ESTADÍSTICA MULTIVARIABLE ANÁLISIS DE COMPONENTES PRINCIPALES

Dr. Cs. Ruiz Aranibar, Gustavo[1]

ruizaranibargustavo@gmail.com.bo

RESUMEN

El análisis de componentes principales es probablemente la transformación más comúnmente utilizada para realizar varias tareas en muchas aplicaciones, se utiliza generalmente para reducir la dimensionalidad de los datos multivariados de modo que los modelos de rasgos univariados puedan ajustarse a los componentes principales individuales. Convierte un conjunto de observaciones de variables posiblemente correlacionadas en un conjunto de valores de variables sin correlación lineal llamadas componentes principales, cuyos resultados se los debe interpretar.

PALABRAS CLAVE

Ecuación característica, matriz de correlación, valor propio, vector propio, ejes principales, mutivariante, análisis de componentes principales.

1. INTRODUCCIÓN.

En el Análisis de los Componente Principales (ACP), se dispone de datos estadísticos que se deberían representar en un espacio de varias dimensiones, lo que significa que si se tiene un número bastante grande de variables y observaciones; por ejemplo, 10 características económicas (variables) medidas en 50 regiones (observaciones), u 80 caracteres (variables) medidas en 2.000 individuos (observaciones); esta información no se la podría representar. Por lo tanto, se tiene una matriz rectangular de valores numéricos cuyas dimensiones son tales que son un obstáculo para la rápida asimilación por el estadístico de la información contenida en este conjunto de números, lo que se pretende en la presente investigación, es efectuar un estudio de datos estadísticos representándolos en un espacio pequeño.

El objetivo principal que persigue el ACP es la representación de las medidas numéricas de varias variables en un espacio de pocas dimensiones donde nuestros sentidos puedan percibir relaciones, porque de otra manera permanecerían ocultas en dimensiones superiores. Dicha representación debe ser tal que al desechar dimensiones superiores (de cuarta dimensión en adelante) se realiza con la mínima pérdida de información.

El ACP consiste en una transformación lineal de m variables originales en m nuevas variables, donde cada nueva variable es una combinación lineal de los valores originales. El proceso se realiza de una manera que requiere que cada nueva variable explique, sucesivamente, la mayor cantidad posible de las varianzas totales. Cuando se han calculado las nuevas variables de las diferentes muestras, se tendrá en cuenta toda la variante original.

El ACP es una técnica estadística que fue propuesta por Karl Pearson (1901) como parte del análisis de factores y Hotelling

1 Se agradece a la UAGRM por la beca otorgada con fondos del IDH, para cursar y culminar exitosamente el Doctorado en Ciencias en Educación Superior. Especializado en Estadística e Informática.

(1933) quienes introdujeron el concepto de ACP. Sin embargo, la complejidad de los cálculos retrasó su desarrollo y aplicación hasta la aparición de los computadores.

2. DESCRIPCIÓN DEL ANÁLISIS DE COMPONENTES PRINCIPALES.

El ACP es una técnica estadística descriptiva, que tiene como punto de partida una matriz de datos con una serie de individuos, a los que se les ha medido varias variables. Por eso se la clasifica como una técnica multivariante. Dada una matriz de n filas (muestras) y m columnas (variables; $n \geq m$), se puede efectuar un gráfico considerando dos variables para todas las muestras, según sus valores en el eje de abscisas y de ordenadas, también se puede hacer una representación de tres de esas variables en un gráfico tridimensional, pero si el número de variables es mayor a 3, ya no se puede visualizar una representación a m dimensiones, siendo imposible ver en un gráfico una representación respecto a todas las variables al mismo tiempo.

El ACP tiene como objetivo básico inicial suplir este déficit, realizando la representación de una nube de puntos multidimensional (de más de tres dimensiones), en dos o tres dimensiones. En definitiva, se trata de visualizar lo que no se ve. El ACP trata de hacer una representación de las n muestras en dos o tres dimensiones, pero contemplando todas las variables, sin prescindir de ninguna de ellas en el análisis.

Como consecuencia de su propio procedimiento, se consigue crear expresiones matemático-estadísticos (los componentes) en realidad, también podrían considerarse objetivos de la técnica porque en muchas ocasiones permiten establecer relaciones entre las variables, ver cómo se asocian, relacionan, distancian, etc.

3. CONCEPTOS PRELIMINARES.

Se entenderá la esencia de la técnica del ACP de la manera siguiente: si se representa en forma bidimensional dos variables x_1 y x_2 , en un sistema de coordenadas rectangulares X_1 y X_2, luego sin mover los datos, se traza desde el origen un eje Y_1 que se ajuste mejor a los datos donde se proyecta cada punto a este eje, luego un segundo eje Y_2 desde el origen y perpendicular a Y_1, lo que se realizó es una rotación con un ángulo θ, del nuevo sistema de coordenadas, la relación existente entre las variables originales y las nuevas variables son expresadas como una combinación lineal de las variables originales, a esta técnica se la denomina como componentes, observando la expresión de esta combinación de esta forma:

$$y_1 = a_{11}x_1 + a_{12}x_2$$
$$y_2 = a_{21}x_1 + a_{22}x_2$$

Estos coeficientes que multiplican a las variables originales son los vectores propios de la matriz de correlaciones, es la expresión de la transformación lineal realizada. Ahora si en lugar de estar trabajando con dos variables originales, se trabaja con m variables originales la expresión de las m componentes sería:

$$y_1 = a_{11}x_1 + a_{12}x_2 + \ldots\ldots + a_{1m}x_m$$
$$y_2 = a_{21}x_1 + a_{22}x_2 + \ldots\ldots + a_{2m}x_m$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$y_m = a_{m1}x_1 + a_{m2}x_2 + \ldots\ldots + a_{mm}x_m$$

Para mayor comprensión, se tienen tres puntos A, B y C (muestras) y dos variables X_1 y X_2, los cuales pueden representarse en un sistema de coordenadas rectangulares (variables), obteniendo luego los promedios $\overline{X}_1$ y $\overline{X}_2$, para determinar:

$$x_1 = X_1 - \overline{X}_1$$

$$x_1 = X_1 - \overline{X}_2$$

Se rotan con un ángulo Θ, los ejes con coordenadas $\overline{X}_1$ y $\overline{X}_2$, como nuevo origen, proyectándose los valores de los tres puntos sobre Y_1 , de tal forma que la variación a lo largo del nuevo eje Y_1 es grande, pero la variación a lo largo del eje Y_2 es pequeña, porque a lo largo del eje Y_2 es igual a cero. En este caso, el fenómeno indicado por las dos variables (variables explicativas) X_1 y X_2, puede ser explicado por la nueva variable Y_1, cuando se emplea el nuevo sistema de coordenadas Y_1 y Y_2.

Además, si se puede estadísticamente hacer independientes a las variables Y_1 y Y_2, entonces los efectos de Y_1 y Y_2pueden ser diferenciados pudiendo utilizarse esas variables para ulteriores investigaciones.

Conociendo las coordenadas de tres puntos A, B y C, en el nuevo sistema de coordenadas, considerando el ángulo θ, se las determina efectuando operaciones geométricas, algebraicas y trigonométricas, de manera que se encuentra para estos puntos:

Primeras coordenadas en Y_1:

$$\begin{aligned} y_{11} &= x_{11}\cos\theta + x_{12}sen\theta \\ y_{21} &= x_{21}\cos\theta + x_{22}sen\theta \\ y_{31} &= x_{31}\cos\theta + x_{32}sen\theta \end{aligned}$$

Segundas coordenadas en Y_2:

$$\begin{aligned} y_{12} &= -x_{11}sen\theta + x_{12}\cos\theta \\ y_{22} &= -x_{21}sen\theta + x_{22}\cos\theta \\ y_{32} &= -x_{31}sen\theta + x_{32}\cos\theta \end{aligned}$$

Una vez determinada estas nuevas coordenadas de los tres puntos en X_1 y X_2, se tendrá en el nuevo sistema de coordenadas Y_1 y Y_2; trasladando los resultados a notación matricial:

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} = [X_1, X_2]$$

$$\begin{bmatrix} y_{11} \\ y_{21} \\ y_{31} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} \begin{bmatrix} \cos\theta \\ sen\theta \end{bmatrix} = Y_1 \qquad \begin{bmatrix} y_{2} \\ y_{22} \\ y_{32} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} \begin{bmatrix} -sen\theta \\ \cos\theta \end{bmatrix} = Y_2$$

$$Y = [Y_1, Y_2] = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ y_{31} & y_{32} \end{bmatrix}$$

Este cambio de coordenadas expresada en forma matricial para los tres puntos, permite generalizar el análisis para m variables, el cual es utilizado en el programa computacional.

4. FUNDAMENTOS DEL ACP.

La técnica del ACP construye una transformación lineal que escoge un nuevo sistema de coordenadas para el conjunto original de datos, en el cual la varianza de mayor tamaño del conjunto de datos es capturada en el primer eje (llamado el Primer Componente Principal), la segunda varianza más grande es el segundo eje, y así sucesivamente. Para esta transformación lineal debe construirse primero la matriz de varianza-covarianza o la matriz de coeficientes de correlación. Debido a la simetría de estas matrices, existe una base completa de vectores propios de las mismas. La transformación que lleva de las antiguas coordenadas a las coordenadas de la nueva base es precisamente la transformación lineal necesaria para reducir la dimensionalidad de datos. Además, las coordenadas en la nueva base dan la composición en factores subyacentes de los datos iníciales.

Una de las ventajas del ACP para reducir la dimensionalidad de un grupo de datos, es que retiene aquellas características del conjunto de datos que contribuyen más a su varianza, manteniendo un orden de bajo nivel de los componentes principales e ignorando los de alto nivel. El objetivo es que esos componentes de bajo orden a veces contienen el aspecto más importante de esa información.

5. CÁLCULOS Y FÓRMULAS MATEMÁTICO – ESTADÍSTICOS EN EL ACP.

Por definición, un conjunto de datos constituye una muestra aleatoria multivariada, si cada individuo ha sido extraído al azar de una población de n individuos y en él se han medido u observado m características se tiene:

Matriz de datos multivariados:

$$X = \left[x_{ij}\right] = \begin{bmatrix} x_{11} \;\dots\dots\dots\; x_{1m} \\ x_{21} \;\dots\dots\dots\; x_{2m} \\ \dots\dots\dots\dots\dots\dots \\ x_{n1} \;\dots\dots\dots\; x_{nm} \end{bmatrix}$$

$$i = 1,2,3,\dots\dots, n \qquad j = 1,2,3,\dots\dots, m \qquad n \geq m \quad (\text{Si } n < m \text{ se trabaja con la } X_{ij}^{t})$$

A partir de esta matriz, que contiene toda la información estadística de la muestra, es posible calcular las funciones que permitan extraer conclusiones de los datos.

Suma de columnas.

$$X = x_j = \sum_{i=1}^{n} x_{ij}$$

Matriz de suma de cuadrados y productos.

$$C = \left[c_{ij}\right] = \sum_{i=1}^{n} x_{il} * x_{lj}$$

Matriz de suma de cuadrados y suma de productos de desviaciones.

$$A = \left[a_{ii}\right] = SCE_i = \sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i)^2 = \sum_{\alpha=1}^{n} x_{i\alpha}^2 - \frac{1}{n}\left(\sum_{\alpha=1}^{n} x_{1\alpha}\right)^2$$

$$A = \left[a_{ij}\right] = SPE_{ij} = \sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j) = \sum_{\alpha=1}^{n} x_{i\alpha} x_{j\alpha} - \frac{1}{n}\left(\sum_{\alpha=1}^{n} x_{i\alpha}\right)\left(\sum_{\alpha=1}^{n} x_{j\alpha}\right)$$

Matriz de varianza – covarianza o matriz de covarianza.

$$S=\left[s_{ij}\right]=\begin{bmatrix} s_{11}\ \ldots\ldots\ldots\ s_{1m} \\ \ldots\ldots\ldots\ldots \\ s_{m1}\ \ldots\ldots\ldots\ s_{mm}\end{bmatrix}=\begin{bmatrix} a_{11}/n\ \ldots\ldots\ldots\ a_{1m}/n \\ \ldots\ldots\ldots\ldots \\ a_{m1}/n\ \ldots\ldots\ldots\ s_{mm}/n\end{bmatrix}$$

Vector de promedios.

$$\overline{X}=\left[\overline{x}_j\right]=\left(\sum_{i=1}^{n}x_{ij}/n\right)=\begin{bmatrix}\overline{x}_1\\ \overline{x}_2\\ \ldots\\ \overline{x}_m\end{bmatrix}$$

Vector de desviaciones estándar.

$$D=\left[d_j\right]=\left\lfloor\sqrt{a_{jj}}\right\rfloor$$

Matriz estandarizada.

$$X=x_{ij}=\frac{\sum_{j=1}^{n}(x_{ji}-\overline{x}_j)}{d_j}$$

Matriz de coeficientes de correlación.

$$R=\left[r_{ij}\right]=\begin{bmatrix}1\quad r_{12}\ \ldots\ldots\ r_{1m}\\ r_{21}\ 1\ldots\ldots\ldots r_{2m}\\ \ldots\ldots\ldots\ldots\ldots\\ r_{m1}\ r_{m2}\ \ldots\ldots 1\end{bmatrix}$$

Para dos columnas cualesquiera (identificadas por los subíndices), el coeficiente de correlación se determina por:

$$r=\frac{\operatorname{cov}(x,y)}{s_x s_y}=\frac{n\sum XY-\sum X\sum Y}{\sqrt{n\sum X^2-(\sum X)^2}*\sqrt{n\sum Y^2-(\sum Y)^2}}\qquad -1\le r\le +1$$

6. VECTORES Y VALORES PROPIOS.

El ACP se basa en una transformación lineal de las observaciones originales, esta transformación lineal debe satisfacer las exigencias del análisis que se tiene en el campo de álgebra vectorial como generación de vectores y valores propios.

Se procede a encontrar de la matriz de correlación, las raíces características λ (escalar) donde I es una matriz unitaria:

$$R-\lambda I=\begin{bmatrix}1-\lambda & r_{12}\ \ldots\ldots\ldots\ldots r_{1m}\\ r_{21} & 1-\lambda\ldots\ldots\ldots r_{2m}\\ \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots\\ r_{m1} & r_{m2}\ \ldots\ldots\ldots 1-\lambda\end{bmatrix}$$

El determinante de esta matriz es:

$$|R-\lambda I|\qquad\qquad |R-\lambda I|=0$$

$$|R-\lambda I|=\begin{vmatrix}r_{11}-\lambda & r_{12}\ldots\ldots\ldots\ldots r_{1m}\\ r_{21} & r_{22}-\lambda\ldots\ldots r_{2m}\\ \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots\\ r_{m1} & r_{m2}\ldots\ldots r_{mm}-\lambda\end{vmatrix}=0$$

Igualando a cero este determinante, y desarrollando permite encontrar un polinomio de grado m en λ, que se designa por $f(\lambda)$, denominado polinomio característico de la matriz; considerando el teorema fundamental del algebra lineal, que indica que todo polinomio de orden m posee m raíces. El polinomio característico de la matriz R de orden m puede escribirse de una manera general como:

$$f(\lambda) = c_m\lambda^m + c_{n-1}\lambda^{n-1} + \ \ldots\ldots\ c_1\lambda + c_0$$
$$c_m = (-1)^m \qquad c_0 = |A| \qquad c_{m-1} = (-1)^{m-1}tr(A)$$

Donde los coeficientes c_j dependen de los elementos de la matriz R.

La ecuación: $|R - \lambda I| = 0$ como origina un polinomio con respecto a la incógnita λ, el cual puede ser resuelto con respecto a λ, y obtener, entonces los vectores característicos.

Considerando los casos a), b) y c):

a) Teniéndose una matriz R de orden 2, se obtendrá:

$$(r_{11} - \lambda)(r_{22} - \lambda) - r_{12}r_{21} = 0$$

Es decir:

$$\lambda^2 - (r_{11} + r_{22})\lambda + (r_{11}r_{22} - r_{12}r_{21}) = 0$$

Cuyas raíces son:

$$\lambda_1 = \frac{1}{2}\left[(r_{11} + r_{22}) + \sqrt{(r_{11} + r_{22})^2 - 4(r_{11}r_{22} - r_{12}r_{21})}\right]$$
$$\lambda_2 = \frac{1}{2}\left[(r_{11} + r_{22}) - \sqrt{(r_{11} + r_{22})^2 - 4(r_{11}r_{22} - r_{12}r_{21})}\right]$$

En el caso particular de una matriz simétrica, $r_{12} = r_{21}$, las raíces valen:

$$\lambda = \frac{1}{2}\left[(r_{11} + r_{22}) \pm \sqrt{(r_{11} + r_{22})^2 + 4r_{12}^2}\right]$$

Y dado que el radicando es la suma de dos cuadrados, las raíces λ_1 y λ_2 son necesariamente reales para una matriz simétrica.

O también si:

$$R = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}$$

Entonces se tiene:

$$f(\lambda) = \lambda^2 - (r_{11} + r_{22})\lambda + \begin{vmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{vmatrix} = \lambda^2 - (tr\,R)\lambda + |R|$$

Donde tr R denota la traza de R, esto es, la suma de sus elementos diagonales.

b) Teniéndose una matriz R de orden 3, se obtendrá:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{333} \end{bmatrix}$$

Entonces:

$$f(\lambda) = \lambda^3 - (r_{11} + r_{22} + r_{33})\lambda^2 + \left(\begin{vmatrix} r_{22} & r_{23} \\ r_{32} & r_{33} \end{vmatrix} + \begin{vmatrix} r_{11} & r_{13} \\ r_{31} & r_{33} \end{vmatrix} + \begin{vmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{vmatrix}\right)t - \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix}$$
$$f(\lambda) = \lambda^3 - (tr\,R)\lambda^2 + (R_{11} + R_{22} + R_{33})\lambda - |R|$$

Aquí R_{11}, R_{22}, R_{33} denotan, los cofactores de los elementos diagonales r_{11}, r_{22}, r_{33}.

Considerando de nuevo la matriz cuadrada R de orden 3, $R=[r_{ij}]$, como se indicó anteriormente,

$$S_1 = tr\, R \qquad S_2 = R_{11} + R_{22} + R_{33} \qquad S_3 = |R|$$

Son los coeficientes de su polinomio característico, con signos alternantes. Por otra parte, cada S_k es la suma de los menores principales de orden k de R.

c) En el caso general, si R es una matriz cuadrada de orden m, su polinomio característico es:

$$f(\lambda) = \lambda^m - S_1 t^{m-1} + S_2 t^{m-2} - \ldots\ldots + (-1)^m S_m$$

Donde S_k es la suma de los menores principales de orden k.

Sabiendo que la traza de una matriz cuadrada es igual a la suma de sus valores característicos y el determinante de una matriz es igual al producto de sus valores característicos. Para calcular el determinante de una matriz, se emplea la expansión del determinante por los cofactores.

De una manera general, para la matriz R de orden m, se tendrá:

$$|R| = \sum_{j=1}^{m} r_{ij} R_{ij} \qquad |R| = \sum_{i=1}^{m} r_{ij} R_{ij}$$

La primera expresión corresponde a la expansión del determinante por los cofactores de los elementos de la fila i; la segunda corresponde a la expansión del determinante por los cofactores de los elementos de la columna j.

De manera que:

f(λ)=0 Ecuación característica de la matriz, que admite m soluciones:

$\lambda_1, \lambda_2, \ldots, \lambda_m$ no todas necesariamente diferentes.

Obteniendo las m soluciones que constituyen los valores característicos o valores propios, cada valor propio podrá generar un vector propio con una dimensión igual a la de la matriz original, y en función de estos valores característicos, ordenados de mayor a menor, se encuentran los m vectores característicos, los cuales multiplicados por la matriz original proporcionan los componentes principales.

La técnica del ACP, transforma a las variables originales en nuevas variables (componentes), las cuales tiene desigualdad en cuanto a la información explicada, lo que significa que se tiene componentes muy informativos y otros no. Por eso se tiene unos componentes principales, que son los que se usan para hacer la representación. Esta desigualdad generada al crear los componentes permite elegir, entre ellos, los principales y eliminar los poco importantes, lo que no sucede con los originales porque todos ellos eran principales, todos eran importantes, no prescindiendo de ninguno de ellos.

Lo importante es que se tiene la nube de puntos desde unos ejes, donde uno es mucho más importante que el otro, ahora los ejes son Y_1 y Y_2. Si x_1 y x_2 eran dos variables que tenían la misma cantidad de información, ahora y_1 y y_2 no tienen la misma cantidad de información, y_1 tiene mucha más información que y_2. (en estadística información es equivalente a dispersión). Una variable que no varía no tiene información, una variable que varía mucho, es muy informativa.

El objetivo de la técnica ACP es conseguir girar los ejes de tal forma que exista la mayor desigualdad posible entre la varianza de la nube de puntos originales en las proyecciones en cada uno de los respectivos nuevos ejes, además, estos ejes, de estas nuevas variables, sean independientes entre sí, o sea, que tengan correlación cero.

La búsqueda de estos nuevos ejes se hace mediante el cálculo de los llamados valores propios y vectores propios de la matriz de correlaciones entre todas las variables del estudio. Puede hacerse también a partir de la matriz de varianzas-covarianzas, pero ésta tiene el problema de que cuando las variables tienen unidades de escala muy diferentes introduce un exceso de influencia por parte de las variables con mayor varianza. Por esto, suele trabajarse con la matriz de correlaciones; de esta forma se unifica el peso de las variables originales. Suele hablarse de variables estandarizadas cuando se trabaja con la matriz de correlaciones. Una variable es estandarizada cuando la muestra se transforma de manera que tiene un promedio cero y la varianza o la desviación estándar tiene por valor la unidad. Esto se consigue restando a cada valor de los valores de la columna considerada de la muestra el promedio y dividiendo por la desviación estándar de dicha columna. De esta forma todas las variables del estudio tienen la misma media y la misma desviación estándar y ninguna pesa más que otra.

7. METODOS DEL ACP.

Existen dos formas básicas de aplicar el ACP:

1. Método basado en la matriz de correlación, cuando la información no es homogénea o las medidas empleadas para las diferentes variables no son las mismas, o la magnitud de las variables aleatorias es diferente y se las estandariza previamente,

2. Método basado en la matriz de covarianzas, que se usa cuando los datos son dimensionalmente homogéneos y presentan valores medios similares, este método parte de la determinación de la matriz de varianza-covarianza sin estandarizar la información.

   Posteriormente en cada uno de estos métodos se procede a efectuar los diferentes cálculos en forma similar.

   Las conclusiones que se sacará con estos cálculos efectuados, es el de interpretar los resultados de la matriz de los coeficientes de correlación, de la matriz de los ejes principales, observando los valores propios generados y habrá que ser muy cuidadoso al interpretar los vectores propios, ya que el método no es independiente de la escala de medición de las variables originales; para mayor confiabilidad de la interpretación, será el profesional entendido en los problemas de su especialidad, donde se tomará en cuenta las variables consideradas en los resultados de las matrices quien interpretará y sacará conclusiones en forma correcta.

8. TRABAJO COMPUTACIONAL.

Siendo el número de cálculos a efectuar por lo general numerosos, especialmente cuando el número de muestras *n* y el número de variables *m*, son grandes este se vuelve muy complicado, razón por la cual, se ha desarrollado un programa computacional

(12), para los dos métodos mencionados, que facilita enormemente este trabajo, cuyos resultados se ilustrará para su comprensión con aplicaciones en algunas áreas como ser:

Aplicación 1: Análisis de calificaciones por asignatura en una universidad. En las universidades, el progreso en las asignaturas realizado por los estudiantes se evalúa con una calificación de 0 a 100, y a partir de una calificación mayor o igual a 51 el estudiante se considera aprobado y por debajo de este valor es reprobado. Se presenta en esta aplicación del ACP, las notas finales de 8 asignaturas de 20 estudiantes de un determinado curso de acuerdo al pensum de una carrera y se busca la forma de sintetizar la información contenida en estas asignaturas a fin de obtener relaciones entre materias, como una medida sintética del aprovechamiento estudiantil que pueda ser utilizada con fines descriptivos.

Las variables X_j, designan las asignaturas, identificadas como:

X_1 = Cálculo I

X_2 = Inglés

X_3 = Metodología de la Investigación

X_4 = Física

X_5 = Álgebra lineal

X_6 = Química

X_7 = Geología

X_8 = Geometría descriptiva.

La información disponible de acuerdo a lo especificado, es la siguiente:

Nº est.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	50,00	50,00	50,00	50,00	50,00	50,00	50,00	50,00
2	73,00	39,00	32,00	83,00	54,00	65,00	33,00	82,00
3	83,00	49,00	42,00	93,00	64,00	75,00	43,00	92,00
4	49,00	79,00	75,00	64,00	49,00	65,00	69,00	55,00
5	68,00	24,00	45,00	84,00	72,00	67,00	28,00	64,00
6	85,00	89,00	95,00	79,00	83,00	75,00	92,00	42,00
7	38,00	92,00	84,00	43,00	32,00	43,00	74,00	52,00
8	59,00	39,00	42,00	64,00	22,00	51,00	32,00	76,00
9	45,00	75,00	76,00	32,00	31,00	22,00	79,00	29,00
10	52,00	51,00	46,00	49,00	62,00	51,00	42,00	12
11	72,00	44,00	55,00	76,00	82,00	81,00	44,00	65,00
12	76,00	83,00	78,00	72,00	69,00	63,00	75,00	91,00
13	86,00	93,00	88,00	82,00	79,00	73,00	85,00	96,00
14	66,00	73,00	68,00	62,00	59,00	53,00	65,00	76,00
15	43,00	31,00	33,00	42,00	35,00	19,00	12,00	39,00
16	76,00	42,00	44,00	75,00	82,00	71,00	45,00	55,00
17	28,00	48,00	49,00	21,00	30,00	29,00	49,00	53,00
18	47,00	59,00	58,00	49,00	52,00	51,00	63,00	62,00
19	60,00	82,00	84,00	60,00	60,00	60,00	70,00	65,00
20	62,00	73,00	67,00	70,00	65,00	68,00	76,00	77,00

Cálculos. Cada uno de los cálculos efectuados por el programa de ACP son los siguientes:

Totales	1218,00	1215,00	1216,00	1250,00	1132,00	1132,00	1126,00	1253,00

Matriz de suma de cuadrados y productos.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	79396	74686	75038	81664	73726	73505	69671	79485
2	74686	82737	81589	75601	69708	69631	76758	76881
3	75038	81589	81462	76043	70451	70015	76066	76297
4	81664	75601	76043	85060	75773	76454	70349	82392
5	73726	69708	70451	75773	70884	69298	65521	72455
6	73505	69631	70015	76454	69298	69960	65189	74200
7	69671	76758	76066	70349	65521	65189	72178	71301
8	79485	76881	76297	82392	72455	74200	71301	87409

Matriz de suma de cuadrados y productos de las desviaciones.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	5219,797	692,500	983,602	5539,000	4787,203	4566,203	1097,602	3177,297
2	692,500	8925,750	7717,000	-336,500	939,000	862,000	8353,500	761,250
3	983,602	7717,000	7529,203	43,000	1625,399	1189,399	7605,203	114,602
4	5539,000	-336,500	43,000	6935,000	5023,000	5504,000	-26,000	4079,500
5	4787,203	939,000	1625,399	5023,000	6812,801	5226,801	1789,399	1535,203
6	4566,203	862,000	1189,399	5704,000	5226,801	5888,801	1457,399	3280,203
7	1097,602	8353,500	7605,203	-26,000	1789,399	1457,399	8784,199	757,102
8	3177,297	761,250	114,602	4079,500	1535,203	3280,203	757,102	8908,547

Matriz de varianza-covarianza.

	1,000	**2,000**	**3,000**	**4,000**	**5,000**	**6,000**	**7,000**	**8,000**
1	260,990	34,625	49,180	276,950	239,360	228,310	54,880	158,865
2	34,625	446,283	385,850	-16,825	46,950	43,100	417,675	38,063
3	49,180	385,850	376,460	2,150	81,270	59,470	380,260	5,730
4	276,950	-16,825	2,150	346,750	251,150	285,200	-1,300	203,975
5	239,360	46,950	81,270	251,150	340,640	261,340	89,470	76,760
6	228,310	43,100	59,470	285,200	261,340	294,440	72,870	164,010
7	54,880	417,675	380,260	-1,300	89,470	72,870	439,210	37,855
8	158,865	38,063	5,730	203,975	76,760	164,010	37,855	445,427

Promedios.

	60,9	60,75	60,8	62,5	56,6	56,6	56,3	62,65

Desviaciones estándar.

	16,16	21,13	19,4	18,62	18,46	17,16	20,96	21,11

Datos estandarizados.

	1	2	3	4	5	6	7	8
1	-0,6576	-0,4960	-0,5425	-0,6543	-0,3485	-0,3749	-0,2930	-0,5842
2	0,7300	-1,0035	-1,4468	1,0730	-0,1373	0,4771	-1,0836	0,8936
3	1,3333	-0,5421	-0,9444	1,5964	0,3908	1,0452	-0,6186	1,3554
4	-0,7180	0,8420	0,7133	0,0785	-0,4014	0,4771	0,5907	-0,3533
5	0,4284	-1,6956	-0,7937	1,1254	0,8133	0,5907	-1,3162	0,0624
6	1,4540	1,3034	1,9692	0,8637	1,3942	1,0452	1,6603	-0,9537
7	-1,3816	1,4418	1,1654	-1,0207	-1,2991	-0,7725	0,8232	-0,4918
8	-0,1146	-1,0035	-0,9444	0,0785	-1,8272	0,3181	-1,1301	0,6165
9	-0,9593	0,6575	0,7636	-1,5964	-1,3519	-1,9657	1,0557	-1,5540
10	-0,5370	-0,4498	-0,7435	-0,7435	-0,7066	0,2852	-0,6651	-2,3391
11	0,6697	-0,7728	-0,2914	0,7066	1,3414	1,3860	-0,5721	0,1085
12	0,9110	1,0266	0,8640	0,4973	0,6548	0,3635	0,8697	1,3093
13	1,5143	1,4879	1,3664	1,0207	1,1829	0,9316	1,3345	1,5402
14	0,3077	0,5652	0,3617	-0,0262	0,1267	-0,2045	0,4046	0,6165
15	-1,0800	-1,3726	-1,3965	-1,0730	-1,1407	-2,1358	-2,0603	-1,0922
16	0,9110	-0,8651	-0,8439	0,6543	1,3414	0,8180	-0,5255	-0,3533
17	-1,9849	-0,5883	-0,5928	-2,1722	-1,4047	-1,5677	-0,3395	0,4780
18	-0,8386	-0,0807	-0,1407	-0,7066	-0,2429	-0,3181	0,3116	-0,0300
19	-0,0664	0,9804	1,1654	0,13086	0,1796	0,1931	0,6372	0,1085
20	0,0664	0,5652	0,3115	0,3926	0,4436	0,6475	0,9162	0,6627

Matriz de coeficientes de correlación.

	1	2	3	4	5	6	7	8
1	1,000	0,101	0,157	0,921	0,803	0,824	0,162	0,466
2	0,101	1,000	0,941	-0,043	0,120	0,119	0,943	0,085
3	0,157	0,941	1,000	0,006	0,227	0,179	0,935	0,014
4	0,921	-0,043	0,006	1,000	0,731	0,893	-0,003	0,519
5	0,803	0,120	0,227	0,731	1,000	0,825	0,231	0,197
6	0,824	0,119	0,179	0,893	0,825	1,000	0,203	0,453
7	0,162	0,943	0,935	-0,003	0,231	0,203	1,000	0,086
8	0,466	0,085	0,014	0,519	0,197	0,453	0,086	1,000

Valores propios y porcentajes de contribución.

	Valores Propios	% valores Propios	% acumulados valores propios
	1	2	3
1	3,92797	49,09965	49,09965
2	2,73479	34,18494	83,28458
3	0,84449	10,55614	93,84074
4	0,20224	2,52805	96,36879
5	0,16320	2,03995	98,40874
6	0,06228	0,77849	99,18772
7	0,04355	0,54434	99,73157
8	0,02147	0,26843	100,00000

Ejes principales o componentes principales.

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7	Vector 8
1	0,45492	-0,17748	-0,06237	-0,52625	-0,45226	0,22511	0,10870	-0,46225
2	0,20206	0,54049	0,10107	-0,14360	0,06183	0,00310	-0,78849	-0,10323
3	0,22742	0,52566	-0,04902	-0,10187	0,01887	-0,67339	0,44251	-0,09776
4	0,43221	-0,27286	0,01655	-0,43593	0,28339	-0,13126	-0,06062	0,66874
5	0,42389	-0,11097	-0,43495	0,56042	-0,45145	-0,16135	-0,18740	0,19949
6	0,45811	-0,15840	-0,09232	0,30228	0,68519	0,10718	0,05185	-0,42600
7	0,23529	0,52168	0,02670	0,12451	-0,02348	0,65547	0,35847	0,31239
8	0,26225	-0,12669	0,88589	0,28474	-0,19268	-0,10712	0,02582	0,00547

Resultados de los componentes principales.

Columnas = Vectores propios

Filas = Observaciones

	1	2	3	4	5	6	7	8
1	134,808	37,071	19,578	3,215	-3,530	-4,113	-2,458	4,842
2	166,178	-7,165	43,230	-6,091	-2,894	-4,788	-6,489	8,447
3	193,139	0,249	47,145	-5,447	-3,600	-5,611	-6,981	9,415
4	164,179	69,258	25,564	-1,310	12,474	-9,233	-7,314	8,602
5	166,910	-10,466	17,312	1,879	-4,202	-18,885	4,960	12,783
6	215,742	85,610	-3,262	-9,537	-4,999	-7,847	0,924	0,152
7	137,877	97,057	33,676	-5,559	5,533	-10,989	-9,744	4,948
8	132,084	11,768	53,167	-15,457	4,275	-8,530	2,323	0,470
9	116,153	94,382	13,863	-14,026	-11,576	1,016	1,845	7,076
10	128,275	34,577	-19,456	-1,926	-1,954	-3,837	-10,784	2,798
11	186,267	11,966	14,132	11,484	-2,343	-13,339	-0,852	3,576
12	199,825	62,688	47,808	1,716	-14,643	-9,578	-7,461	7,678
13	225,475	70,736	47,294	0,935	-14,385	-9,864	-8,082	8,619
14	171,552	55,907	39,463	-0,351	-12,973	-8,219	-7,099	6,683
15	88,075	9,436	17,422	-10,794	-15,584	-17,882	-7,977	4,634
16	177,780	8,038	6,294	5,477	-9,715	-4,251	-4,065	6,858
17	99,332	49,392	51,303	16,692	-11,709	-6,735	0,938	0,692
18	144,161	51,827	30,281	9,581	-4,584	-2,996	-1,633	7,949
19	175,337	73,575	29,237	0,923	-3,621	-15,009	185,962	4,345
20	185,228	56,479	37,079	7,920	-1,798	-1,754	-4,839	12,228

Los valores propios de los componentes indican la cantidad de varianza, la cantidad de información que tiene cada componente. Como se puede observar en la tabla de valores propios y porcentaje de contribución del primer componente tiene un 49.099 % de información y el segundo un 34.184 %. Los dos juntos tienen un 83.284 %. Por lo tanto, haciendo una representación en dos dimensiones con esas dos primeras componentes se pierde un 16.716% de información únicamente.

En este problema se considera lo siguiente:

1) Observar en los resultados de los ejes principales o componentes, el valor absoluto de los coeficientes distinguiendo los que tienen un valor grande y un valor pequeño. En este caso en el primer componente se observa que en las asignaturas X_1, X_4, X_5 y X_6, tienen coeficientes con valor absoluto grande, cercano en todos los casos a 0.5, los demás son más pequeños, pesan mucho menos en este componente. En el segundo componente el peso principal se lo tiene en las asignaturas X_2, X_3, y X_7, con coeficientes cercanos a 0.54. Las demás asignaturas pesan poco.

2) Observar entre los coeficientes con valor absoluto grande el juego de signos que hay. En este problema el signo es el mismo, por lo tanto, las variables que pesan en un componente y el otro todos van en la misma dirección. Pero en otro caso se podría encontrar con valores de signos contrarios; por lo tanto, se tiene que interpretar el juego de fuerzas de los signos.

   La interpretación es más clara, si se analiza de esta manera, en el primer componente se tiene reunidas las asignaturas de ciencias exactas, donde existe mayor correlación entre ellas, en el segundo componente está relacionado con las asignaturas de humanidades, la asignatura de geometría descriptiva no pesa ni en uno ni en el otro, porque no tiene ninguna relación ni con las materias de ciencias exactas ni con las de letras. Así parece que existe un conjunto de habilidades cognitivas relacionadas con las ciencias exactas y un segundo relacionado con las de humanidades, estos dos conjuntos de habilidades son estadísticamente independientes. Interpretación realizada por un docente universitario.

Aplicación 2: Análisis comercial. Venta de calzados por una empresa. En la venta de calzados se utilizó 8 muestras y se determinó 5 factores, que afectan las ventas de zapatos en una ciudad, excluyendo el precio de venta (variable dependiente), siendo estos:

Nº	X_1	X_2	X_3	X_4	X_5	Y
	1	2	3	4	5	
1	4,00	5,00	1,00	8,00	105,00	20,00
2	5,00	5,00	2,00	10,00	104,00	24,00
3	5,00	7,00	2,00	11,00	101,00	24,00
4	10,00	7,00	2,00	12,00	95,00	24,00
5	15,00	8,00	5,00	14,00	100,00	32,00
6	16,00	9,00	6,00	15,00	103,00	38,00
7	18,00	9,00	7,00	17,00	110,00	45,00
8	24,00	12,00	7,00	19,00	117,00	55,00

Donde:

X_1 = Número de salidas o ventas de pares del producto

X_2 = Número de vendedores

X_3 = Gastos de propaganda en unidades monetarias (en miles de $us, Bs, etc.)

X_4 = Ingresos en u. m. (en miles de $us, Bs, etc.)

X_5 = Índice de las condiciones generales del negocio

Y = Ventas de calzados en u. m. (en cientos de miles de $us, Bs, etc.)

Cálculos. La determinación de los diferentes cálculos efectuados por el programa computacional de ACP son:

Suma de las columnas:

	97	62	32	106	835

Matriz de suma de cuadrados y productos:

	1	2	3	4	5
1	1547,00	864,00	509,00	1469,00	10331,00
2	864,00	518,00	284,00	879,00	6538,00
3	509,00	284,00	172,00	406,00	3412,00
4	1469,00	879,00	486,00	1500,00	11169,00
5	10331,00	6538,00	3412,00	11169,00	87465,00

Matriz de suma de cuadrados y productos de las desviaciones:

	1	2	3	4	5
1	370,875	112,250	121,000	183,750	206,625
2	112,250	37,500	36,000	57,500	66,750
3	121,000	36,000	44,000	62,000	72,000
4	183,750	57,500	62,000	95,500	105,250
5	206,625	66,750	72,000	105,250	311,875

Matriz de varianza – covarianza:

	1	2	3	4	5
1	46,359	14,031	15,125	22,969	25,828
2	14,031	4,688	4,500	7,188	8,344
3	15,125	4,500	5,500	7,750	9,000
4	22,969	7,188	7,750	11,938	13,156
5	25,828	8,344	9,000	13,156	38,984

Promedios

	12,125	7750,000	4,000	13,250	104,375

Desviaciones estándar

	6,809	2,165	2,345	3,455	6,244

Datos estandarizados

	1	2	3	4	5
1	-1,11624	-1,18814	-1,19659	-1,42137	0,09364
2	-0,97886	-1,18814	-0,79772	-0,87989	-0,05618
3	-0,97886	-0,32404	-0,79772	-0,60916	-0,50563
4	-0,29194	-0,32404	-0,79772	-0,33842	-1,40453
5	0,39498	0,10801	0,39886	0,20305	-0,65545
6	0,53236	0,54006	0,79772	0,47379	-0,20600
7	0,80713	0,54006	1,19659	1,01526	0,84272
8	1,63143	1,83621	1,19659	1,55674	1,89143

Matriz de suma de cuadrados y productos:

	1	2	3	4	5
1	7,000	6,663	6,630	6,835	4,253
2	6,663	7,000	6,204	6,726	4,321
3	6,630	6,204	7,000	6,695	4,302
4	6,835	6,726	6,695	7,000	4,269
5	4,253	4,321	4,302	4,269	7,000

Matriz de suma de cuadrados y productos de las desviaciones

	1	2	3	4	5
1	7,000	6,663	6,630	6,835	4,253
2	6,663	7,000	6,204	6,726	4,321
3	6,630	6,204	7,000	6,695	4,302
4	6,835	6,726	6,695	7,000	4,269
5	4,253	4,321	4,302	4,269	7,000

Matriz de varianza – covarianza

	1	2	3	4	5
1	0,875	0,833	0,829	0,854	0,532
2	0,833	0,875	0,775	0,841	0,540
3	0,829	0,775	0,875	0,837	0,538
4	0,854	0,841	0,837	0,875	0,534
5	0,532	0,540	0,538	0,534	0,875

Matriz de coeficientes de correlación:

	1	2	3	4	5
1	1,00000	0,95182	0,94721	0,97636	0,60755
2	0,95182	1,00000	0,88626	0,96084	0,61723
3	0,94721	0,88626	1,00000	0,95645	0,61463
4	0,97636	0,96084	0,95645	1,00000	0,60986
5	0,60755	0,61723	0,61463	0,60986	1,00000

Valores propios y porcentajes de contribución.

	Valores Propios	% valores Propios	% acumulados valores propios
	1	2	3
1	4,29505	85,90107	85,90107
2	0,54577	10,91538	96,81644
3	0,11385	2,27697	99,09334
4	0,02927	0,58536	99,67878
5	0,01606	0,32122	100,00000

Matriz de los ejes principales

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5
1	0,47222	-0,19470	-0,01947	-0,85434	-0,09396
2	0,46463	-0,15160	-0,68869	0,35150	-0,40410
3	0,46340	-0,15316	0,72466	0,31527	-0,37050
4	0,47439	-0,19587	-0,01401	0,21503	0,83055
5	0,34834	0,93664	0,00005	-0,02401	0,02817

Resultado de los componentes principales:

Columnas = Vectores propios = Componentes
Filas = Observaciones = Variables

	1	2	3	4	5
1	45,04660	95,08986	-2,90340	-2,13899	6,83499
2	46,58266	93,41362	-2,22628	-2,22239	8,00346
3	46,94128	90,10464	-3,61781	-1,23153	7,94130
4	47,68672	83,31547	-3,72946	-5,14334	8,13306
5	54,59313	86,02236	-2,36925	-7,80611	7,94957
6	57,51281	88,13694	-2,36660	-7,84987	7,99605
7	62,30783	93,75909	-1,70854	-8,97969	9,29589
8	69,92223	98,30081	-3,91906	-12,78763	9,37809

Como se puede observar en la tabla de valores propios y porcentaje de contribución del primer componente tiene un 85.901 % de información y el segundo un 10.915 %. Los dos juntos tienen un 96.816 %. Por lo tanto, haciendo una representación en dos dimensiones con esos dos primeros componentes se pierde un 3.184 % de información únicamente.

En la interpretación de estos componentes se observa que:

1) En los resultados de los ejes principales o componentes, el valor absoluto de los coeficientes distinguiendo los que tienen un valor grande y un valor pequeño. En este caso en el primer componente se observa que en las variables X_1, X_2, X_3 y X_4, tienen coeficientes con valor absoluto más grande, cercano en todos los casos a 0.465, el restante es más pequeño, pesan mucho menos en este componente. En el segundo componente el peso principal se lo tiene en la variable X_5, con coeficiente den 93.967.

2) Entre los coeficientes con valor absoluto grande el juego de signos que se obtuvo, en este caso el signo es el mismo, por lo tanto, las variables que pesan en un componente y el otro todos van en la misma dirección. En otros casos se podría encontrar con valores de signos contrarios; por lo tanto, se tiene que interpretar el juego de fuerzas de los signos.

   La interpretación es más clara, si se analiza de esta manera, en el primer componente se tiene reunidas las variables de salidas, vendedores, gastos de propaganda e ingresos, donde existe mayor correlación entre ellas, el segundo componente está relacionado con el índice de las condiciones generales del negocio que no tiene mucha correlación con las otras variables.

3) La función del primer componente principal que contribuye con un 85.901 % de la variación es:

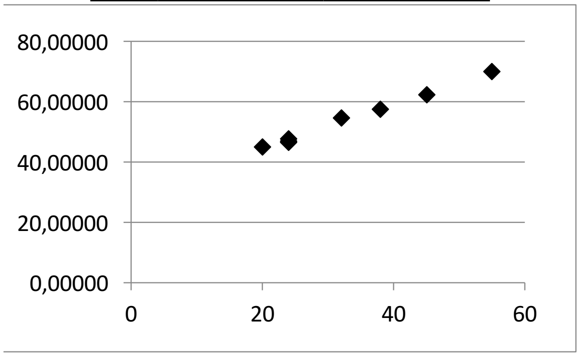
$$Z_i = \sum_{j=1}^{n} V_j Y_j \qquad i = 1,\ldots\ldots, m$$

Donde V = vector i de la matriz de ejes principales.

$$Z_1 = 0.47222Y_1 + 0.46463Y_2 + 0.46340Y_3 + 0.47439Y_4 + 0.34834_5$$

Sustituyendo los valores de Y_i se encuentra:

	Y	Z
1	20	45,04660
2	24	46,58266
3	24	46,94128
4	24	47,68672
5	32	54,59313
6	38	57,51281
7	45	62,30783
8	55	69,92223



Efectuando un gráfico con los valores de Y y Z, se observa la relación que existe entre Z_1 y Y. Como puede observarse esta relación tiende a ser lineal en consecuencia, se puede ajustar una línea de regresión por el método de los mínimos cuadrados, de acuerdo a su diagrama de dispersión se ajustará a una recta, a una parábola, cúbica, etc., y determinar su coeficiente de correlación, en el caso de ajustarse a una recta, su ecuación seria:

$$Y = 30.1538 + 0.7228Z$$

Los coeficientes de correlación y de determinación son:

$$r = 0.997 \qquad r^2 = 0.994$$

Que muestran que existe una buena correlación entre estas variables. Para una mejor interpretación de estos resultados se aconseja a un profesional del área de administración de empresas, economía o ingeniero comercial.

CONCLUSIONES.

- Mediante el ACP se facilita la comprensión de la información multivariante, reduciendo el número de variables en todos los casos considerados, manteniendo la máxima cantidad de información, presentada en una tabla de datos de variables cuantitativas. Además, la interpretación de las fórmulas, resultados y aplicaciones se facilita, cuando se tiene un conocimiento acerca de: estadística, álgebra lineal y algebra matricial.

- Cabe señalar que en un gráfico de dos dimensiones se observa una buena fotografía de las posiciones relativas de los puntos, en cambio en la representación de ocho dimensiones originales no se puede ver. Se denomina fotografía porque la metáfora es apropiada, puesto que cuando se observa una fotografía en realidad se está viendo una representación bidimensional de una realidad que es tridimensional. En el ACP se está realizando algo similar, se observa una fotografía bidimensional, para que se pueda observar una realidad constituida por varias dimensiones pero que no son visualizadas; por lo tanto, se está observando una fotografía bidimensional de una realidad de ocho dimensiones.

- El objetivo principal que persigue el ACP es la representación de las medidas numéricas de varias variables en un espacio de pocas dimensiones donde nuestros sentidos puedan percibir relaciones que de otra manera permanecerían ocultas en dimensiones superiores. Dicha representación debe ser tal que al desechar dimensiones superiores (generalmente de la tercera o cuarta en adelante) la pérdida de información sea mínima.

- La mejor interpretación de los resultados, la dará el profesional especializado en cada uno de los temas abordados en los diferentes problemas.

COLABORACIÓN

- Prof. M. Nilda Avilés de Ruiz

  Lic. en Idiomas. Universidad Autónoma Gabriel René Moreno

  Santa Cruz – Bolivia (agosto, 2005)

- M. Cs. Astrid Keitel Ruiz Avilés

  Lic. en Adm. Emp. Universidad Católica Boliviana, Santa Cruz – Bolivia (octubre,1999)

  M. Cs. Academia Diplomática Boliviana ¨Rafael Bustillo¨, La Paz – Bolivia (abril, 2004)

- Dr. Osman Miranda Lira

  Universidad Mayor de San Andrés, La Paz – Bolivia (Julio, 1991).

  Especialidad: Radiologia e Imagenologia. (agosto, 1993)

BIBLIOGRAFÍA

BALESTRA Pietro, Calcul Matricial pour Économistes. Éditions Castella, Lausanne, Suisse, 1972, pp. 241 – X.

BERTIER P., BOUROCHE J. M., Analyse des Données Multidimensionnelles. Presses Universitaires de France, Vendóme, France, 1975, pp. 270 – XI.

DAGNELIE Pierre, Analyse Statistique á Plusieurs Variables. Les Presses Agronomiques de Gembloux, Gembloux, Bélgica, 1975 (2da. Edición), pp. 362 – XIV.

DAVIS C. John, Statistics and Data Analisis in Geology. John Wiley & Sons, New York, Estados Unidos, 1973, pp. 550 – VII.

GANTMACHER F. R., Théorie des Matrices (Tomo I y II), Dunod, Paris, France, 1966, pp. 638 – XV.

JOHNSON A. Richard, Probabilidad y Estadística para Ingenieros. Prentice Hall Hispanoamericana, S.A., México D.F. 1997 (5ta. edición), pp. 630 – XV.

JOHNSTON J., Métodos de Econometría. Editorial Vicens- Vives, Barcelona, España (3ra. edición), pp. 464 – XIII.

JORESKOG J.K., KLOVAN J. E. REYMENT R. A., Geological Factor Analysis. Elsevier Scientific Publishing Company. Amsterdam, Holanda, 1976, pp 178 – VII.

KRUMBEIN W. C., GRAYBILL Franklin A., An Introduction to Statistical Models in Geology. Editorial McGraw-Hill, New York, Estados Unidos, 1965, pp. 475 – XV.

LIPSCHUTZ Seymour, Álgebra Lineal. Editorial McGraw-Hill, Madrid, España, 1998, pp. 553 – XV.

RUIZ Aranibar Gustavo. Análisis de las Componentes Principales. U.T.O. – F.C.E.F. Revista del I. I. E., Nº 26-27. Oruro – Bolivia. 1988, pp. 261 – XVI.

RUIZ Aranibar Gustavo. Libreria Científica de Programas Informáticos, La Paz -Bolivia.

13. YAMANE Taro, Estadística. Ed. Harla, México, D, F, México, 1979 (3ra. Edición), pp. 771 – XXIV.

WAYNE W. Daniel, Bioestadística. Ed. Limusa Wiley, México, D, F, México, 2004 (4ta. Edición), pp. 755 – XIV.

PENSAMIENTO.

No es lo que comes, sino lo que digieres, lo que te hace fuerte y saludable.

No es lo que ganas, sino lo que ahorras, lo que te hace rico y solvente.

No es lo que aprendes, sino lo que recuerdas lo que te hace sabio y juicioso.

No es lo que investigas, sino lo que creas, desarrollas y das a conocer a tus semejantes, lo que te hace científico.

Gustavo Ruiz Aranibar