} = \mathbf{B}'\mathbf{X}$ como ya ha sido presentada. Si $\text{Det}(\Sigma_Y)$ denota al determinante de la matriz de covarianzas de \mathbf{y} , entonces $\text{Det}(\Sigma_Y)$ es maximizado cuando $\mathbf{B} = \mathbf{A}_q$.

P5 Supóngase que se desea predecir cada variable aleatoria x_j en \mathbf{X} con una función lineal de \mathbf{Y} , donde $\mathbf{Y} = \mathbf{B}'\mathbf{X}$, como se ha presentado. Si σ_j^2 es la varianza residual al predecir x_j con \mathbf{y} , entonces $\sum_{j=1}^p \sigma_j^2$ es mínima cuando $\mathbf{B} = \mathbf{A}_q$

La propiedad **P1** enfatiza que los CPs explican, sucesivamente, tanto como sea posible de $\text{tr}(\boldsymbol{\Sigma})$, es decir la varianza de las variables originales.

La implicación estadística de **P2** es que los últimos CPs no son sólo remanentes sin estructura tras remover los CPs más importantes. Dado que estos últimos CPs tienen varianzas tan pequeñas como sean posibles, son útiles por derecho propio. Éstos pueden ayudar a detectar relaciones lineales insospechadas, cercanas a lo constante, entre los elementos de \mathbf{X} , ser útiles para fines de regresión, para la selección de subconjuntos de interés de \mathbf{X} , o para la detección de *outliers* al identificar valores grandes en el caso de usar muestras.

La ecuación (2.6) sugiere a su vez que la matriz de covarianza o correlación puede ser construida exactamente, dando los coeficientes y las varianzas de los primeros r CPs, donde r es el rango de la matriz, en este caso consideramos que la matriz es de rango completo. Lo que esta propiedad implica es que si los valores $\lambda_{q+1}, \dots, \lambda_p$ son pequeños al retener solamente los primeros q componentes, podemos recuperar no solo la varianza sino también la correlación de las variables originales.

La importancia estadística de **P4** radica en que el determinante de una matriz de covarianza, o *varianza generalizada*, puede ser usado como una medida simple de dispersión para una variable aleatoria multivariada. La raíz cuadrada de una varianza generalizada para una distribución normal multivariada, por ejemplo, es proporcional al “volumen” en el espacio p -dimensional que contiene una proporción fija de la distribución de probabilidad de \mathbf{X} , los primeros q CPs son, por lo tanto, y como consecuencia de **P4**, q funciones lineales de \mathbf{X} cuya distribución de probabilidad conjunta tiene contornos de probabilidad fija conteniendo el volumen máximo.

Finalmente, la implicación estadística de la propiedad **P5** consiste en que, si se desea obtener el mejor predictor lineal de \mathbf{X} en un sub-espacio q -dimensional, en el sentido de minimizar la suma sobre los elementos de \mathbf{X} de las varianzas residuales, entonces el subespacio óptimo está definido por los primeros q CPs. Esta propiedad algebraica se encuentra también en íntima correlación con cierta propiedad geométrica de CPs muestrales que se presentará más adelante.

2.2. Componentes principales usando una matriz de correlación

Las propiedades de los CPs consideradas anteriormente están basadas en vectores propios y valores propios de la matriz de covarianza. Sin embargo, en la práctica es común definir CPs como

$$\mathbf{Z} = \mathbf{A}'\mathbf{X}^*, \quad (2.7)$$

donde \mathbf{A} tiene columnas consistentes en vectores propios de la matriz de correlación de \mathbf{X} , y \mathbf{X}^* consiste de variables estandarizadas. La intención de adoptar esta representación es encontrar los CPs de \mathbf{X}^* , una versión estandarizada de \mathbf{X} , donde \mathbf{X}^* tiene j -ésimo elemento $x_j/\sqrt{\sigma_{jj}}$, $j = 1, \dots, p$, x_j es el j -ésimo elemento de \mathbf{X} , y σ_{jj} es la varianza de x_j . De esta manera la matriz de covarianza de \mathbf{X}^* es la matriz de correlación de \mathbf{X} , y los CPs de \mathbf{X}^* están dados por (2.7).

Otra opción es utilizar las covarianzas de x_j/w_j , donde los pesos w_j son elegidos de tal manera que reflejen alguna idea previa acerca de la importancia relativa de las variables. El caso especial donde $w_j = \sigma_{jj}^{1/2}$ lleva a \mathbf{X}^* , y a CPs basados en la matriz de correlación. Sin embargo tal elección de w_j puede ser igual de arbitraria que cualquier otra y, aunque es un caso que se da de manera inusual, en ocasiones algunos pesos w_j son sugeridos de manera intuitiva por las cualidades del conjunto de datos que se esté manejando.

De esta manera, las propiedades presentadas para los CPs siguen siendo válidas para matrices de correlaciones, o matrices de covarianzas “ponderadas” por un conjunto de pesos w_j , con la excepción de que se consideran entonces los CPs de \mathbf{X}^* , es decir, de alguna transformación de \mathbf{X} , en vez de los relativos a sus elementos originales.

Podría parecer que los CPs de una matriz de correlación pudieran ser obtenidos de manera más o menos sencilla a partir de aquellos para la matriz de covarianza correspondiente, dado que \mathbf{X}^* está relacionado con \mathbf{X} por una transformación relativamente simple. Sin embargo éste no es el caso, los valores

y vectores propios de la matriz de correlación no tienen relación con aquellos de la matriz de covarianza correspondiente. En particular, si los CPs obtenidos de la matriz de correlación están expresados en términos de \mathbf{X} por una transformación de \mathbf{X}^* a \mathbf{X} , entonces estos CPs casi nunca son los mismos que aquellos encontrados a partir de Σ . Esto se debe a que los CPs son invariantes bajo transformaciones ortogonales, pero no necesariamente bajo otro tipo de transformaciones. La transformación de \mathbf{X} a \mathbf{X}^* no es ortogonal, por lo tanto, los CPs obtenidos a partir de las matrices de correlación y de covarianza no son necesariamente equivalentes en información, ni pueden ser derivados directamente los unos de los otros.

Un argumento importante para usar matrices de correlación en vez de matrices de covarianza para definir CPs radica en la facilidad para comparar directamente resultados obtenidos de diferentes conjuntos de individuos. Una gran desventaja del ACP basado en matrices de covarianza es la sensibilidad de los CPs a las unidades de medida usadas para cada elemento de \mathbf{X} . Si existen grandes diferencias entre las varianzas de los elementos de \mathbf{X} , aquellos con mayor varianza tenderán a dominar los primeros CPs, lo cual podría ser adecuado cuando todas las variables utilizan la misma escala, sin embargo, cuando se usan distintas unidades entre ellas, los resultados pueden ser engañosos.

Al trabajar con matrices de correlaciones para obtener CPs se puede observar que estos cumplen tres propiedades interesantes. La primera es que los CPs no dependen de los valores absolutos de las correlaciones sino solamente de sus razones. Esto se debe a que la multiplicación de todos los elementos no diagonales de la matriz por una misma constante no genera cambios en los vectores propios de la matriz.

La segunda propiedad señala que si, en vez de la normalización $\alpha'_k \alpha_k = 1$, se usa

$$\tilde{\alpha}'_k \tilde{\alpha}_k = \lambda_k \quad (2.8)$$

con $k = 1, \dots, p$, entonces $\tilde{\alpha}_{kj}$, el j -ésimo elemento de $\tilde{\alpha}_k$ es la correlación entre la j -ésima variable estandarizada x_j^* y el k -ésimo CP. Para observar lo

anterior nótese que para $k = 1, \dots, p$,

$$\tilde{\alpha}_k = \lambda_k^{1/2} \alpha_k, \mathbf{Var}(z_k) = \lambda_k,$$

y que el vector p -dimensional $\Sigma \alpha_k$ tiene como j -ésimo elemento la covarianza entre x_j^* y z_k . Pero $\Sigma \alpha_k = \lambda_k \alpha_k$, entonces la covarianza entre x_j^* y z_k es $\lambda_k \alpha_{kj}$. Asimismo $\mathbf{Var}(x_j^*) = 1$, y la correlación entre x_j^* y z_k es por tanto

$$\frac{\lambda_k \alpha_{jk}}{\sqrt{\mathbf{Var}(x_j^*) \mathbf{Var}(z_k)}} = \lambda_k^{1/2} \alpha_{kj} = \tilde{\alpha}_{kj},$$

como era requerido. Debido a esta propiedad, la normalización (2.8) es usada frecuentemente en paquetes computacionales, sin embargo tiene la desventaja de que es menos fácil hacer interpretaciones informales y comparar un conjunto de CPs cuando cada CP tiene una normalización diferente en sus coeficientes.

La tercera propiedad es lo suficientemente importante para ponerle una etiqueta y agregarla al listado de propiedades algebraicas ya presentadas.

P6 Para cualquier entero q , $1 \leq q \leq p$ considere la transformación ortonormal

$$\mathbf{Y} = \mathbf{B}'\mathbf{X}, \quad (2.9)$$

como fuera definida en **P1**. Sea $R_{j;q}^2$ la correlación múltiple cuadrática entre x_j y las q variables y_1, \dots, y_q , definidas por los elementos de \mathbf{y} . El criterio

$$\sum_{j=1}^p R_{j;q}^2$$

es maximizado cuando y_1, \dots, y_q son los primeros q CPs de la matriz de correlación. El valor maximizado del criterio es igual a la suma de los q valores propios más grandes de la matriz de correlación

Dado que los componentes principales no están correlacionados, el criterio en **P6** se reduce a

$$\sum_{j=1}^p \sum_{k=1}^q r_{jk}^2,$$

donde r_{jk}^2 es la correlación cuadrática entre la j -ésima variable y el k -ésimo CP. El criterio será maximizado por cualquier matriz \mathbf{B} que de \mathbf{y} extendiéndose en el mismo espacio q -dimensional que los primeros q CPs. Sin embargo, los CPs de las matrices de correlaciones son especiales, dado que sucesivamente maximizan el criterio para $q = 1, \dots, p$.

La propiedad **P6** puede ser fácilmente modificada para dar una nueva propiedad para el ACP de las matrices de covarianza. Los primeros q CPs de una matriz de covarianza maximizan, entre todas las transformaciones ortonormales de \mathbf{X} , la suma de las covarianzas cuadráticas entre x_1, \dots, x_p y las variables derivadas y_1, \dots, y_p . Las covarianzas, a diferencia de las correlaciones, no son invariantes ante cambios en la escala, y por lo tanto tampoco lo es el ACP de las matrices de covarianza.

2.3. Propiedades algebraicas óptimas de los componentes principales muestrales

Supóngase que se tienen n observaciones independientes en el vector p -dimensional \mathbf{X} ; denótense estas n observaciones $\mathbf{x}_1, \dots, \mathbf{x}_n$. Sea $\tilde{z}_{i1} = \mathbf{a}'_1 \mathbf{x}_i$, $i = 1, \dots, n$; y elíjase el vector de coeficientes \mathbf{a}_1 de tal manera que maximice la varianza muestral

$$\frac{1}{n-1} \sum_{i=1}^n (\tilde{z}_{i1} - \bar{z}_1)^2,$$

sujeta a la restricción de normalización $\mathbf{a}'_1 \mathbf{a}_1 = 1$. Después, sea $\tilde{z}_{i2} = \mathbf{a}'_2 \mathbf{x}_i$, $i = 1, \dots, n$; y elíjase el vector de coeficientes \mathbf{a}_2 de tal manera que maximice la varianza muestral, sujeta a la restricción $\mathbf{a}'_2 \mathbf{a}_2 = 1$, y a que \tilde{z}_{i2} sea independiente de \tilde{z}_{i1} . Continuando este proceso de igual manera se obtiene una versión muestral de la definición de CPs dada para poblaciones. De esta manera $\mathbf{a}'_k \mathbf{x}$ queda definido como el k -ésimo CP muestral, con $k = 1, \dots, p$, y z_{ik} es el puntaje para i -ésima observación en el k -ésimo CP. Continuando la analogía de esta definición con el caso poblacional, resulta que la varianza muestral de los puntajes para el k -ésimo CP muestral es l_k , el k -ésimo valor propio más grande de la matriz de covarianza muestral \mathbf{S} para $\mathbf{x}_1, \dots, \mathbf{x}_n$; y \mathbf{a}_k es el vector propio correspondiente para $k = 1, \dots, p$.

Defínanse las matrices $(n \times p)$ -dimensionales $\tilde{\mathbf{X}}$ y $\tilde{\mathbf{Z}}$ de tal manera que tengan como (i, k) -ésimos elementos el valor del k -ésimo elemento \tilde{x}_{ik} de \mathbf{x}_i , y el de \tilde{z}_{ik} , respectivamente. Entonces $\tilde{\mathbf{Z}}$ y $\tilde{\mathbf{X}}$ están relacionados por $\tilde{\mathbf{Z}} = \tilde{\mathbf{X}}\mathbf{A}$, donde \mathbf{A} es la matriz ortogonal $(p \times p)$ -dimensional cuya k -ésima columna es \mathbf{a}_k .

Si se sabe que la media de cada elemento de $\tilde{\mathbf{X}}$ es cero, entonces $\mathbf{S} = \frac{1}{n} \tilde{\mathbf{X}}' \tilde{\mathbf{X}}$. Sin embargo, es más usual que la media sea desconocida, en cuyo caso el (j, k) -ésimo elemento de \mathbf{S} es

$$\frac{1}{n-1} \sum_{i=1}^n (\tilde{x}_{ij} - \bar{x}_j)(\tilde{x}_{ik} - \bar{x}_k),$$

donde

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n \tilde{x}_{ij},$$

para $j = 1, \dots, p$.

La matriz \mathbf{S} puede ser entonces escrita como

$$\mathbf{S} = \frac{1}{n-1} \mathbf{X}' \mathbf{X}, \quad (2.10)$$

donde \mathbf{X} es una matriz de dimensiones $(n \times p)$ con (i, j) -ésimo elemento $(\tilde{x}_{ij} - \bar{x}_j)$.

Para fines prácticos será conveniente definir la matriz de puntajes de los CPs como

$$\mathbf{Z} = \mathbf{X}\mathbf{A} \quad (2.11)$$

Estos puntajes tendrán exactamente las mismas varianzas y covarianzas que las dadas por $\tilde{\mathbf{Z}}$, pero tendrán media cero, en vez de medias \tilde{z}_k , $k = 1, \dots, p$. Otro punto a notar es que los vectores propios de $\frac{1}{n-1} \mathbf{X}' \mathbf{X}$ y $\mathbf{X}' \mathbf{X}$ son idénticos, y que los valores propios de $\frac{1}{n-1} \mathbf{X}' \mathbf{X}$ son simplemente $\frac{1}{n-1}$ veces los de $\mathbf{X}' \mathbf{X}$.

Entonces, definiendo de manera similar que para la propiedades **P1-P5**, se tiene

$$\mathbf{y}_i = \mathbf{B}'\mathbf{X}_i \quad (2.12)$$

para $i = 1, \dots, n$. De esta manera, la propiedades **P1**, **P2**, **P4** y **P5** se cumplen al reemplazar Σ_Y con la matriz de covarianza muestral de las observaciones \mathbf{y}_i , $i = 1, \dots, n$, y con la matriz \mathbf{A} definida ahora con k -ésima columna \mathbf{a}_k , y con \mathbf{A}_q , \mathbf{A}_q^* representando, respectivamente, sus primera y última q columnas.

La descomposición espectral presentada en la propiedad **P3**, también se cumple para muestras en la forma

$$\mathbf{S} = l_1 \mathbf{a}_1 \mathbf{a}_1' + \dots + l_p \mathbf{a}_p \mathbf{a}_p' \quad (2.13)$$

Las implicaciones estadísticas de ésta y las anteriores propiedades algebraicas **P1**, **P2**, **P4** y **P5** son las mismas que las correspondientes para poblaciones, con la excepción de que deben ser interpretadas en un contexto muestral.

Otra propiedad, que concierne al uso de CPs en modelos de regresión se presenta a continuación.

P7 *Supóngase ahora que \mathbf{X} consiste de n observaciones de p variables predictoras x medidas por sus medias muestrales, y que la ecuación de regresión correspondiente es*

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \quad (2.14)$$

donde \mathbf{Y} es el vector de n observaciones de la variable dependiente, medida de nuevo por la media muestral. Supóngase que \mathbf{X} es transformada por la ecuación $\mathbf{Z} = \mathbf{X}\mathbf{B}$, donde \mathbf{B} es una matriz ortogonal de $(p \times p)$ elementos. Entonces, la ecuación de regresión puede ser reescrita como

$$\mathbf{Y} = \mathbf{Z}\gamma + \epsilon,$$

donde $\gamma = \mathbf{B}^{-1}\beta$. El estimador de mínimos cuadrados usual para γ es $\hat{\gamma} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$. Entonces los elementos de $\hat{\gamma}$ tienen, sucesivamente, las varianzas más pequeñas posibles si $\mathbf{B} = \mathbf{A}$, la matriz cuya k -ésima columna es el k -ésimo vector propio de $\mathbf{X}'\mathbf{X}$, y por tanto, el k -ésimo vector propio de \mathbf{S} . Esto es, \mathbf{Z} consiste de valores de los CPs muestrales de \mathbf{X} .

Esta propiedad permite considerar que una manera de resolver el problema de la colinealidad en regresión lineal es usar componentes principales en vez de los regresores originales, ignorando aquellos componentes cuyas varianzas son cercanas a cero. Weisberg (1980) asegura que la matriz de covarianzas de los nuevos regresores es diagonal, y que todos los elementos de esta diagonal son distintos de cero, evitando problemas numéricos. Este método puede ser visto como una forma de regresión robusta donde, al reescribir \mathbf{Z} con las combinaciones lineales de \mathbf{X} se obtiene $\mathbf{Y} = \mathbf{X}\mathbf{B}\gamma + \epsilon$, donde $\mathbf{B}\gamma$ sería un estimador sesgado de β pero con menor varianza.

En caso de que se hagan suposiciones sobre la distribución de las \mathbf{X} , en particular en el caso que asuma una distribución dentro de la familia exponencial existen resultados distribucionales para l_k y a_k , sin embargo los resultados presentados hasta aquí son válidos para cualquier distribución de \mathbf{X} , incluso distribuciones discretas.

Capítulo 3

Análisis de Componentes Principales y Análisis Factorial

En muchas ocasiones el ACP suele presentarse equivocadamente como un caso particular de análisis factorial. Aun cuando tanto el ACP como las distintas técnicas de análisis factorial pretenden la reducción de dimensionalidad de un conjunto de datos, ambos procedimientos tienen grandes diferencias entre sí. A continuación se presentan brevemente los fundamentos del análisis factorial para poder ilustrar las similitudes y discrepancias entre ambos tipos de análisis.

3.1. Modelos para Análisis Factorial

La idea básica subyacente al análisis factorial es que si \mathbf{X} es el vector de p variables observadas, éstas pueden ser expresadas como funciones lineales de $m (< p)$ variables aleatorias hipotéticas o factores comunes, más un término de error, en ocasiones llamado factor específico por existir uno distinto para cada variable original. Si x_1, \dots, x_p son las variables, f_1, \dots, f_m los factores

para una m dada, y e_1, \dots, e_p los errores, entonces

$$\begin{aligned} x_1 &= \lambda_{11}f_1 + \lambda_{12}f_2 + \dots + \lambda_{1m}f_m + e_1 \\ x_2 &= \lambda_{21}f_1 + \lambda_{22}f_2 + \dots + \lambda_{2m}f_m + e_2 \\ &\vdots = \vdots \\ x_p &= \lambda_{p1}f_1 + \lambda_{p2}f_2 + \dots + \lambda_{pm}f_m + e_p \end{aligned} \tag{3.1}$$

donde λ_{jk} , $j = 1, \dots, p$, $k = 1, \dots, m$ son constantes desconocidas llamadas cargas factoriales, y e_j , $j = 1, \dots, p$ son términos de error, llamados comúnmente “error específico”. La ecuación (3.1) puede ser escrita también en forma matricial:

$$\mathbf{X} = \mathbf{\Lambda f} + \mathbf{e} \tag{3.2}$$

El primer contraste entre el ACP y el análisis factorial se hace evidente. Un análisis factorial intenta reducir p a m dimensiones usando un modelo que relaciona x_1, \dots, x_p con m variables hipotéticas no observadas. El ACP no tiene un modelo explícito.

La ecuación (3.2) puede ser modificada haciendo el modelo no lineal, o permitiendo la inclusión de varios errores de medida y otros residuales específicos (Thurstone, 1969), los cuales se encuentran representados por el término de error único \mathbf{e} del modelo presentado.

Existen varios supuestos asociados con el modelo factorial

(i)

$$\mathbf{E}(\mathbf{e}) = \mathbf{0}, \mathbf{E}(\mathbf{f}) = \mathbf{0}, \mathbf{E}(\mathbf{X}) = \mathbf{0}.$$

De estos tres supuestos, el primero es un supuesto estándar para los términos de error común a la mayoría de los modelos estadísticos, el segundo es conveniente y no causa pérdida de generalidad. El tercero podría no ser cierto, sin embargo (3.2) puede ser modificado para convertirse en $\mathbf{X} = \boldsymbol{\mu} + \mathbf{\Lambda f} + \mathbf{e}$, donde $\mathbf{E}(\mathbf{X}) = \boldsymbol{\mu}$, o bien, trabajar con variables centradas. Lo cual no causa mayores complicaciones algebraicas ni pérdida de generalidad.

(ii)

$$\mathbf{E}(\mathbf{e}\mathbf{e}') = \boldsymbol{\Psi}, \mathbf{E}(\mathbf{f}\mathbf{e}') = \mathbf{O}, \mathbf{E}(\mathbf{f}\mathbf{f}') = \mathbf{I}_m$$

donde Ψ es una matriz diagonal, \mathbf{O} es una matriz de ceros y \mathbf{I}_m es una matriz identidad de orden m .

El primer supuesto indica que los términos de error no están correlacionados, lo cual es un supuesto básico del modelo factorial, el que todo en \mathbf{X} que pueda ser atribuido a influencias comunes esté contenido en $\Lambda\mathbf{f}$, y $e_j, e_k, j \neq k$ no estén relacionados. El segundo supuesto, que los factores comunes no tengan relación con los factores específicos también es fundamental. Sin embargo el tercer supuesto puede ser relajado, permitiendo cierta correlación entre los factores comunes.

(iii) Para fines de inferencia estadística, es necesario hacer supuestos acerca de la distribución de los datos. Usualmente, tanto en el análisis factorial como en ACP, se asume que los datos son normales multivariados, aun cuando en la práctica no siempre se cumplen estos supuestos.

(iv) Para evitar el problema de multiplicidad de soluciones de Λ , se requieren varias restricciones de esta matriz dependiendo del método de estimación a utilizar.

A diferencia del modelo de regresión estándar, donde Λ sería conocido y los únicos valores desconocidos estarían contenidos por \mathbf{f} , en el modelo factorial tanto Λ como \mathbf{f} son desconocidos. Lo anterior implica que las técnicas de estimación a utilizar serán diferentes, que existe indeterminación en las soluciones, y que la solución de “mejor ajuste” no será única.

La estimación en este modelo es usual e inicialmente hecha en términos de Λ y Ψ , mientras que la estimación de \mathbf{f} se realizan en una etapa posterior. Dados los supuestos presentados, la matriz de covarianza puede ser calculada para ambos lados de (3.2) dando

$$\Sigma = \Lambda\Lambda' + \Psi \quad (3.3)$$

En la práctica, suele utilizarse la matriz de covarianza o correlación \mathbf{S} en vez de Σ , y Λ y Ψ son encontrados tales que satisfagan

$$\mathbf{S} = \Lambda\Lambda' + \Psi,$$

tan precisamente como sea posible. La indeterminación de la solución es ahora evidente; si $\mathbf{\Lambda}$, $\mathbf{\Psi}$ es una solución de (3.3), y \mathbf{T} es una matriz ortogonal, entonces $\mathbf{\Lambda}^*$ y $\mathbf{\Psi}$ son también una solución, dado $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{T}$. Esto se debe a que

$$\mathbf{\Lambda}^* \mathbf{\Lambda}^{*'} = (\mathbf{\Lambda}\mathbf{T})(\mathbf{\Lambda}\mathbf{T})' = \mathbf{\Lambda}\mathbf{T}\mathbf{T}'\mathbf{\Lambda}' = \mathbf{\Lambda}\mathbf{\Lambda}',$$

dado que \mathbf{T} es ortogonal.

Debido a esta situación, la estimación de $\mathbf{\Lambda}$ y $\mathbf{\Psi}$ procede típicamente en dos etapas. En la primera, algunas restricciones son impuestas sobre $\mathbf{\Lambda}$ de tal manera que se encuentre una solución única. Habiendo encontrado esta solución inicial, otras soluciones son halladas mediante rotación de $\mathbf{\Lambda}$, esto es, se exploran otras soluciones mediante multiplicación por una matriz ortogonal \mathbf{T} . Un criterio particular permite decidir entonces cuál de esas soluciones es “la mejor”. Existen varios criterios para esta elección, en general se busca hacer la estructura de $\mathbf{\Lambda}$ tan simple como sea posible, ya sea haciendo que la mayor parte de sus elementos sean lo más cercanos o lejanos de cero, y con tan pocos como sean posibles tomando valores intermedios. Asimismo, algunos criterios permiten que \mathbf{T} sea oblicua y no ortogonal, en casos donde se suponga cierta relación entre factores.

3.2. Estimación de una solución inicial única

En la primera etapa de la estimación de $\mathbf{\Lambda}$ y $\mathbf{\Psi}$, en ocasiones existen problemas con la identificabilidad, cuando el tamaño del conjunto de datos es demasiado pequeño comparado con el número de parámetros para permitir que estos parámetros sean estimados. Asumiendo que la identificabilidad no sea un problema, el desarrollo de la tecnología de cómputo ha permitido hacer estimaciones iniciales por métodos con adecuados sustentos matemáticos que antes resultaban impensables por la complejidad de los cálculos algebraicos implicados.

El método más usado para la estimación de una solución única es el llamado análisis del factor principal. Éste es un método iterativo que parte de la estimación inicial de una matriz $\mathbf{\Psi}_i$ ($i = 1$), $\hat{\mathbf{\Sigma}}_i := \mathbf{\Sigma} - \mathbf{\Psi}_i$, con $\hat{\mathbf{\Sigma}}_i \approx \mathbf{\Lambda}\mathbf{\Lambda}'$, a partir de donde se obtiene $\mathbf{\Lambda}$ como matriz de factores principales. Dada

ésta estimación se procede a obtener $\Psi_{i+1} = \text{Diag}(\Sigma - \hat{\Sigma}_i)$ y se repiten los pasos hasta convergencia de las iteraciones.

Otro de los métodos más usados implica la estimación máximo verosímil de Ψ y Λ , asumiendo normalidad multivariada de \mathbf{f} y \mathbf{e} , y por tanto de \mathbf{X} . El método busca estimadores máximo verosímiles (EMV) para las cargas factoriales y varianzas del modelo $\Sigma = \Lambda\Lambda' + \Psi$, donde $\Sigma = \mathbf{E}(\mathbf{X}\mathbf{X}')$. Para una muestra de n observaciones $\mathbf{x}_1, \dots, \mathbf{x}_n$ independientes p -dimensionales de una distribución normal multivariada se tiene que $\mathbf{E}(\mathbf{X}) = \mu$ y $\mathbf{E}(\mathbf{X}\mathbf{X}' - \mu\mu') = \Sigma$. La verosimilitud conjunta para estas n observaciones está dada por el producto

$$L = \prod_{i=1}^n \frac{|\Sigma|^{-\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} \exp \left[-\frac{1}{2}(\mathbf{x}_i - \mu)' \Sigma^{-1}(\mathbf{x}_i - \mu) \right].$$

Aunque es más conveniente la expresión usando su logaritmo natural:

$$\ln L = -\frac{1}{2} \left[np \ln(2\pi) + n \ln |\Sigma| + \sum_{i=1}^n (\mathbf{x}_i - \mu)' \Sigma^{-1}(\mathbf{x}_i - \mu) \right].$$

A partir de la cual se estiman simultáneamente μ y Σ . Morrison (1967) demuestra que el EMV para μ es: $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$, y sustituyendo este valor, Jöreskog (1976) propone un método basado en el algoritmo de Fletcher-Powell y descomposición de Cholesky para estimar Σ .

Es de llamar la atención que las cargas factoriales halladas por máxima verosimilitud para una matriz de correlación son equivalentes a aquellas de la correspondiente matriz de covarianza, esto es, son invariantes ante cambios en la escala. Lo anterior contrasta con lo que sucede con el ACP.

Un problema potencial de la estimación por máxima verosimilitud es que se basa en el supuesto de normalidad multivariada, lo cual podría no ser justificable en muchas ocasiones. Sin embargo Morrison (1976) y Rao (1955) han demostrado que los EMV también optimizan dos criterios no basados en supuestos distribucionales.

Morrison señala que si el modelo factorial (3.2) se cumple exactamente, entonces las correlaciones parciales entre los elementos de \mathbf{X} , dado \mathbf{f} , son cero, ya que \mathbf{f} aporta toda la variación común en los elementos de \mathbf{X} . Así, la determinante de esta matriz de correlaciones obtendrá su valor máximo cuando

todos los elementos fuera de la diagonal sean cero, por lo tanto, maximizar este determinante es una manera de intentar minimizar los valores absolutos de las correlaciones parciales. Este proceso de maximización lleva a los EMV, los cuáles se obtendrían se cumpla o no el supuesto de normalidad multivariada.

Por otro lado, Rao ha sugerido un proceso basado en análisis de correlaciones canónicas entre \mathbf{X} y \mathbf{f} . Este autor sugiere buscar, sucesivamente, pares de funciones lineales $\{\mathbf{a}'_{k1}\mathbf{X}, \mathbf{a}'_{k2}\mathbf{f}\}$ que tengan correlación máxima tras ser sujetas a estar no correlacionadas con los pares previos. De esta manera las cargas factoriales son proporcionales a los elementos de \mathbf{a}_{k2} , $k = 1, \dots, m$, lo cual lleva a obtener los mismos EMV basados en el supuesto de normalidad multivariada. Sin embargo, el procedimiento de Rao, al estar basado en correlaciones parciales y análisis canónico, no requiere supuesto alguno sobre la distribución de las variables.

En cierta manera, el comportamiento de las correlaciones parciales y el criterio de correlaciones canónicas son similares al fenómeno en regresión donde el criterio de mínimos cuadrados es válido sin importar la distribución de los términos de error, pero si los errores son normalmente distribuidos entonces los estimadores de mínimos cuadrados tienen la virtud añadida de maximizar la función de verosimilitud, siendo posible a su vez encontrar una distribución para ello.

Una forma alternativa de obtener estimadores iniciales para $\mathbf{\Lambda}$ es utilizar los m primeros CPs. Si $\mathbf{Z} = \mathbf{A}'\mathbf{X}$ es el vector consistente de todos los p CPs, con \mathbf{A} definida con α_k , el k -ésimo valor propio de $\mathbf{\Sigma}$, y su k -ésima columna como en (2.1), entonces $\mathbf{X} = \mathbf{A}\mathbf{z}$ debido a la ortogonalidad de \mathbf{A} . Si \mathbf{A} es particionada en sus primeras m y sus últimas $(p - m)$ columnas, con una partición similar de las filas de \mathbf{z} , entonces

$$\begin{aligned} \mathbf{X} &= (\mathbf{A}_m \mid \mathbf{A}_{p-m}^*) \begin{pmatrix} \mathbf{Z}_m \\ \mathbf{Z}_{p-m}^* \end{pmatrix} \\ &= \mathbf{A}_m \mathbf{Z}_m + \mathbf{A}_{p-m}^* \mathbf{Z}_{p-m}^* \\ &= \mathbf{\Lambda} \mathbf{f} + \mathbf{e} \end{aligned} \tag{3.4}$$

donde

$$\mathbf{\Lambda} = \mathbf{A}_m, \mathbf{f} = \mathbf{Z}_m, \mathbf{e} = \mathbf{A}_{p-m}^* \mathbf{Z}_{p-m}^*.$$

La ecuación (3.4) es muy parecida al modelo factorial (3.2), sin embargo viola un supuesto básico del modelo: los elementos de \mathbf{e} en (3.4) no se encuentran usualmente sin correlacionar. Así, a pesar de la sofisticación aparente de usar la versión muestral de \mathbf{A}_m como estimador inicial, su justificación teórica no es en realidad muy fuerte.

3.3. Rotación de factores

Uno de los criterios más populares de rotación, opción automática en los paquetes estadísticos más conocidos, es la rotación varimax, una rotación ortogonal del eje de factores que maximiza la varianza de las cargas cuadráticas de un factor (columna) en todas las variables (filas) de la matriz de factores. Este método se ilustra a continuación.

Supóngase que $\mathbf{B} = \mathbf{AT}$ y que \mathbf{B} tiene elementos b_{jk} ; $j = 1, \dots, p$, $k = 1, \dots, m$. Entonces para rotación varimax la matriz de rotación ortogonal \mathbf{T} es elegida de tal manera que maximice

$$Q = \sum_{k=1}^m \left[\sum_{j=1}^p b_{jk}^4 - \frac{1}{p} \left(\sum_{j=1}^p b_{jk}^2 \right)^2 \right] \quad (3.5)$$

Donde los términos entre corchetes son proporcionales a las varianzas de las cargas cuadráticas para cada factor rotado. En las implementaciones usuales del análisis factorial las cargas están necesariamente entre -1 y 1, así el criterio genera cargas cuadráticas hacia los extremos del rango entre 0 y 1, y por tanto cargas que tienden hacia el -1, el 1 o el cero, alejándose de los valores intermedios, como es requerido. La cantidad Q en la ecuación (3.5) es el criterio varimax crudo. Se puede obtener una versión normalizada reemplazando b_{jk} por

$$\frac{b_{jk}}{\sqrt{\sum_{k=1}^m b_{jk}^2}}$$

en (3.5) (Kaiser, 1958).

A pesar de la popularidad de la rotación varimax, en la mayoría de los paquetes computacionales actuales se cuenta con otras opciones de rotación ortogonal como la cuartimax y la equamax.

Distintos analistas han propuesto diferentes criterios equivalentes involucrando los elementos b_{jk} , $j = 1, \dots, p$, $k = 1, \dots, m$ de \mathbf{B} (Neuhaus & Wrigley, 1954; Carrol, 1953; Saunders, 1953; Ferguson, 1954), mismos que son conocidos como criterio cuartimax. El criterio cuartimax es una opción que minimiza el número de factores necesarios para explicar cada variable. Este tipo de rotación suele generar un factor general en el cual la mayoría de las variables están cargadas en un nivel alto o medio.

Una de las expresiones más simples para estos fines es la de Ferguson, quien propone maximizar la suma de los coeficientes b_{jk} , cada uno de ellos elevados a la cuarta potencia, propuesta con base en algunos resultados de teoría de la información. Así, para este autor, el criterio cuartimax consiste en encontrar simultáneamente todos los coeficientes de la matriz de transformación ortonormal \mathbf{T} cuando, multiplicada por \mathbf{A} , produce \mathbf{B} de tal manera que se hace máxima

$$Q = \sum_{j=1}^p \sum_{k=1}^m b_{jk}^4$$

El criterio equamax, por su parte, consiste en encontrar el valor óptimo para

$$K = \frac{\sum_{j=1}^r \sum_{i=1}^n a_{ij}^4}{\sum_{j=1}^r \left(\sum_{i=1}^n a_{ij}^2 \right)^2}$$

donde los elementos a_{ij} son los correspondientes a la matriz de estructura factorial normalizada $\mathbf{H}^{-1}\mathbf{A}$, donde \mathbf{H}^2 es la matriz diagonal de $n \times n$ de las communalidades. Este factor K es una constante que aparece al intentar integrar los criterios varimax y cuartimax en una sola ecuación que se resuelva para ambos. En la rotación varimax la constante K toma el valor de 1, mientras que en la cuartimax, toma un valor de 0. Saunders (1953), sin embargo, notó que estos intentos de integración no siempre daban resultados satisfactorios por la complejidad de las expresiones necesarias, en especial cuando

se hubieran elegido muchos factores para rotación. Sin embargo, también encontró que este método daba los resultados más satisfactorios cuando K era igual a la mitad del número de factores rotados, esto es, cuando $K = \frac{m}{2}$. Saunders decidió llamar a la solución con $K = \frac{m}{2}$ solución equamax.

Además de éstas y otras técnicas de rotación ortogonales, es posible encontrar rotaciones de tipo oblicuo como la oblimin directa, la cuatrimin y la pro-max. El hecho de que los métodos ortogonales produzcan factores más fáciles de interpretar ha favorecido su elección en la práctica del usuario común. Sin embargo en muchos casos se esperaría teóricamente que los factores a obtener se encuentren correlacionados entre si, por lo que los métodos de rotación oblicua generarían soluciones más realistas. Además, si los factores se encontraran verdaderamente sin correlacionar, ambos tipos de rotaciones producirían resultados similares.

Por ejemplo, como contraparte oblicua a la rotación varimax, Kaiser (1958) propuso la minimización a través de todos los pares de los ejes de referencia de la suma de las covarianzas de las cargas factoriales cuadráticas (sumando cada par a través de las variables), es decir, minimizar

$$Q = \sum_{j=1}^m \sum_{k=1}^m \left[\sum_{i=1}^p b_{ij}^2 b_{ik}^2 - \frac{1}{p} \left(\sum_{i=1}^p b_{ij}^2 \right) \left(\sum_{i=1}^p b_{ik}^2 \right) \right]; \forall j \neq k \quad (3.6)$$

Este criterio alternativo es conocido como oblimin directo. La rotación oblimin directa es el método más usado cuando uno desea una solución no ortogonal (Abdi, 2003).

Así como la rotación es una parte integral de hacer un análisis factorial, haciendo los factores rotados tan fáciles de interpretar como sea posible, la misma idea puede ser utilizada para simplificar CPs. Un componente principal es una función lineal de todas la p variables originales. Si los coeficientes o cargas para un CP son todos de tamaño similar, o si unos pocos son grandes y los restantes pequeños, el componente se puede interpretar fácilmente. Sin embargo, si existen cargas grandes, pequeñas e intermedias, el componente puede ser difícil de interpretar.

Supóngase que se ha decidido que los primeros m componentes aportan la mayor parte de la variación en un conjunto de datos p -dimensional. Podría

entonces argumentarse que es más importante interpretar simplemente el espacio m -dimensional definido por estos componentes que interpretar cada componente individual. Una manera de conseguirlo es rotar los ejes dentro de su espacio m -dimensional de tal manera que simplifique la interpretación de los ejes tanto como sea posible.

Supóngase que \mathbf{A}_m es la matriz de $(p \times m)$ cuya k -ésima columna es el vector de coeficientes para el k -ésimo CP. De manera similar al procedimiento usado en el análisis factorial, los CPs rotados ortogonalmente tendrán coeficientes dados por las columnas de \mathbf{B}_m , donde $\mathbf{B}_m = \mathbf{A}_m \mathbf{T}$, y \mathbf{T} es una matriz ortogonal de $(m \times m)$. La matriz \mathbf{T} se elige de tal manera que optimice uno o varios criterios de simplicidad usados en el análisis factorial. La elección de cual criterio de simplicidad suele ser menos importante que la elección de m . Rotaciones oblicuas también son posibles y suelen dar un poco más de flexibilidad para algunos tipos de datos en particular.

El tipo de simplicidad que es favorecido por la mayoría de los criterios de rotación intenta llevar los coeficientes a cero o hacia su máximo valor absoluto posible, el cual con la mayor parte de las escalas usadas para los datos es uno. La idea es que debería ser entonces claro cuáles variables son importantes en un componente rotado -aquellas con grandes valores absolutos en sus coeficientes- y cuáles no son importantes -aquellos con coeficientes cercanos a cero.

En el contexto del ACP, es importante señalar algunos aspectos del proceso de rotación:

- Tras rotación, ya no se cuenta con los CPs obtenidos durante el análisis, sino con proyecciones de los puntos originales sobre un subespacio m -dimensional.
- Un criterio de rotación debe ser tomado de entre un amplio número posible de ellos. Frecuentemente se elige arbitrariamente el que se use por *default* (usualmente varimax) en algún paquete de computación especializado. aportada por los CPs. Cuando la rotación es llevada a cabo, el total de varianza en el subespacio m -dimensional permanece constante; es aun la máxima que puede ser obtenida, pero redistribuida entre los componentes de manera más equitativa. Esto significa que la información acerca de la naturaleza de algún componente dominante

podría perderse, a menos que éste ya sea simple en el sentido definido por el criterio de rotación utilizado.

- La elección de m puede tener grandes efectos en los resultados tras rotación. Interpretar las “dimensiones” más importantes para un conjunto de datos es más difícil cuando estas dimensiones pudieran ser observadas o no según el valor elegido para m .
- La elección de la restricción de normalización usada para las columnas de la matriz \mathbf{A}_m cambia las propiedades de los coeficientes rotados. En los paquetes computacionales se suele rotar una matriz $\tilde{\mathbf{A}}_m$ cuyas columnas tienen longitudes iguales a sus valores propios correspondientes, en vez de haber sido normalizadas para tener longitud unitaria. Lo anterior se debe a que por fines prácticos, estos paquetes toman prestado el algoritmo del análisis factorial, donde suele estar categorizado el ACP.

A pesar de la popularidad de las técnicas de rotación, éstas no son la única manera para buscar representaciones útiles de espacios de variables en espacios de menor dimensión. Otras representaciones son consideradas en el capítulo 5 de este trabajo.

Capítulo 4

Elección de un Subconjunto de Componentes Principales

El principal objetivo en muchas aplicaciones del ACP es reemplazar p elementos de \mathbf{X} por un número mucho más pequeño m de CPs, los cuales sin embargo conserven la mayor parte de la varianza. Así, es crucial saber qué tan pequeño puede ser m sin una pérdida importante de varianza. A continuación se presentan varias de las propuestas conocidas más comunes para seleccionar los subconjuntos de CPs que satisfagan estas condiciones.

4.1. Porcentaje acumulado de varianza total

Probablemente el criterio más obvio para seleccionar m sea elegir un porcentaje acumulado de varianza mínimo con el que se desea que los CPs seleccionados contribuyan. El número mínimo de CPs es entonces el valor de m para el cual el porcentaje elegido es igualado o excedido. Basta recordar que los CPs son elegidos sucesivamente para tener la varianza más grande que sea posible, y la varianza del k -ésimo CP es l_k . Además, $\sum_{k=1}^p l_k = \sum_{j=1}^p s_{jj}$, la suma de las varianzas de los CPs es igual a la suma de las varianzas de los elementos de \mathbf{X} . De esta manera el “porcentaje de varianza con la que los primeros m CPs contribuyen” es

$$t_m = 100 \frac{\sum_{k=1}^m l_k}{\sum_{j=1}^p s_{jj}} = 100 \frac{\sum_{k=1}^m l_k}{\sum_{k=1}^p l_k},$$

lo cual se reduce a

$$t_m = \frac{100}{p} \sum_{k=1}^m l_k$$

en el caso del ACP usando una matriz de correlación.

Elegir un punto de corte t^* en el rango de 70 % a 90 % y reteniendo m CPs, donde m es el entero más pequeño para el cual $t_m \geq t^*$, da una regla que en la práctica preserva en los primeros m CPs la mayor parte de la información en \mathbf{X} . El valor óptimo para t^* se hará generalmente más pequeño según p o n , el número de observaciones, se vuelvan más grandes. Y aunque un punto de corte sensible entre el 70 % y el 90 % sea lo más común, en ocasiones un valor mayor o menor puede ser más apropiado según los detalles prácticos del conjunto particular de datos con los que se esté trabajando, y la razón para disminuir la dimensión.

4.2. Tamaño de la varianza de los componentes principales

Esta regla está construida específicamente para su uso con matrices de correlación, sin embargo puede ser adaptada para algunos tipos de matrices de covarianza. La idea detrás de la regla es que, si todos los elementos de \mathbf{X} son independientes, entonces los CPs son los mismos que las variables originales y tendrán todas varianzas unitarias en el caso de una matriz de correlación. Así, cualquier CP con varianza menor que uno contendrá menos información que cualquiera de las variables originales y por tanto no valdrá la pena conservarla. La regla, en su forma más simple, es llamada *Regla de Kaiser* (Kaiser, 1960) y retiene solo aquellos CPs cuyas varianzas l_k sean mayores que la unidad. Si el conjunto de datos contiene grupos de variables con correlaciones intragrupalas elevadas, pero correlaciones bajas entre grupos, entonces existirá un CP asociado con cada grupo cuya varianza es mayor que uno, y cualquier otro CP asociado al grupo tendrá varianza menor que uno. Es decir, la regla conservará un, y sólo un, CP asociado con cada uno de estos grupos de variables, lo cual parece un curso de acción razonable para datos de esta naturaleza.

Se puede argumentar que un punto de corte en $l_k = 1$ conserva muy pocos CPs. Considérese una variable que, en la población, es relativamente independiente de las demás. En una muestra, tal variable tendrá pequeños coeficientes en $(p - 1)$ de los CPs pero dominará uno de los CPs, cuya varianza será cercana a $l_k = 1$ en una matriz de correlación. Como esta variable provee información independiente a la de las demás variables no sería recomendable eliminarla. Sin embargo, usando la Regla de Kaiser, esta eliminación podría ocurrir dado que, por variaciones muestrales, $l_k < 1$. De esta manera sería recomendable elegir un punto de corte l^* algo menor que 1, para permitir variaciones muestrales. Jolliffe (2002) sugiere que $l^* = 0,7$ podría ser un valor adecuado.

La Regla de Kaiser fue específicamente diseñada para matrices de correlación, sin embargo puede ser adaptada para matrices de covarianza eligiendo como punto de corte $l^* = \bar{l}$, la media de los valores propios, en vez de la unidad, o bien, $l^* = 0,7\bar{l}$, en lugar del criterio de Jolliffe. Sin embargo, para matrices de covarianza con varianzas muy diferentes entre si, este método y el del porcentaje de varianza total retienen muy pocos CPs.

Otra manera de observar los tamaños de las varianzas individuales es usar el modelo de la varita rota. Si se rompe una varita de manera aleatoria en p segmentos, entonces puede ser demostrado que la longitud esperada del k -ésimo segmento es

$$l_k^* = \frac{1}{p} \sum_{j=k}^p \frac{1}{j}.$$

Una manera de decidir si la proporción de varianza con la que contribuye el k -ésimo CP es lo suficientemente grande para conservarlo es comparar esta proporción con l_k^* . Los CPs para los cuales la proporción exceda l_k^* son conservados, los demás serán eliminados.

4.3. La gráfica de sedimentación y el diagrama log - valor propio

Las dos primera reglas presentadas involucran cierto grado de subjetividad en la elección de los puntos de corte t^* y l^* . La gráfica de sedimentación (*scree graph*, en inglés) discutida y bautizada por Cattell (1966), aún cuando ya era previamente de uso común, es aún más subjetiva en su forma usual, ya que involucra la observación de una gráfica de l_k contra k para decidir en cuál valor de k las pendientes de la líneas que unen los puntos graficados tienen una pendiente pronunciada a la izquierda de k , y no la tienen a la derecha. Este valor de k , definiendo un “codo” en la gráfica, es tomado entonces como el número de CPs que deben ser conservados. Una alternativa, cuando los valores propios de los datos decaen en progresión geométrica, es graficar $\log(l_k)$, en vez de l_k , contra k ; la gráfica resultante es conocida como el Diagrama Log - Valor Propio (LVP).

La gráfica de sedimentación fue formulada por Cattell presentando argumentos poderosos de que cuando es usada en análisis factorial es enteramente objetiva y debe sugerir el número “correcto” de factores. De hecho, presentaba la regla como una manera de elegir una cota superior al número de factores tras rotación. Y a pesar de que Cattell no parece haber considerado su uso para ACP, este método ha sido adoptado ampliamente para estos fines.

La forma en que Cattell formula la regla de elección de k va más allá de un simple cambio de pendiente pronunciada a no pronunciada. Él propone buscar el punto a partir del cual la gráfica define una línea más o menos recta, no necesariamente horizontal. El primer punto en la línea recta es tomado como el último factor (o componente) a ser conservado. Si hay dos o más líneas rectas formadas por los valores propios menores, entonces el punto de corte se toma del extremo superior (el de la izquierda) de la línea recta más a la izquierda. Cattell discute acerca de si este punto debe ser el primer factor excluido o el último conservado, y concluye que es preferible incluir este factor, a pesar de que ambas variantes son usadas en la práctica.

Mientras que la primera regla se basa en t^* , el porcentaje mínimo de varianza aportada; la segunda en l^* , la varianza mínima, este tercer método se basa en el tamaño de la diferencia $l_{k-1} - l_k$. De hecho, juzgar cuándo

$l_{k-1} - l_k$ deja de ser grande o pronunciada dependerá tanto de los valores relativos de $l_{k-1} - l_k$ y $l_k - l_{k+1}$, así como del valor absoluto de $l_{k-1} - l_k$, por lo tanto es difícil formular una regla numérica formal y el procedimiento se realiza puramente por análisis gráfico. La formulación de Cattell de buscar el punto en el cual $l_{k-1} - l_k$ se vuelve relativamente constante para varios valores sucesivos de k puede resultar un poco menos subjetivo, sin embargo aun involucra algún grado de juicio.

Las formulaciones propuestas para el uso de la gráfica de sedimentación suelen funcionar bastante bien en la práctica mientras exista un “codo” visible en ella, un cambio en la pendiente discernible. Sin embargo en muchas ocasiones el cambio en la pendiente es gradual haciendo casi imposible usar este procedimiento.

4.4. Número de componentes con valores propios desiguales

Existen varias pruebas de hipótesis concernientes a patrones en Σ y sus vectores y valores propios. La más conocida es aquella que prueba

$$H_{0,q} : \lambda_{q+1} = \lambda_{q+2} = \cdots = \lambda_p = 0,$$

esto es, el caso cuando los últimos $(p - q)$ valores propios son iguales e iguales a cero, contra la alternativa $H_{1,q}$, el caso cuando al menos dos de los últimos $(p - q)$ son diferentes. Ésta es conocida como *Prueba de Bartlett*. Se ha argumentado que usando esta prueba para varios valores de q , pueden ser descubiertos cuántos CPs contribuyen con cantidades sustanciales de variación, y cuántos son simplemente “ruido”. Si m , el número requerido de CPs que han de ser conservados, es definido como el número de CPs que no son ruido, esta prueba puede ser usada secuencialmente para encontrar m .

Siguiendo este procedimiento, se prueba primero $H_{0,p-2}$, esto es $\lambda_{p-1} = \lambda_p = 0$, y si $H_{0,p-2}$ no es rechazada entonces se prueba $H_{0,p-3}$, y si $H_{0,p-3}$ no se rechaza, se prueba $H_{0,p-4}$, y así consecutivamente hasta que la prueba de $H_{0,q}$ es rechazada para $q = q^*$. El valor de m es entonces $q^* + 1$, o $q^* + 2$ en

el caso de que $q^* = p - 2$.

Existen varios problemas para este procedimiento, el primero es que el estadístico de prueba utilizado asume normalidad multivariada de \mathbf{X} . El segundo problema consiste en el hecho de que a menos que $H_{0,p-2}$ sea rechazada, aun se requiere realizar otras pruebas y, en ese caso, hay que tomar en cuenta que el nivel de significancia global del total de pruebas no es el mismo que el nivel de significancia de cada una de ellas. Más aun, el nivel de significancia global no puede ser calculado ya que el número de pruebas no es fijo sino aleatorio, y a que las pruebas no son independientes entre si. Por lo anterior, a pesar de que el procedimiento puede ser usado para estimar m , no puede ser considerado como una pieza formal de inferencia estadística ya que sus niveles de significancia son usualmente desconocidos. Un tercer problema práctico consiste en que esta prueba tiende a conservar más CPs de los que son realmente necesarios. La secuencia inversa de pruebas $H_{0,0}, H_{0,1}, \dots$ también puede ser utilizada, pero tiene las desventajas similares. A fin de cuentas, a pesar de que esta propuesta es mucho más formal que el uso de una gráfica de sedimentación, en esencia ambos métodos son muy parecidos.

4.5. Correlaciones parciales

Para ACP basado en matrices de correlación, Velicer (1976) ha sugerido que las correlaciones parciales entre las p variables, dados los valores de los primeros m CPs, podrían ser usadas para determinar cuántos CPs conservar. El criterio propuesto es el promedio de las correlaciones parciales cuadráticas

$$V = \sum_{i=1}^p \sum_{j=1}^p \frac{(r_{ij}^*)}{p(p-1)},$$

para $i \neq j$, y donde r_{ij}^* es la correlación parcial entre la i -ésima y j -ésima variable, dados los primeros m CPs. El estadístico r_{ij}^* es definido como la correlación entre los residuales de la regresión lineal de la i -ésima variable en los primeros m CPs y los residuales de la regresión correspondiente a la j -ésima variable en los mismos m CPs. Por lo tanto este estadístico lo que hace es medir la fuerza de la relación lineal entre las i -ésima y j -ésima variables tras remover los efectos comunes de los primeros m CPs. El criterio V

decrece para luego incrementarse, según m va creciendo. Velicer sugiere que el valor óptimo de m corresponde al valor mínimo del criterio V .

Sin embargo, supóngase una vez más que existe una variable única relativamente independiente de las otras $(p - 1)$ variables y que tenga dominio sobre uno de los CPs. Esta variable sería omitida usando el criterio de Velicer, a pesar de que en un ACP la información de esta variable es de importancia. Otras propuestas similares han sido presentadas por Jackson (1993) y Reddon (1984), las cuales son desarrolladas en un contexto de análisis factorial, donde la eliminación de variables de este tipo no afectan la determinación del número de factores, sin embargo no son plausibles para un ACP.

Capítulo 5

Componentes Principales e Información

Dado que un objetivo del ACP es identificar combinaciones lineales de las variables originales con medidas importantes de su varianza, entonces se tiene un proceso que crea conglomerados de dichas variables. Para muchos fines, estas combinaciones mantienen gran cantidad de la información contenida en el conjunto mayor de variables originales, sin embargo en otras ocasiones el ACP puede perder gran parte de la información relevante, encontrando conglomerados de datos que resultarían mejor explicados desde una óptica diferente a la de las medidas de los CPs, como por ejemplo el método conocido como *projection pursuit*, y cuya conformación pudiera ser predicha a través de las variables originales mediante modelos lineales generalizados para datos categóricos.

5.1. “Projection Pursuit”

La mayor parte de los métodos que buscan proyecciones de datos p -dimensionales en un número pequeño m de dimensiones, usualmente $m=2$ ó 3 ; confían en una medida del grado en el cuál las soluciones obtenidas reflejan el aspecto de interés de realizar tal proyección, y a partir de estos criterios se elige aquella que optimice esta medida. El ACP es un ejemplo de este proceso, con una medida cuantitativa de la varianza. Sin embargo, una vez que

vamos más allá de los objetivos de maximizar el total de la varianza, las medidas cuantitativas pueden volverse demasiado complicadas para su cálculo con lápiz y papel, o difíciles de programar para su cálculo automatizado en una computadora. Por este motivo, se buscan proyecciones de a lo más dos o tres dimensiones, y en ocasiones sólo una proyección unidimensional resulta viable. La estrategia general para este tipo de descripción de datos ha sido llamada *projection pursuit*, una “búsqueda de la proyección” (Friedman & Tukey, 1974).

El algoritmo de búsqueda de proyecciones mencionado utiliza las distancias interpuntuales, así como la varianza de la nube de puntos, en su búsqueda de proyecciones óptimas. Este algoritmo asocia a cada dirección en el espacio multidimensional, un índice que mide su utilidad como eje de proyección, y luego varía la dirección de la proyección hasta maximizar tal índice. Se procura que el índice usado sea una función con la continuidad suficiente para utilizar algoritmos iterativos de aproximación a su valor máximo. Asimismo, este método asume la existencia de múltiples soluciones, y el usuario deberá analizar visualmente cada una de ellas para juzgar su utilidad e interpretabilidad.

Originalmente se propuso la *projection pursuit* por su mayor simplicidad al momento de usar computadoras menos potentes y veloces que las actuales; sin embargo, el uso de estos procedimientos alternativos también pueden ser considerados cuando se dificulte la obtención de una matriz de varianza-covarianza, cuando la dimensionalidad de los datos sea muy grande, o cuando explicaciones alternativas pudieran proveer mayor información que los CPs tradicionales.

Tómese por ejemplo la **Figura 1**, en ella se observa como los datos se agrupan en dos conglomerados que parecen dos muestras de una distribución elíptica normal bivariada con una pequeña correlación positiva. Estos dos conjuntos se separan por un desplazamiento de la media sobre el segundo CP. Como se espera, el primer CP está en la dirección de la mayor extensión de los datos, mientras que el segundo es ortogonal a éste primero. Esto es, el primer CP no contiene información sobre los conglomerados. Se supondría que si la mayor fuente de variación en los datos es la variación entre conglomerados, entonces el ACP identificará los conglomerados; sin embargo, en la práctica éste no siempre es el caso y en ocasiones el primer CP no estará en la dirección más

informativa.

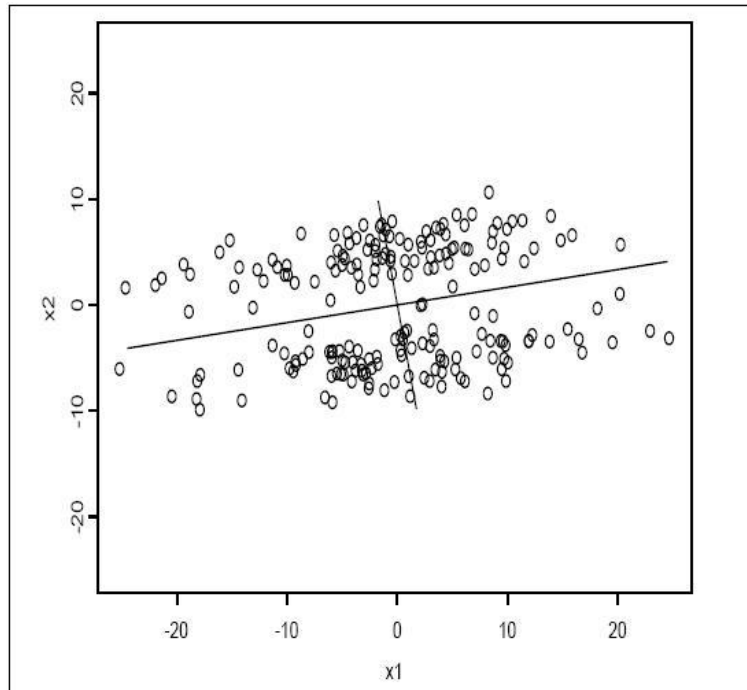


Figura 1. Componentes principales de un conjunto de datos bimodal con dos conglomerados

El método de *projection pursuit* busca proyecciones de los datos en direcciones que exhiben otras estructuras interesantes, como es el caso de la bimodalidad sobre el segundo CP en el ejemplo presentado. Una “estructura interesante” suele hallarse analizando proyecciones de los datos en espacios de menor dimensionalidad como los CPs. Las proyecciones pueden ser usadas para tener una mejor visualización de los conglomerados de los datos, estimaciones de su densidad, e incluso con fines de análisis de regresión.

Una proyección seleccionada al azar de un conjunto de datos de dimensionalidad alta en un espacio de dimensionalidad mucho menor (2 ó 3 dimensiones), tenderá a verse similar a una muestra de una distribución normal multivariada. Este resultado, el cual podría verse como un Teorema Central del Límite para proyecciones, implica que un conjunto de datos normales multivariados es el menos “interesante”. Una proyección específica de dichos datos, sin

embargo, podría revelar patrones no normales, como sesgos, “agujeros”, o multimodalidad en proyecciones lineales de los datos.

5.2. Búsqueda de estructuras de interés

Considérese un variable aleatoria p -dimensional \mathbf{Y} . En general el interés se centrará en una proyección m -dimensional de \mathbf{Y} con $m \ll p$, digamos $\mathbf{A}'\mathbf{Y}$, tales que la variable aleatoria $\mathbf{A}'\mathbf{Y}$ difiere evidentemente de una distribución normal m -variada. Debido a que todas las marginales univariadas de una normal multivariada son normales, y a que el producto cruz de normales es normal multivariado, es necesario encontrar esa $\mathbf{Z} = \mathbf{A}'\mathbf{Y}$ que se distribuya lo más diferente posible de una normal.

Con el fin anterior, algunos autores como Huber (1985), Hall (1989) o Friedman (1987) han propuesto índices de normalidad para medir que tanto se alejan de esta distribución las distintas proyecciones; aunque en muchas ocasiones, como en el ejemplo presentado, la falta de normalidad es evidente. En el ejemplo presentado la bimodalidad de los datos es clara. Al realizar el ACP, si solamente observamos los parámetros típicos que se obtienen del procedimiento, el primer CP ocultará la existencia de los dos grupos, mientras que el segundo CP la mostrará con claridad.

Por ejemplo, una manera de cuantificar la no normalidad es considerar la función de densidad probabilística de la variable proyectada y compararla con la función de densidad probabilística ϕ de una variable aleatoria normal estándar. Si p es la densidad de \mathbf{Z} , se desea que ésta sea muy diferente de ϕ . Éste es un problema opuesto al problema de la función de aproximación, sin embargo las soluciones propuestas son similares. Mientras que para las funciones de aproximación la norma de Chevishev centra su interés en hallar una función que es “diferente”, una norma L_2 ,

$$\int_{-\infty}^{\infty} (p(z) - \phi(z))^2 dz \quad (5.1)$$

tiene más sentido como medida de la diferencia.

El índice (5.1) es llamado *Índice de Hermite* porque los polinomios de Hermite son adecuados para aproximarse sobre el dominio no acotado. También se le llama *Índice de Hall*, porque Hall (1989) lo estudió extensamente.

Entonces, el objetivo en la búsqueda de proyecciones es encontrar una \mathbf{A} que maximice esta norma. Por ejemplo Friedman (1987) propone primero el mapeo $\mathbf{Z} = \mathbf{A}'\mathbf{Y}$ en $[-1,1]$ mediante la transformada

$$\mathbf{R} = 2\Phi(\mathbf{Z}) - 1,$$

donde Φ es la función de densidad acumulada de una distribución normal estándar. Si p_Z es la función de densidad probabilística de \mathbf{Z} , entonces la densidad probabilística de \mathbf{R} es

$$p_R(r) = \frac{1}{2} \cdot \frac{p_Z(\Phi^{-1}(\frac{r+1}{2}))}{\phi(\Phi^{-1}(\frac{r+1}{2}))}.$$

Si \mathbf{Z} tuviera una distribución normal con media cero y varianza uno, \mathbf{R} tendría distribución uniforme sobre $(-1, 1)$, y por tanto tendría densidad constante $\frac{1}{2}$. Luego, el problema consiste en hallar \mathbf{A} tal que p_R sea lo más diferente de $\frac{1}{2}$. La norma L_2 relevante es:

$$L(\mathbf{A}) = \int_{-1}^1 \left(p_R(r) - \frac{1}{2} \right)^2 dr,$$

lo cual se simplifica

$$L(\mathbf{A}) = \int_{-1}^1 p_R^2(r) dr - \frac{1}{2} \tag{5.2}$$

La norma (5.2), la cual es una función escalar de \mathbf{A} y una funcional de p_R , se llama algunas veces *Índice de Legendre* porque los polinomios de Legendre son series de polinomios ortogonales para funciones sobre dominios finitos que se aproximan naturalmente.

Cuando no se encuentren estructuras evidentemente no normales, el primer

paso es generalmente centrar y esferizar los datos, es decir, hacer las varianzas en todas las direcciones casi iguales. Los datos de la **Figura 1** una vez centrados y esferizados se muestran en la **Figura 2**. Ahora los componentes principales se encuentran sobre los ejes, la aparente normalidad ya no es evidente y, sin embargo, la bimodalidad se mantiene y es más fácil de manejar para otros análisis de la estructura de los datos.

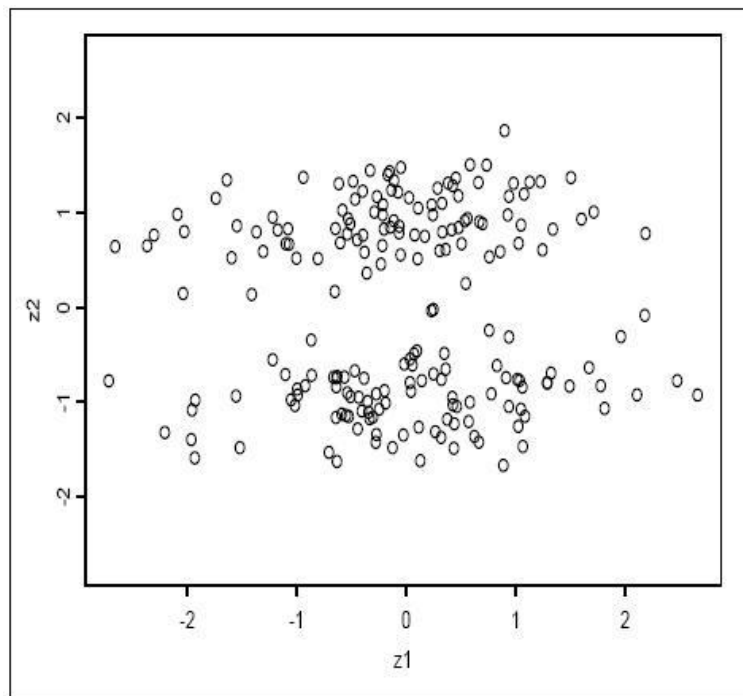


Figura 2. Conjunto de datos esferizados con dos conglomerados

En ocasiones, cuando las proyecciones con las que se está trabajando son tridimensionales, es común que los conglomerados de puntos que se forman en su gráfica pudieran parecer una sola nube de puntos que cumple con las características de normalidad esperadas, sin embargo, a través del uso de software estadístico reciente como SAS, JMP o SPSS es posible “rotar” virtualmente estas nubes de puntos encontrando de manera sencilla nuevas estructuras de interés, simplificando de esta manera el análisis de la información resumida a través de un ACP. Este procedimiento, a pesar de su gran simplicidad, pudiera evitar el tener que enfrentarse a los problemas estadísticos relacionados con la búsqueda numérica de patrones no normales.

5.3. Modelos Lineales Generalizados

Ahora supóngase que sólo existen casos que pertenecen a una de las dos nubes de puntos de la **Figura 2**, cualquier otro caso será considerado una anomalía y desechado, y que se desea predecir la pertenencia de un participante a una de estas dos nubes. Entonces lo que se desea es predecir una variable dicotómica a la cuál se le pueden asignar los valores uno, si el participante se encuentra en la nube, o cero, si no se encuentra en ella y por tanto se encuentra en la otra.

En este caso, ajustar un modelo de regresión lineal tiene varios inconvenientes. Debido a que la variable respuesta Y toma solo dos valores, cero y uno, los errores no son normalmente distribuidos, así se viola un supuesto necesario para hacer inferencia. Además, los errores son heteroscedásticos: $\text{Var}(\varepsilon) = Y(x)(1 - Y(x))$, donde Y es la probabilidad del evento; como Y depende de la matriz de datos \mathbf{X} , se viola otro de los supuestos de la regresión lineal. Finalmente, los valores predichos podrían ser mayores a uno o menores a cero, lo cual es un problema si se deseaba usar estos valores para análisis posteriores.

Como una solución a estos inconvenientes se utilizan los modelos lineales generalizados (MLG), los cuales extienden los modelos de regresión ordinarios para abarcar distribuciones de respuesta no normales y modelar funciones de la media. Los MLG constan de tres componentes:

- 1 Un componente aleatorio que identifica la variable respuesta Y y cuya distribución de probabilidad, pertenece a la familia exponencial:

$$f(y_i, \theta_i) = a(\theta_i)b(y_i)\exp[y_id(\theta_i)],$$

donde θ_i y y_i denotan: el parámetro del modelo y la variable respuesta para la i -ésima observación respectivamente y a , b y d son funciones reales.

- 2 Un componente sistemático que especifica las variables explicativas usadas en una función lineal. Esto es, un vector (η_1, \dots, η_n) de variables

explicativas a través de un modelo lineal. Dado x_{ij} el valor de la predictora j para el sujeto i entonces

$$\eta_i = \sum_{j=0}^p \beta_j x_{ij}, \quad (5.3)$$

con $x_{i0} = 1$ para toda i .

- 3 Una función de enlace que conecta los componentes aleatorio y sistemático. Dada $\mu_i = \mathbf{E}(\mathbf{Y}_i)$, $i = 1, \dots, n$. El modelo enlaza a μ_i con η_i por $\eta_i = g(\mu_i)$, donde la función de enlace g es monótona y diferenciable.

En el modelo de regresión logística se tienen n variables aleatorias binomiales y_i , $i = 1, \dots, n$. Para cada y_i se conoce la cantidad de ensayos m_i y un vector asociado de p predictoras $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$. La probabilidad de éxito $\pi(\mathbf{x}_i)$ depende de \mathbf{x}_i , entonces $y_i \mid \mathbf{x}_i \sim \text{Bin}(m_i, \pi(\mathbf{x}_i))$, $i = 1, \dots, n$.

Un caso común es que $m_i = 1$, es decir, y_i es una variable Bernoulli que toma el valor uno si la característica de interés está presente y cero si no. Para estimar la probabilidad $\pi(\mathbf{x}_i)$ se utiliza una función kernel M aplicada a η_i . La función kernel para la regresión logística debe estar entre cero y uno. La más frecuentemente utilizada es la función logística dada por:

$$\pi(\mathbf{x}_i) = M(\eta_i) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} = \frac{1}{1 + \exp(-\eta_i)}.$$

Entonces, para el modelo de regresión logística el componente sistemático está dado por la ecuación 5.3, y la función de enlace (también denominada *logit*) que hace al modelo lineal es:

$$g(\pi(\mathbf{x}_i)) = \ln \frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)}$$

De esta manera ya se puede contar con un modelo para predecir la pertenencia a uno de los conglomerados como era deseado, a través de asignar un uno al conglomerado de mayor probabilidad. Más aun queda por resolver aquellas posibles situaciones donde se hallen más de dos conglomerados como resultados de nuestra exploración. Suponga que la variable respuesta Y tiene J

categorías: $0, 1, 2, \dots, J - 1$ conglomerados. De manera similar al caso binomial, en el cual la variable respuesta es parametrizada en términos del logit de $Y = 1$ vs $Y = 0$, en el caso multinomial se tienen $J - 1$ funciones logit comparando: $Y = 1$ vs $Y = 0$, $Y = 2$ vs $Y = 0$, \dots , $Y = J - 1$ vs $Y = 0$; los demás logit pueden derivarse de combinaciones de los anteriores, por ejemplo el logit de $Y = 2$ vs $Y = 1$ se obtiene de la diferencia entre el logit de $Y = 2$ vs $Y = 0$ y el logit de $Y = 1$ vs $Y = 0$. Dado $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$ el vector de p predictoras asociado a y_i , y dado $\pi_{ji} = \pi_j(\mathbf{x}_i) = \mathbf{P}(Y = j \mid \mathbf{x}_i)$ para $j = 0, \dots, J - 1$, entonces las $J - 1$ funciones logit son

$$\begin{aligned} g_1(\mathbf{x}_i) &= \ln \left(\frac{\pi_{1i}}{\pi_{0i}} \right) = \beta_{10} + \beta_{11}x_{i1} + \dots + \beta_{1p}x_{ip} \\ g_2(\mathbf{x}_i) &= \ln \left(\frac{\pi_{2i}}{\pi_{0i}} \right) = \beta_{20} + \beta_{21}x_{i1} + \dots + \beta_{2p}x_{ip} \\ &\vdots \\ g_{(J-1)}(\mathbf{x}_i) &= \ln \left(\frac{\pi_{(J-1)i}}{\pi_{0i}} \right) = \beta_{(J-1)0} + \beta_{(J-1)1}x_{i1} + \dots + \beta_{(J-1)p}x_{ip} \end{aligned}$$

Ahora, dado que $\sum_{j=0}^{J-1} \pi_{ji} = 1$, es posible demostrar que:

$$\pi_{ji} = \frac{\exp g_j(\mathbf{x}_i)}{\sum_{j=0}^{J-1} \exp g_j(\mathbf{x}_i)}, \quad j = 0, 1, \dots, J - 1, \quad \exp g_0(\mathbf{x}_i) = 1$$

De esta manera, el problema de hallar un modelo para predecir la pertenencia a alguno de los J conglomerados queda resuelto, asignando al individuo al conglomerado de mayor probabilidad ajustada. Una presentación extensa de la racional para obtener las ecuaciones de los MLG que fueron presentadas anteriormente puede ser encontrada en Prieto-Castellanos (2005).

La estimación de los parámetros para estos modelos puede realizarse por medio de máxima verosimilitud. A partir de la función log-verosímil para la familia exponencial,

$$\ell = \sum_{i=1}^n \frac{y_i \theta_i - d(\theta_i)}{a(\phi)} - \sum_{i=1}^n c(y_i, \phi)$$

es posible obtener el estimador máximo verosímil

$$\frac{\partial \ell}{\partial \beta} = \frac{1}{a(\phi)} \mathbf{X}' \mathbf{W} \Delta (\mathbf{y} - \mu),$$

donde $\mathbf{W}, \mathbf{\Delta}, \mu$ son funciones del parámetro β desconocido. Específicamente

$$\mathbf{W} = \text{diag} \left[\frac{\partial^2 d(\theta_i)}{\partial \theta_i^2} \cdot \left(\frac{\partial g(\mu_i)}{\partial \mu} \right)^2 \right]^{-1}, \quad \mathbf{\Delta} = \text{diag} \left[\frac{\partial g(\mu_i)}{\partial \mu} \right].$$

Igualando a cero $\frac{\partial \ell}{\partial \beta}$ se obtiene

$$\mathbf{X}'\mathbf{W}\mathbf{\Delta}\mathbf{y} = \mathbf{X}'\mathbf{W}\mathbf{\Delta}\mu. \quad (5.4)$$

Sin embargo, típicamente las funciones $\mathbf{W}, \mathbf{\Delta}, \mu$ no son funciones lineales de β , impidiendo la resolución analítica de la ecuación 5.4, resolviéndose usualmente por métodos iterativos como el de mínimos cuadrados ponderados (McCullach & Searle, 2001).

En la aplicación que se presenta en este trabajo se realizará un ACP tanto con una rotación ortogonal (varimax) como una oblicua (oblimin directa) y se realizará una exploración posterior buscando estructuras interesantes en los CPs generados a partir de los datos de la Consulta Infantil y Juvenil 2000 del Instituto Federal Electoral, y se pondrán a prueba MLGs para predecir la pertenencia de los participantes a distintos conglomerados de CPs, de ser halladas estructuras adecuadas para esta práctica.

Capítulo 6

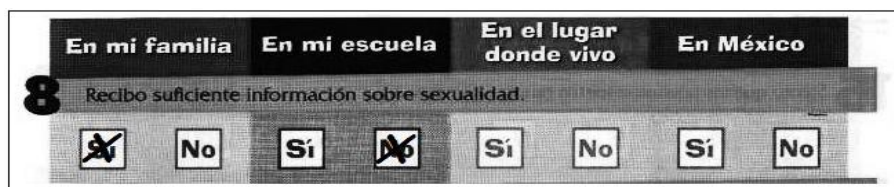
Aplicación

6.1. Descripción de la base de datos

Los datos utilizados para este trabajo corresponden a las respuestas de 150,718 adolescentes entre 14 y 17 años a la Consulta Infantil y Juvenil 2000 convocada por el Instituto Federal Electoral. La base recupera 113 variables mediante el instrumento presentado en el **Anexo I**. Se eligieron estos cuestionarios, de una total superior a los 750,000, debido a que fueron aquellos que cumplieron con el criterio de que se tuvieran claramente registradas las respuestas positivas con un número 1, así como las negativas con un 0. Los demás cuestionarios contaban con valores perdidos que no permiten determinar si la celda vacía se debe a una respuesta negativa o a la ausencia de respuesta.

A través de este instrumento se recopila información explorando percepciones de los adolescentes relacionadas con la formación de valores ciudadanos como la democracia, ciudadanía, cultura política y participación, en el escenario de la familia, la escuela, la comunidad y el país. Los reactivos de la consulta con este fin presentan frases ante las cuales los adolescentes debieron marcar con una “palomita” (✓) recuadros etiquetados con las palabras “*Sí*” o “*No*”, según el acuerdo que ellos tuvieran con cada frase. Por ejemplo, ante el reactivo “*Recibo suficiente información sobre sexualidad*”, en caso de que el adolescente considere que sí la recibe en su familia, pero no en la escuela, deberá marcar el recuadro etiquetado con “*Sí*” bajo la columna “*En la Familia*”, y el marcado con “*No*” bajo la columna “*En mi escuela*”, como se

ilustra en la **Figura 3**.



The image shows a survey question interface. At the top, there are four dark grey boxes with white text: "En mi familia", "En mi escuela", "En el lugar donde vivo", and "En México". Below these is a question number "8" in a circle, followed by the text "Recibo suficiente información sobre sexualidad." Below the question, there are eight response options in a row: a crossed-out "Sí" button, a "No" button, a "Sí" button, a crossed-out "No" button, a "Sí" button, a "No" button, a "Sí" button, and a "No" button. The "Sí" buttons are highlighted with a lighter background.

Figura 3. Ejemplo de respuesta a los reactivos de la consulta

Estas respuestas fueron codificadas para el ACP con un valor de 1 para cada “Sí” marcado, y con un 0 para cada “No”, contando así con un conjunto de 64 variables dicotómicas sobre las cuales se realizaron los análisis. Estas variables se presentarán en los resultados identificando los aspectos conceptuales que la consulta pretendió medir, el ámbito en el que es medido y la frase usada en el reactivo.

Por ejemplo, la pregunta acerca de si los adolescentes consideran que reciben suficiente información sobre sexualidad en la escuela se presentará como “*Información (Escuela): Recibí suficiente información sobre sexualidad*”. Lo anterior tiene la intención de facilitar la interpretación de los resultados obtenidos.

El objetivo del ACP, es encontrar un grupo pequeño de variables que resulten trascendentes para los adolescentes que respondieron la consulta, y con las cuáles se pudieran realizar posteriormente otros análisis e investigaciones que ayuden a caracterizar la población adolescente en cuanto a su formación ciudadana. Para el procesamiento de los datos se utilizaron los paquetes especializados “*Statistical Package for the Social Sciences for Windows, Release 13.0*” (SPSS, 2004) y “*JMP ®, Release 7.0*” (SAS Institute Inc., 2007). En general, aun cuando los análisis fueron realizados de manera paralela en ambos paquetes, se eligió usar las gráficas en tonos de gris y las tablas de las salidas del SPSS por las opciones de este programa para generar las primeras y ordenar las segundas; de la misma forma, las gráficas a color provienen del JMP por la sencillez de visualización y edición de gráficos de este paquete.

6.2. Análisis y resultados

De manera inicial se realizó un ACP de las 64 variables, utilizando la matriz de correlación de las variables. Nótese que al encontrarse todas las variables en la misma escala de medida, la matriz de covarianza también pudo haber sido utilizada como ya se comentara en la sección 2.2.

Como salidas se eligieron las gráficas de sedimentación, las cargas de las variables originales en los CPs, los valores propios, las varianzas explicadas y las varianzas acumuladas de los CPs. Asimismo, con la intención de obtener una mejor interpretabilidad de estos resultados, se eligió la realización de una rotación ortogonal de los componentes (varimax), y una oblicua (direct oblimin), a manera de comparación.

La gráfica de sedimentación resultante de este primer análisis se presenta en la **Fig 4**.

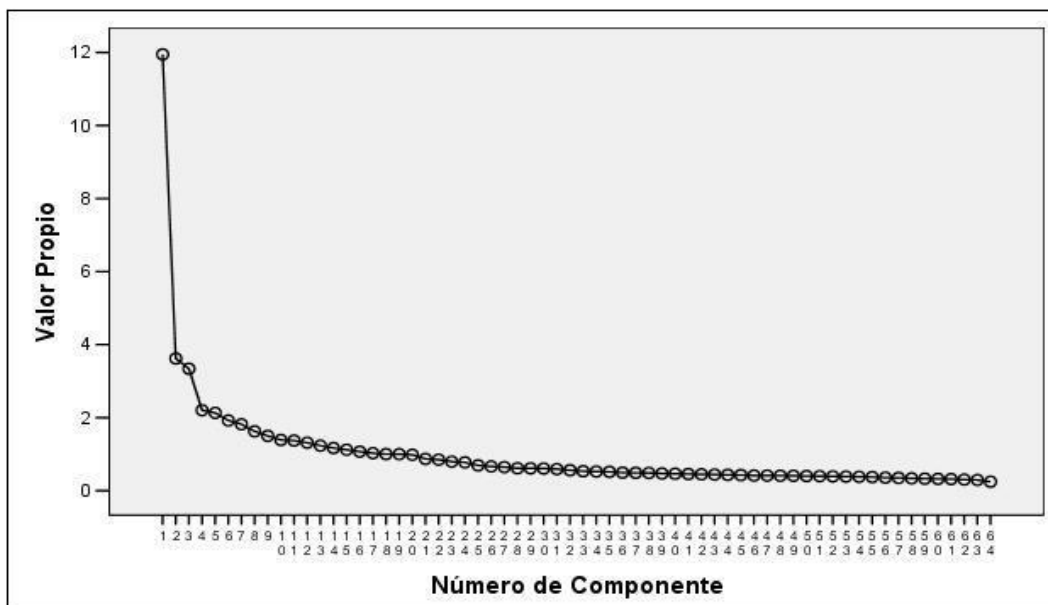


Figura 4. Gráfica de sedimentación del ACP con todos los reactivos de la encuesta

Componente	Suma de Cargas Cuadráticas (Extracción)			Suma de Cargas Cuadráticas (Rotación)		
	Total	% de Varianza	% acumulado	Total	% de Varianza	% acumulado
1	11.948	18.669	18.669	7.985	12.476	12.476
2	3.619	5.654	24.324	5.754	8.991	21.467
3	3.338	5.216	29.540	5.166	8.072	29.540

Tabla 1. Varianza explicada de los componentes principales

En este caso, tomar la aparición del primer componente con valor propio menor o igual a uno como criterio de corte para el número de CPs a considerar puede ser una mala idea, ya que, ante una gran cantidad de variables similares, nos encontramos con una veintena de los primeros componentes extraídos que tienen valor propio superior a la unidad. Esta situación se observa claramente en la gráfica de sedimentación, misma que sugiere tomar los tres primeros CPs antes del codo, y que presentan una diferencia “visible” entre sus valores propios y los del componente siguiente. De esta manera se conservan para el análisis tres CPs de 64 que explican cerca del 30% de la varianza total (**Tabla 1**); además al ser uno de los objetivos explorar la estructura de los datos, no es posible visualizar la estructura de mayor número de variables. La matriz de componentes extraídos, presentada en la **Tabla 2**, muestra los coeficientes de las variables para cada CP. Estos valores han sido ordenados de mayor a menor absoluto, del primer al último CP considerado, valores menores a 0.3 no fueron presentados, como criterio convencional de su importancia y para una mejor lectura e interpretación de la información.

	Componente		
	1	2	3
Equidad y Justicia (comunidad): Hay abuso de autoridad	-.559		
Respeto (comunidad): Me quieren y estoy contento	.558		
Democracia (comunidad): Tengo espacio para discutir los problemas y sus soluciones	.546		
Equidad y Justicia (comunidad): Todos respetan las reglas	.541		
Respeto (México): Me quieren y estoy contento	.539		
Participación (comunidad): Lo que digo se toma en cuenta	.535		
Respeto (escuela): Me quieren y estoy contento	.532		
Equidad y Justicia (comunidad): Se trata igual a las niñas y a los niños	.531		
Equidad y Justicia (escuela): Hay abuso de autoridad	-.530		
Información (comunidad): Recibí suficiente información sobre alcohol y drogas	.519		
Oportunidades (comunidad): Hay oportunidades de desarrollo para jóvenes	.515		
Equidad y Justicia (escuela): Todos respetan las reglas	.512		
Información (México): Recibí suficiente información sobre alcohol y drogas	.510		
Democracia (México): Tengo espacio para discutir los problemas y sus soluciones	.509		
Información (comunidad): Recibí información sobre el Sida	.504		
Paz Vs Violencia (familia): Hay violencia		.555	
Equidad y Justicia (familia) Hay discriminación		.549	
Paz Vs Violencia (comunidad): Hay violencia		-.529	
Paz Vs Violencia (escuela): Hay violencia		.526	
Información (familia): Recibí información sobre el Sida			.502

Tabla 2. Matriz de componentes extraídos

Una primera mirada a los 3 primeros CPs no permite descubrir algún patrón claro de las variables que los conforman, por lo cual se procede a rotar esta solución, primero a través de rotación ortogonal varimax. Los resultados de este procedimiento se observan en la **Tabla 3**. Donde ya se observa cómo el segundo componente se encuentra relacionado con variables que miden la información que los adolescentes perciben que están recibiendo, mientras que el primero y el tercero comparten en este sentido variables de los conjuntos de “Equidad y Justicia” y de “Paz vs Violencia”. En la **Figura 5** se presentan los componentes extraídos representados en el espacio rotado ortogonalmente, sin embargo, aun es difícil encontrar un patrón claro para el primer y el tercero de ellos, o conjuntos de variables que espacialmente tengan un comportamiento definido.

	Componente		
	1	2	3
Equidad y Justicia (México): Todos respetan las reglas	.669		
Equidad y Justicia (México): Hay abuso de autoridad	-.665		
Participación (México): Lo que digo se toma en cuenta	.661		
Equidad y Justicia (comunidad): Hay abuso de autoridad	-.596		
Equidad y Justicia (México): Se trata igual a las niñas y a los niños	.586		
Participación (comunidad): Lo que digo se toma en cuenta	.585		
Participación (México): Todos ayudamos	.585		
Respeto (México): Me quieren y estoy contento	.568		
Equidad y Justicia (comunidad): Todos respetan las reglas	.554		
Democracia (México): Tengo espacio para discutir los problemas y sus soluciones	.513		
Respeto (comunidad): Me quieren y estoy contento	.510		
Participación (comunidad): Todos ayudamos	.501		
Información (familia): Recibí información sobre el Sida		.658	
Información (familia): Recibí suficiente información sobre alcohol y drogas		.633	
Información (comunidad): Recibí información sobre el Sida		.625	
Información (escuela): Recibí información sobre el Sida		.616	
Información (México): Recibí información sobre el Sida		.614	
Información (comunidad): Recibí suficiente información sobre alcohol y drogas		.606	
Información (escuela): Recibí suficiente información sobre alcohol y drogas		.598	
Información (familia): Recibí suficiente información sobre sexualidad		.595	
Información (México): Recibí suficiente información sobre alcohol y drogas		.594	
Información (México): Recibí suficiente información sobre sexualidad		.560	
Información (escuela): Recibí suficiente información sobre sexualidad		.553	
Información (comunidad): Recibí suficiente información sobre sexualidad		.543	
Democracia (familia): Tengo espacio para discutir los problemas y sus soluciones		.543	
Paz Vs Violencia (familia): Hay violencia			.572
Equidad y Justicia (familia) Hay discriminación			.556
Paz Vs Violencia (comunidad): Hay violencia			-.553
Paz Vs Violencia (escuela): Hay violencia			.545
Equidad y Justicia (escuela) Hay discriminación			-.527

Tabla 3. Matriz de componentes tras rotación varimax

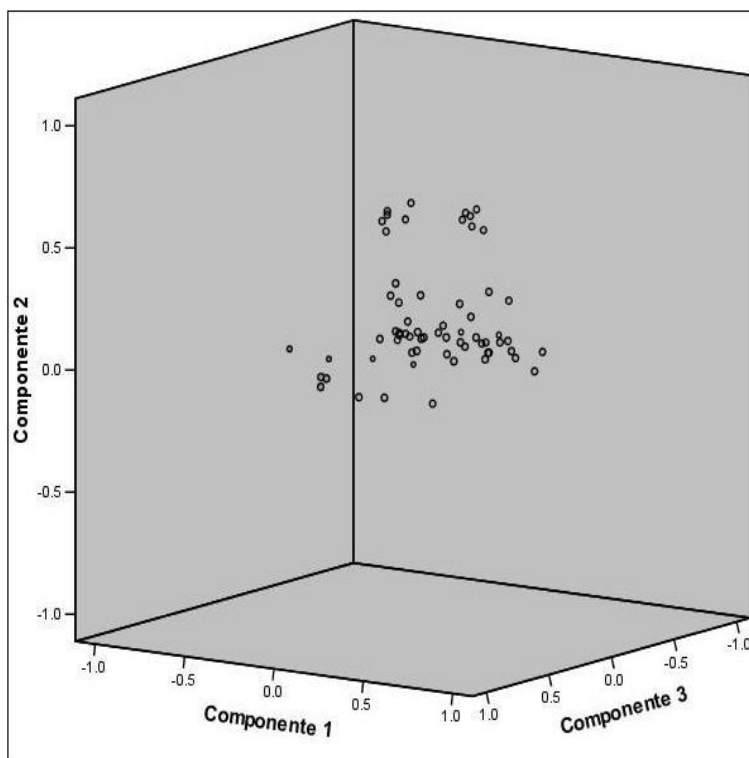


Figura 5. Componentes en el espacio tridimensional tras rotación varimax

Dado que se comparten variables clasificadas originalmente en los mismos grupos del instrumento, se puede sospechar que una rotación que permita cierta correlación entre estos componentes pueda ser más útil para su interpretación. Una rotación oblicua oblimin de los datos permite observar cómo los tres primeros componentes pueden transformarse para estar relacionados con variables “Abuso de Autoridad”, “Violencia”, y la ya observada previamente “Información”. Más allá, en cuanto a las variables de “Violencia”, se observa en la matriz de componentes rotados oblicuamente como al ser medidas en el contexto social micro: Escuela y Familia, operan en sentido contrario que al ser medidas en el contexto social macro: Comunidad y Nación. Lo anterior queda ilustrado en la **Tabla 4**, la matriz patrón, también llamada de configuración, que presenta las correlaciones parciales entre el reactivo y el componente rotado (Jennrich & Sampson, 1966).

De la misma manera, al observar en la **Tabla 5** las correlaciones simples de las variables con estas transformaciones, también es posible observar como

las variables de “Abuso de Autoridad”, operan en sentido contrario a otros aspectos positivos de “Equidad y Justicia”. Esta tabla es llamada matriz estructura, o simplemente de correlaciones, ilustrando la correlación completa entre reactivo y componente, incluyendo aquellas porciones de la correlación aportadas indirectamente a través de las correlaciones con los otros componentes. En la **Figura 6**, estas relaciones se hacen más evidentes, encontrando que las variables de “Violencia” se comportan de manera independiente a las de “Abuso de Autoridad” y las de “Información”. Las variables señaladas como *v6e* y *v6f* cuestionan a los adolescentes con respecto a si encuentran violencia en su escuela y su familia; las variables marcadas como *y3m* y *x3l*, si la encuentran en México y en su comunidad, respectivamente. La codificación completa de las variables incluidas en el análisis se encuentran en el **Anexo II**.

	Componente		
	1	2	3
Equidad y Justicia (comunidad): Hay abuso de autoridad	-.694		
Equidad y Justicia (escuela): Hay abuso de autoridad	-.682		
Equidad y Justicia (familia): Hay abuso de autoridad	-.652		
Equidad y Justicia (México): Hay abuso de autoridad	-.582		
Paz Vs Violencia (comunidad): Hay violencia		-.832	
Paz Vs Violencia (escuela): Hay violencia		.808	
Paz Vs Violencia (familia): Hay violencia		.732	
Paz Vs Violencia (México): Hay violencia		-.728	
Información (escuela): Recibí suficiente información sobre alcohol y drogas			.791
Información (México): Recibí suficiente información sobre alcohol y drogas			.642
Información (escuela): Recibí información sobre el Sida			.597
Información (familia): Recibí suficiente información sobre alcohol y drogas			.510
Información (comunidad): Recibí suficiente información sobre alcohol y drogas			.447
Información (México): Recibí información sobre el Sida			.437
Información (escuela): Recibí suficiente información sobre sexualidad			.381

Tabla 4. Matriz patrón tras rotación oblimin (correlaciones parciales)

Con el fin de explorar la estructura de los datos para las variables de “Equidad y Justicia”, “Paz vs Violencia” e “Información” como CPs, se realizaron ACPs independientes para cada uno de los conjuntos de reactivos relacionados con ellas; asimismo se realizó una búsqueda de patrones o conglomerados por inspección gráfica de sus primeros tres componentes, para finalizar con la realización de regresiones logísticas multinomiales para predecir la pertenencia

cia de los participantes a dichos conglomerados, en caso de que éstos hayan sido encontrados.

EQUIDAD Y JUSTICIA. Realizando un ACP con rotación oblimin para las variables de la categoría “Equidad y Justicia” se encuentran cinco CPs, tomando como criterio que los CPs tengan un valor propio superior o igual a la unidad y explicando el 66.2% de la varianza. Estos CPs quedan conformados por reactivos que preguntan los mismos aspectos específicos a diferentes contextos, por ejemplo, el primer CP estuvo conformado por los reactivos que cuestionaban si “*Todos respetan las reglas*” en la escuela, la familia, la comunidad y el país, validando de esta manera que los reactivos evalúan los aspectos considerados teóricamente por la encuesta. Estos resultados se observan en la **Tabla 6**. Cabe notar en la tabla que el primer y el tercer componentes tienen cargas indirectas del uno para el otro, así los reactivos que hablan de respeto a las reglas también presentan cargas considerables en el componente de trato con equidad de género, y viceversa.

	Componente		
	1	2	3
Equidad y Justicia (comunidad): Hay abuso de autoridad	-.780		
Equidad y Justicia (escuela): Hay abuso de autoridad	-.750		
Equidad y Justicia (familia): Hay abuso de autoridad	-.707		
Equidad y Justicia (México): Hay abuso de autoridad	-.667		
Equidad y Justicia (comunidad): Se trata igual a las niñas y a los niños	.326		
Equidad y Justicia (México): Se trata igual a las niñas y a los niños	.320		
Participación (comunidad): Lo que digo se toma en cuenta	.307		
Equidad y Justicia (comunidad): Todos respetan las reglas	.347		
Equidad y Justicia (escuela): Todos respetan las reglas	.312		
Equidad y Justicia (México): Todos respetan las reglas	.337		
Paz Vs Violencia (comunidad): Hay violencia		-.819	
Paz Vs Violencia (escuela): Hay violencia		.798	
Paz Vs Violencia (familia): Hay violencia		.750	
Paz Vs Violencia (México): Hay violencia		-.710	
Información (escuela): Recibí suficiente información sobre alcohol y drogas			.800
Equidad y Justicia (comunidad): Se trata igual a las niñas y a los niños			.691
Equidad y Justicia (México): Se trata igual a las niñas y a los niños			.677
Participación (comunidad): Lo que digo se toma en cuenta			.611
Equidad y Justicia (comunidad): Todos respetan las reglas			.562
Equidad y Justicia (escuela): Todos respetan las reglas			.549
Equidad y Justicia (México): Todos respetan las reglas			.545
Información (familia): Recibí información sobre el Sida			.406
Información (comunidad): Recibí información sobre el Sida			.400
Información (México): Recibí suficiente información sobre sexualidad			.413

Tabla 5. Matriz estructura tras rotación oblimin (correlaciones simples)

Al graficar los tres primeros CPs para los participantes de la encuesta en una gráfica de puntos, y tras realizar algunas rotaciones de la misma para buscar patrones en esta representación, se nota como se forman cinco conglomerados de puntos, es decir, cinco conjuntos de participantes, entre planos definidos por el espacio tridimensional de los tres primeros CPs. Más allá, tras una segunda rotación de la gráfica se observa como intersectando estos planos puede observarse otro conjunto de planos que separarían a los participantes en 13 conglomerados de puntos, distintos de los cinco primeros (**Figuras 7 y 8**). A partir de esta observación se puede suponer que los tres primeros CPs definen la pertenencia de los participantes a uno de los grupos definidos por cualquiera de los conglomerados mencionados.

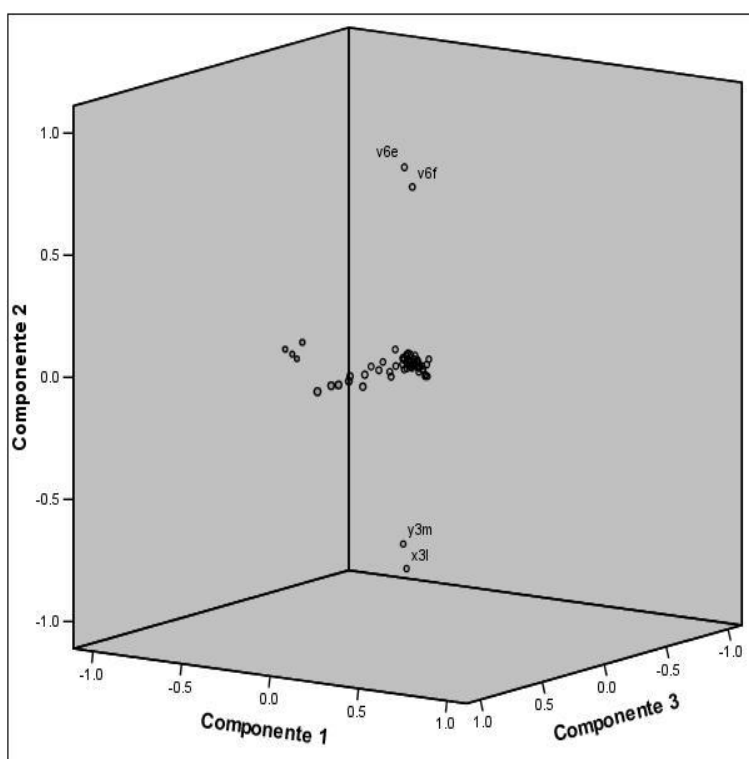


Figura 6. Componentes en el espacio tridimensional tras rotación oblimin

	Componente				
	1	2	3	4	5
Equidad y Justicia (escuela): Todos respetan las reglas	.791		.312		-.351
Equidad y Justicia (comunidad): Todos respetan las reglas	.774		.300		-.375
Equidad y Justicia (familia): Todos respetan las reglas	.745		.344		-.339
Equidad y Justicia (México): Todos respetan las reglas	.665			.566	-.322
Equidad y Justicia (comunidad) Hay discriminación		.815			
Equidad y Justicia (escuela) Hay discriminación		.811			
Equidad y Justicia (familia) Hay discriminación		.726		-.319	
Equidad y Justicia (México) Hay discriminación		.643		.411	
Equidad y Justicia (escuela): Se trata igual a las niñas y a los niños	.329		.814		
Equidad y Justicia (comunidad): Se trata igual a las niñas y a los niños	.318		.813		-.353
Equidad y Justicia (familia): Se trata igual a las niñas y a los niños			.788		-.315
Equidad y Justicia (México): Se trata igual a las niñas y a los niños	.312		.624	.561	-.313
Equidad y Justicia (comunidad): Hay abuso de autoridad	-.358		-.308		.784
Equidad y Justicia (escuela): Hay abuso de autoridad	-.395		-.334		.776
Equidad y Justicia (familia): Hay abuso de autoridad	-.342		-.349		.761
Equidad y Justicia (México): Hay abuso de autoridad	-.326			-.608	.639

Tabla 6. Matriz estructura tras rotación oblimin para los reactivos de Equidad y Justicia

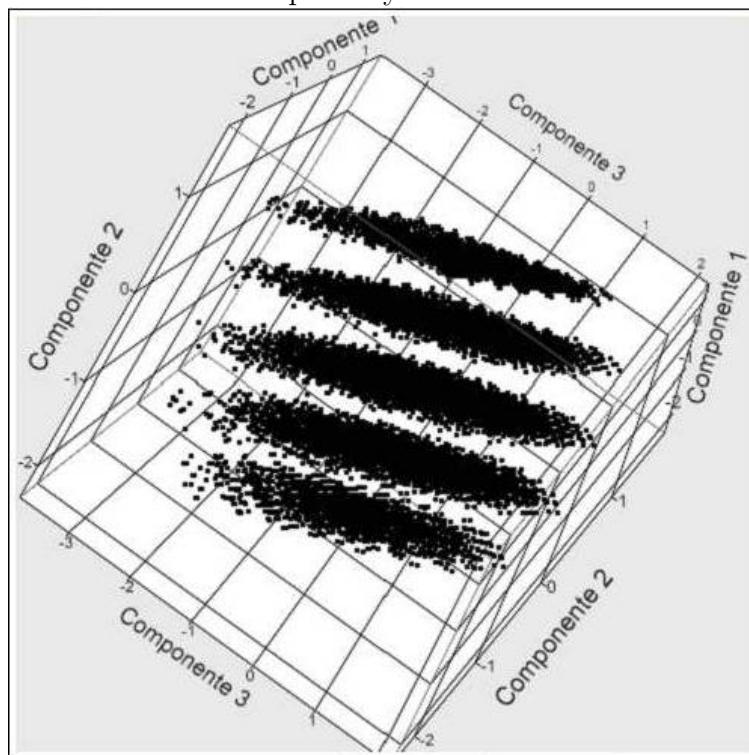


Figura 7. Conjunto de conglomerados al graficar los tres primeros CPs de los participantes para los reactivos de “Equidad y justicia”

Por ejemplo, un conjunto de planos que separa los primeros cinco conglomerados observados sería el siguiente:

$$C_1 + 2,491C_2 - 0,705C_3 - 2,358 = 0$$

$$C_1 + 2,883C_2 - 0,790C_3 - 0,521 = 0$$

$$C_1 + 3,147C_2 - 0,813C_3 + 1,176 = 0$$

$$C_1 + 3,046C_2 - 0,723C_3 + 4,043 = 0$$

Donde C_i es la coordenada sobre el i -ésimo CP. Los planos presentados fueron hallados por ensayo y error mediante la elección de tres puntos que, por inspección visual simple de la gráfica, parecieran encontrarse en un plano entre conglomerados. Nótese que, si asignamos un color a cada punto definido en el espacio de los tres primeros CPs según la sección definida por los planos donde se encuentre, se observa como cada conglomerado queda conformado sólo por puntos del mismo color (**Figura 9**).

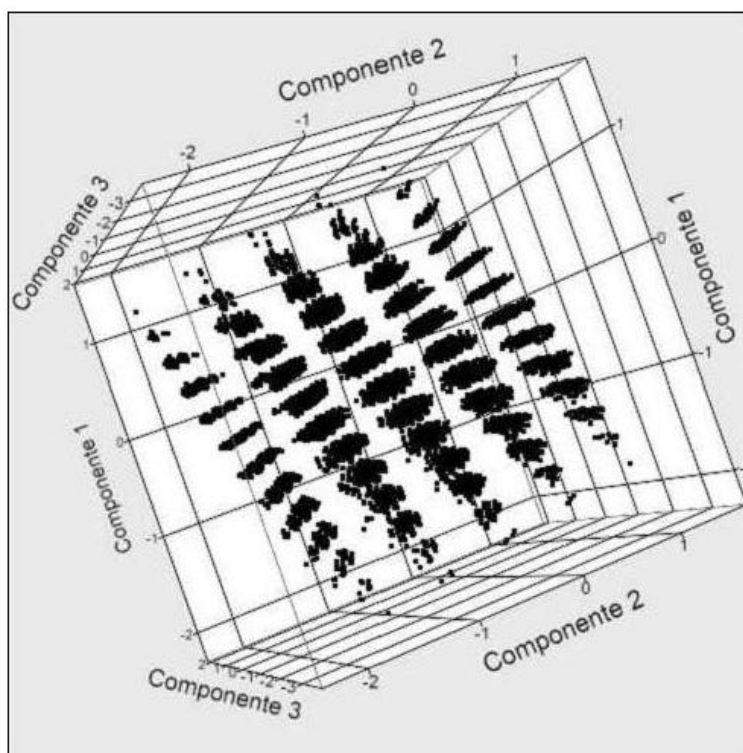


Figura 8. Conjunto de conglomerados al graficar los tres primeros CPs de los participantes para los reactivos de “Equidad y justicia”

Sin embargo no hay que perder de vista que dentro de estos conglomerados aun existen subgrupos que no se alcanzan a diferenciar de esta manera, y que tal vez podrían ser explorados posteriormente.

Si en vez de utilizar la sección definida por los planos para colorear la gráfica utilizamos las variables originales, se observa como con las variables acerca de si hay discriminación en la familia, la escuela, la comunidad y en México, las respuestas van cambiando de “Sí” a “No” a través de los conglomerados, quedando los extremos perfectamente definidos por sujetos que respondieron exclusivamente una respuesta o la otra. Cuando sucede que uno o más conglomerados quedan perfectamente separados por una sola variable como en este caso, se dice que existe separación completa (Albert & Anderson, 1984). Este hecho se ilustra en la **Figura 10**.

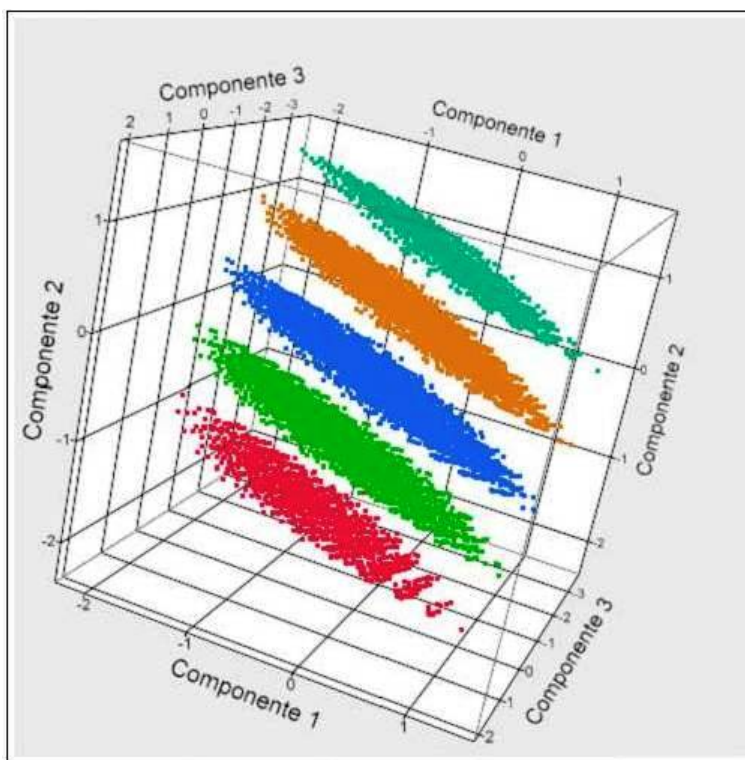


Figura 9. Conjunto de conglomerados al graficar los tres primeros CPs de los participantes para los reactivos de “Equidad y justicia”, coloreados a partir de las secciones definidas por los planos

Continuando con el análisis, las diferentes regiones delimitadas por los planos presentados, y por lo tanto los conglomerados, fueron numeradas del cero al cuatro. De esta manera, sin tomar en cuenta las variables sobre discriminación; e incluyendo las variables sociodemográficas sexo, edad y tipo de comunidad como predictores; se ajustó el siguiente modelo de regresión logística multinomial (RLM):

$$\pi_{ji} = \frac{e^{g_j(\mathbf{x}(i))}}{1 + e^{g_1(\mathbf{x}(i))} + \dots + e^{g_{J-1}(\mathbf{x}(i))}};$$

$$g_j(\mathbf{x}(i)) = \ln \frac{\pi_{ji}}{\pi_{0i}} = \beta_{j0} + \sum_{k=1}^K \beta_{jk} x_{ik}$$

Donde:

π_{ji} es la probabilidad de que el sujeto i pertenezca al conglomerado j , $j = 0, \dots, J - 1$.

$g_j(\mathbf{x}(i))$ es una función de enlace que hace al modelo lineal (el logaritmo natural de la razón de probabilidades de que el i -ésimo sujeto esté en el conglomerado j con respecto a que se encuentre en el conglomerado 0).

$\mathbf{x}(i)$ es el valor de los predictores para el i -ésimo sujeto.

x_{ik} es el valor del k -ésimo predictor para el i -ésimo sujeto.

β_{jk} es el coeficiente del k -ésimo predictor para el j -ésimo conglomerado.

En el caso de la consulta, los predictores fueron codificados de la siguiente manera: un valor de cero correspondió a las comunidades rurales y un uno a las urbanas; para el sexo, las mujeres fueron codificadas con un cero y los hombres con un uno; finalmente, las edad de los sujetos pudo ser 14, 15, 16 ó 17 años. Los reactivos de ciudadanía se mantuvieron codificadas como ya fue comentado.

Las variables de discriminación fueron excluidas debido a la separación completa, al quedar definida la pertenencia a un conglomerado por una de las variables ya no es necesaria su inclusión pues se está seguro de poder predecir la pertenencia por el valor que tomen estas variables: los primeros conglomerados (según el orden de la numeración) contendrán a quienes no perciben discriminación en ningún contexto; los últimos, a quienes sí la perciben. Además, al ocurrir esta separación los algoritmos para realizar las estimaciones propias de este tipo de modelos de regresión fallan, ya que el logaritmo de la función

de verosimilitud converge a cero pero el estimador de máxima verosimilitud no existe, tendiendo a infinito.

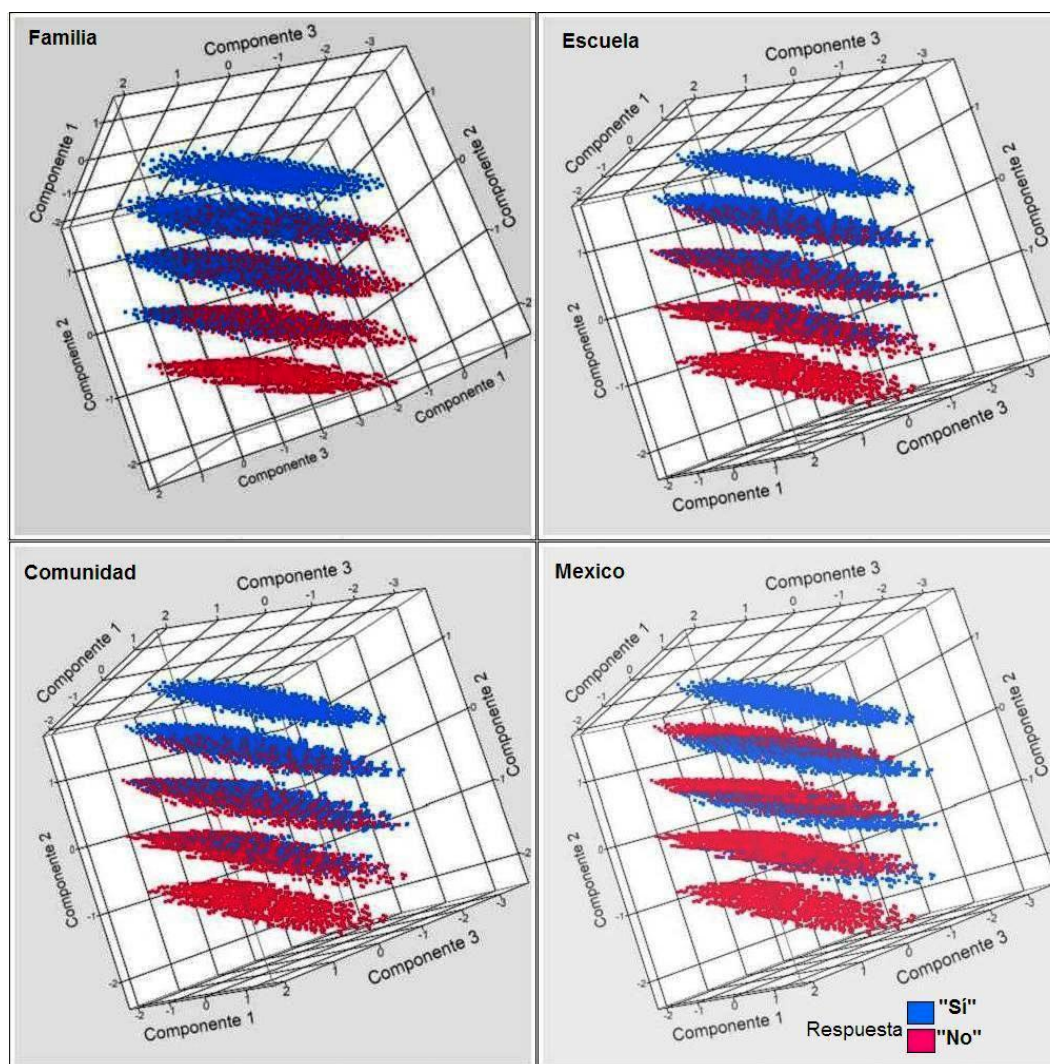


Figura 10. Conjunto de conglomerados al graficar los tres primeros CPs de los participantes para los reactivos de “Equidad y justicia”, coloreados a partir de las respuestas sobre si hay discriminación en la familia, la escuela, la comunidad y México

Se eligió el algoritmo *Custom Forward Stepwise* del SPSS para la estimación de los coeficientes de las K variables predictoras con mayor efecto en el

modelo; las variables edad, sexo y tipo de comunidad fueron introducidas de manera obligada al modelo. Se encontró que, con excepción del tipo de comunidad, todas las variables aportan un efecto significativo para la predicción de la pertenencia de los sujetos a un conglomerado particular. Los valores de los coeficientes b_{kj} , así como de su significancia y del ajuste de la prueba se presentan en el **Anexo III**.

Con respecto a la edad se encontró que aquellos participantes de menor edad tendían a encontrarse con mayor probabilidad en los primeros conglomerados, mientras que en el cuarto se encontrarían con mayor probabilidad los participantes de 17 años y los cercanos a esa edad. Las mujeres también tienen la tendencia a encontrarse en los conglomerados 3 y 4, mientras que los hombres tienen mayor probabilidad de encontrarse en los primeros. Los adolescentes que consideran que no se respetan las reglas en la familia ni en la escuela, y que creen que en la escuela no se trata igual a niños y a niñas tienen mayor probabilidad de pertenecer al conglomerado 0, lo mismo sucede con aquellos sujetos que creen que no hay abuso de autoridad en México. Esta tendencia a pertenecer con mayor probabilidad a los primeros conglomerados se mantiene para quienes consideran que no se trata igual a los niños y las niñas en la comunidad. Quienes creen que no se respetan las reglas y que no se trata igual a niños y niñas en México, a su vez, tienen mayor probabilidad de estar en los últimos conglomerados. Estos resultados son congruentes con la definición de los primeros conglomerados por aquellos sujetos que no perciben discriminación y los últimos por aquellos que sí la perciben.

Así es posible notar que existe una tendencia en el mismo sentido de los varones y los sujetos más jóvenes de la muestra y la percepción de injusticia y falta de equidad en los ambientes más cercanos, la familia y la escuela; mientras que en sentido opuesto se encuentran los mayores, las mujeres, y la tendencia a percibir estos problemas a niveles más lejanos, notando el problema a nivel nacional, sin importar la pertenencia a comunidades rurales o urbanas.

INFORMACIÓN. Al realizar el ACP de las variables de “Información” se pueden encontrar resultados similares. En su caso, se encuentran tres CPs tras el análisis que explican el 61.8% de la varianza total. Más, al analizar la gráfica de las variables originales en el espacio de estos componentes, es posible observar como se forman dos conjuntos claramente definidos: el que

se refiere a la información en el contexto más íntimo de la familia y la escuela, y el que contiene las variables en el contexto comunitario y nacional. Lo cual resulta congruente con lo hallado al analizar las variables de “Equidad y Justicia”. Estos resultados se ilustran en la **Tabla 7** y la **Figura 11**, respectivamente.

Como en el caso del análisis previo, al graficar los tres primeros CPs de los participantes de la encuesta, y tras realizar una rotación, se observa como se forman siete conglomerados que pueden ser separados por seis planos, por ejemplo, los planos que se presentan a continuación:

$$\begin{aligned}
 C_1 + 0,518C_2 + 0,144C_3 - 0,527 &= 0 \\
 C_1 + 0,457C_2 + 0,126C_3 + 0,141 &= 0 \\
 C_1 + 0,479C_2 + 0,283C_3 + 0,672 &= 0 \\
 C_1 + 0,503C_2 + 0,304C_3 + 1,356 &= 0 \\
 C_1 + 0,482C_2 + 0,267C_3 + 2,008 &= 0 \\
 C_1 + 0,486C_2 + 0,247C_3 + 2,708 &= 0
 \end{aligned}$$

Una vez más, si asignamos un color a cada punto definido en el espacio de los tres primeros CPs según la sección definida por los planos donde se encuentre, se observa como cada conglomerado queda conformado sólo por puntos del mismo color (**Figura 12**).

	Componente		
	1	2	3
Información (México): Recibí información sobre el Sida	.793	.391	
Información (México): Recibí suficiente información sobre alcohol y drogas	.773	.472	
Información (México): Recibí suficiente información sobre sexualidad	.737		
Información (comunidad): Recibí información sobre el Sida	.724		.561
Información (comunidad): Recibí suficiente información sobre alcohol y drogas	.704	.331	.376
Información (comunidad): Recibí suficiente información sobre sexualidad	.646		.624
Información (escuela): Recibí suficiente información sobre alcohol y drogas	.362	.842	
Información (escuela): Recibí información sobre el Sida	.374	.771	
Información (escuela): Recibí suficiente información sobre sexualidad	.316	.690	
Información (familia): Recibí suficiente información sobre alcohol y drogas	.352	.616	.506
Información (familia): Recibí suficiente información sobre sexualidad	.321	.347	.817
Información (familia): Recibí información sobre el Sida	.406	.437	.763

Tabla 7. Matriz estructura tras rotación oblimin para los reactivos de “Información”

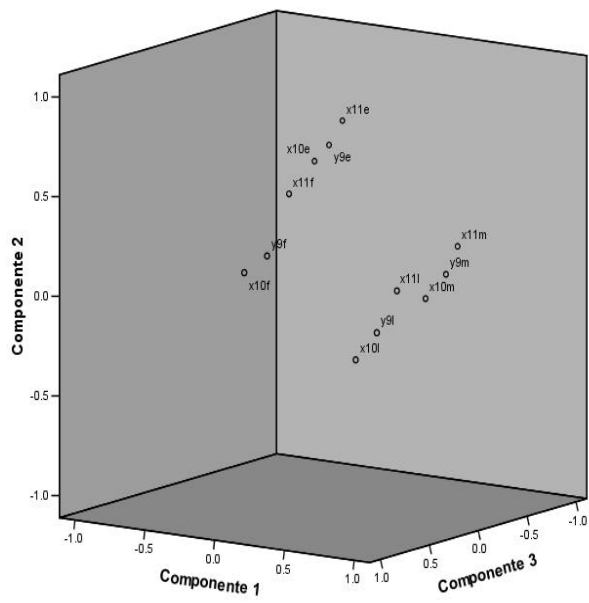


Figura 11. Componentes en el espacio tridimensional tras rotación obliqua para los reactivos de “Información”

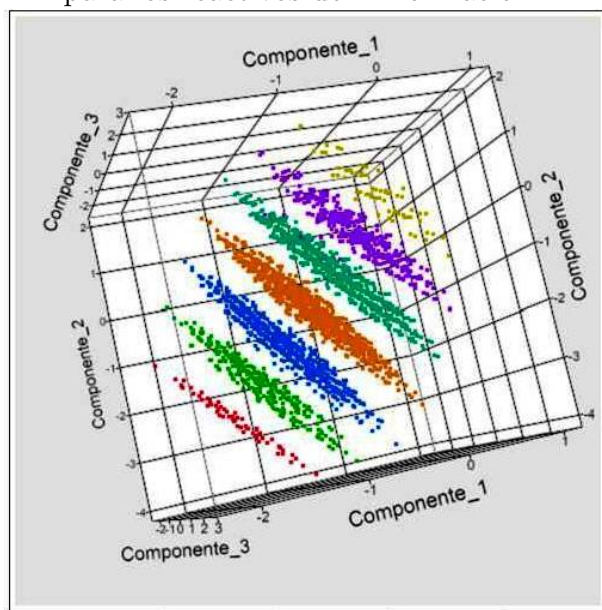


Figura 12. Conjunto de conglomerados al graficar los tres primeros CPs de los participantes para los reactivos de “Información”, coloreados a partir de las secciones definidas por los planos

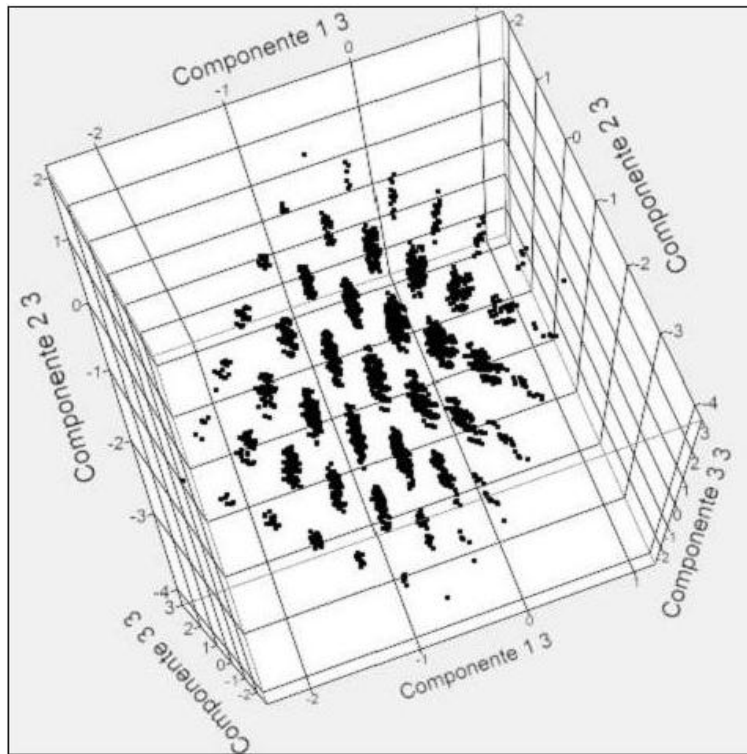


Figura 13. Conjunto de conglomerados al graficar los tres primeros CPs de los participantes para los reactivos de “Información”

Asimismo, también sería posible encontrar otro conjunto de planos que intersectan a los presentados y que definirían conglomerados en otra dirección de la gráfica, cada grupo conformando subgrupos dentro de los conglomerados observados en la otra dirección. Esto se observa en la **Figura 13**. Además, en este caso, al intentar colorear con base en los valores de las variables originales se encuentra que, para toda variable original, siempre existen conglomerados a los extremos conformados por un solo valor de la variable, sin importar la dirección elegida. Como ejemplo de esta situación está la **Figura 14**, coloreada a partir de la variable x_{10f} : “*Recibí suficiente información sobre sexualidad (familia)*”. De rojo cuando no se recibió suficiente información, de azul cuando sí se recibió.

Con base en estas observaciones y de manera similar al análisis de las variables de “Equidad y Justicia”, se excluyeron como predictoras de la pertenencia a un conglomerado a las variables de “Información”. Sin embargo, en

esta ocasión se encontró que sólo las variables sociodemográficas sexo, edad y tipo de comunidad contribuyen significativamente al modelo, confirmando la independencia del constructo “Información recibida” de los otros que se exploraron a través de la consulta. Las medidas de ajuste y parámetros estimados para esta regresión se presentan en el **Anexo IV**.

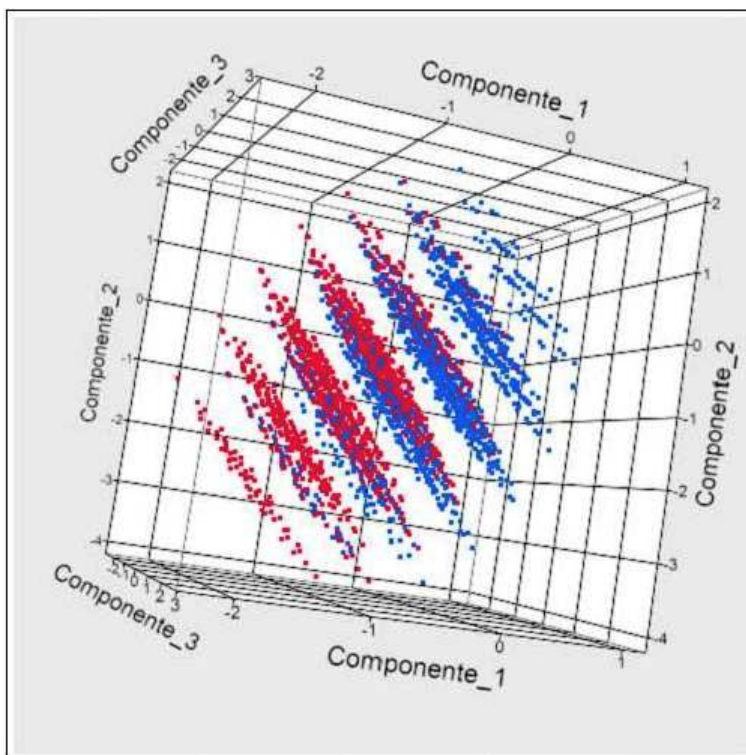


Figura 14. Conjunto de conglomerados al graficar los tres primeros CPs de los participantes para los reactivos de “Información”, coloreados a partir de los valores de la variable x_{10f}

En estos resultados es posible observar una tendencia de que los participantes de poblaciones rurales tienen menor probabilidad de pertenecer a los últimos conglomerados. Asimismo es posible ver como el hecho de ser mujer disminuye la probabilidad de pertenencia a los primeros conglomerados, mientras que los sujetos de mayor edad tenderán a encontrarse en los últimos. Al parecer las variables sociodemográficas sexo y edad siguen definiendo la pertenencia a los conglomerados de la misma manera en que lo hicieron para las variables de “Equidad y Justicia”. La tendencia a percibir que sí se recibe información se da en el mismo sentido que el hecho de ser mujer, pertenecer

a una comunidad urbana y ser de mayor edad.

PAZ VS. VIOLENCIA. Finalmente, al realizar el ACP a las variables de “Paz vs Violencia”, y a diferencia de lo ocurrido en los dos análisis previos, es posible observar como los tres primeros CPs definen 16 microcúmulos en su espacio, observados como puntos en la gráfica. Para cada uno de estos CPs existen ocho cúmulos cerca del cero, cuatro en valores positivos cercanos al tres, y cuatro en los valores cercanos al tres negativo. Este patrón se ilustra en la **Figura 15**. La varianza explicada por los tres primeros componentes es el 88.9% del total.

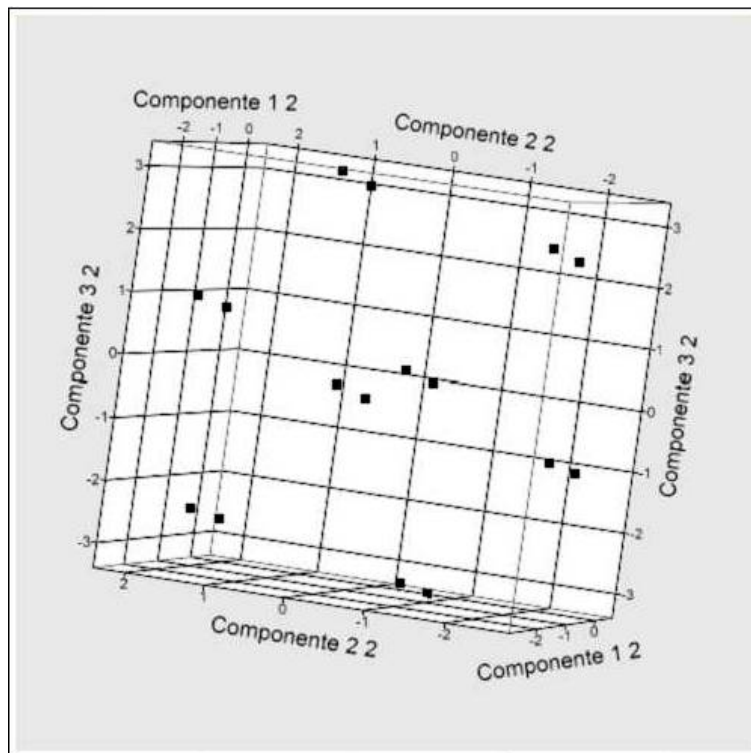


Figura 15. Conjuntos de puntos encontrados tras el análisis para los reactivos de “Paz vs Violencia” en el espacio de los tres primeros CPs

Este resultado no es tan extraño si consideramos que el análisis fue realizado con una categoría de la consulta conformada por sólo cuatro reactivos, con dos opciones de respuesta cada uno, $2^4 = 16$ combinaciones posibles de respuesta; y a que sólo el primer CP tiene un valor propio mayor a la unidad (2.412), mientras que los siguientes se encuentran muy por debajo de este

valor (**Tabla 8**).

Componente	Valores propios iniciales		
	Total	% of Varianza	% Acumulado
1	2.412	60.293	60.293
2	.662	16.543	76.837
3	.483	12.084	88.920
4	.443	11.080	100.000

Tabla 8. Valores propios y varianzas que explican los CPs de los reactivos de “Paz vs Violencia”

6.3. Conclusiones

Como puede observarse, el ACP confirma que los reactivos incluidos en el instrumento analizado pueden ser reducidos a variables congruentes con los contenidos que se pretendan medir: equidad y justicia, información recibida, y paz vs violencia. Además, es posible comprobar que, en casos como los de las variables de “Equidad y Justicia”, existe un conjunto reducido de los reactivos de la consulta que pudieran predecir la pertenencia de los participantes a categorías definidas (que habría que estudiar de manera teórica desde la perspectiva de las ciencias sociales), en esta ocasión, las experiencias y percepciones de discriminación, a través de las respuestas a los reactivos específicos de este tema. En otros casos, como al analizar la porción del instrumento acerca de la información recibida, no es posible encontrar a primera instancia un tema, o un conjunto de reactivos, que pueda usarse para este tipo de predicciones. Asimismo, es posible ver que variables sociodemográficas como el sexo, la edad, o la pertenencia a una comunidad rural o indígena, también influyen en la respuesta que se da a los reactivos de la consulta, y por tanto a la pertenencia a las categorías teóricas sugeridas.

Al señalar de esta manera cuáles son los grupos de la sociedad que se separan de los otros en cuanto a su experiencia con los temas de la consulta, se sugiere la posibilidad, a nivel descriptivo, de crear índices de desarrollo ciudadano con sólo recolectar un conjunto menor de información al recolectado en esta ocasión.

A nivel evaluativo y de toma de decisiones, el análisis permite seleccionar aquellos grupos que pudieran estar en desventaja con respecto a otros, para

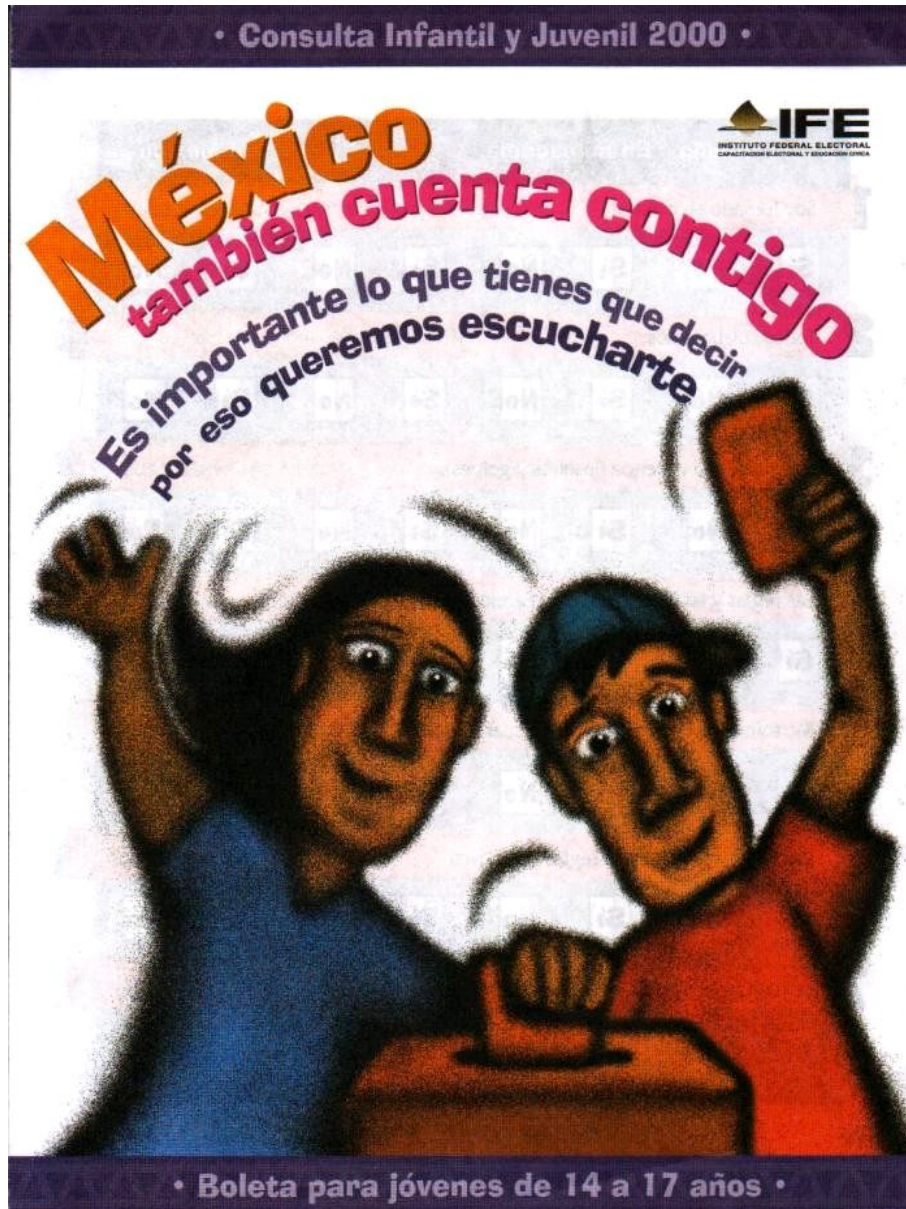
así llevar a cabo programas de intervención que busquen el ideal de equidad de oportunidades entre los individuos deseable en toda sociedad.

Por otro lado, a nivel metodológico, también ha quedado ilustrado cómo, tras profundizar un poco en los análisis, el instrumento no solamente es respondido sobre los temas esperados, sino que también los contextos de experiencia de estos temas -escuela, familia, comunidad, nación- se responden con cierta consistencia interna, quedando agrupados aun tras transformaciones como las realizadas en un ACP. Esto habla de la validez y las limitaciones que como instrumento de medida tiene el analizado, además de sugerir posibles mejoras del mismo al señalar cuáles de sus reactivos son los que están funcionando de mejor manera para describir las diferencias en la población. Vale la pena aclarar que, aunque se espera encontrar estos patrones al hacer análisis de validez de los instrumentos, en la práctica no siempre sucede así.

Por todo lo anterior, es evidente como el ACP, y el análisis posterior de los resultados que éste arroje, son de gran utilidad para la validación de instrumentos de medición de fenómenos sociales, el avance de las teorías al sugerir nuevas posibles relaciones entre variables, la comprobación de hipótesis teóricas sobre los grupos de interés de las investigaciones sociales, y la toma de decisiones sobre las acciones a realizar con estos grupos.

ANEXO I

Instrumento utilizado para la Consulta Infantil y Juvenil 2000 del Instituto Federal Electoral



Instrucciones: Marca con una ✓ en el **Sí** o en el **No** según lo que tú vivas

	En mi familia	En mi escuela	En el lugar donde vivo	En México
1	Soy tomado en cuenta.			
	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input checked="" type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No
2	Todos colaboramos.			
	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input checked="" type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No
3	Me tratan con violencia (insultos y golpes).			
	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input checked="" type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No
4	Las reglas y las leyes se aplican parejo.			
	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input checked="" type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No
5	Me toman en cuenta para elaborar las reglas.			
	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input checked="" type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No
6	Los adultos respetan las reglas y las leyes.			
	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input checked="" type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No
7	Se trata igual a las mujeres y a los hombres.			
	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input checked="" type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No

Instrucciones: Marca con una ✓ en el **Sí** o en el **No** según lo que tú vivas

	En mi familia	En mi escuela	En el lugar donde vivo	En México
8	Recibo suficiente información sobre sexualidad.			
	Sí No	Sí No	Sí No	Sí No
9	Recibo suficiente información sobre SIDA.			
	Sí No	Sí No	Sí No	Sí No
10	Recibo suficiente información sobre alcohol y drogas.			
	Sí No	Sí No	Sí No	Sí No
11	Tengo espacios y oportunidades para discutir problemas y soluciones.			
	Sí No	Sí No	Sí No	Sí No
12	Cuento con lugares para hacer deportes y divertirme.			
		Sí No	Sí No	
13	Hay discriminación.			
	Sí No	Sí No	Sí No	Sí No
14	Hay oportunidades de desarrollo para los jóvenes.			
	Sí No	Sí No	Sí No	Sí No

Instrucciones: Marca con una en el **Sí** o en el **No** según lo que tú vivas

	En mi familia	En mi escuela	En el lugar donde vivo	En México
15	Siento que los adultos respetan nuestra forma de ser.			
	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input checked="" type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No
16	Para mí las votaciones son importantes.			
		<input checked="" type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No
17	Los partidos políticos toman en cuenta los intereses de los jóvenes.			
				<input type="checkbox"/> Sí <input type="checkbox"/> No
18	Los partidos políticos toman en cuenta los intereses de toda la población.			
				<input type="checkbox"/> Sí <input type="checkbox"/> No
19	Creo que quienes organizan las elecciones de los gobernantes...			
	<input type="text"/>			
20	Yo quiero que en México nunca más...			
	<input type="text"/>			
21	Quiero que en las elecciones de los adultos siempre haya una consulta como ésta.			
				<input type="checkbox"/> Sí <input type="checkbox"/> No
Datos Generales:				
	Voy a la escuela... <input type="checkbox"/> Sí <input type="checkbox"/> No		Soy... <input type="checkbox"/> Mujer <input type="checkbox"/> Hombre	
	Estoy en <input type="text"/> grado de <input type="text"/>	Tengo <input type="text"/> años.		

ANEXO II

Claves de identificación de los reactivos

Clave	Reactivo
v1f	Respeto (familia): Me quieren y estoy contento
v2f	Participación (familia): Todos ayudamos
v3f	Participación (familia): Lo que digo se toma en cuenta
v4f	Equidad y Justicia (familia): Todos respetan las reglas
v5f	Equidad y Justicia (familia): Se trata igual a las niñas y a los niños
v6f	Paz Vs Violencia (familia): Hay violencia
v7f	Paz Vs Violencia (familia): Cuando los grandes no están de acuerdo se insultan y se golpean
v8f	Respeto (familia): Me escuchan y me dejan hablar
v9f	Equidad y Justicia (familia): Hay abuso de autoridad
γ9f	Información (familia): Recibí información sobre el Sida
v10f	Juego y recreación (familia): Tengo lugares donde jugar
x10f	Información (familia): Recibí suficiente información sobre sexualidad
x11f	Información (familia): Recibí suficiente información sobre alcohol y drogas
γ11f	Democracia (familia): Tengo espacio para discutir los problemas y sus soluciones
γ13f	Equidad y Justicia (familia) Hay discriminación
γ14f	Oportunidades (familia): Hay oportunidades de desarrollo para jóvenes
v1e	Respeto (escuela): Me quieren y estoy contento
v2e	Participación (escuela): Todos ayudamos
v3e	Participación (escuela): Lo que digo se toma en cuenta
v4e	Equidad y Justicia (escuela): Todos respetan las reglas
v5e	Equidad y Justicia (escuela): Se trata igual a las niñas y a los niños
v6e	Paz Vs Violencia (escuela): Hay violencia
v7e	Paz Vs Violencia (escuela): Cuando los grandes no están de acuerdo se insultan y se golpean
v8e	Respeto (escuela): Me escuchan y me dejan hablar
v9e	Equidad y Justicia (escuela): Hay abuso de autoridad
γ9e	Información (escuela): Recibí información sobre el Sida
v10e	Juego y recreación (escuela): Tengo lugares donde jugar
x10e	Información (escuela): Recibí suficiente información sobre sexualidad
x11e	Información (escuela): Recibí suficiente información sobre alcohol y drogas
γ11e	Democracia (escuela): Tengo espacio para discutir los problemas y sus soluciones
γ13e	Equidad y Justicia (escuela) Hay discriminación
γ14e	Oportunidades (escuela): Hay oportunidades de desarrollo para jóvenes
γ16e	Democracia (escuela): Para mí las votaciones son importantes
v12	Democracia: Quiero que en las elecciones haya encuestas

Clave	Reactivo (Continua)
x11	Respeto (comunidad): Me quieren y estoy contento
x21	Participación (comunidad): Todos ayudamos
x31	Paz Vs Violencia (comunidad): Hay violencia
x41	Respeto (comunidad): Me escuchan y me dejan hablar
x71	Equidad y Justicia (comunidad): Se trata igual a las niñas y a los niños
x81	Participación (comunidad): Lo que digo se toma en cuenta
x91	Equidad y Justicia (comunidad): Todos respetan las reglas
γ91	Información (comunidad): Recibí información sobre el Sida
γ91	Equidad y Justicia (comunidad): Hay abuso de autoridad
γ101	Juego y recreación (comunidad): Tengo lugares donde jugar
x101	Información (comunidad): Recibí suficiente información sobre sexualidad
x111	Información (comunidad): Recibí suficiente información sobre alcohol y drogas
γ111	Democracia (comunidad): Tengo espacio para discutir los problemas y sus soluciones
γ131	Equidad y Justicia (comunidad) Hay discriminación
γ141	Oportunidades (comunidad): Hay oportunidades de desarrollo para jóvenes
γ161	Democracia (comunidad): Para mí las votaciones son importantes
γ1m	Respeto (México): Me escuchan y me dejan hablar
γ2m	Participación (México): Todos ayudamos
γ3m	Paz Vs Violencia (México): Hay violencia
γ5m	Participación (México): Lo que digo se toma en cuenta
γ6m	Equidad y Justicia (México): Hay abuso de autoridad
γ7m	Equidad y Justicia (México): Se trata igual a las niñas y a los niños
x9m	Equidad y Justicia (México): Todos respetan las reglas
γ9m	Información (México): Recibí información sobre el Sida
x10m	Información (México): Recibí suficiente información sobre sexualidad
x11m	Información (México): Recibí suficiente información sobre alcohol y drogas
γ11m	Democracia (México): Tengo espacio para discutir los problemas y sus soluciones
γ13m	Equidad y Justicia (México) Hay discriminación
γ14m	Oportunidades (México): Hay oportunidades de desarrollo para jóvenes
γ15m	Respeto (México): Me quieren y estoy contento
γ16m	Democracia (México): Para mí las votaciones son importantes
γ17	Democracia: Los partidos políticos toman en cuenta los intereses de los jóvenes
γ18	Democracia: Los partidos políticos toman en cuenta los intereses de toda la población

ANEXO III

Salidas de SPSS 13.0 para la regresión multinomial logística de los conglomerados de “Equidad y Justicia”

Step Summary

Model	Action	Effect(s)	Model Fitting Criteria	Effect Selection Tests		
			-2 Log Likelihood	Chi-Square ^{a,b}	df	Sig.
Step 0	0	Entered	135119.731	.		
Step 1	1	Entered	132401.039	2718.692	4	.000
Step 2	2	Entered	130513.003	1888.036	4	.000
Step 3	3	Entered	127388.039	3124.964	4	.000
Step 4	4	Entered	125491.330	1896.709	4	.000
Step 5	5	Entered	122475.693	3015.637	4	.000
Step 6	6	Entered	115104.810	7370.883	4	.000
Step 7	7	Entered	112099.014	3005.796	4	.000
Step 8	8	Entered	109385.533	2713.482	4	.000
Step 9	9	Entered	108425.833	959.699	4	.000
Step 10	10	Entered	107576.472	849.361	4	.000
Step 11	11	Entered	107192.380	384.092	4	.000

Stepwise Method: Forward Stepwise

a. The chi-square for entry is based on the likelihood ratio test.

b. The chi-square for removal is based on the likelihood ratio test.

Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	135977.890			
Final	107192.380	28785.509	64	.000

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	107192.380 ^a	.000	0	.
edad	107376.217	183.837	12	.000
sexo	107311.892	119.512	4	.000
ciudad	107196.901	4.520	4	.340
v4f	107626.182	433.802	4	.000
v5f	107946.755	754.375	4	.000
v4e	107530.252	337.872	4	.000
v9e	107722.358	529.978	4	.000
x7l	108107.641	915.261	4	.000
y6m	109041.393	1849.013	4	.000
x9m	108957.889	1765.509	4	.000
y7m	109918.937	2726.557	4	.000
v9f	108170.815	978.435	4	.000
x9l	108075.513	883.133	4	.000
v5e	107576.472	384.092	4	.000

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Pseudo R-Square

Cox and Snell	.187
Nagelkerke	.196
McFadden	.066

Parameter Estimates

	Cúmulo ^a	B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
Intercept	0	-.012	.036	.117	1	.732			
	1	-1.001	.040	611.759	1	.000			
	2	-.793	.038	445.249	1	.000			
	3	-.489	.035	189.618	1	.000			
[edad=14]	0	-.274	.024	126.893	1	.000	.760	.725	.798
	1	-.253	.026	93.000	1	.000	.776	.737	.817
	2	-.114	.024	22.023	1	.000	.892	.850	.936
	3	-.080	.023	12.401	1	.000	.923	.882	.965
[edad=15]	0	-.179	.026	45.752	1	.000	.836	.794	.881
	1	-.195	.029	46.841	1	.000	.823	.778	.870
	2	-.077	.026	8.528	1	.003	.926	.879	.975
	3	-.068	.025	7.421	1	.006	.934	.890	.981
[edad=16]	0	-.132	.028	21.902	1	.000	.876	.829	.926
	1	-.116	.030	14.983	1	.000	.890	.839	.944
	2	-.065	.028	5.327	1	.021	.937	.887	.990
	3	-.030	.026	1.299	1	.254	.970	.921	1.022
[sexo=0]	0	-.146	.017	72.842	1	.000	.864	.836	.894
	1	-.106	.019	31.438	1	.000	.899	.867	.933
	2	-.043	.017	6.308	1	.012	.958	.926	.991
	3	.021	.016	1.706	1	.192	1.021	.990	1.054
[ciudad=0]	0	.029	.025	1.413	1	.234	1.030	.981	1.081
	1	-.019	.028	.466	1	.495	.981	.929	1.036
	2	.029	.025	1.309	1	.253	1.029	.980	1.082
	3	-.001	.024	.001	1	.973	.999	.954	1.047
[v4f=0]	0	.371	.024	235.920	1	.000	1.450	1.383	1.520
	1	-.075	.025	8.726	1	.003	.928	.883	.975
	2	-.089	.024	13.838	1	.000	.915	.873	.959
	3	-.044	.023	3.666	1	.056	.957	.915	1.001
[v5f=0]	0	.274	.026	112.399	1	.000	1.315	1.250	1.384
	1	-.374	.028	183.602	1	.000	.688	.651	.726
	2	-.337	.026	164.649	1	.000	.714	.678	.752
	3	-.252	.026	96.131	1	.000	.777	.739	.817
[v4e=0]	0	.218	.023	86.178	1	.000	1.243	1.188	1.302
	1	.392	.024	267.577	1	.000	1.481	1.413	1.552
	2	.093	.022	17.675	1	.000	1.098	1.051	1.147
	3	.044	.021	4.406	1	.036	1.045	1.003	1.090
[v9e=0]	0	-.218	.024	84.925	1	.000	.804	.768	.842
	1	-.493	.024	421.523	1	.000	.611	.583	.640
	2	-.318	.022	203.204	1	.000	.728	.697	.760
	3	-.088	.021	17.133	1	.000	.915	.878	.955
[x7l=0]	0	.296	.026	129.998	1	.000	1.344	1.278	1.414
	1	.438	.026	282.876	1	.000	1.550	1.473	1.631
	2	.321	.024	178.918	1	.000	1.379	1.316	1.446
	3	-.193	.023	68.342	1	.000	.825	.788	.863
[y6m=0]	0	.399	.022	327.725	1	.000	1.491	1.427	1.556
	1	-.320	.025	161.609	1	.000	.726	.691	.763
	2	-.429	.022	366.969	1	.000	.651	.623	.681
	3	-.477	.020	556.332	1	.000	.621	.597	.646
[x9m=0]	0	-.431	.023	350.879	1	.000	.650	.621	.680
	1	.236	.026	80.251	1	.000	1.266	1.202	1.333
	2	.359	.023	241.859	1	.000	1.433	1.369	1.499
	3	.508	.021	601.893	1	.000	1.663	1.596	1.732
[y7m=0]	0	-.003	.023	.013	1	.909	.997	.953	1.044
	1	.663	.025	705.029	1	.000	1.940	1.847	2.037
	2	.745	.022	1141.042	1	.000	2.107	2.018	2.200
	3	.837	.020	1776.154	1	.000	2.308	2.220	2.400
[v9f=0]	0	-.566	.025	509.206	1	.000	.568	.541	.597
	1	.022	.027	.701	1	.402	1.023	.970	1.078
	2	.129	.026	25.677	1	.000	1.138	1.083	1.196
	3	.144	.025	33.378	1	.000	1.154	1.100	1.212
[x9l=0]	0	.165	.024	49.481	1	.000	1.180	1.127	1.236
	1	.455	.024	357.120	1	.000	1.576	1.503	1.652
	2	.315	.022	207.121	1	.000	1.370	1.312	1.430
	3	-.147	.021	50.201	1	.000	.863	.829	.899
[v5e=0]	0	.152	.025	35.810	1	.000	1.164	1.108	1.224
	1	.130	.026	25.167	1	.000	1.139	1.082	1.198
	2	-.113	.025	21.359	1	.000	.893	.851	.937
	3	-.273	.024	130.619	1	.000	.761	.726	.797

a. The reference category is: 4.

ANEXO IV

Salidas de SPSS 13.0 para la regresión multinomial logística de los conglomerados de “Información”

Model Fitting Information					Likelihood Ratio Tests				
Model	Model Fitting Criteria	Likelihood Ratio Tests			Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.		-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept Only	2691.015				Intercept	857.499 ^a	.000	0	.
Final	857.499	1833.516	30	.000	ciudad	1594.045	736.545	6	.000
					sexo	1501.443	643.944	6	.000
					edad	1317.685	460.186	18	.000

Pseudo R-Square	
Cox and Snell	.012
Nagelkerke	.013
McFadden	.004

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Parameter Estimates									
	Nube	B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
Intercept	0	-1.843	.030	3707.509	1	.000			
	1	-2.449	.040	3838.458	1	.000			
	2	-1.724	.028	3729.970	1	.000			
	3	-1.361	.024	3338.788	1	.000			
	4	-1.302	.023	3120.290	1	.000			
	5	-.946	.020	2143.917	1	.000			
[ciudad=0]	0	.665	.033	408.841	1	.000	1.944	1.823	2.074
	1	.529	.044	144.233	1	.000	1.696	1.556	1.849
	2	.487	.031	250.559	1	.000	1.628	1.533	1.729
	3	.347	.026	180.725	1	.000	1.415	1.345	1.488
	4	.342	.025	182.961	1	.000	1.408	1.340	1.479
	5	.275	.023	148.645	1	.000	1.317	1.260	1.376
[sexo=0]	0	-.374	.025	220.774	1	.000	.688	.655	.723
	1	-.294	.033	80.383	1	.000	.746	.699	.795
	2	-.292	.022	170.819	1	.000	.746	.714	.780
	3	-.126	.018	49.052	1	.000	.881	.851	.913
	4	-.247	.017	199.959	1	.000	.781	.755	.808
	5	-.273	.015	321.639	1	.000	.761	.739	.784
[edad=14]	0	-.491	.034	212.281	1	.000	.612	.573	.654
	1	-.463	.044	112.035	1	.000	.629	.578	.686
	2	-.239	.030	61.676	1	.000	.788	.742	.836
	3	-.151	.025	36.958	1	.000	.859	.818	.902
	4	-.001	.025	.003	1	.956	.999	.952	1.048
	5	-.004	.022	.037	1	.847	.996	.955	1.039
[edad=15]	0	-.462	.037	154.596	1	.000	.630	.586	.677
	1	-.431	.048	79.718	1	.000	.650	.591	.714
	2	-.275	.034	66.869	1	.000	.760	.711	.812
	3	-.150	.027	30.556	1	.000	.860	.816	.908
	4	-.093	.027	11.802	1	.001	.911	.864	.961
	5	-.053	.024	4.962	1	.026	.949	.906	.994
[edad=16]	0	-.277	.038	52.470	1	.000	.758	.703	.817
	1	-.253	.050	26.103	1	.000	.776	.704	.855
	2	-.271	.036	56.773	1	.000	.763	.711	.818
	3	-.098	.029	11.702	1	.001	.907	.857	.959
	4	-.123	.029	17.994	1	.000	.884	.835	.936
	5	-.042	.025	2.847	1	.092	.959	.913	1.007

Bibliografía

- [1] Abdi, H. (2003). Factor Rotations in Factor Analysis, en Lewis-Beck, M., Bryman, A. & Futing, T. (Eds.) *Encyclopedia of Social Sciences Research Methods*. Sage.
- [2] Albert, A. and Anderson, J. A. (1984). On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. *Biometrika*, **71**, 110.
- [3] Carrol, J. B. (1953). An Analytical Solution for Approximating Simple Structure in Factor Analysis. *Psychometrika*, **18**, 23-38.
- [4] Cattell, R. B. (1966). The Scree Test for the Number of Factors. *Journal of Multivariate Behavior Research*, **1**, 245-276.
- [5] Ferguson, G. A. (1954). The Concept of Parsimony in Factor Analysis. *Psychometrika*, **19**, 281-290.
- [6] Friedman, J. H. (1987). Exploratory Projection Pursuit. *Journal of the American Statistical Association*, **82 (397)**, 249-266.
- [7] Friedman, J. H. & Tukey, J. W. (1974). A Projection Pursuit Algorithm for Exploratory Factor Analysis. *IEEE Trans, on Computers*, **C-23**, 887-890.
- [8] Hall, P. (1989). Polynomial Projection Pursuit. *The Annals of Statistics*, **17**, 589-605.
- [9] Huber, P. (1985). Projection Pursuit. *The Annals of Statistics*, **13 (2)**, 435-475.

- [10] Jackson, D. (1993). *A Users Guide to Principal Components*. Wiley.
- [11] Jennrich, R. I. & Sampson, P. F. (1966). Rotation for Simple Loadings. *Psychometrika*, **31**, 313-323.
- [12] Jolliffe, I. T. (2002). *An Introduction to Principal Component Analysis*. Springer-Verlag.
- [13] Jöreskog, K. G. (1967) Some Contributions to Maximum Likelihood Factor Analysis. *Psychometrika*, **32**, 443-482.
- [14] Kaiser, H. F. (1958). The Varimax Criterion for Analytic Rotation in Factor Analysis. *Psychometrika*, **23**, 187-200.
- [15] Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational Psychology Measuring*, **20**, 141-151.
- [16] McCullach, C. E. & Searle, S. R. (2001). *Generalized, Linear and Mixed Models*. Wiley Interscience.
- [17] Morrison, D. F. (1976). *Multivariate Statistical Methods*, (2a. Ed.). McGraw-Hill.
- [18] Neuhaus, J. O. & Wrigley, C. (1954). The Quartimax Method: An Analytical Approach to Orthogonal Simple Structure. *British Journal of Statistical Psychology*, **7**, 81-91.
- [19] Prieto-Castellanos, K. A. (2005). *Regresión Logística con Penalidad Ridge Aplicada a Datos de Expresión Genética*. Tesis de Maestría. Universidad de Puerto Rico.
- [20] Rao, C. R. (1955). Estimation and Tests of Significance in Factor Analysis. *Psychometrika*, **20**, 93-111.
- [21] Reddon, J. R. (1984). *The Number of principal Components Problem: A Monte Carlo Study*. Tesis Doctoral. University of Western Ontario.
- [22] Saunders, D. R. (1953). An Analytic Method for Rotation to Orthogonal Simple Structure. *Educational Testing Service Research Bulletin*, 53-10.
- [23] SAS Institute Inc. (2007). *JMP®*, Release 7.0.

- [24] SPSS Inc. (2004). *Statistical Package for the Social Sciences for Windows*, Release 13.0.
- [25] Thurstone, L. L. (1969). *Multiple-Factor Analysis*. University of Chicago Press.
- [26] Weisberg, S. (1980). *Applied Linear Regression*. Wiley.
- [27] Velicer, W. F. (1976). Determining the Number of Components from the Matrix of Partial Correlations. *Psychometrika*, **41**, 321-327.