4.6 Análisis de componentes principales

El análisis de componentes principales es una técnica multivariada, que se utiliza con el objetivo explicar la estructura de la matriz de varianzas y covarianzas de un grupo de p variables aleatorias observables, para lo cual se construyen k variables aleatorias no observables, como combinaciones lineales de las p variables.

Los propósitos de la aplicación de esta técnica son dos. El primero es la reducción de datos, para la cual supongamos que se tienen n unidades de investigación y p variables aleatorias observables X_1, X_2, ... ,X_p con las cuales se construyen k componentes principales, la reducción de datos se produce cuando $k<p$, el valor de k se lo selecciona de tal forma que las k componentes principales contengan la mayor cantidad de información de las p variables originales en el menor número de estas, entonces la matriz original de datos de tamaño n x p se reduce a una de tamaño n x k.

El segundo propósito es la interpretación de las variables no observables, con el fin de identificar alguna representación real de estas.

4.6.1 Determinación de las componentes principales

Sea $\mathbf{X}^T = (X_1, X_2, \ . \ . \ . \ , X_p)$ un vector aleatorio p-variado observable, es decir que $\mathbf{X} \in R^p$, sean la matriz de varianzas y covarianzas y el vector de medias correspondientes a este vector, S y $\boldsymbol{\mu}$ respectivamente y siendo $\lambda_1 \geq \lambda_2 \geq . . . \geq \lambda_p \geq 0$, los valores propios correspondientes a la matriz Σ, se construyen las siguientes combinaciones lineales.

$$
\begin{aligned}
Y_1 &= \mathbf{a_1}^T \mathbf{X} = a_{11} X_1 + a_{21} X_2 + \ . \ . \ . + a_{p1} X_p \\
Y_2 &= \mathbf{a_2}^T \mathbf{X} = a_{12} X_1 + a_{22} X_2 + \ . \ . \ . + a_{p2} X_p \\
&\vdots \\
Y_p &= \mathbf{a_p}^T \mathbf{X} = a_{1p} X_1 + a_{2p} X_2 + \ . \ . \ . + a_{pp} X_p
\end{aligned}
$$

Donde cada combinación lineal tiene:

$$\mathrm{Var}\,(Y_i) = \mathbf{a_i}^T S\ \mathbf{a_i} \qquad \text{para } i = 1, 2, \ldots, p$$

$$\mathrm{Cov}\,(Y_i\ ,\ Y_j) = \mathbf{a_i}^T S\ \mathbf{a_j} \qquad \text{para } i, j = 1, 2, \ldots, p$$

Las componenetes principales son las combinaciones lineales no correlacionadas $Y_1, Y_2, \ . \ . \ . \ , Y_p$ cuyas varianzas son tan grandes como es posible.

La primera componente principal es la combinación lineal con máxima varianza. Esto es, maximiza Var (Y_1) = $\mathbf{a_1}^T$ S $\mathbf{a_1}$. Esta claro que este valor puede ser puede incrementado, multiplicando $\mathbf{a_1}$ por alguna constante. Para eliminar este inconveniente, es conveniente restringir la norma de los vectores a uno. Entonces definimos:

Y_1 es la primera componente principal y es la combinación lineal $\mathbf{a_1}^T \mathbf{X}$ que máxima la varianza Var ($\mathbf{a_1}^T \mathbf{X}$), sujeta a la condición de que:

$$\| \mathbf{a_1} \| = 1$$

Y_2 es la segunda componente principal y es la combinación lineal $\mathbf{a_2}^T \mathbf{X}$ que máxima la varianza Var ($\mathbf{a_2}^T \mathbf{X}$), sujeta a las condiciones de que:

$$\| \mathbf{a_2} \| = 1$$

$$\text{Cov}(\mathbf{a_1}^T \mathbf{X}, \mathbf{a_2}^T \mathbf{X}) = 0$$

Generalizando tenemos que:

Y_i es la i-ésima componente principal y es la combinación lineal $\mathbf{a_i}^T \mathbf{X}$ que máxima la varianza Var ($\mathbf{a_i}^T \mathbf{X}$), sujeta a la condiciones de que:

$$\| \mathbf{a_i} \| = 1$$

$$\text{Cov}(\mathbf{a_i}^T \mathbf{X}, \mathbf{a_j}^T \mathbf{X}) = 0, \text{ para } i< j, \text{ donde } i, j = 1, 2, , \ldots , p$$

Tenemos S la matriz de varianzas y covarianzas del vector aleatorio $\mathbf{X}^T$ = $(X_1, X_2, \ldots, X_p)$. S tiene los pares de valores y vectores propios $(\lambda_1, \mathbf{e_1})$, $(\lambda_2, \mathbf{e_2})$, ... $(\lambda_p, \mathbf{e_p})$; donde $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$. Entonces el i esimo componente principal es:

$$Y_1 = \mathbf{e_1}^T \mathbf{X} = e_{11} X_1 + e_{21} X_2 + \ldots + e_{p1} X_p, \text{ para } i = 1, 2, \ldots, p$$

Sujeto a:

$$\text{Var}(Y_i) = \mathbf{e_i}^T S\ \mathbf{e_i} = \lambda_i \text{ para } i = 1, 2, \ldots, p$$

$$\text{Cov}(Y_i, Y_j) = \mathbf{e_i}^T S\ \mathbf{e_j} = 0 \text{ para } i \neq j$$

Además se tiene que:

- La suma de las varianzas de las variables aleatorias observables X_i es igual a la suma de las varianzas de las componentes principales (variables artificiales) Y_i, para i= 1, 2, ... , $k \leq p$

$$s_{11} + s_{22} + \cdots + s_{pp} = \sum_{i=1}^{p} \text{Var}(X_i) = ?_1 + ?_2 + \cdots + ?_p = \sum_{i=1}^{p} \text{Var}(Y_i)$$

- La correlación entre la componente principal Y_i y la variable aleatoria observable X_k es igual a

$$?_{y_i\, x_k} = \frac{a_{ik} \sqrt{?_i}}{\sqrt{s_{kk}}} \qquad i, k = 1, 2, \cdots, p$$

Una vez que se han construido las componentes principales, el objetivo es obtener la mayor proporción del total de la varianza de la población, explicada en el menor número de componentes principales, así el aporte de cada componente principal está dado por:

$$\frac{l_k}{l_1 + l_2 + \cdots + l_p} \qquad k = 1, 2, \cdots, p$$

que es la proporción del total de la varianza de la población explicada por la k-ésima componente, y λ_i es el i-ésimo valor propio de la matriz S.

El número de componentes principales escogidas, dependerá del porcentaje de la varianza que se desea que estas expliquen, este porcentaje se determina en base a las necesidades del estudio que se realiza.

4.6.2 Construcción de componentes principales de las variables correspondientes a *directores y rectores*

- **Construcción de componentes principales aplicadas a los datos originales**

En esta sección se construirán los componentes principales utilizando las variables observables de *directores y rectores*. Los resultados de los componentes principales, las varianzas de los componentes y su porcentaje de explicación se los obtuvo utilizando el software de aplicaciones estadísticas SPSS (Social Pourpose Statistical System) versión 8.0, el procedimiento es el siguiente se abre la matriz de datos, se va al menú de estadísticas (Statistics), se avanza hasta la opción de reducción de datos (data reduction), seleccionándola. Aparece la ventana de análisis de factores, y seleccionamos el grupo de variables aleatorias del cual queremos fabricar los factores.

Luego de esto recorremos los menús para escoger la forma de construir los componentes principales. En la sección descriptivas (descriptives), seleccionamos las estadísticas que necesitamos para el análisis, la cuales pueden ser estadísticas descriptivas univariadas, o mostrar la solución inicial. Además podemos seleccionar operaciones sobre la matriz de correlación, como calcular determinante, inversa, coeficientes de la matriz de correlación y realizar la prueba de especificidad de Barttlet.

Regresamos a la ventana análisis de factores y escogemos la opción extracción (Extraction). Dentro de este menú podemos elegir el método de extracción de los componentes principales, si trabajar con la matriz de correlación o la matriz de covarianzas, el número de componentes a extraer y graficar las varianzas de los componentes. De vuelta en la ventana análisis de factores, escogemos la opción rotación (Rotation), si deseamos rotar los componentes obtenidas, luego de lo cual salimos de esta ventana y aceptamos, obteniendo así las componentes principales.

Es necesario determinar si es factible aplicar componentes principales en la matriz de directores y rectores, por lo que se realizó en el software SPSS 8.0 la prueba Barttlet 1950, la cual nos indica si la matriz de covarianzas no es una matriz diagonal. La prueba de hipótesis es:

$$H_o : \sigma_{ij} = 0, \text{ para } i \neq j$$

$$\text{vs}$$

$$H_1 = \neg H_o$$

Estadístico de prueba: 4135.41

Grados de libertad: 190

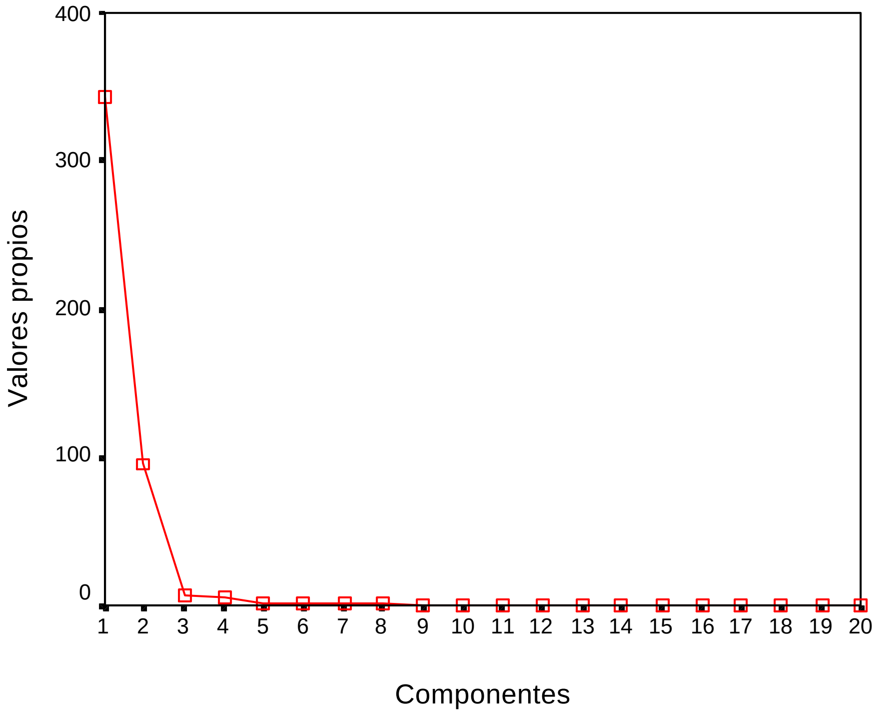
Valor p: 0.000

Por lo tanto no existe evidencia estadística para afirmar que la matriz de covarianzas es una matriz diagonal.

Luego de efectuar el análisis de componentes principales en el software SPSS, se obtuvo como resultado que el porcentaje de explicación de la varianza de la población de los dos primeros principales es muy alto, en relación al total de varianza explicada por todos los componentes obtenidos. Este resultado produce una reducción de datos.

GRÁFICO 4.5

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VALORES PROPIOS DE LAS COMPONENTES PRINCIPALES, DE LOS DATOS ORIGINALES DE LA MATRIZ DE *DIRECTORES Y RECTORES*



FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

En el gráfico 4.5 se pueden observar los valores propios obtenidos de la matriz Σ, con respecto a los 20 componentes principales, en este gráfico se puede apreciar que alrededor del tercer componente principal se forma un codo, el cual indica que los valores propios de ahí en adelante son bastante pequeños (varianzas de los componentes), por lo que no influirán en la proporción de explicación del total de la varianza de la población.

En la tabla CLXXVIII se muestran las varianzas individuales de los dos primeros componentes principales, con los que al calcular el porcentaje de explicación del total de la varianzas de la población, de estos componentes se obtiene como resultado 96.07%., la elección de los componentes principales construidos a partir de las variables aleatorias no estandarizadas con su respectiva matriz de varianzas y covarianzas de directores y rectores se basa en la formación del codo en el gráfico 4.25, así obteniéndose el mayor porcentaje de explicación de estas y la considerable reducción de datos al escoger las dos primeras componentes.

TABLA CLXXVIII

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VARIANZAS Y PORCENTAJE DE EXPLICACIÓN DE LAS 2 PRIMERAS COMPONENTES PRINCIPALES DE LOS DATOS ORIGINALES DE LA MATRIZ DE *DIRECTORES Y RECTORES*

Componente	1	**2**
Varianza	342.844	94.982
Porcentaje	75.224	20.840

Varianza Total = 455.76

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

El porcentaje de explicación de los dos primeros componentes principales se lo obtiene utilizando la siguiente expresión:

$$\frac{\sum_{i=1}^{2} ?_i}{\sum_{i=1}^{20} ?_i} = \frac{437.83}{455.76} = 0.9607$$

Los coeficientes de los dos primeros componentes obtenidos en base a las variables aleatorias observables no estandarizadas, son mostrados en el anexo 19.

- **Construcción de componentes principales aplicadas a las variables estandarizadas de directores y rectores**

Utilizando el procedimiento descrito en la sección 4.4.1 para la determinación de los componentes principales de un vector p-variado observable, donde p = 20, se realizó el cálculo sobre la base de las variables estandarizadas, es decir que se utilizó la matriz de correlación de directores y rectores, en vez de la de varianzas y covarianzas.

La estandarización de las variables aleatorias se la realiza cuando, las escalas de los valores que toman las variables de estudio difieren considerablemente de variable a variable. La estandarización de las variables aleatorias observables se la realiza de la siguiente manera:

$$Z_1 = \frac{(X_1 - \mu_1)}{\sqrt{s_{11}}}$$

$$Z_2 = \frac{(X_2 - \mu_2)}{\sqrt{s_{22}}}$$

$$\vdots$$

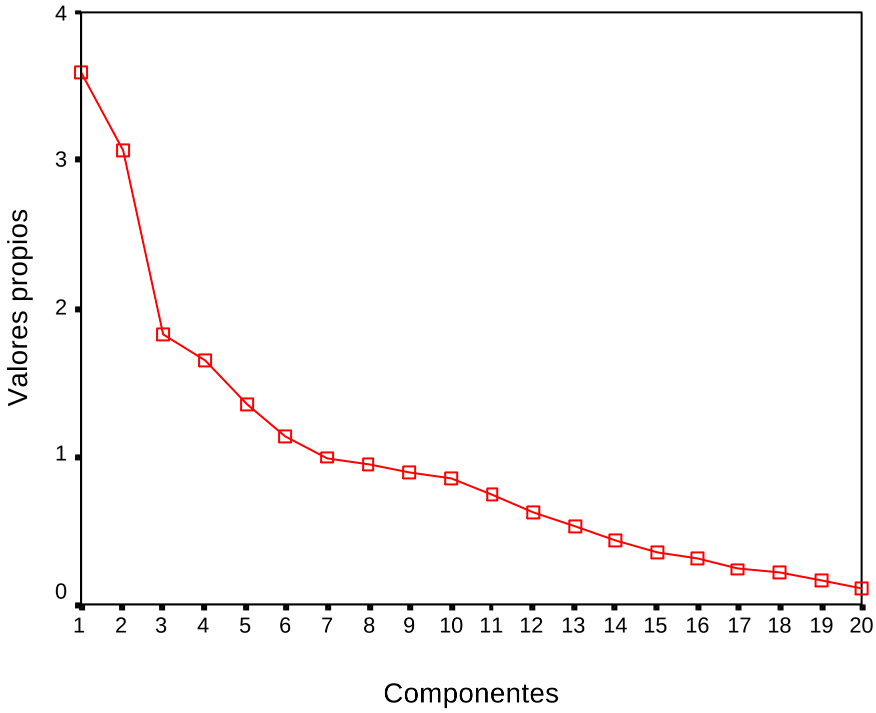
$$Z_p = \frac{(X_p - \mu_p)}{\sqrt{s_{pp}}}$$

Donde X_i e la i-ésima variable aleatoria observable, μ_p es el valor esperado de X_i y $\sqrt{\sigma_{pp}}$ representa la desviación estándar de X_i, entonces Z_i es la i-ésima variable aleatoria estandarizada, para i = 1, 2, ... ,p.

Una vez que se estandarizaron las variables de estudio, las cuales son representadas por el vector $\mathbf{Z}^T = (Z_1, Z_2, \ldots, Z_p)$, donde p= 20, se calculó la matriz de varianzas y covarianzas de **Z** que es igual a la matriz de correlación ? de directores y rectores, lo cual se denota como Cov(**Z**)= ?.

GRÁFICO 4.6

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VALORES PROPIOS DE LAS COMPONENTES PRINCIPALES, DE LOS DATOS ESTANDARIZADOS DE MATRIZ *DIRECTORES Y RECTORES*



FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

La determinación del número apropiado de componentes, por medio del gráfico de la magnitud de los valores propios de la matriz de correlación de directores y rectores versus el número de componentes, indica que, se deben elegir solamente los seis primeros componentes, debido a que desde λ_7 hasta λ_{20} todos los valores propios son muy pequeños y están alrededor de un mismo valor. En el gráfico 4.6 se puede observar esta situación, pues alrededor del séptimo componente se forma un codo, luego del cual los valores propios son pequeños, de acuerdo a esto el número de componentes elegidos es seis. Los porcentajes de explicación individuales de cada uno de los seis componentes elegidos se muestran en la tabla CLXXIX. Los coeficientes de los seis componentes principales de la matriz de correlación ? se muestran en el anexo 20.

El porcentaje de explicación del total de la varianza de la población, de las seis primeras componentes se lo obtiene mediante el cálculo de la siguiente expresión:

$$\frac{\sum_{i=1}^{6} ?_i}{\sum_{i=1}^{20} ?_i} = \frac{12.602}{20.000} = 0.6301$$

Este resultado indica que el 63.01% del total de la varianza de la población de directores y rectores es explicada con los seis primeros componentes principales.

TABLA CLXXIX

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VARIANZAS Y PORCENTAJE DE EXPLICACIÓN DE LAS 6 PRIMERAS COMPONENTES PRINCIPALES DE LAS VARIABLES ESTANDARIZADAS DE MATRIZ *DIRECTORES Y RECTORES*

Componente	**1**	**2**	**3**
Varianza	3.589	3.054	1.825
Porcentaje	17.944	15.270	9.126
Componente	**4**	**5**	**6**
Varianza	1.644	1.349	1.141
Porcentaje	8.219	6.747	5.706

Varianza Total = 20

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

Una vez que se calcularon los coeficientes de los vectores propios e_i correspondientes a los valores propios λ_i, para i = 1, 2, ... , k, donde k=6, a continuación se tabulan las componentes principales:

$$\mathbf{Y}_1 = a_1^T X = -0.140IP_1 + 0.521IP_2 + \ldots - 0.787\,IL_8$$

$$\mathbf{Y}_2 = a_2^T X = -0.034IP_1 - 0.567IP_2 + \ldots + 0.213\,IL_8$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$\mathbf{Y}_6 = a_6^T X = 0.185\,IP_1 + 0.043IP_2 + \ldots + 0.012\,IL_8$$

Si se estandarizan las variables aleatorias observables de directores y rectores, se logra reducir la matriz de datos original de tamaño 537 x 20 a otra de 537 x 6.

Aunque las dos primeras componentes principales obtenidas a partir de las variables aleatorias originales, explican un mayor porcentaje del total de la varianza de la población, que las seis componentes obtenidas en base a las variables aleatorias estandarizadas de la matriz de datos de directores y rectores, las componentes que se utilizarán para remplazar a las variables originales son las seis primeras componentes principales obtenidos a partir de las variables estandarizadas. La selección se basa en el hecho de que al estandarizar las variables, se reduce el efecto de las variables que son medidas en escalas con rangos muy diferentes o si las unidades de medida no son proporcionales.

- **Rotación de las componentes principales obtenidas en base de las variables aleatorias estandarizadas de directores y rectores**

Como se mencionó en la sección anterior las componentes principales seleccionadas fueron las que se obtuvieron a partir de las variables aleatorias observables estandarizadas, con esta selección se obtuvo una reducción de datos de una matriz original de datos de directores y rectores de tamaño 537 x 20 a una de tamaño 537 x 6.

El siguiente paso que se realizó en este análisis es el de rotar los componentes principales. Las combinaciones lineales geométricamente representan la selección de un nuevo sistema de coordenadas, luego de rotar el sistema original con axisas IP_1, IP_2, ... , IL_8, las nuevas coordenadas representan las direcciones con máxima variación de las observaciones, que proveen una simple y detallada descripción de la estructura de la covarianza del vector **X**. Para rotar los componentes se utilizó el método varimax, mediante el cual el sistema de coordenadas obtenido es ortogonal. La rotación se la realiza con el objetivo de interpretar las combinaciones lineales (componentes principales) para obtener una representación real de lo que significan.

TABLA CLXXX

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VARIANZAS Y PORCENTAJE DE EXPLICACIÓN DE LAS 6 PRIMERAS COMPONENTES PRINCIPALES DE VARIABLES ESTANDARIZADAS Y ROTADAS DE *DIRECTORES Y RECTORES*

Componente	**1**	**2**	**3**
Varianza	3.456	3.046	1.718
Porcentaje	17.282	15.232	8.588
Componente	**4**	**5**	**6**
Varianza	1.556	1.550	1.275
Porcentaje	7.782	7.751	6.377

Varianza Total = 20

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

El porcentaje del total de la varianza de la población de directores y rectores explicada por los componentes principales rotados es del 63.01%. Calculado de la siguiente forma:

$$\frac{\sum_{i=1}^{6} ?_i}{\sum_{i=1}^{20} ?_i} = \frac{12.602}{20.000} = 0.6301$$

En la tabla CLXXX se puede observar los porcentajes individuales de la explicación del total de la varianza de las seis primeras componentes principales obtenidas a partir de las variables aleatorias estandarizadas y rotadas de directores y rectores. Si bien no aumento el porcentaje de explicación de las seis componentes, este se distribuyó entre los componentes seleccionados. Los coeficientes de las seis combinaciones lineales obtenidas luego de estandarizar y rotar la variables aleatorias originales, se muestran en el anexo 21. El programa que se utilizó para calcular las componentes principales es el software estadístico SPSS, versión 8.0.

- **Rotulación de las componentes principales aplicadas a las variables aleatorias estandarizadas y rotadas**

La rotulación es el proceso que se realiza para nominar las componentes principales, de tal forma que el nombre corresponda a alguna representación real del componente. Para este efecto se considera la magnitud de los coeficientes que tienen las variables aleatorias observables de la matriz de directores y rectores (IP_1, IP_2, ... , IL_8) en cada una de las combinaciones lineales.

Para que los coeficientes de las combinaciones lineales sean considerados como significativos estos deben ser mayores a 0.5, en este caso se utiliza la variable para rotular el componente y si son menores a 0.3 se la omite.

- **Rotulación de la primera componente principal**

  Los coeficientes de la primera componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IP_2 (Edad), IE_2 (Título docente), IE_6 (Años de experiencia), IE_7 (Categoría nominal), IE_8 (Categoría económica), IL_6 (Zona de institución), , por lo tanto el nombre con el que se rotularía esta componente sería el de **factor experiencia del director**.

- **Rotulación de la segunda componente principal**

  Los coeficientes de la segunda componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IE_3 (título no docente), IE_4 (Clase de título), IE_5 (Tipo de nombramiento), IL_7 (Relación laboral), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor información laboral del director**.

- **Rotulación de la tercera componente principal**

Los coeficientes de la tercera componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IP_5 (Provincia que habita), IP_6 (Cantón que habita), IL_2 (Cantón de institución), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor movilización del director**.

- **Rotulación de la cuarta componente principal**

Los coeficientes de la cuarta componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IE_1 (Último nivel de instrucción), IL_4 (Nivel de institución), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor nivel donde labora el director según instrucción**.

- **Rotulación de la quinta componente principal**

Los coeficientes de la quinta componente principal que son mayores a 0.5 son los correspondientes a la variable aleatoria IL_5 (Sostenimiento de institución), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor sostenimiento de institución del director**.

- **Rotulación de la sexta componente principal**

Los coeficientes de la sexta componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IP_1 (Provincia de nacimiento), IP_5 (Provincia que habita), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor migración del director**.

4.6.3 Construcción de componentes principales de las variables correspondientes a profesores

- **Construcción de componentes principales aplicadas a los datos originales de la matriz de profesores**

En esta sección se construirán los componentes principales utilizando las variables observables no estandarizadas de la matriz de profesores, bs resultados de los componentes principales, las varianzas de los componentes y su porcentaje de explicación con respecto a la varianza total se los obtuvo utilizando el software de aplicaciones estadísticas SPSS (Social Pourpose Statistical System) versión 8.0 con el procedimiento explicado en la sección 4.5.2.

Es necesario determinar si es factible aplicar componentes principales en la matriz de profesores, por lo que se realizó en el software SPSS 8.0 la prueba Barttlet 1950, la cual nos indica si la matriz de covarianzas de profesores no es una matriz diagonal. La prueba de hipótesis es:

$$H_o : \sigma_{ij} = 0, \text{ para } i \neq j$$

vs

$$H_1 = \neg H_o$$

Estadístico de prueba: 36604.84

Grados de libertad: 210.000

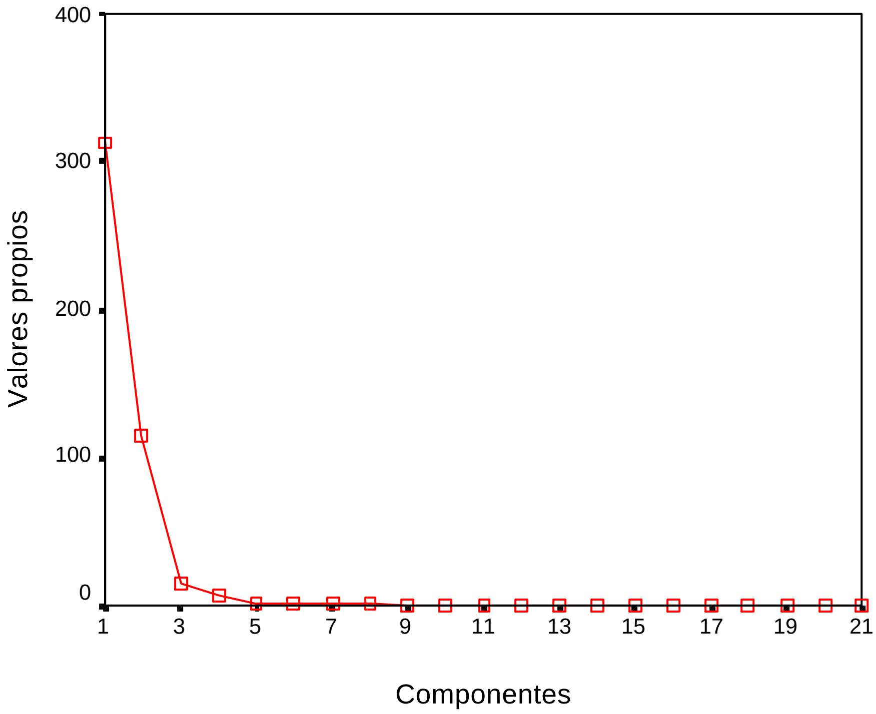
Valor p: 0.000

Por lo tanto no existe evidencia estadística para afirmar que la matriz de covarianzas de profesores es una matriz diagonal.

Efectuando este análisis se obtuvo como resultado que el porcentaje de explicación de la varianza de la población es muy alto, en relación con el de los componentes determinados sobre las variables estandarizadas. Otro resultado es que se produce una mayor reducción de datos.

GRÁFICO 4.7

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VALORES PROPIOS DE LAS COMPONENTES PRINCIPALES, DE LOS DATOS ORIGINALES DE LA MATRIZ DE *PROFESORES*



FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

En el gráfico 4.7 se pueden observar los valores propios obtenidos de la matriz de profesores Σ, con respecto a los 21 componentes principales, en este gráfico se puede apreciar que alrededor del segundo componente principal se forma un codo, el cual indica que los valores propios de ahí en adelante son bastante pequeños (varianzas de los componentes), por lo que no influirán en la proporción de explicación del total de la varianza de la población.

En la tabla CLXXXI se muestran las varianzas individuales de los dos primeros componentes principales obtenidos a partir de la matriz de profesores, con los que al calcular el porcentaje de explicación del total de la varianzas de la población, de estos componentes se obtiene como resultado 93.586%.

TABLA CLXXXI

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VARIANZAS Y PORCENTAJE DE EXPLICACIÓN DE LAS 2 PRIMERAS COMPONENTES PRINCIPALES DE LOS DATOS ORIGINALES DE LA MATRIZ DE *PROFESORES*

Componente	**1**	**2**
Varianza	313.092	114.710
Porcentaje	68.492	25.094

Varianza Total = 457.122

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

El porcentaje de explicación de los dos primeros componentes principales de la matriz de profesores se lo obtiene utilizando la siguiente expresión:

$$\frac{\sum_{i=1}^{2} ?_i}{\sum_{i=1}^{21} ?_i} = \frac{427.80}{457.12} = 0.9358$$

Este resultado indica que el 93.586% del total de la varianza de la población es explicada por las dos primeras componentes principales.

Los coeficientes de los dos primeros componentes obtenidos en base a las variables aleatorias observables no estandarizadas, son mostrados en el anexo 22.

- **Construcción de componentes principales aplicadas a las variables estandarizadas de la matriz de *profesores***

Utilizando el procedimiento descrito en la sección 4.4.1 para la determinación de los componentes principales de un vector p-variado observable, donde p = 21, se realizó el cálculo sobre la base de las variables estandarizadas, es decir que se utilizó la matriz de correlación de profesores, en vez de la de varianzas y covarianzas.

La estandarización de las variables aleatorias se la realiza cuando, las escalas de los valores que toman las variables de estudio difieren considerablemente de variable a variable.

En la tabla CLXXXII se muestran los 6 valores propios de la matriz de correlación de profesores ?, que corresponden a las varianzas de las primeras seis componentes principales.

TABLA CLXXXII

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VARIANZAS Y PORCENTAJE DE EXPLICACIÓN DE LAS 6 PRIMERAS COMPONENTES PRINCIPALES DE LAS VARIABLES ESTANDARIZADAS DE LA MATRIZ DE *PROFESORES*

Componente	**1**	**2**	**3**
Varianza	5.320	2.205	1.890
Porcentaje	25.333	10.500	9.000
Componente	**4**	**5**	**6**
Varianza	1.693	1.447	1.145
Porcentaje	8.061	6.891	5.450

Varianza Total = 21

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

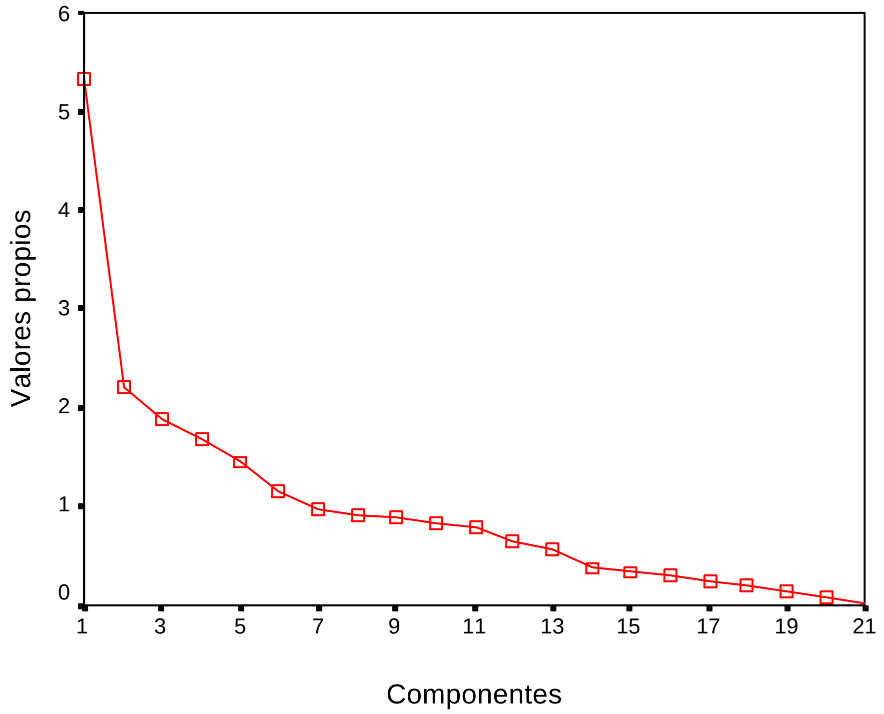
El porcentaje de explicación del total de la varianza de la población, de las seis primeras componentes se lo obtiene mediante el cálculo de la siguiente expresión:

$$\frac{\sum_{i=1}^{6} ?_i}{\sum_{i=1}^{21} ?_i} = \frac{13.699}{21.000} = 0.6523$$

Este resultado indica que el 65.235% del total de la varianza de la población de los profesores es explicada con los seis primeros componentes principales.

GRÁFICO 4.8

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VALORES PROPIOS DE LAS COMPONENTES PRINCIPALES, DE LOS DATOS ESTANDARIZADOS DE LA MATRIZ DE *PROFESORES*



FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

Por otro lado, la determinación del número apropiado de componentes, por medio del gráfico de la magnitud de los valores propios versus el número de componentes, concuerda que, se deben elegir solamente los seis primeros componentes, debido a que desde λ_7 hasta λ_{21} todos los valores propios de la matriz de profesores son muy pequeños y están alrededor de un mismo valor.

En el gráfico 4.8 se puede observar esta situación, pues alrededor del séptimo componente se forma un codo, luego del cual los valores propios obtenidos de la matriz de correlación de los profesores son pequeños. Los coeficientes de los seis componentes principales se muestran en el anexo 23.

Una vez que se calcularon los coeficientes de los vectores propios e_i correspondientes a los valores propios λ_i, para i = 1, 2, ... , k, donde k=6, a continuación se tabulan las componentes principales:

$$\mathbf{Y}_1 = a_1^T X = 0.077IP_1 + 0.522IP_2 + \ldots - 0.787\, IL_8$$

$$\mathbf{Y}_2 = a_2^T X = -0.009IP_1 - 0.561IP_2 + \ldots - 0.050\, IL_8$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$\mathbf{Y}_6 = a_6^T X = 0.117\, IP_1 + 0.487IP_2 + \ldots + 0.050\, IL_8$$

Si se estandarizan las variables aleatorias observables, se logra reducir la matriz de datos original de profesores de tamaño 2815 x 21 a otra de 2815 x 6.

Aunque las dos primeras componentes principales obtenidas a partir de las variables aleatorias originales de la matriz de datos de profesores, explican un mayor porcentaje del total de la varianza de la población, que las seis componentes obtenidas en base a las variables aleatorias estandarizadas, las componentes que se utilizarán para remplazar a las variables originales son las seis primeras componentes principales obtenidos a partir de las variables estandarizadas de profesores. La selección se basa en el hecho de que al estandarizar las variables, se reduce el efecto de las variables que son medidas en escalas con rangos muy diferentes o si las unidades de medida no son proporcionales.

- **Rotación de las componentes principales obtenidas en base de las variables aleatorias estandarizadas de la matriz de profesores**

Como se mencionó en la sección anterior las componentes principales seleccionadas fueron las que se obtuvieron a partir de las variables aleatorias observables estandarizadas de profesores, con esta selección se obtuvo una reducción de datos de una matriz original de tamaño 2815 x 21 a una de tamaño 2815 x 6.

El siguiente paso que se realizó en este análisis es el de rotar los componentes principales. Las combinaciones lineales geométricamente representan la selección de un nuevo sistema de coordenadas, luego de rotar el sistema original con axisas IP_1, IP_2, ... , IL_8, las nuevas coordenadas representan las direcciones con máxima variación de las observaciones, que proveen una simple y detallada descripción de la estructura de la covarianza del vector **X**.

Para rotar los componentes se utilizó el método varimax, mediante el cual el sistema de coordenadas obtenido es ortogonal. La rotación se la realiza con el objetivo de interpretar las combinaciones lineales (componentes principales) para obtener una representación real de lo que significan.

TABLA CLXXXIII

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VARIANZAS Y PORCENTAJE DE EXPLICACIÓN DE LAS 6 PRIMERAS COMPONENTES PRINCIPALES DE LAS VARIABLES ESTANDARIZADAS Y ROTADAS DE LA MATIZ DE *PROFESORES*

Componente	**1**	**2**	**3**
Varianza	3.997	2.982	1.874
Porcentaje	19.032	14.201	8.923
Componente	**4**	**5**	**6**
Varianza	1.799	1.749	1.299
Porcentaje	8.565	8.328	6.187

Varianza Total = 21

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

En la tabla CLXXXIII se puede observar los porcentajes individuales de la explicación del total de la varianza de las seis primeras componentes principales obtenidas a partir de las variables aleatorias estandarizadas y rotadas de la matriz de datos de profesores. Los coeficientes de las seis combinaciones lineales obtenidas luego de estandarizar y rotar la variables aleatorias originales de la matriz de profesores, se muestran en el anexo 24.

El programa que se utilizó para calcular las componentes principales es el software estadístico SPSS, versión 8.0.

- **Rotulación de las componentes principales aplicadas a las variables aleatorias estandarizadas y rotadas de profesores**

La rotulación es el proceso que se realiza para nominar las componentes principales, de tal forma que el nombre corresponda a alguna representación real del componente. Para este efecto se considera la magnitud de los coeficientes que tienen las variables aleatorias observables de la matriz de datos de profesores (IP_1, IP_2, ... , IL_8) en cada una de las combinaciones lineales.

Para que los coeficientes de las combinaciones lineales sean considerados como significativos estos deben ser mayores a 0.5, en este caso se utiliza la variable para rotular el componente y si son menores a 0.3 se la omite.

- **Rotulación de la primera componente principal**

Los coeficientes de la primera componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IP_2 (Edad), IE_5 (Tipo de nombramiento), IE_6 (Años de experiencia), IE_7 (Categoría nominal), IE_8 (Categoría económica), IL_7 (Relación laboral), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor información laboral de profesor**.

- **Rotulación de la segunda componente principal**

Los coeficientes de la segunda componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IE_1 (Último nivel de instrucción), IE_2 (Título docente), IE_3 (título no docente), IE_4 (Clase de título), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor instrucción de profesor**.

- **Rotulación de la tercera componente principal**

Los coeficientes de la tercera componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IP_3 (Sexo), IP_4 (Estado civil) y IE_1 (Último nivel de instrucción), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor instrucción según sexo de profesor**.

- **Rotulación de la cuarta componente principal**

Los coeficientes de la cuarta componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IP_6 (Cantón que habita) y IL_2 (Cantón de institución), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor ubicación de hogar y trabajo de profesor**.

- **Rotulación de la quinta componente principal**

Los coeficientes de la quinta componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IL_6 (Zona de institución) y IL_8 (Vivienda rural), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor residencia de profesor según zona de institución**.

- **Rotulación de la sexta componente principal**

Los coeficientes de la sexta componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IP_1 (Provincia de nacimiento) y IP_5 (Provincia que habita), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor movilidad geográfica de profesor**.

4.6.4 Construcción de componentes principales de las variables correspondientes a *otros funcionarios*

- **Construcción de componentes principales aplicadas a los datos originales *otros funcionarios***

En esta sección se construirán los componentes principales utilizando las variables observables no estandarizadas de la matriz de otros funcionarios, los resultados de los componentes principales, las varianzas de los componentes y su porcentaje de explicación con respecto a la varianza total se los obtuvo utilizando el software de aplicaciones estadísticas SPSS (Social Pourpose Statistical System) versión 8.0 con el procedimiento explicado en la sección 4.5.2.

Es necesario determinar si es factible aplicar componentes principales en la matriz de otros funcionarios, por lo que se realizó en el software SPSS 8.0 la prueba Barttlet 1950, la cual nos indica si la matriz de covarianzas de profesores no es una matriz diagonal. La prueba de hipótesis es:

$$H_o : \sigma_{ij} = 0, \text{ para } i \neq j$$

$$\text{vs}$$

$$H_1 = \neg H_o$$

Estadístico de prueba: 4871.94

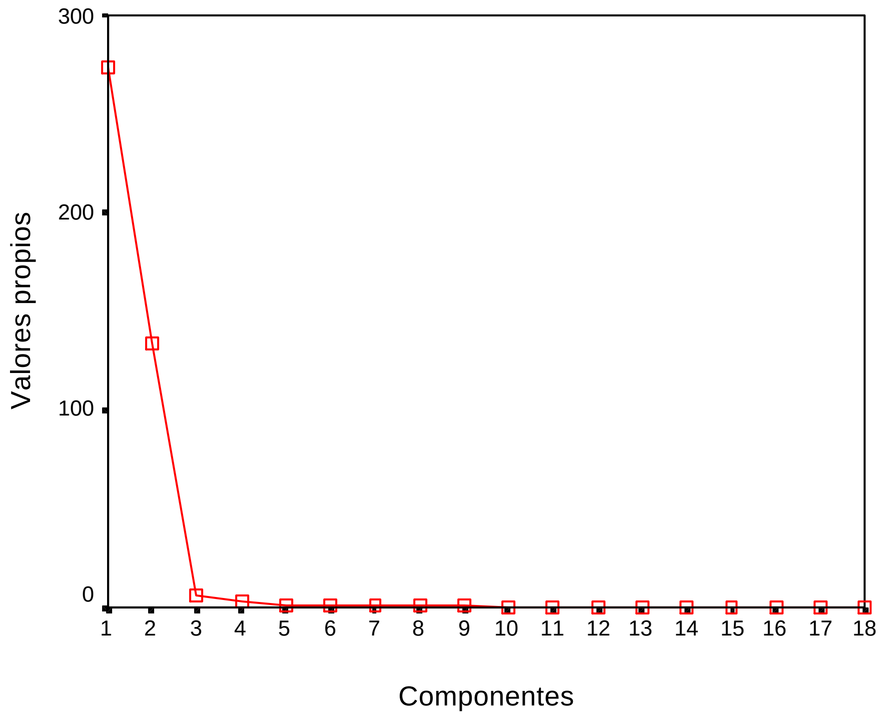
Grados de libertad: 153

Valor p: 0.000

Por lo tanto no existe evidencia estadística para afirmar que la matriz de correlaciones de otros funcionarios es una matriz diagonal, por ende es aplicable la técnica de componentes principales.

GRÁFICO 4.9

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VALORES PROPIOS DE LAS COMPONENTES PRINCIPALES, DE LOS DATOS ORIGINALES DE LA MATRIZ DE *OTROS FUNCIONARIOS*



FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

En el gráfico 4.9 se pueden observar los valores propios obtenidos de la matriz de covarianzas de otros funcionarios Σ, con respecto a los 18 componentes principales, en este gráfico se puede apreciar que alrededor del segundo componente principal se forma un codo, el cual indica que los valores propios de la matriz de covarianzas de otros funcionarios de ahí en adelante son bastante pequeños (varianzas de los componentes), por lo que no influirán en la proporción de explicación del total de la varianza de la población de otros funcionarios.

En la tabla CLXXXIV se muestran las varianzas individuales de los dos primeros componentes principales, con los que al calcular el porcentaje de explicación del total de la varianzas de la población, de estos componentes se obtiene como resultado 96.403%.

TABLA CLXXXIV

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VARIANZAS Y PORCENTAJE DE EXPLICACIÓN DE LAS 2 PRIMERAS COMPONENTES PRINCIPALES DE LOS DATOS ORIGINALES DE LA MATRIZ *OTROS FUNCIONARIOS*

Componente	**1**	**2**
Varianza	273.528	133.817
Porcentaje	64.734	31.670

Varianza Total = 422.541

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

En la Tabla CLXXXIV se muestran los porcentajes de explicación individuales de los dos primeros componentes principales, la elección de los componentes principales construidos a partir de las variables aleatorias no estandarizadas de la matriz de datos de otros funcionarios con su respectiva matriz de varianzas y covarianzas se basa en el mayor porcentaje de explicación de estas y la considerable reducción de datos al escoger las dos primeras componentes.

El porcentaje de explicación de los dos primeros componentes principales se lo obtiene utilizando la siguiente expresión:

$$\frac{\sum_{i=1}^{2} ?_i}{\sum_{i=1}^{18} ?_i} = \frac{407.34}{422.54} = 0.964$$

Este resultado indica que el 96.403% del total de la varianza de la población es explicada por las dos primeras componentes principales. Los porcentajes de explicación individuales de cada componente se muestran en la tabla CLXXXIV.

Los coeficientes de los dos primeros componentes obtenidos en base a las variables aleatorias observables no estandarizadas de la matriz de datos de otros funcionarios, son mostrados en el anexo 25.

- **Construcción de componentes principales aplicadas a las variables estandarizadas de la matriz de *otros funcionarios***

Utilizando el procedimiento descrito en la sección 4.4.1 para la determinación de los componentes principales de un vector p-variado observable, donde p = 18, se realizó el cálculo sobre la base de las variables estandarizadas, es decir que se utilizó la matriz de correlación de otros funcionarios, en vez de la de varianzas y covarianzas.

La estandarización de las variables aleatorias se la realiza cuando, las escalas de los valores que toman las variables de estudio difieren considerablemente de variable a variable.

En la tabla CLXXXV se muestran los 7 valores propios de la matriz de correlación ? de otros funcionarios, que corresponden a las varianzas de las primeras siete componentes principales. El porcentaje de explicación del total de la varianza de la población, de las siete primeras componentes se lo obtiene mediante el cálculo de la siguiente expresión:

$$\frac{\sum_{i=1}^{7} ?_i}{\sum_{i=1}^{18} ?_i} = \frac{12.517}{18.000} = 0.69538$$

Este resultado indica que el 69.538% del total de la varianza de la población es explicada con los siete primeros componentes principales obtenidos de la matriz de datos de otros funcionarios.

TABLA CLXXXV

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VARIANZAS Y PORCENTAJE DE EXPLICACIÓN DE LAS 7 PRIMERAS COMPONENTES PRINCIPALES DE LAS VARIABLES ESTANDARIZADAS A LA MATRIZ DE *OTROS FUNCIONARIOS*

Componente	**1**	**2**	**3**
Varianza	3.386	2.229	1.900
Porcentaje	18.810	12.381	10.556
Componente	**4**	**5**	**6**
Varianza	1.507	1.318	1.170
Porcentaje	8.370	7.321	6.498
Componente		**7**	
Varianza		1.008	
Porcentaje		5.601	

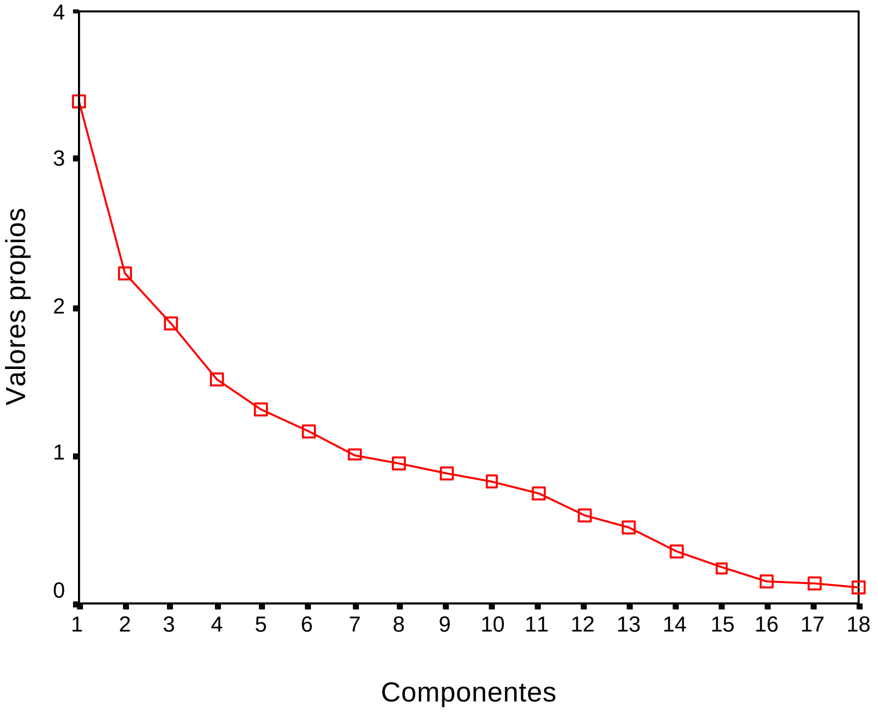
Varianza Total = 18

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

Por otro lado, la determinación del número apropiado de componentes, por medio del gráfico de la magnitud de los valores propios versus el número de componentes, concuerda que, se deben elegir solamente los siete primeros componentes, debido a que desde λ_8 hasta λ_{18} todos los valores propios de la matriz de correlación de otros funcionarios son muy pequeños y están alrededor de un mismo valor. En el gráfico 4.10 se puede observar esta situación, pues alrededor del décimo componente se forma un codo, luego del cual los valores propios son pequeños.

GRÁFICO 4.10

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VALORES PROPIOS DE LAS COMPONENTES PRINCIPALES, DE LOS DATOS ESTANDARIZADOS DE LA MATRIZ DE *OTROS FUNCIONARIOS*



FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

La elección del número de componentes principales construidos en base a las variables estandarizadas, se la realizó utilizando el criterio del porcentaje de la varianza de la población de los otros funcionarios que estos explican, de acuerdo a esto el número de componentes elegidos es siete. Los porcentajes de explicación individuales de cada uno de los siete componentes elegidos se muestran en la tabla CLXXXV. Los coeficientes de los siete componentes principales se muestran en el anexo 26.

Una vez que se calcularon los coeficientes de los vectores propios e_i correspondientes a los valores propios λ_i, de la matriz de correlacion de otros funcionarios para i = 1, 2, ... , k, donde k=7, a continuación se tabulan las componentes principales:

$$\mathbf{Y}_1 = a_1^T X = 0.021IP_1 - 0.128IP_2 + \ldots - 0.276\,IL_7$$

$$\mathbf{Y}_2 = a_2^T X = 0.117IP_1 + 0.814IP_2 + \ldots - 0.491\,IL_7$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$\mathbf{Y}_7 = a_7^T X = -0.604\,IP_1 + 0.238IP_2 + \ldots + 0.009\,IL_7$$

Si se estandarizan las variables aleatorias observables, se logra reducir la matriz de datos original de otros funcionarios de tamaño 712 x 18 a otra de 712 x 7.

Aunque las dos primeras componentes principales obtenidas a partir de las variables aleatorias originales, explican un mayor porcentaje del total de la varianza de la población de otros funcionarios, que las siete componentes obtenidas en base a las variables aleatorias estandarizadas, las componentes que se utilizarán para remplazar a las variables originales son las siete primeras componentes principales obtenidos a partir de las variables estandarizadas. La selección se basa en el hecho de que al estandarizar las variables, se reduce el efecto de las variables de la matriz de datos de otros funcionarios son medidas en escalas con rangos muy diferentes o si las unidades de medida no son proporcionales.

- **Rotación de las componentes principales obtenidas en base de las variables aleatorias estandarizadas de la matriz de *otros funcionarios***

Como se mencionó en la sección anterior las componentes principales seleccionadas fueron las que se obtuvieron a partir de las variables aleatorias observables estandarizadas, con esta selección se obtuvo una reducción de datos de una matriz de otros funcionarios original de tamaño 712 x 18 a una de tamaño 712 x 7.

El siguiente paso que se realizó en este análisis es el de rotar los componentes principales. Las combinaciones lineales geométricamente representan la selección de un nuevo sistema de coordenadas, luego de rotar el sistema original con axisas IP_1, IP_2, ... , IL_7, las nuevas coordenadas representan las direcciones con máxima variación de las observaciones, que proveen una simple y detallada descripción de la estructura de la covarianza del vector **X**. Para rotar los componentes se utilizó el método varimax, mediante el cual el sistema de coordenadas obtenido es ortogonal. La rotación se la realiza con el objetivo de interpretar las combinaciones lineales (componentes principales) para obtener una representación real de lo que significan.

TABLA CLXXXVI

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VARIANZAS Y PORCENTAJE DE EXPLICACIÓN DE LAS 7 PRIMERAS COMPONENTES PRINCIPALES DE VARIABLES ESTANDARIZADAS Y ROTADAS DE MATRIZ *OTROS FUNCIONARIOS*

Componente	**1**	**2**	**3**
Varianza	2.957	2.008	1.939
Porcentaje	16.429	11.155	10.770
Componente	**4**	**5**	**6**
Varianza	1.788	1.434	1.207
Porcentaje	9.933	7.965	6.705
Componente		**7**	
Varianza		1.185	
Porcentaje		6.581	

Varianza Total = 18

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

En la tabla CLXXXVI se puede observar los porcentajes individuales de la explicación del total de la varianza de las siete primeras componentes principales obtenidas a partir de las variables aleatorias estandarizadas y rotadas de la matriz de otros funcionarios. Los coeficientes de las siete combinaciones lineales obtenidas luego de estandarizar y rotar la variables aleatorias originales, se muestran en el anexo 27.

El programa que se utilizó para calcular las componentes principales es el software estadístico SPSS, versión 8.0.

- **Rotulación de las componentes principales aplicadas a las variables aleatorias estandarizadas y rotadas de *otros funcionarios***

La rotulación es el proceso que se realiza para nominar las componentes principales, de tal forma que el nombre corresponda a alguna representación real del componente. Para este efecto se considera la magnitud de los coeficientes que tienen las variables aleatorias observables (IP_1, IP_2, ... , IL_7) en cada una de las combinaciones lineales.

Para que los coeficientes de las combinaciones lineales sean considerados como significativos estos deben ser mayores a 0.5, en este caso se utiliza la variable para rotular el componente y si son menores a 0.3 se la omite.

- **Rotulación de la primera componente principal**

Los coeficientes de la primera componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IE_1 (Último nivel de instrucción), IE_2 (Título docente), IE_4 (Clase de título), IE_5 (Tipo de nombramiento), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor instrucción de funcionario**.

- **Rotulación de la segunda componente principal**

Los coeficientes de la segunda componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IP_2 (Edad) y IE_6 (Años de experiencia), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor edad y experiencia de funcionario**.

- **Rotulación de la tercera componente principal**

Los coeficientes de la tercera componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IP_6 (Cantón que habita) y IL_2 (Cantón de institución), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor ubicación de hogar y trabajo de funcionario**.

- **Rotulación de la cuarta componente principal**

Los coeficientes de la cuarta componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IE_3 (título no docente), IE_5 (Tipo de nombramiento), IL_4 (Nivel de institución), IL_7 (Relación laboral), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor información laboral de funcionario según nivel de institución**.

- **Rotulación de la quinta componente principal**

Los coeficientes de la quinta componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IP_5 (Provincia que habita), IL_1 (Institución), IL_6 (Zona de institución), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor institución que labora y zona de institución según provincia que habita de funcionario**.

- **Rotulación de la sexta componente principal**

Los coeficientes de la sexta componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias IP_3 (Sexo) y IL_5 (Sostenimiento de institución), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor sexo de funcionario según sostenimiento de institución**.

- **Rotulación de la séptima componente principal**

Los coeficientes de la séptima componente principal que son mayores a 0.5 es el correspondientes a la variables aleatoria IP_1 (Provincia de nacimiento), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor provincia de nacimiento de funcionario**.

4.6.5 Construcción de componentes principales de las variables correspondientes a planteles

- **Construcción de componentes principales aplicadas a los datos originales de *planteles***

En esta sección se construirán los componentes principales utilizando las variables observables no estandarizadas de la matriz de planteles, los resultados de los componentes principales, las varianzas de los componentes y su porcentaje de explicación con respecto a la varianza total se los obtuvo utilizando el software de aplicaciones estadísticas SPSS (Social Pourpose Statistical System) versión 8.0 con el procedimiento explicado en la sección 4.5.2.

Es necesario determinar si es factible aplicar componentes principales en la matriz de planteles, por lo que se realizó en el software SPSS 8.0 la prueba Barttlet 1950, la cual nos indica si la matriz de covarianzas de profesores no es una matriz diagonal. La prueba de hipótesis es:

$$H_o : \sigma_{ij} = 0, \text{ para } i \neq j$$

$$\text{vs}$$

$$H_1 = \neg H_o$$

Estadístico de prueba: 10 603.735

Grados de libertad: 378

Valor p: 0.000

Por lo tanto no existe evidencia estadística para afirmar que la matriz de correlaciones de planteles es una matriz diagonal, por ende es aplicable la técnica de componentes principales.

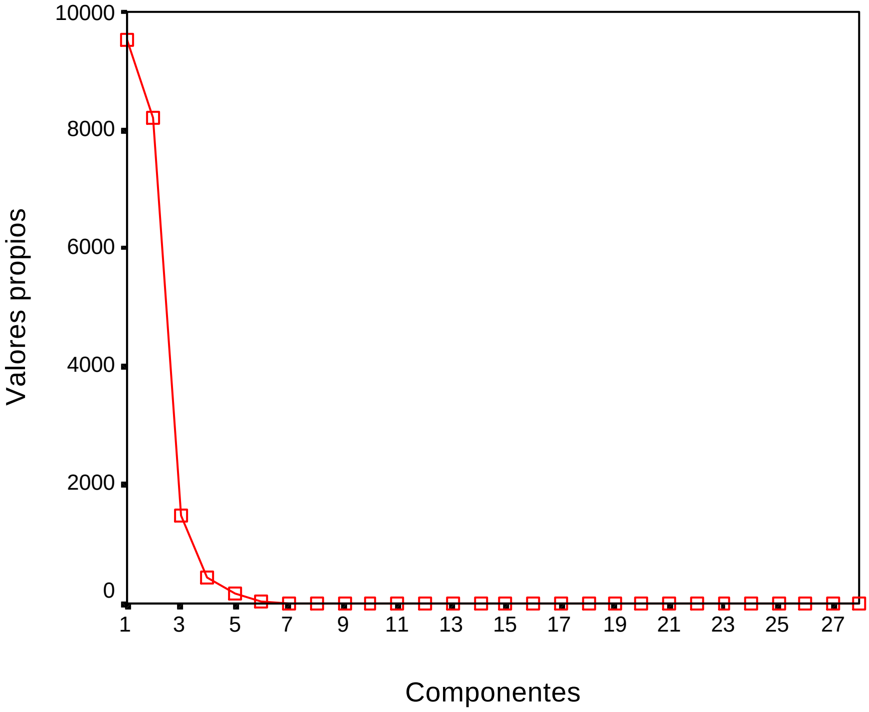
Luego de efectuar este análisis se obtuvo como resultado que el porcentaje de explicación de la varianza de las dos primeras componentes principales de la población es muy alto, en relación a la varianza total de la matriz de covarianzas de planteles.

En el gráfico 4.11 se pueden observar los valores propios obtenidos de la matriz de covarianzas de planteles Σ, con respecto a los 28 componentes principales, en este gráfico se puede apreciar que alrededor del segundo componente principal se forma un codo, el cual indica que los valores propios de ahí en adelante son bastante pequeños (varianzas de los componentes), por lo que no influirán en la proporción de explicación del total de la varianza de la población.

En la tabla CLXXXVII se muestran las varianzas individuales de los dos primeros componentes principales obtenidos a partir de los datos originales de la matriz de datos de planteles, con los que al calcular el porcentaje de explicación del total de la varianzas de la población, de estos componentes se obtiene como resultado 89.281%.

GRÁFICO 4.11

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VALORES PROPIOS DE LAS COMPONENTES PRINCIPALES, DE LOS DATOS ORIGINALES DE LA MATRIZ DE PLANTELES



FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

La elección de los componentes principales construidos a partir de las variables aleatorias no estandarizadas con su respectiva matriz de varianzas y covarianzas se basa en el mayor porcentaje de explicación de estas y considerando la reducción de datos escogemos las dos primeras componentes.

TABLA CLXXXVII

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VARIANZAS Y PORCENTAJE DE EXPLICACIÓN DE LAS 2 PRIMERAS COMPONENTES PRINCIPALES DE LOS DATOS ORIGINALES DE LA MATRIZ DE *PLANTELES*

Componente	**1**	**2**
Varianza	9497.494	8216.691
Porcentaje	47.868	41.413

Varianza Total = 19840.82

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

El porcentaje de explicación de los dos primeros componentes principales se lo obtiene utilizando la siguiente expresión:

$$\frac{\sum_{i=1}^{2} ?_i}{\sum_{i=1}^{28} ?_i} = \frac{17714.18}{19840.82} = 0.89281$$

Este resultado indica que el 89.281% del total de la varianza de la población de planteles es explicada por las dos primeras componentes principales. Los porcentajes de explicación individuales de cada componente se muestran en la tabla CLXXXVII.

Los coeficientes de los dos primeros componentes obtenidos en base a las variables aleatorias observables no estandarizadas de la matriz de datos de planteles, son mostrados en el anexo 28.

- **Construcción de componentes principales aplicadas a las variables estandarizadas de la matriz de planteles**

Utilizando el procedimiento descrito en la sección 4.4.1 para la determinación de los componentes principales de un vector p-variado observable, donde p = 28, se realizó el cálculo sobre la base de las variables estandarizadas, es decir que se utilizó la matriz de correlación de planteles, en vez de la de varianzas y covarianzas.

La estandarización de las variables aleatorias se la realiza cuando, las escalas de los valores que toman las variables de estudio difieren considerablemente de variable a variable.

En la tabla CLXXXVIII se muestran los 9 valores propios de la matriz de correlación ? de planteles, que corresponden a las varianzas de las primeras nueve componentes principales. El porcentaje de explicación del total de la varianza de la población, de las nueve primeras componentes se lo obtiene mediante el cálculo de la siguiente expresión:

$$\frac{\sum_{i=1}^{9} ?_i}{\sum_{i=1}^{28} ?_i} = \frac{20.022}{28.000} = 0.71509$$

Este resultado indica que el 71.509% del total de la varianza de la población planteles es explicada con los nueve primeros componentes principales.

TABLA CLXXXVIII

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VARIANZAS Y PORCENTAJE DE EXPLICACIÓN DE LAS 9 PRIMERAS COMPONENTES PRINCIPALES DE LAS VARIABLES ESTANDARIZADAS DE LA MATRIZ DE *PLANTELES*

Componente	**1**	**2**	**3**
Varianza	6.720	2.920	2.307
Porcentaje	24.000	10.430	8.240
Componente	**4**	**5**	**6**
Varianza	1.878	1.525	1.305
Porcentaje	6.709	5.448	4.661
Componente	**7**	**8**	**9**
Varianza	1.202	1.096	1.068
Porcentaje	4.292	3.916	3.813

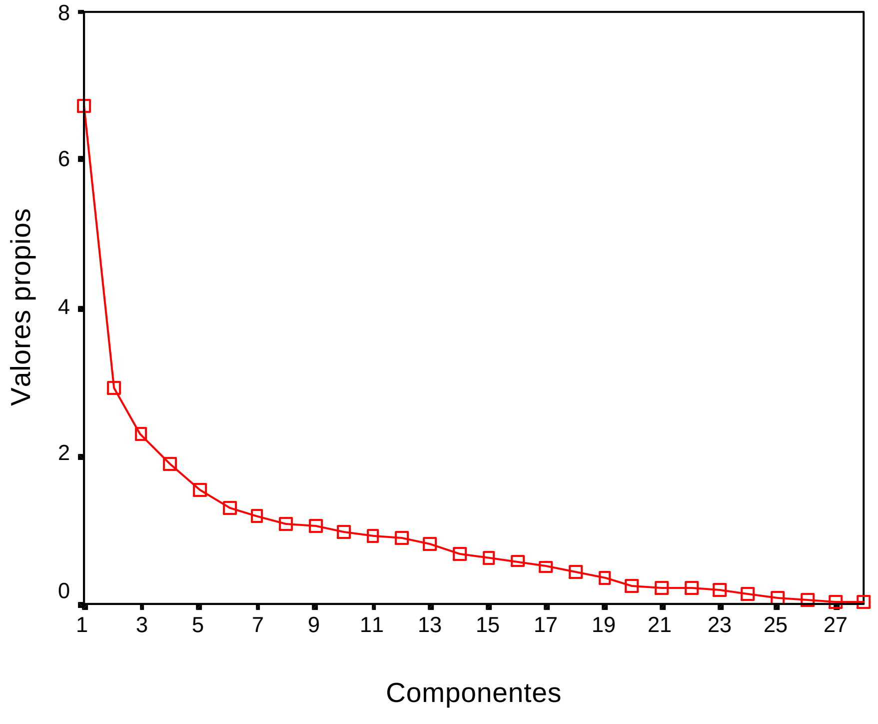
Varianza Total = 28

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

Por otro lado, la determinación del número apropiado de componentes, por medio del gráfico de la magnitud de los valores propios versus el número de componentes, concuerda que, se deben elegir solamente los nueve primeros componentes, debido a que desde λ_{10} hasta λ_{28} todos los valores propios son muy pequeños y están alrededor de un mismo valor. En el gráfico 4.12 se puede observar esta situación, pues alrededor del décimo componente se forma un codo, luego del cual los valores propios son pequeños.

GRÁFICO 4.12

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VALORES PROPIOS DE LAS COMPONENTES PRINCIPALES, DE LOS DATOS ESTANDARIZADOS DE LA MATRIZ DE *PLANTELES*



FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

La elección del número de componentes principales construidos en base a las variables estandarizadas de la matriz de planteles, se la realizó utilizando el criterio del porcentaje de la varianza de la población que estos explican, de acuerdo a esto el número de componentes elegidos es nueve. Los porcentajes de explicación individuales de cada uno de los nueve componentes elegidos se muestran en la tabla CLXXXVIII. Los coeficientes de los nueve componentes principales se muestran en el anexo 29.

Una vez que se calcularon los coeficientes de los vectores propios e_i correspondientes a los valores propios λ_i, para i = 1, 2, ... , k, donde k=9, a continuación se tabulan las componentes principales:

$$\mathbf{Y}_1 = a_1^T X = 0.021 ID_1 - 0.128 ID_2 + \ldots - 0.276\, ID_{28}$$

$$\mathbf{Y}_2 = a_2^T X = 0.117 ID_1 + 0.814 ID_2 + \ldots - 0.491\, ID_{28}$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$\mathbf{Y}_9 = a_9^T X = -0.604\, ID_1 + 0.238 ID_2 + \ldots + 0.009\, ID_{28}$$

Si se estandarizan las variables aleatorias observables, se logra reducir la matriz de datos original de tamaño 537 x 28 a otra de 537 x 9.

Aunque las dos primeras componentes principales obtenidas a partir de las variables aleatorias originales, explican un mayor porcentaje del total de la varianza de la población de los planteles, que las nueve componentes obtenidas en base a las variables aleatorias estandarizadas, las componentes que se utilizarán para remplazar a las variables originales son las nueve primeras componentes principales obtenidos a partir de las variables estandarizadas. La selección se basa en el hecho de que al estandarizar las variables de los planteles, se reduce el efecto de las variables que son medidas en escalas con rangos muy diferentes o si las unidades de medida no son proporcionales.

- **Rotación de las componentes principales obtenidas en base de las variables aleatorias estandarizadas de *planteles***

Como se mencionó en la sección anterior las componentes principales seleccionadas fueron las que se obtuvieron a partir de las variables aleatorias observables estandarizadas de planteles, con esta selección se obtuvo una reducción de datos de una matriz original de tamaño 537 x 28 a una de tamaño 537 x 9.

El siguiente paso que se realizó en este análisis es el de rotar los componentes principales. Las combinaciones lineales geométricamente representan la selección de un nuevo sistema de coordenadas, luego de rotar el sistema original con axisas ID_1, ID_2, ... , ID_{28}, las nuevas coordenadas representan las direcciones con máxima variación de las observaciones, que proveen una simple y detallada descripción de la estructura de la covarianza del vector **X**. Para rotar los componentes se utilizó el método varimax, mediante el cual el sistema de coordenadas obtenido es ortogonal. La rotación se la realiza con el objetivo de interpretar las combinaciones lineales (componentes principales) para obtener una representación real de lo que significan.

TABLA CIXC

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, VARIANZAS Y PORCENTAJE DE EXPLICACIÓN DE LAS 9 PRIMERAS COMPONENTES PRINCIPALES DE LAS VARIABLES ESTANDARIZADAS Y ROTADAS DE LA MATRIZ DE *PLANTELES*

Componente	**1**	**2**	**3**
Varianza	4.719	4.203	2.442
Porcentaje	16.854	15.011	8.721
Componente	**4**	**5**	**6**
Varianza	1.876	1.626	1.401
Porcentaje	6.698	5.807	5.005
Componente	**7**	**8**	**9**
Varianza	1.262	1.252	1.242
Porcentaje	4.507	4.470	4.435

Varianza Total = 28

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

En la tabla CIXC se puede observar los porcentajes individuales de la explicación del total de la varianza de las nueve primeras componentes principales obtenidas a partir de las variables aleatorias estandarizadas y rotadas de la matriz de planteles. Los coeficientes de las nueve combinaciones lineales obtenidas luego de estandarizar y rotar la variables aleatorias originales, se muestran en el anexo 30.

El programa que se utilizó para calcular las componentes principales es el software estadístico SPSS, versión 8.0.

- **Rotulación de las componentes principales aplicadas a las variables aleatorias estandarizadas y rotadas de la matriz de planteles**

La rotulación es el proceso que se realiza para nominar las componentes principales, de tal forma que el nombre corresponda a alguna representación real del componente. Para este efecto se considera la magnitud de los coeficientes que tienen las variables aleatorias observables (ID_1, ID_2, ... , ID_{28}) en cada una de las combinaciones lineales.

Para que los coeficientes de las combinaciones lineales sean considerados como significativos estos deben ser mayores a 0.5, en este caso se utiliza la variable para rotular el componente y si son menores a 0.3 se la omite.

- **Rotulación de la primera componente principal**

  Los coeficientes de la primera componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias Nivel de institución (ID_5), Personal docente del plantel (ID_{14}), Personal administrativo y de servicio del plantel (ID_{15}), Personal con nombramiento del plantel (ID_{17}), Personal con otra relación laboral de la institución (ID_{20}), Alumnos matriculados en ciclo básico de la institución (ID_{23}), Alumnos matriculados en ciclo diversificado de la institución (ID_{24}), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor personal y alumnos según nivel de institución**

- **Rotulación de la segunda componente principal**

  Los coeficientes de la segunda componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias Zona de la institución (ID_7), Vivienda para profesores rurales (ID_{26}), Servicio de agua (ID_{27}), Servicio de luz (ID_{28}), Servicio de alcantarillado (ID_{29}), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor servicios para profesores por zona de institución**

- **Rotulación de la tercera componente principal**

  Los coeficientes de la tercera componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias Otro personal de la institución (ID_{16}), Personal bonificado de la institución (ID_{19}), Alumnos matriculados en ciclo básico de la institución (ID_{23}), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor personal con otro nombramiento por alumnos de ciclo básico de institución.**

- **Rotulación de la cuarta componente principal**

  Los coeficientes de la cuarta componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias Subsistema (ID_1), Modalidad (ID_2), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor sistema educativo de institución**.

- **Rotulación de la quinta componente principal**

  Los coeficientes de la quinta componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias Completitud del plantel (ID_{13}), Alumnos matriculados en primaria de la institución (ID_{22}), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor completitud de institución según alumnos matriculados en primaria.**

- **Rotulación de la sexta componente principal**

Los coeficientes de la sexta componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias Cantón de institución (ID_3), Tipo de plantel (ID_{10}), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor tipo de institución por cantón.**

- **Rotulación de la séptima componente principal**

Los coeficientes de la séptima componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias Sostenimiento de institución (ID_6), Clase de plantel (ID_{12}), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor clase de institución según sostenimiento.**

- **Rotulación de la octava componente principal**

Los coeficientes de la octava componente principal que son mayores a 0.5 son los correspondientes a las variables aleatorias Jornada del plantel (ID_9), Alumnos matriculados en preprimaria de la institución (ID_{21}), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor alumnos de preprimaria según la jornada de institución**

- **Rotulación de la novena componente principal**

Los coeficientes de la novena componente principal que son mayores a 0.5 son los correspondientes a la variable aleatoria Alumnos matriculados en otros niveles educativos de la institución (ID_{25}), por lo tanto el nombre con el que se rotularía esta componente sería el de **factor alumnos matriculados en otros niveles educativos de la institución.**

4.7 Análisis de correlación canónica

Las correlaciones canónicas constituyen una generalización de las correlaciones simples y múltiples. Las correlaciones simples estiman la relación lineal existente entre dos variables aleatorias X y Y. Las correlaciones canónicas estiman la relación lineal existente entre dos conjuntos de variables aleatorias artificiales U_k y V_k las cuales se construyen a partir de dos grupos de variables aleatorias observables.

Sean $\mathbf{X}^{(1)} = (X_1, X_2, \ldots, X_p)$ y $\mathbf{X}^{(2)} = (X_1, X_2, \ldots, X_q)$ dos vectores observables aleatorios p-variado y q-variado respectivamente con p<q, y siendo Σ_{12} la matriz de varianzas y covarianzas de estos vectores, se construyen las variables artificiales U y V de la siguiente manera:

$$U = \mathbf{a}^{\mathbf{T}} \mathbf{X}^{(1)}$$

$$V = \mathbf{b}^{\mathbf{T}} \mathbf{X}^{(2)}$$

Las variables artificiales U y V deben satisfacer las siguientes condiciones:

$$\text{Var}(U) = \mathbf{a}^{T} \text{Cov}(\mathbf{X}^{(1)})\, \mathbf{a} = \mathbf{a}^{T} S_{11}\, \mathbf{a}$$

$$\text{Var}(V) = \mathbf{a}^{T} \text{Cov}(\mathbf{X}^{(2)})\, \mathbf{a} = \mathbf{b}^{T} S_{11}\, \mathbf{b}$$

$$\text{Cov}(U, V) = \mathbf{a}^{T} \text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})\, \mathbf{b} = \mathbf{a}^{T} S_{12}\, \mathbf{b}$$

Los vectores a y b se deben determinar de tal forma que maximicen la correlación entre las variables canónicas U y V.

$$-1 \leq r_{U,V} = \frac{\mathbf{a}^{\mathbf{T}} \Sigma_{12} \mathbf{b}}{\sqrt{\mathbf{a}^{\mathbf{T}} \Sigma_{11} \mathbf{a}} \sqrt{\mathbf{b}^{\mathbf{T}} \Sigma_{22} \mathbf{b}}} \leq 1$$

Entonces en base a lo descrito anteriormente se puede definir a la variables canónicas como sigue:

El primer par de variables canónicas, es el par de combinaciones lineales U_1 y V_1 que tienen varianza igual a uno y que maximizan la correlación Corr (U_1, V_1).

El segundo par de variables canónicas es el par de combinaciones lineales U_2 y V_2 que tienen varianza igual a uno y que maximizan la correlación Corr (U_2, V_2) y que no están correlacionadas con el primer par de variables aleatorias canónicas

El k-ésimo par de variables canónicas es el par de combinaciones lineales U_k y V_k que tienen varianza igual a uno y que maximizan la correlación Corr (U_k, V_k) y que no están correlacionadas con ninguno de los pares anteriores.

4.7.1 Determinación de las variables canónicas

Dados los vectores aleatorios $\mathbf{X}^{(1)} \in R^p$ y $\mathbf{X}^{(2)} \in R^q$, donde $p < q$, entonces se define lo siguiente:

- Las matrices de varianza y covarianza

$$\text{Cov}\ (\mathbf{X}^{(1)}) = S_{11}$$

$$\text{Cov}\ (\mathbf{X}^{(2)}) = S_{22}$$

- Los vectores de media

$$E\ (\mathbf{X}^{(1)}) = m^{(1)}$$

$$E\ (\mathbf{X}^{(2)}) = m^{(2)}$$

- La covarianza entre $\mathbf{X}^{(1)}$ y $\mathbf{X}^{(2)}$

$$\text{Cov}\ (\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = S_{12} = S_{21}^T$$

Siendo los vectores de coeficientes $\mathbf{a} \in R^p$ y $\mathbf{b} \in R^q$ de las combinaciones lineales $U = \mathbf{a}^T \mathbf{X}^{(1)}$ y $V = \mathbf{b}^T \mathbf{X}^{(2)}$ entonces,

$$\max_{a\ b} \text{Corr}\ (U, V) = \rho_1^*$$

a partir de lo cual se construye el primer par de variables canónicas

$$U_1 = \mathbf{e}_1^T \mathbf{S}_{11}^{-1/2}\ \mathbf{X}^{(1)} \text{ y } V_1 = \mathbf{f}_1^T \mathbf{S}_{22}^{-1/2}\ \mathbf{X}^{(2)},$$

El k-ésimo par de variables canónicas para k = 2, 3, ... , p es

$$U_k = \mathbf{e}_k^T \mathbf{S}_{11}^{-1/2}\ \mathbf{X}^{(1)} \text{ y } V_k = \mathbf{f}_k^T \mathbf{S}_{22}^{-1/2}\ \mathbf{X}^{(2)}$$

que maximiza Corr (U_k , V_k) = ρ_k^*, donde las k-ésimas variables canónicas no están correlacionadas con las anteriores k-1 variables canónicas.

El propósito es encontrar un par de vectores U_k y V_k, tal que el coeficiente de correlación entre estos vectores sea el mayor posible para k, l = 1,2, ... ,p. Donde $r_1^{*2} \geq r_2^{*2} \geq \ldots \geq r_p^{*2}$, representan los valores propios de la matriz $S_{11}^{-1/2} S_{12} S_{22}^{-1} S_{21} S_{11}^{-1/2}$ con vectores propios $e_1, e_2, \ldots, e_p$ de tamaño (px1), los cuales son iguales a los valores propios de la matriz $S_{22}^{-1/2} S_{21} S_{11}^{-1} S_{12} S_{22}^{-1/2}$ con vectores propios asociados $f_1, f_2, \ldots, f_p$. de tamaño (qxp).

Donde $\boldsymbol{a}_k^T = \mathbf{e}_k^T \mathbf{S}_{11}^{-1/2}$ y $\mathbf{b}_k^T = \mathbf{f}_k^T \mathbf{S}_{22}^{-1/2}$; los $\boldsymbol{e}_k^T$ y $\boldsymbol{f}_k^T$ son los vectores propios de la matriz $\Sigma \in M_{(p+q)*(p+q)}$ que como se explica posteriormente se forma por la unión de las matrices S_{11}, S_{22}, S_{12} y S_{21}^T, y $S_{11}^{-1/2}$ y $S_{22}^{-1/2}$

son las matrices de varianza y covarianza de cada uno de los grupos de variables observables elevados a la -1/2.

El objetivo de la correlación canónica es medir la fuerza de la relación entre dos conjuntos de variables. Para el análisis es conveniente definir el vector **X** que esta formado por la unión de dos vectores de dos grupos de variables observables consideradas en el análisis como se muestra a continuación:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \cdots \\ \mathbf{X}^{(2)} \end{bmatrix} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_p^{(1)} \\ \cdots \\ X_{p+1}^{(2)} \\ X_{p+2}^{(2)} \\ \vdots \\ X_q^{(2)} \end{bmatrix} \in \Re^{p+q}$$

El vector de medias del vector **X** $\in$ R^{p+q} está formada por el valor esperado de cada uno de los vectores componentes como se muestra a continuación:

$$\boldsymbol{\mu} = \mathrm{E}[\mathbf{X}] = \begin{bmatrix} \mathrm{E}\left[\mathbf{X}^{(1)}\right] \\ \cdots \\ \mathrm{E}\left[\mathbf{X}^{(2)}\right] \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \cdots \\ \boldsymbol{\mu}^{(2)} \end{bmatrix} \in \Re^{p+q}$$

La matriz de varianzas y covarianzas del vector **X** se calcula de la siguiente manera:

$$\mathbf{S} = E\left[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^{T}\right] \in M_{(p+q)x(p+q)}$$

La expresión anterior se puede expresar como la unión de las matrices de varianzas y covarianzas de los vectores $\mathbf{X}^{(1)}$ y $\mathbf{X}^{(2)}$ respectivamente y la matriz de covarianzas entre estos mismos vectores como se muestra a continuación:

$$\Sigma = \begin{bmatrix} E\left[\left(X^{(1)}-m^{(1)}\right)\left(X^{(1)}-m^{(1)}\right)^{T}\right] & \vdots & E\left[\left(X^{(1)}-m^{(1)}\right)\left(X^{(2)}-m^{(2)}\right)^{T}\right] \\ \cdots\cdots\cdots\cdots\cdots\cdots & \vdots & \cdots\cdots\cdots\cdots\cdots\cdots \\ E\left[\left(X^{(2)}-m^{(2)}\right)\left(X^{(1)}-m^{(1)}\right)^{T}\right] & \vdots & E\left[\left(X^{(2)}-m^{(2)}\right)\left(X^{(1)}-m^{(1)}\right)^{T}\right] \end{bmatrix}$$

La matriz $S \in M_{(p+q)\ x(\ p+q)}$ es simétrica y definida positiva, se puede abreviar de la siguiente manera:

$$\mathbf{S} = \left[\begin{array}{c|c} \underset{(pxp)}{\Sigma_{11}} & \underset{(pxq)}{\Sigma_{12}} \\ \hline \underset{(qxp)}{\Sigma_{21}} & \underset{(qxq)}{\Sigma_{22}} \end{array}\right] \in M_{(p+q)x(p+q)}$$

4.7.2 Correlación canónica para las variables de la matriz de *directores y rectores*

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se los aplicará a los conjuntos de variables de la matriz de *directores y rectores*, denominando a estos conjuntos vectores.

Definiremos el vector $\mathbf{X}^{(1)} \in R^p$ a las variables que corresponden a la sección I de la boleta censal, información personal del director o rector, cuyas variables son: IP_1 (Provincia de nacimiento), IP_2 (Edad), IP_3 (Sexo), IP_4 (Estado civil), IP_5 (Provincia que habita), IP_6 (Cantón que habita), donde p=6.

Luego definimos al vector $\mathbf{X}^{(2)} \in R^q$ a las variables que corresponden a la sección II de la boleta censal, instrucción y experiencia del director o rector, siendo las variables investigadas durante el censo de funcionarios públicos del Ministerio de Educación y Cultura las siguientes: IE_1 (Último nivel de instrucción), IE_2 (Título docente), IE_3 (título no docente), IE_4 (Clase de título), IE_5 (Tipo de nombramiento), IE_6 (Años de experiencia), IE_7 (Categoría nominal), IE_8 (Categoría económica), donde q=8.

Y por último definimos al vector $\mathbf{X}^{(3)} \in R^n$, donde tendremos a las variables aleatorias de la sección III, teniendo en el listado a las variables

IL_2 (Cantón de institución), IL_4 (Nivel de institución), IL_5 (Sostenimiento de institución), IL_6 (Zona de institución), IL_7 (Relación laboral), IL_8 (Vivienda rural), donde n=6; las cuales corresponden a la información laboral de directores y rectores.

Se obtendrán las correlaciones canónicas y los coeficientes de las variables canónicas realizan las combinaciones de dos en tres con los vectores $\mathbf{X}^{(1)} \in R^p$, $\mathbf{X}^{(2)} \in R^q$, y $\mathbf{X}^{(3)} \in R^n$.

Estos resultados se los obtuvo utilizando el software de aplicaciones estadísticas SPSS (Social Pourpose Statistical System) versión 8.0. Como esta versión del software no contiene este análisis dentro de su menú, el procedimiento para realizar análisis de correlación canónica es el siguiente se abre la matriz de datos, se va al menú principal, se abre la ventana de New y se entra a la ventana Syntax en la que se realiza la siguiente función:

INCLUDE 'Canonical Correlation.sps'

CANCORR SET 1 = Nombres de las variables del (ó códigos de las variables) /Set 2 = Nombres de las variables de j esimo supravector (ó códigos de las variables).

Luego desde el menú inicial se corre esta función con la que se obtiene los coeficientes de las variables canónicas y los coeficientes de correlación canónica.

Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... ,i, donde i representa la menor dimensión de los vectores, los cuales corresponden a cada una de las secciones de la boleta censal. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

- **Correlación canónica para los conjuntos de variables de secciones información personal, e instrucción y experiencia, de la matriz de *directores y rectores***

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se aplicó a los conjuntos de variables de la sección información personal y la sección instrucción y experiencia de la matriz de directores y rectores. Siendo el vector $\mathbf{X}^{(1)} \in R^p$ el que corresponde a las variables de información personal, donde p=6 y el vector $\mathbf{X}^{(2)} \in R^q$ el que corresponde a las variables de instrucción y experiencia, donde q=8.

Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... ,6 son mostradas en la tabla CXC. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

TABLA CXC

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, COEFICIENTES DE CORRELACIONES CANÓNICAS DE LOS CONJUNTOS PERSONAL Y EXPERIENCIA DE *DIRECTORES Y RECTORES*

1	0.815	**4**	0.098
2	0.193	**5**	0.078
3	0.109	**6**	0.049

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

El primer par de variables canónicas es U_1 y V_1

Si se analizan los coeficientes (de las variables observables del conjunto de la sección información personal y la sección instrucción y experiencia de la matriz de directores y rectores) canónicos asociados a la primera correlación canónica resulta que IP_2 (Edad)=0.992, y IE_6 (Años de experiencia)=0.601 son las variables que más aportación ofrecen a la correlación existente ρ^*_1 =0.815. Tal que la Var (U_1) = 1, Var (V_1) = 1 y la Cov (U_1 , V_1) = 0.815.

El segundo par de variables canónicas es U_2 y V_2

Los coeficientes del segundo par de variables canónicas de la sección información personal y la sección experiencia y instrucción más significativos son IP_3 (Sexo)=0.879, IE_5 (Tipo de nombramiento)= -0.599, y IE_8 (Categoría económica)= -0.564. Estas son las variables que más aportación ofrecen a la correlación existente entre U_2 y V_2, donde ρ^*_2 =0.201. Tal que la Var (U_2) = 1, Var (V_2) = 1 y la Cov (U_2 , V_2) = 0.193.

- **Correlación canónica para los conjuntos de variables de secciones información personal, e información laboral de la matriz de *directores y rectores***

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se aplicó a los conjuntos de variables de la sección información personal y la sección información laboral de la matriz de directores y rectores. Siendo el vector $\mathbf{X}^{(1)} \in R^p$ el que corresponde a las variables de información personal, donde p=6 y el vector $\mathbf{X}^{(3)} \in R^n$ el que corresponde a las variables de información laboral, donde n=6.

Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... ,6 son mostradas en la tabla CXCI. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

TABLA CXCI

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, COEFICIENTES DE CORRELACIONES CANÓNICAS DE LOS CONJUNTOS PERSONAL Y LABORAL DE *DIRECTORES Y RECTORES*

1	0.701	**4**	0.198
2	0.392	**5**	0.051
3	0.294	**6**	0.023

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

El primer par de variables canónicas es U_1 y V_1

Si se analizan los coeficientes (de las variables observables del conjunto de la sección información personal y la sección información laboral de la matriz de directores y rectores) canónicos asociados a la primera correlación canónica resulta IP_6 (Cantón que habita)= 1.085 y IL_2 (Cantón de institución)= 1.003, son las variables que más aportación ofrecen a la correlación existente ρ^*_1 =0.701. Tal que la Var (U_1) = 1, Var (V_1) = 1 y la Cov (U_1 , V_1) = 0.701.

El segundo par de variables canónicas es U_2 y V_2

Los coeficientes del segundo par de variables canónicas de la sección instrucción y experiencia y la sección información laboral más significativos son $\mathbb{L}_6$ (Zona de institución)= -1.049, y $\mathbb{P}_2$ (Edad)=-0.915. Estas son las variables que más aportación ofrecen a la correlación existente entre U_2 y V_2, donde ρ^*_2 =0.392. Tal que la Var (U_2) = 1, Var (V_2) = 1 y la Cov (U_2 , V_2) = 0.392.

- **Correlación canónica para los conjuntos de variables de secciones instrucción y experiencia, e información laboral de la matriz de directores y rectores**

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se aplicó a los conjuntos de variables de la sección instrucción y experiencia y la sección información laboral de la matriz de directores y rectores. Siendo el vector $\mathbf{X}^{(2)} \in R^q$ el que corresponde a las variables de instrucción y experiencia, donde q=8 y el vector $\mathbf{X}^{(3)} \in R^n$ el que corresponde a las variables de información laboral, donde n=6.

TABLA CXCII

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, COEFICIENTES DE CORRELACIONES CANÓNICAS DE LOS CONJUNTOS EXPERIENCIA Y LABORAL DE *DIRECTORES Y RECTORES*

1	0.816	**4**	0.170
2	0.420	**5**	0.113
3	0.313	**6**	0.036

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... ,6 son mostradas en la tabla CXCII. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

El primer par de variables canónicas es U_1 y V_1

Si se analizan los coeficientes (de las variables observables del conjunto de la sección instrucción y experiencia y la sección información laboral de la matriz de directores y rectores) canónicos asociados a la primera correlación canónica resulta que IL_7 (Relación laboral)=0.983 y IE_5 (Tipo de nombramiento)=-1.010, son las variables que más aportación ofrecen a la correlación existente ρ^*_1 =0.816. Tal que la Var (U_1) = 1, Var (V_1) = 1 y la Cov (U_1 , V_1) = 0.816.

El segundo par de variables canónicas es U_2 y V_2

Los coeficientes del segundo par de variables canónicas de la sección instrucción y experiencia y la sección información laboral más significativos son IL_6 (Zona de institución)= 0.843, IE_1 (Último nivel de instrucción)= -0.598, IE_6 (Años de experiencia)= -0.606. Estas son las variables que más aportación ofrecen a la correlación existente entre U_2

y V_2, donde ρ^*_2 =0.420. Tal que la Var (U_2) = 1, Var (V_2) = 1 y la Cov (U_2, V_2) = 0.420.

4.7.3 Correlación canónica para las variables de la matriz de *profesores*

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se los aplicará a los conjuntos de variables de la matriz de profesores, denominando a estos conjuntos vectores.

Definiremos el vector $\mathbf{X}^{(1)} \in R^p$ a las variables que corresponden a la sección I de la boleta censal, información personal del profesor, cuyas variables son: IP_1 (Provincia de nacimiento), IP_2 (Edad), IP_3 (Sexo), IP_4 (Estado civil), IP_5 (Provincia que habita), IP_6 (Cantón que habita), donde p=6.

Luego definimos al vector $\mathbf{X}^{(2)} \in R^q$ a las variables que corresponden a la sección II de la boleta censal, instrucción y experiencia del profesor, siendo las variables investigadas durante el censo de funcionarios públicos del Ministerio de Educación y Cultura las siguientes: IE_1 (Último nivel de instrucción), IE_2 (Título docente), IE_3 (título no docente), IE_4 (Clase de título), IE_5 (Tipo de nombramiento), IE_6 (Años de experiencia), IE_7 (Categoría nominal), IE_8 (Categoría económica), donde q=8.

Y por último definimos al vector $\mathbf{X}^{(3)} \in R^n$, donde tendremos a las variables aleatorias de la sección III, teniendo en el listado a las variables IL_1 (Institución), IL_2 (Cantón de institución), IL_4 (Nivel de institución), IL_5 (Sostenimiento de institución), IL_6 (Zona de institución), IL_7 (Relación laboral), IL_8 (Vivienda rural), donde n=7; las cuales corresponden a la información laboral de profesores.

Se obtendrán las correlaciones canónicas y los coeficientes de las variables canónicas realizan las combinaciones de dos en tres con los vectores $\mathbf{X}^{(1)} \in R^p$, $\mathbf{X}^{(2)} \in R^q$, y $\mathbf{X}^{(3)} \in R^n$.

Estos resultados se los obtuvo utilizando el software de aplicaciones estadísticas SPSS (Social Pourpose Statistical System) versión 8.0. Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k, V_k para k = 1, 2, ... ,i, donde i representa la menor dimensión de los vectores, los cuales corresponden a cada una de las secciones de la boleta censal. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

- **Correlación canónica para los conjuntos de variables de secciones información personal, e instrucción y experiencia, de la matriz de *profesores***

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se aplicó a los conjuntos de variables de la sección información personal y la sección instrucción y experiencia de la matriz de profesores. Siendo el vector $\mathbf{X}^{(1)} \in R^p$ el que corresponde a las variables de información personal, donde p=6 y el vector $\mathbf{X}^{(2)} \in R^q$ el que corresponde a las variables de instrucción y experiencia, donde q=8.

Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... 6 son mostradas en la tabla CXCIII. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

TABLA CXCIII

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, COEFICIENTES DE CORRELACIONES CANÓNICAS DE LOS CONJUNTOS PERSONAL Y EXPERIENCIA DE *PROFESORES*

1	0.731	**4**	0.092
2	0.235	**5**	0.046
3	0.143	**6**	0.023

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

El primer par de variables canónicas es U_1 y V_1

Si se analizan los coeficientes (de las variables observables del conjunto de la sección información personal y la sección instrucción y experiencia de la matriz de profesores) canónicos asociados a la primera correlación canónica resulta que IP_2 (Edad)= -0.978, IE_6 (Años de experiencia)= -0.643, y IE_8 (Categoría económica)= -0.555, son las variables que más aportación ofrecen a la correlación existente ρ^*_1 =0.731. Tal que la Var (U_1) = 1, Var (V_1) = 1 y la Cov (U_1 , V_1) = 0.731.

El segundo par de variables canónicas es U_2 y V_2

Los coeficientes del segundo par de variables canónicas de la sección información personal y la sección información laboral más significativos son IP_3 (Sexo)=0.912, IE_1 (Último nivel de instrucción)= -0.726, IE_4 (Clase de título)= 0.525, IE_5 (Tipo de nombramiento)= -0.933 y IE_8 (Categoría económica)= 0.642. Estas son las variables que más aportación ofrecen a la correlación existente entre U_2 y V_2, donde ρ^*_2 =0.235. Tal que la Var (U_2) = 1, Var (V_2) = 1 y la Cov (U_2 , V_2) = 0.235.

- **Correlación canónica para los conjuntos de variables de secciones información personal, e información laboral de la matriz de *profesores***

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se aplicó a los conjuntos de variables de la sección información personal y la sección información laboral de la matriz de profesores. Siendo el vector $\mathbf{X}^{(1)} \in R^p$ el que corresponde a las variables de información personal, donde p=6 y el vector $\mathbf{X}^{(3)} \in R^n$ el que corresponde a las variables de información laboral, donde n=7.

Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... ,6 son mostradas en la tabla CXCIV. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

TABLA CXCIV

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, COEFICIENTES DE CORRELACIONES CANÓNICAS DE LOS CONJUNTOS PERSONAL Y LABORAL DE *PROFESORES*

1	0.834	**4**	0.113
2	0.437	**5**	0.025
3	0.249	**6**	0.018

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

El primer par de variables canónicas es U_1 y V_1

Si se analizan los coeficientes (de las variables observables del conjunto de la sección información personal y la sección información laboral de la matriz de profesores) canónicos asociados a la primera correlación canónica resulta IP_6 (Cantón que habita)= -1.090 y IL_2 (Cantón de institución)= -0.994, son las variables que más aportación ofrecen a la correlación existente ρ^*_1 =0.834. Tal que la Var (U_1) = 1, Var (V_1) = 1 y la Cov (U_1 , V_1) = 0.834.

El segundo par de variables canónicas es U_2 y V_2

Los coeficientes del segundo par de variables canónicas de la sección instrucción y experiencia y la sección información personal más significativos son IL_7 (Relación laboral)= 0.834, IP_2 (Edad)= 0.834. Estas son las variables que más aportación ofrecen a la correlación existente entre U_2 y V_2, donde ρ^*_2 =0.437. Tal que la Var (U_2) = 1, Var (V_2) = 1 y la Cov (U_2 , V_2) = 0.437.

- **Correlación canónica para los conjuntos de variables de secciones instrucción y experiencia, e información laboral de la matriz de *profesores***

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se aplicó a los conjuntos de variables de la sección instrucción y experiencia y la sección información laboral de la matriz de profesores. Siendo el vector $\mathbf{X}^{(2)} \in R^q$ el que corresponde a las variables de instrucción y experiencia, donde q=8 y el vector $\mathbf{X}^{(3)} \in R^n$ el que corresponde a las variables de información laboral, donde n=7.

Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... ,7 son mostradas en la tabla CXCV. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

TABLA CXCV

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, COEFICIENTES DE CORRELACIONES CANÓNICAS DE LOS CONJUNTOS EXPERIENCIA Y LABORAL DE *PROFESORES*

1	0.935	**5**	0.054
2	0.426	**6**	0.030
3	0.310	**7**	0.013
4	0.096		

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

El primer par de variables canónicas es U_1 y V_1

Si se analizan los coeficientes (de las variables observables del conjunto de la sección instrucción y experiencia y la sección información laboral de la matriz de profesores) canónicos asociados a la primera correlación canónica resulta que $\mathbb{IL}_7$ (Relación laboral)=0.96 y IE_5 (Tipo de nombramiento)= -0.824, son las variables que más aportación ofrecen a la correlación existente ρ^*_1 =0.935. Tal que la Var (U_1) = 1, Var (V_1) = 1 y la Cov (U_1 , V_1) = 0.935.

El segundo par de variables canónicas es U_2 y V_2

Los coeficientes del segundo par de variables canónicas de la sección instrucción y experiencia y la sección información laboral más significativos son $\mathbb{IL}_4$ (Nivel de institución)= -0.999, $\mathbb{IE}_1$ (Último nivel de instrucción)= -1.175, estas son las variables que más aportación ofrecen a la correlación existente entre U_2 y V_2, donde ρ^*_2 =0.426. Tal que la Var (U_2) = 1, Var (V_2) = 1 y la Cov (U_2 , V_2) = 0.426.

4.7.4 Correlación canónica para las variables de la matriz de *otros funcionarios*

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se los aplicará a los conjuntos de variables de la matriz de otros funcionarios, denominando a estos conjuntos vectores.

Definiremos el vector $\mathbf{X}^{(1)} \in R^p$ a las variables que corresponden a la sección I de la boleta censal, información personal del funcionario, cuyas variables son: IP_1 (Provincia de nacimiento), IP_2 (Edad), IP_3 (Sexo), IP_4 (Estado civil), IP_5 (Provincia que habita), IP_6 (Cantón que habita), donde p=6.

Luego definimos al vector $\mathbf{X}^{(2)} \in R^q$ a las variables que corresponden a la sección II de la boleta censal, instrucción y experiencia del funcionario, siendo las variables investigadas durante el censo de funcionarios públicos del Ministerio de Educación y Cultura las siguientes: IE_1 (Último nivel de instrucción), IE_2 (Título docente), IE_3 (título no docente), IE_4 (Clase de título), IE_5 (Tipo de nombramiento), IE_6 (Años de experiencia), donde q=6.

Y por último definimos al vector $\mathbf{X}^{(3)} \in R^n$, donde tendremos a las variables aleatorias de la sección III, teniendo en el listado a las variables IL_1 (Institución), IL_2 (Cantón de institución), IL_4 (Nivel de institución), IL_5 (Sostenimiento de institución), IL_6 (Zona de institución), IL_7 (Relación laboral), donde n=6; las cuales corresponden a la información laboral de los funcionarios.

Se obtendrán las correlaciones canónicas y los coeficientes de las variables canónicas realizan las combinaciones de dos en tres con los vectores $\mathbf{X}^{(1)} \in R^p$, $\mathbf{X}^{(2)} \in R^q$, y $\mathbf{X}^{(3)} \in R^n$.

Estos resultados se los obtuvo utilizando el software de aplicaciones estadísticas SPSS (Social Pourpose Statistical System) versión 8.0. Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... ,i, donde i representa la menor dimensión de los vectores, los cuales corresponden a cada una de las secciones de la boleta censal. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

- **Correlación canónica para los conjuntos de variables de secciones información personal, e instrucción y experiencia, de la matriz de *otros funcionarios***

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se aplicó a los conjuntos de variables de la sección información personal y la sección instrucción y experiencia de la matriz de otros funcionarios. Siendo el vector $\mathbf{X}^{(1)} \in R^p$ el que corresponde a las variables de información personal, donde p=6 y el vector $\mathbf{X}^{(2)} \in R^q$ el que corresponde a las variables de instrucción y experiencia, donde q=6.

Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... ,6 son mostradas en la tabla CXCVI. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

TABLA CXCVI

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, COEFICIENTES DE CORRELACIONES CANÓNICAS DE LOS CONJUNTOS PERSONAL Y EXPERIENCIA DE *OTROS FUNCIONARIOS*

1	0.682	**4**	0.104
2	0.181	**5**	0.056
3	0.112	**6**	0.010

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

El primer par de variables canónicas es U_1 y V_1

Si se analizan los coeficientes (de las variables observables del conjunto de la sección información personal y la sección instrucción y experiencia de la matriz de otros funcionarios) canónicos asociados a la primera correlación canónica resulta que IP_2 (Edad)=0.961, y IE_6 (Años de experiencia)=0.751, son las variables que más aportación ofrecen a la correlación existente ρ^*_1 =0.682. Tal que la Var (U_1) = 1, Var (V_1) = 1 y la Cov (U_1 , V_1) = 0.682.

El segundo par de variables canónicas es U_2 y V_2

Los coeficientes del segundo par de variables canónicas de la sección información personal y la sección información laboral más significativos son IP_1 (Provincia de nacimiento)= 0.894, IE_1 (Último nivel de instrucción)= -1.346, IE_2 (Título docente)= 0.568, IE_5 (Tipo de nombramiento)= 0.688, IE_6 (Años de experiencia)= -0.663. Estas son las variables que más aportación ofrecen a la correlación existente entre U_2

y V_2, donde ρ^*_2 =0.181. Tal que la Var (U_2) = 1, Var (V_2) = 1 y la Cov (U_2, V_2) = 0.181.

- **Correlación canónica para los conjuntos de variables de secciones información personal, e información laboral de la matriz de *otros funcionarios***

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se aplicó a los conjuntos de variables de la sección información personal y la sección información laboral de la matriz de otros funcionarios. Siendo el vector $\mathbf{X}^{(1)} \in R^p$ el que corresponde a las variables de información personal, donde p=6 y el vector $\mathbf{X}^{(3)} \in R^n$ el que corresponde a las variables de información laboral, donde n=6.

TABLA CXCVII

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, COEFICIENTES DE CORRELACIONES CANÓNICAS DE LOS CONJUNTOS PERSONAL Y LABORAL DE *OTROS FUNCIONARIOS*

1	0.834	**4**	0.113
2	0.391	**5**	0.06
3	0.183	**6**	0.021

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... ,6 son mostradas en la tabla CXCVII. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

El primer par de variables canónicas es U_1 y V_1

Si se analizan los coeficientes (de las variables observables del conjunto de la sección información personal y la sección información laboral de la matriz de otros funcionarios) canónicos asociados a la primera correlación canónica resulta IP_6 (Cantón que habita)= -1.045, IL_2 (Cantón de institución)= -1.018, son las variables que más aportación ofrecen a la correlación existente ρ^*_1 =0.834. Tal que la Var (U_1) = 1, Var (V_1) = 1 y la Cov (U_1 , V_1) = 0.834.

El segundo par de variables canónicas es U_2 y V_2

Los coeficientes del segundo par de variables canónicas de la sección instrucción y experiencia y la sección información laboral más significativos son IP_2 (Edad)= -0.787 y IL_7 (Relación laboral)= 0.728. Estas son las variables que más aportación ofrecen a la correlación existente entre U_2 y V_2, donde ρ^*_2 =0.391. Tal que la Var (U_2) = 1, Var (V_2) = 1 y la Cov (U_2 , V_2) = 0.391.

- **Correlación canónica para los conjuntos de variables de secciones instrucción y experiencia, e información laboral de la matriz de *otros funcionarios***

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se aplicó a los conjuntos de variables de la sección instrucción y experiencia y la sección información laboral de la matriz de otros funcionarios. Siendo el vector $\mathbf{X}^{(2)} \in R^q$ el que corresponde a las variables de instrucción y experiencia, donde q=6 y el vector $\mathbf{X}^{(3)} \in R^n$ el que corresponde a las variables de información laboral, donde n=6.

TABLA CXCVIII

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, COEFICIENTES DE CORRELACIONES CANÓNICAS DE LOSCONJUNTOS EXPERIENCIA Y LABORAL DE *OTROS FUNCIONARIOS*

1	0.668	**4**	0.109
2	0.318	**5**	0.043
3	0.263	**6**	0.032

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... ,6 son mostradas en la tabla CXCVIII. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

El primer par de variables canónicas es U_1 y V_1

Si se analizan los coeficientes (de las variables observables del conjunto de la sección instrucción y experiencia y la sección información laboral de la matriz de otros funcionarios) canónicos asociados a la primera correlación canónica resulta que IL_7 (Relación laboral)=0.929, y IE_5 (Tipo de nombramiento)= -1.158, son las variables que más aportación ofrecen a la correlación existente ρ^*_1 =0.668. Tal que la Var (U_1) = 1, Var (V_1) = 1 y la Cov (U_1 , V_1) = 0.668.

El segundo par de variables canónicas es U_2 y V_2

Los coeficientes del segundo par de variables canónicas de la sección instrucción y experiencia y la sección información laboral más significativos son IL_4 (Nivel de institución)=-0.617, IE_1 (Último nivel de instrucción)=-1.031, estas son las variables que más aportación ofrecen a la correlación existente entre U_2 y V_2, donde ρ^*_2 =0.318. Tal que la Var (U_2) = 1, Var (V_2) = 1 y la Cov (U_2 , V_2) = 0.318.

4.7.5 Correlación canónica para las variables de la matriz de *planteles*

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se los aplicará a los conjuntos de variables de la matriz de planteles, denominando a estos conjuntos vectores. Esta sección fue llenada exclusivamente por los directores, constituyéndose en la sección IV de la boleta censal, la cual recopilaba información del plantel que estaba en concordancia con el formulario FR1, el cual debía estar entregado y certificado por la Dirección Provincial de Educación. Se procedió a separar la sección IV en los siguientes conjuntos de variables.

Definiremos el vector $\mathbf{X}^{(1)} \in R^p$ a las variables que corresponden a la información del plantel del director o rector, cuyas variables son: Subsistema (ID_1), Modalidad (ID_2), Cantón de institución (ID_3) ,Nivel de institución (ID_5), Sostenimiento de institución (ID_6), Zona de la institución (ID_7), Régimen del plantel (ID_8), Jornada del plantel (ID_9) , Tipo de plantel (ID_{10}), Genero de plantel (ID_{11}), Clase de plantel (ID_{12}), Completitud del plantel (ID_{13}), donde p=12.

Luego definimos al vector $\mathbf{X}^{(2)} \in R^q$ a las variables que corresponden a las preguntas de la boleta censal que contestaron los directores y rectores durante el primer censo de funcionarios públicos del Ministerio de Educación y Cultura, que se realizó el 14 de diciembre del 2000, con respecto al personal que labora en las instituciones del sistema educativo nacional en la provincia de Bolívar, siendo las variables investigadas las siguientes: Personal docente del plantel (ID_{14}), Personal administrativo y de servicio del plantel (ID_{15}), Otro personal de la institución (ID_{16}), Personal con nombramiento del plantel (ID_{17}), Personal con contratos de la institución (ID_{18}), Personal bonificado de la institución (ID_{19}), Personal con otra relación laboral de la institución (ID_{20}), donde q=7.

Y por último definimos al vector $\mathbf{X}^{(3)} \in R^n$, donde tendremos a las variables aleatorias relacionadas con el alumnado por sección de la institución, teniendo en el listado a las variables Alumnos matriculados en preprimaria de la institución (ID_{21}), Alumnos matriculados en primaria de la institución (ID_{22}), Alumnos matriculados en ciclo básico de la institución (ID_{23}), Alumnos matriculados en ciclo diversificado de la institución (ID_{24}), Alumnos matriculados en otros niveles educativos de la institución (ID_{25}), donde n=5.

Se descarto crear una cuarta sección con relación a la información de los servicios en planteles rurales, en vista de que hay un alto porcentaje de no respuesta, ya sea porque los directores o rectores no informaron las condiciones de su plantel, o eran no debían responder esas preguntas por ser directores o rectores de planteles urbanos, lo cual podría haber sesgado las variables canónicas y sus correlaciones. Es importante anotar que durante el análisis univariado se demostró la falencia de servicios básicos en los planteles rurales.

Se obtendrán las correlaciones canónicas y los coeficientes de las variables canónicas realizan las combinaciones de dos en tres con los vectores $\mathbf{X}^{(1)} \in R^p$, $\mathbf{X}^{(2)} \in R^q$, y $\mathbf{X}^{(3)} \in R^n$.

Estos resultados se los obtuvo utilizando el software de aplicaciones estadísticas SPSS (Social Pourpose Statistical System) versión 8.0. Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... ,i, donde i representa la menor dimensión de los vectores, los cuales corresponden a cada una de las secciones de la boleta censal. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

- **Correlación canónica para los conjuntos de variables de secciones personal, y características de la matriz de *planteles***

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se aplicó a los conjuntos de variables de la sección personal del plantel y la sección características del plantel. Siendo el vector $\mathbf{X}^{(2)} \in R^q$ el que corresponde a las variables de personal del plantel, donde q=7 y el vector $\mathbf{X}^{(1)} \in R^p$ el que corresponde a las variables de características del plantel, donde p=12.

TABLA CIC

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, COEFICIENTES DE CORRELACIONES CANÓNICAS DE LOSCONJUNTOS CARACTERISTICAS Y PERSONAL DE *PLANTELES*

1	0.723	**5**	0.199
2	0.544	**6**	0.167
3	0.319	**7**	0.102
4	0.284		

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... ,7 son mostradas en la tabla CIC. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

El primer par de variables canónicas es U_1 y V_1

Si se analizan los coeficientes (de las variables observables del conjunto de la sección personal del plantel y la sección características del plantel) canónicos asociados a la primera correlación canónica resulta que Personal administrativo y de servicio del plantel (ID_{15})=0.97, Otro personal de la institución (ID_{16})=0.574, Nivel de institución (ID_5)=0.581 son las variables que más aportación ofrecen a la correlación existente ρ^*_1 =0.723. Tal que la Var (U_1) = 1, Var (V_1) = 1 y la Cov (U_1 , V_1) = 0.723.

El segundo par de variables canónicas es U_2 y V_2

Los coeficientes del segundo par de variables canónicas de la sección personal del plantel y la sección características del plantel más significativos son Personal docente del plantel (ID_{14})=2.211, Personal administrativo y de servicio del plantel (ID_{15})=-1.273, Personal bonificado de la institución (ID_{19})=1.294, Zona de la institución (ID_7)=-0.756, Completitud del plantel (ID_{13})=0.551 estas son las variables que más aportación ofrecen a la correlación existente entre U_2 y V_2, donde ρ^*_2 =0.544. Tal que la Var (U_2) = 1, Var (V_2) = 1 y la Cov (U_2 , V_2) = 0.544.

- **Correlación canónica para los conjuntos de variables de secciones alumnos y características de la institución de la matriz de *planteles***

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se aplicó a los conjuntos de variables de la sección alumnos del plantel y la sección características del plantel. Siendo el vector $\mathbf{X}^{(3)} \in R^n$ el que corresponde a las variables de alumnos del plantel, donde n=5 y el vector $\mathbf{X}^{(1)} \in R^p$ el que corresponde a las variables de características del plantel, donde p=12.

TABLA CC

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, COEFICIENTES DE CORRELACIONES CANÓNICAS DE LOSCONJUNTOS CARACTERISTICAS Y ALUMNOS DE *PLANTELES*

1	0.589	**4**	0.201
2	0.526	**5**	0.145
3	0.312		

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... ,5 son mostradas en la tabla CC. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

El primer par de variables canónicas es U_1 y V_1

Si se analizan los coeficientes (de las variables observables del conjunto de la sección alumnos del plantel y la sección características del plantel) canónicos asociados a la primera correlación canónica resulta que Zona de la institución (ID_7)=0.697, y Alumnos matriculados en primaria de la institución (ID_{22})= -0.597, son las variables que más aportación ofrecen a la correlación existente ρ^*_1 =0.589. Tal que la Var (U_1) = 1, Var (V_1) = 1 y la Cov (U_1 , V_1) = 0.589.

El segundo par de variables canónicas es U_2 y V_2

Los coeficientes del segundo par de variables canónicas de la sección alumnos del plantel y la sección características del plantel más significativos son Alumnos matriculados en primaria de la institución (ID_{22})=0.787, Completitud del plantel (ID_{13})=0.69 estas son las variables que más aportación ofrecen a la correlación existente entre U_2 y V_2, donde ρ^*_2 =0.526. Tal que la Var (U_2) = 1, Var (V_2) = 1 y la Cov (U_2 , V_2) = 0.526.

- **Correlación canónica para los conjuntos de variables de secciones alumnos y personal de la institución de la matriz de *planteles***

El procedimiento descrito en la sección 4.7.1 para la determinación de las variables canónicas, se aplicó a los conjuntos de variables de la sección alumnos del *plantel* y la sección personal del *plantel*. Siendo el vector $\mathbf{X}^{(3)} \in R^n$ el que corresponde a las variables de alumnos del *plantel*, donde n=5 y el vector $\mathbf{X}^{(2)} \in R^q$ el que corresponde a las variables de características del *plantel*, donde q=7.

TABLA CCI

PROVINCIA DE BOLÍVAR: CENSO DEL MAGISTERIO NACIONAL, COEFICIENTES DE CORRELACIONES CANÓNICAS DE LOSCONJUNTOS PERSONAL Y ALUMNOS DE *PLANTELES*

1	0.970	**4**	0.275
2	0.926	**5**	0.064
3	0.590		

FUENTE: Base de datos del I Censo de funcionarios públicos del MEC
ELABORACIÓN: por el autor, Carlos Ronquillo Franco

Las correlaciones canónicas ρ_k^* obtenidas entre cada par de las variables canónicas U_k , V_k para k = 1, 2, ... ,5 son mostradas en la tabla CCI. A continuación se analizan los dos primeros pares de variables canónicas, que son los que tienen los mayores coeficientes de correlación canónica.

El primer par de variables canónicas es U_1 y V_1

Si se analizan los coeficientes (de las variables observables del conjunto de la sección alumnos del plantel y la sección personal del plantel) canónicos asociados a la primera correlación canónica resulta que Alumnos matriculados en ciclo básico de la institución (ID_{23})=1.317, Alumnos matriculados en ciclo diversificado de la institución (ID_{24})= -0.670, y Otro personal de la institución (ID_{16})=0.687, son las variables que más aportación ofrecen a la correlación existente ρ^*_1 =0.970. Tal que la Var (U_1) = 1, Var (V_1) = 1 y la Cov (U_1 , V_1) = 0.970.

El segundo par de variables canónicas es U_2 y V_2

Los coeficientes del segundo par de variables canónicas de la sección alumnos del plantel y la sección personal del plantel más significativos son Alumnos matriculados en primaria de la institución (ID_{22})=0.536, Alumnos matriculados en ciclo diversificado de la institución (ID_{24})=

1.158, y Personal docente del plantel (ID_{14})=0.958, estas son las variables que más aportación ofrecen a la correlación existente entre U_2 y V_2, donde ρ^*_2 =0.926. Tal que la Var (U_2) = 1, Var (V_2) = 1 y la Cov (U_2, V_2) = 0.926.