APLICACIÓN DEL ALGEBRA LINEAL EN EL ANÁLISIS ESTADÍSTICO MULTIVARIADO DE UNA SERIE DATOS EN UN PPROYECTO DE APLICACIÓN DE LA LICENCITURA EN INGENIERÍA

Juan Carlos Axotla, Miguel Pineda, Armando Aguilar, Frida León, Omar García.
Universidad Nacional Autónoma de México. Facultad de Estudios Superiores Cuautitlán. México
jc_axotla@fesc.unam.mx; mnazarethp@fesc.cuautitlan2.unam.mx
Área Temática: Articulación y extensión
Palabras clave: Vectores propios, valores propios, matriz de correlación, análisis estadístico multivariado.

Resumen

Hoy en día existen diversas ramas de la ciencia que generan conjuntos complejos de datos que constan de una gran cantidad de variables, las cuales, deben ser analizadas estadísticamente; pero el análisis se torna complejo al utilizar métodos estadísticos comunes. La utilización del análisis estadístico multivariante permite encontrar relaciones entre las variables de respuesta, las unidades experimentales y la interacción de ambas. Presentamos el desarrollo de un proyecto de análisis estadístico con el uso de técnicas PCA (Analisis de Componentes Principales), y se obtuvo un polinomio característico, los valores y vectores propios de la matriz de correlación obtenida de los datos generados para las variables de respuesta utilizando software matemático.

Introducción

El análisis multivariado se refiere a todos los métodos estadísticos que analizan simultáneamente medidas múltiples de cada individuo u objeto sometido a investigación; entonces, cualquier análisis simultáneo de dos o más variables puede ser considerado como un análisis multivariable. El análisis multivariado considera como elemento esencial un valor teórico, el cual es una combinación lineal de variables.

Antecedentes

El análisis multivariado integra un conjunto de técnicas, entre las más conocidas se encuentran: 1) regresión múltiple y correlación múltiple; 2) análisis discriminante múltiple; 3) componentes principales y análisis factorial común ; 4) análisis multivariable de varianza y covarianza; 5) correlación canónica; 6) análisis cluster; 7) análisis multidimensional y 8) análisis conjunto.

El presenta trabajo se refiere al análisis de componentes principales, se presentará el procedimiento para analizar datos experimentales con este método utilizando software matemático.

El análisis de componentes principales es una técnica estadística cuyo el objetivo es reducir las variables provenientes de un conjunto de datos amplio perdiendo la menor cantidad de información posible. Esta síntesis de información genera componentes principales o factores que son una combinación lineal de las variables originales, estos componentes son independientes entre sí.

El análisis de componentes principales (ACP) tiene distintas fases: 1) análisis de la matriz de correlación, 2) selección de los factores, 3) análisis de la matriz factorial, 4) interpretación de los factores y 5) cálculo de las puntuaciones factoriales. El ACP se basa en la obtención de valores y vectores propios de la matriz de correlación.

Métodos y Procedimientos

Los vectores y valores propios son característicos de una matriz ya que contienen información importante acerca de la naturaleza de la misma. En problemas matemáticos que involucran transformaciones lineales T: $V \rightarrow W$, resulta útil encontrar un vector v en V tal que Tv y v son paralelos; es decir, se busca un vector v y un escalar λ tal que $Tv = \lambda v$, si $v \neq 0$ y λ satisface $Tv = \lambda v$, entonces λ se denomina valor propio o característico de T para el valor λ.

Definición: Sea A una matriz de n x n con componentes reales. El número λ (real o complejo) se denomina valor propio de A si existe un vector diferente de cero v en $\mathbb{R}n$ tal que:

$$Av = \lambda v$$

El vector $v \neq 0$ se denomina **vector propio** de A correspondiente al valor característico de λ. El conjunto de todos los vectores propios correspondientes a λ, junto con el vector cero, se denomina **espacio propio o caracteristico** de λ y se puede denotar por E_{λ}.

Para una matriz A de n x n se cumple que λ es un valor característico de A si y sólo si:

$$P(\lambda) = \det (A - \lambda I) = 0$$

La ecuación anterior se denomina ecuación característica de A; $P(\lambda)$ se denomina **polinomio característico** de A.

El ejemplo siguiente presenta la obtención de valores y vectores característicos de una matriz de manera manual.

Ejemplo 1: Determinar polinomio característico, eigenvalores y eigenvectores de la matriz A.

$$A = \begin{bmatrix} 4 & 6 & 6 \\ 1 & 3 & 2 \\ -1 & -5 & -2 \end{bmatrix}$$

a) Obtención del polinomio característico mediante P(λ)= det ($A - \lambda I$)

$$P(\lambda) = \det\left[\begin{bmatrix} 4 & 6 & 6 \\ 1 & 3 & 2 \\ -1 & -5 & -2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right]$$

Calculando el valor del determinante se obtiene el polinomio característico de la matriz A que es:

$$P(\lambda)= - \lambda^3 + 5\lambda^2 - 8\lambda + 4$$

El teorema fundamental del álgebra dice que cualquier polinomio de grado n con coeficientes reales o complejos tiene exactamente n raíces (contando multiplicidades); de modo que, cualquier valor característico de A es una raíz del polinomio característico de dicha matriz.

b) Cálculo de los valores propios

Las raíces del polinomio característico de la matriz representan los valores propios o característicos de la matriz A.

Calculando las raíces del polinomio $P(\lambda)= - \lambda^3 + 5\lambda^2 - 8\lambda + 4$ se obtienen los eigenvalores, para este ejemplo son:

$$P(\lambda)= (\lambda-1)(\lambda-2)(\lambda-2)$$

$$\lambda_1=1,\ \lambda_2=2,\ \lambda_3=2$$

c) Obtención de los vectores propios

En el ejemplo se obtuvieron tres valores característicos: $\lambda_1=1$, $\lambda_2=2$, $\lambda_3=2$, se presenta multiplicidad para el valor 2, por tanto se formarán 2 vectores propios que son linealmente independientes.

Para $\lambda_1=1$.

Establecer un sistema de ecuaciones homogéneo de la forma $(A - \lambda I)v = 0$, con lo cual se obtendrá el vector característico correspondiente a $\lambda_1=1$:

1. Obtener la matriz $(A - \lambda I)$

$$(A - \lambda I) = \begin{bmatrix} (4-\lambda) & 6 & 6 \\ 1 & (3-\lambda) & 2 \\ -1 & -5 & (-2-\lambda) \end{bmatrix}$$

2. Sustituir el valor $\lambda_1=1$

$$[A-(1)I] = \begin{bmatrix} (4-1) & 6 & 6 \\ 1 & (3-1) & 2 \\ -1 & -5 & (-2-1) \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 6 & 6 \\ 1 & 2 & 2 \\ -1 & -5 & -3 \end{bmatrix}$$

3. Generar y resolver el sistema $(A - \lambda I)v = 0$

$$\begin{bmatrix} 3 & 6 & 6 \\ 1 & 2 & 2 \\ -1 & -5 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Resolviendo con Gauss Jordan

$$\left[\begin{array}{ccc|c} 3 & 6 & 6 & 0 \\ 1 & 2 & 2 & 0 \\ -1 & -5 & -3 & 0 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & \frac{4}{3} & 0 \\ 0 & 1 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Así, $x_1 = -4/3x_3$, $x_2 = -1/3x_3$. Un vector característico es

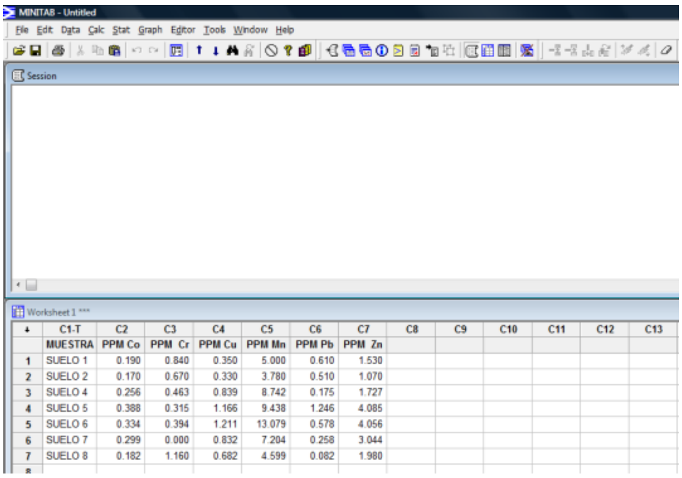
$$v_1 = \begin{bmatrix} -4/3 \\ -1/3 \\ 1 \end{bmatrix}$$

Para λ_2=2, un vector característico es

$$v_2 = \begin{bmatrix} -3/2 \\ -1/2 \\ 1 \end{bmatrix}$$

El ejemplo siguiente presenta el análisis de componentes principales que se realizó en datos experimentales sobre contaminación en suelos, el procedimiento matemático realizado es similar al presentado en el ejemplo 1. La obtención de los componentes principales requiere calcular los valores y vectores característicos de la matriz de correlación, para hacer este procedimiento más rápido se utilizó el paquete estadístico *Minitab 14*.

Los datos obtenidos por el investigador se capturan en la hoja de trabajo de *Minitab*:



↓	C1-T	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
	MUESTRA	PPM Co	PPM Cr	PPM Cu	PPM Mn	PPM Pb	PPM Zn						
1	SUELO 1	0.190	0.840	0.350	5.000	0.610	1.530						
2	SUELO 2	0.170	0.670	0.330	3.780	0.510	1.070						
3	SUELO 4	0.256	0.463	0.839	8.742	0.175	1.727						
4	SUELO 5	0.388	0.315	1.166	9.438	1.246	4.085						
5	SUELO 6	0.334	0.394	1.211	13.079	0.578	4.056						
6	SUELO 7	0.299	0.000	0.832	7.204	0.258	3.044						
7	SUELO 8	0.182	1.160	0.682	4.599	0.082	1.980						
8													

Figura 1. Datos capturados en la Hoja de trabajo de Minitab

Para realizar el análisis estadístico se utiliza la secuencia *Stat* + *Multivariate* + *Principal Components*:

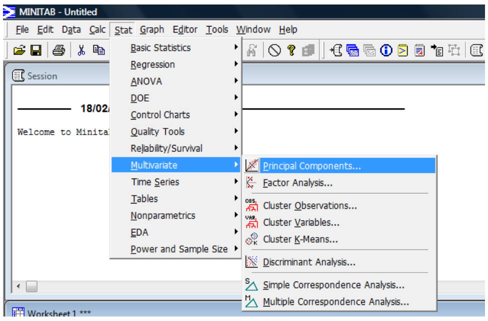


Figura 2. Menú para análisis de componentes principales

Minitab genera salidas gráficas y numéricas útiles para el análisis de los datos. Con los datos capturados se obtiene el ACP, la primera parte consiste en un análisis de los valores propios de la matriz de correlación, en este caso se obtienen seis valores característicos ya que la matriz de correlación es de 6x6. Cada valor característico representa una proporción de la varianza total de los datos, al sumar este porcentaje se explica el 100% de la varianza de los datos, esta información se presenta en la Tabla 1.

Tabla 1. Eigenanálisis de la matriz de correlación

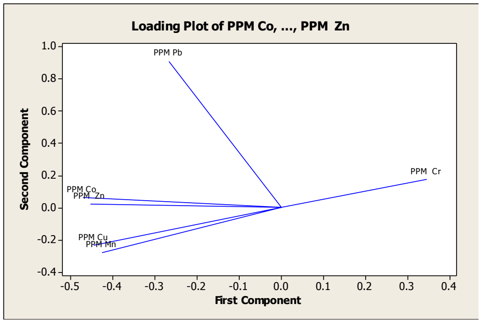
Eigenvalor	4.3865	0.8168	0.5761	0.1615	0.0542	0.0049
Proporción	0.731	0.136	0.096	0.027	0.009	0.001
Acumulado	0.731	0.867	0.963	0.990	0.999	1.000

Como se mencionó anteriormente, cada valor propio genera un vector propio, estos vectores propios se llaman Componentes principales y cada uno presenta una combinación lineal de las variables estudiadas. La tabla 2, muestra los Componentes principales:

Tabla 2. Componentes principales

Variable	PC1	PC2	PC3	PC4	PC5	PC6
PPM Co	-0.472	0.064	-0.075	0.175	0.395	0.762
PPM Cr	0.345	0.176	0.886	0.041	0.064	0.244
PPM Cu	-0.446	-0.234	0.323	0.199	0.545	-0.552
PPM Mn	-0.427	-0.279	0.250	-0.782	-0.242	0.089
PPM Pb	-0.267	0.912	-0.011	-0.221	0.044	-0.215
PPM Zn	-0.454	0.023	0.205	0.518	-0.695	-0.022

Las gráficas permiten comparar la interacción entre los primeros dos componentes principales ya que en ellos se concentra la mayor proporción de la variabilidad en los datos.



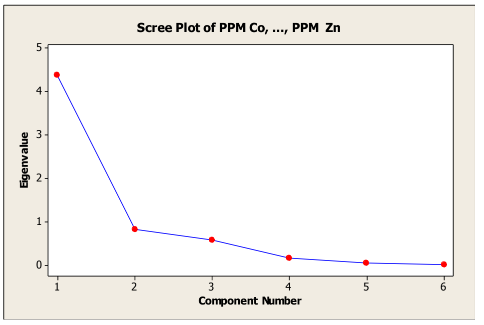


Figura 3. Gráficos de interacción de componentes principales y varianza acumulada

Conclusiones

La enseñanza en asignaturas del área de matemáticas, en muchos casos carece de aplicaciones que puedan ser utilizadas por los estudiantes en otras áreas, esto en ocasiones genera desinterés por parte de los alumnos y afecta significativamente el aprendizaje; el trabajo presentado muestra una aplicación para el tema de valores y vectores característicos que se imparte dentro de la asignatura Álgebra Lineal e integra el manejo de una herramienta tecnológica que facilita el manejo de la información.

Referencias Bibliograficas

[1] Antón, H. (1998). *Introducción al álgebra lineal*. (2ª.ed.). México: Limusa.
[2] Gerber, H. (1990). *Álgebra lineal.* México: Grupo Editorial Iberoamérica.
[3] Grossman, S. (2008). *Álgebra lineal*. (6ª. ed). México: McGraw Hill.
[4] Hair, J.F., Anderson, R.E., Tatham, R.L y Black, W.C. (1999). *Análisis Multivariante*. (5a. ed). Madrid: Prentice Hall Iberia.
[5] Solar, G.E., Speziale, L. (1977). *Apuntes de álgebra lineal*. (3ª. ed). México: Limusa – Facultad de Ingeniería UNAM.
[6] Williams, G. (2002). *Álgebra lineal con aplicaciones*. (4ª. ed). México: McGraw Hill.