Recibido:10/09/2026 Aprobado:16/10/2026

DOI: https://doi.org/10.70208/3007.8245.v6.n3.630

El análisis estadístico multivariado como herramienta didáctica en la formación de estudiantes de ingeniería ambiental

José Luis González Bucio
buciojos@uqroo.edu.mx
https://orcid.org/0000-0002-3515-7003
Universidad de Quintana Roo. División de Ciencias e Ingeniería, Ingeniería Ambiental
Chetumal, Quintana Roo, México

Walter Magaña Landero
walmag@uqroo.edu.mx
https://orcid.org/0009-0009-7912-8341
Universidad de Quintana Roo. División de Ciencias e Ingeniería, Ingeniería Ambiental
Chetumal, Quintana Roo, México

José Manuel Carrión Jiménez
jmcarrion@uqroo.edu.mx
https://orcid.org/0009-0003-7465-6803
Universidad de Quintana Roo. División de Ciencias e Ingeniería, Ingeniería Ambiental
Chetumal, Quintana Roo, México

Víctor Hugo Delgado Blas
blas@uqroo.edu.mx
https://orcid.org/0000-0003-0393-5218
Universidad de Quintana Roo. División de Ciencias e Ingeniería, Ingeniería Ambiental
Chetumal, Quintana Roo, México

Joel Omar Yam Gamboa
oyam@uqroo.edu.mx
https://orcid.org/0000-0001-8478-7872
Universidad de Quintana Roo. División de Ciencias e Ingeniería, Ingeniería Ambiental
Chetumal, Quintana Roo, México

David de Jesús Santillán Sanvicente
iee@itchetumal.edu.mx
https://orcid.org/0009-0000-4575-6758
Instituto Tecnológico de Chetumal, Tecnológico Nacional de México. División de Ingeniería

RESUMEN

El análisis estadístico multivariado constituye una herramienta para organizar, explorar e interpretar conjuntos de datos con múltiples variables. El objetivo de este trabajo fue describir una experiencia didáctica orientada a la aplicación e interpretación de técnicas estadísticas multivariadas por estudiantes de quinto semestre de Ingeniería Ambiental, mediante una base de datos de parámetros fisicoquímicos y químicos obtenidos en 18 puntos de muestreo de la Bahía de Chetumal, Quintana Roo. Se organizaron tres equipos de estudiantes, a quienes se proporcionó la misma base de datos y el programa XLSTAT integrado en Microsoft Excel. La actividad se centró en la correlación de Pearson, el análisis de componentes principales (ACP) y el análisis de conglomerados (Cluster), así como en la elaboración e interpretación de tablas y gráficas. A partir de la base de datos compartida se calcularon nuevamente la matriz de correlaciones y el ACP con variables estandarizadas. Los tres primeros componentes explicaron 73.71 % de la varianza total. La experiencia permitió documentar diferencias en el nivel de interpretación alcanzado por los equipos, desde dificultades para reconocer relaciones estadísticas hasta la identificación de patrones y agrupaciones en los datos. Se destaca el uso de datos ambientales reales como recurso didáctico para favorecer la comprensión de herramientas estadísticas aplicadas.

Palabras clave: Estadística multivariada, enseñanza universitaria, correlación de Pearson, análisis de componentes principales, análisis de conglomerados

Los autores declaran no tener conflicto de intereses.
El editor y los revisores declararon no tener conflicto de intereses.

Licencia: CC BY NC ND

Multivariate statistical analysis as a didactic tool in the education of environmental engineering students

ABSTRACT

Multivariate statistical analysis provides tools to organize, explore, and interpret datasets involving multiple variables. The aim of this study was to describe a teaching experience focused on the application and interpretation of multivariate statistical techniques by fifth-semester Environmental Engineering students, using a dataset of physicochemical and chemical parameters measured at 18 sampling points in Chetumal Bay, Quintana Roo, Mexico. Three student teams received the same dataset and used XLSTAT software integrated into Microsoft Excel. The activity focused on Pearson correlation, principal component analysis (PCA), and cluster analysis, together with the preparation and interpretation of tables and graphs. Pearson correlations and PCA were recalculated from the shared dataset after standardizing the variables. The first three principal components explained 73.71% of the total variance. The educational experience documented differences in the teams' interpretation, ranging from difficulties in recognizing statistical relationships to identifying patterns and groups in the data. The use of real environmental data is highlighted as a teaching resource to support the understanding of applied statistical tools.

Keywords: Multivariate statistics, higher education, Pearson correlation, principal component analysis, cluster analysis

The authors declare no conflict of interest.
The Editor and the Reviewers declared no conflict of interest.

License:

INTRODUCCIÓN

El análisis de conjuntos de datos con múltiples variables requiere herramientas que permitan reconocer relaciones, patrones y similitudes que no siempre son evidentes mediante la revisión aislada de cada parámetro. En el ámbito ambiental, las mediciones suelen presentar variabilidad natural y analítica, por lo que los métodos estadísticos multivariados contribuyen a explorar la información de manera conjunta (Einax, 2003; Grupo de Quimiometría, 2001).

Además de su utilidad en la investigación, estas técnicas representan recursos formativos para estudiantes de Ingeniería Ambiental. Su aprendizaje no se limita a ejecutar procedimientos en un programa informático: implica comprender qué pregunta responde cada análisis, reconocer sus representaciones gráficas y comunicar una interpretación sustentada en los datos.

La correlación de Pearson permite examinar la dirección y la intensidad de la asociación lineal entre pares de variables. El análisis de componentes principales (ACP) transforma un conjunto de variables posiblemente relacionadas en componentes no correlacionados que resumen parte de la variabilidad. Por su parte, el análisis de conglomerados agrupa observaciones de acuerdo con una medida de similitud o distancia. Estas herramientas se han utilizado ampliamente en el análisis de datos químicos y ambientales (Cela, 1994; Carlosena, 1999; Esbensen et al., 2018; Santos et al., 2019).

En este contexto, la experiencia descrita se enfocó en que los estudiantes interpretaran información analítica previamente obtenida, en lugar de repetir los procedimientos de determinación de los parámetros. El propósito fue desarrollar comprensión estadística mediante el análisis de una base de datos ambiental común.

Objetivo

Describir la aplicación de herramientas estadísticas multivariadas como estrategia didáctica para desarrollar habilidades de análisis, interpretación y representación de datos en estudiantes de Ingeniería Ambiental, utilizando como caso de estudio una base de datos de parámetros ambientales de la Bahía de Chetumal, Quintana Roo.

METODOLOGÍA

Contexto de la experiencia didáctica

La actividad se realizó con estudiantes de quinto semestre de la asignatura Análisis Instrumental de la carrera de Ingeniería Ambiental de la Universidad Autónoma del Estado de Quintana Roo. Se organizaron tres equipos de tres integrantes. A todos los equipos se les proporcionó la misma base de datos, correspondiente a 18 puntos de muestreo de la Bahía de Chetumal, Quintana Roo, obtenida durante la primavera de 2026. Los estudiantes no realizaron nuevamente la determinación de los parámetros; el trabajo se concentró en el procesamiento estadístico y la interpretación de los resultados.

Para el procesamiento se utilizó XLSTAT integrado en Microsoft Excel. Se solicitó a los equipos aplicar correlación de Pearson, análisis de componentes principales y análisis de conglomerados, además de elaborar tablas y gráficas que facilitaran la interpretación de las relaciones entre variables y de las similitudes entre puntos de muestreo.

Base de datos

La base de trabajo incluyó 18 observaciones y 13 variables: pH, conductividad eléctrica (CE), temperatura (T°), turbidez, sólidos totales disueltos (STD), demanda bioquímica de oxígeno (DBO_5), demanda química de oxígeno (DQO), fosfato, cloruros, hierro (Fe), cobre (Cu), zinc (Zn) y cadmio (Cd). La información se presenta en la Tabla 1. Las unidades de concentración de los metales no se especifican en la base compartida; por ello, no se asignan unidades en esta tabla y deberán incorporarse cuando se confirme la unidad reportada en los registros analíticos.

Tabla 1. Base de datos utilizada por los tres equipos para los análisis estadísticos

Punto	pH	CE	T°	Turbidez	STD	DBO_5	DQO	Fosfato	Cloruros	Fe	Cu	Zn	Cd
1	7.1	920	32	3	6000	3	30	0.2	7	135	5	45	1
2	7	921	32	4	6500	3	35	0.2	7	146	6	36	1.2
3	7.3	943	32	4	6500	4	38	0.2	7	149	5	39	1
4	7.3	947	32	3	6500	4	39	0.2	7	156	6	43	1.1
5	7	1122	33	4	6500	4	38	0.2	6	145	6	47	0.9
6	7.4	1264	33	5	6500	4	37	0.3	6	163	6	49	0.9
7	7.3	1044	33	5	7000	4	38	0.3	6	145	6	45	0.9
8	7.6	1037	32	4	7000	4	39	0.3	8	164	6	36	0.8
9	7.8	1087	30	4	7000	4	38	0.3	7	146	7	39	0.9
10	7.9	950	30	4	7000	4	38	0.2	7	158	5	43	0.9
11	7.5	984	31	4	6800	4	41	0.2	7	159	5	47	1
12	7.5	1043	32	3	7000	4	42	0.2	7	175	6	49	1.2

13	7.4	1132	31	3	7000	4	41	0.2	8	165	7	49	1
14	7.3	1243	32	3	7000	4	40	0.2	8	173	7	49	1.1
15	7.2	1176	30	4	7000	4	40	0.3	8	175	7	49	1
16	7.3	1279	30	4	9000	4.5	40	0.3	8	173	7	54	1
17	7.2	1236	31	3	9000	4.5	40	0.2	8	175	7	55	1
18	7.1	1198	30	3	9000	4.5	40	0.2	8	175	7	56	1

Procedimientos estadísticos

Correlación de Pearson. Se examinó la asociación lineal entre pares de variables mediante el coeficiente r de Pearson. Los coeficientes varían entre −1 y +1; el signo indica la dirección de la asociación y el valor absoluto expresa su intensidad. La correlación no implica causalidad.

Análisis de componentes principales. Debido a que las variables se expresan en escalas y unidades diferentes, se estandarizaron antes del ACP. El análisis se realizó sobre la matriz de correlaciones. Se calcularon los porcentajes de varianza explicada y las cargas de las variables en los tres primeros componentes.

Análisis de conglomerados. Se conserva en este documento el dendrograma y la interpretación académica consignados en la versión original del trabajo. La versión recibida no especifica de manera suficiente todos los parámetros de cálculo del dendrograma (método de enlace, distancia y criterio de corte); por tanto, no se atribuyen parámetros nuevos ni se presenta el dendrograma como un recálculo independiente.

RESULTADOS

Correlación de Pearson

A partir de las 13 variables de la base de datos se calcularon 78 correlaciones por pares. La Tabla 2 presenta las 20 asociaciones con mayor valor absoluto de r. Los valores p son nominales y no están ajustados por comparaciones múltiples; por ello, deben interpretarse de manera exploratoria.

Variables	r de Pearson	p nominal	Dirección
DBO_5 – DQO	0.774	0.0001649	Positiva
CE – Cu	0.772	0.0001717	Positiva
DQO – Fe	0.757	0.0002733	Positiva
STD – DBO_5	0.738	0.0004778	Positiva
CE – Zn	0.733	0.000535	Positiva
DBO_5 – Fe	0.679	0.001958	Positiva
STD – Zn	0.677	0.002028	Positiva
CE – Fe	0.672	0.002236	Positiva
Turbidez – Fosfato	0.665	0.002626	Positiva
Cloruros – Fe	0.663	0.002722	Positiva

STD – Fe	0.656	0.003113	Positiva
CE – DBO_5	0.650	0.003527	Positiva
Fe – Zn	0.647	0.003734	Positiva
T° – Cloruros	-0.643	0.003986	Negativa
CE – STD	0.626	0.005463	Positiva
STD – Cu	0.604	0.008	Positiva
DBO_5 – Zn	0.601	0.008319	Positiva
Turbidez – Cloruros	-0.587	0.01042	Negativa
Fe – Cu	0.582	0.01128	Positiva
STD – Cloruros	0.575	0.01262	Positiva

Entre las asociaciones positivas de mayor magnitud se encontraron DBO_5–DQO (r = 0.774), CE–Cu (r = 0.772), DQO–Fe (r = 0.757), STD–DBO_5 (r = 0.738) y CE–Zn (r = 0.733). También se observó una asociación negativa entre temperatura y cloruros (r = −0.643). Estos resultados ofrecen ejemplos para que los estudiantes distingan dirección e intensidad de la asociación lineal. No permiten, por sí solos, establecer relaciones causales.

En la experiencia original, el equipo 1 presentó dificultades para identificar e interpretar las correlaciones. El equipo 2 reconoció la asociación positiva entre salinidad y pH que aparecía en la matriz utilizada originalmente. El equipo 3 distinguió asociaciones positivas y negativas y explicó ejemplos de ambas. Dado que la matriz original de la actividad contenía variables distintas a las de la base de 13 variables aquí incorporada, se conserva esta descripción como registro de la experiencia, pero se sustituye la matriz de resultados por el cálculo congruente con la base compartida.

Análisis de componentes principales

El ACP se calculó con las 13 variables estandarizadas. Los tres primeros componentes explicaron 73.71 % de la varianza total: el CP1 explicó 41.89 %, el CP2 19.73 % y el CP3 12.09 %. Esta reducción permite representar una parte importante de la variabilidad de las 13 variables mediante tres dimensiones, sin perder de vista que el resultado depende de la base de datos y del procedimiento de estandarización.

Figura 1. Porcentaje de varianza explicado por los cinco primeros componentes principales.

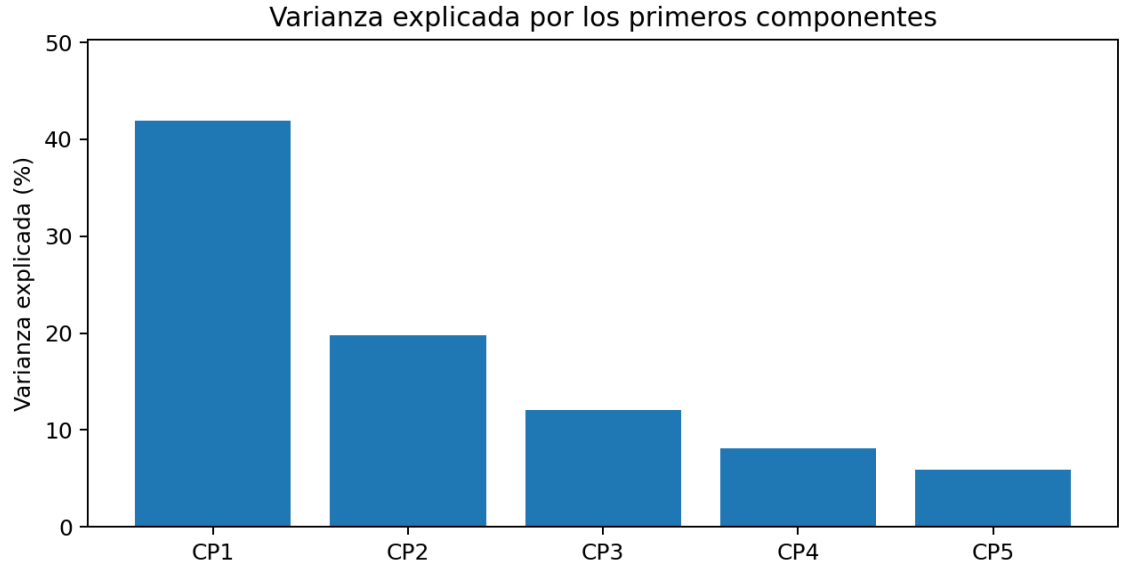


Tabla 3. Cargas de las variables en los tres primeros componentes principales. Las cargas se presentan como correlaciones entre variables estandarizadas y componentes.

Variable	CP1	CP2	CP3
pH	0.017	0.500	0.791
CE	0.811	0.271	-0.470
T°	-0.595	-0.001	-0.584
Turbidez	-0.353	0.840	-0.218
STD	0.881	0.026	-0.048
DBO_5	0.832	0.347	0.039
DQO	0.741	0.100	0.226
Fosfato	0.123	0.861	-0.132
Cloruros	0.720	-0.313	0.354
Fe	0.916	-0.068	0.064
Cu	0.796	0.041	-0.249
Zn	0.776	-0.157	-0.400
Cd	0.057	-0.829	-0.067

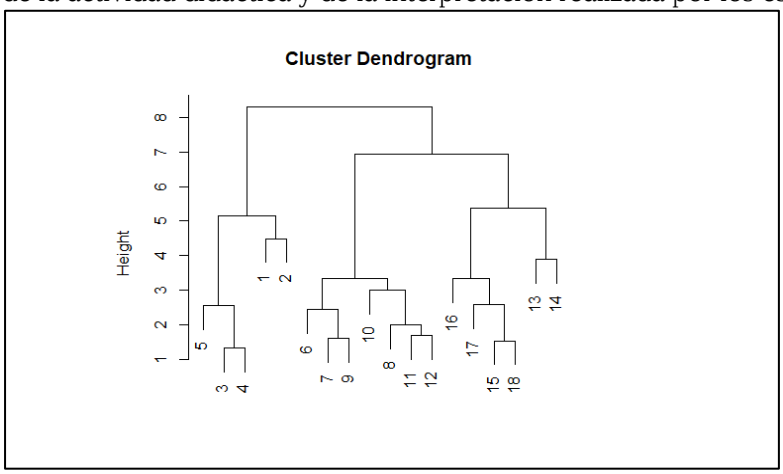
En el CP1 destacan las cargas positivas de Fe (0.916), STD (0.881), DBO_5 (0.832), CE (0.811), Cu (0.796), Zn (0.776), DQO (0.741) y cloruros (0.720), mientras que temperatura presenta una carga negativa (−0.595). En el CP2 destacan fosfato (0.861) y turbidez (0.840), con una carga negativa para Cd (−0.829). En el CP3 destacan pH (0.791), con cargas negativas de temperatura (−0.584), CE (−0.470) y Zn (−0.400). El signo de los componentes puede invertirse matemáticamente sin cambiar la solución; la interpretación debe considerar las variables que cargan conjuntamente y su contribución.

En la experiencia original, el equipo 1 no logró interpretar el ACP. El equipo 2 reconoció la presencia de varianza en los componentes, aunque su explicación fue parcial. El equipo 3 identificó la importancia de las relaciones entre variables, la varianza y la ubicación de las variables en los cuadrantes del gráfico. Esta descripción se mantiene como observación docente de la actividad.

Análisis de conglomerados (Cluster)

En la versión original, los estudiantes interpretaron un dendrograma elaborado a partir de los 18 puntos de muestreo. El equipo 1 presentó dificultades para interpretarlo. El equipo 2 identificó relaciones entre puntos, pero no explicó adecuadamente la estructura completa del dendrograma. El equipo 3 reconoció agrupaciones de puntos con similitud y distinguió dos grupos principales: el primero integrado por los puntos 6 al 18 y el segundo por los puntos 1 al 5.

Figura 2. Dendrograma reproducido de la versión original del manuscrito. Se conserva como evidencia de la actividad didáctica y de la interpretación realizada por los estudiantes



Debido a que no se recuperó el archivo de salida de XLSTAT ni la configuración completa del análisis original, el dendrograma se conserva sin modificar. La agrupación descrita corresponde a la interpretación consignada en el manuscrito original y no debe confundirse con un nuevo cálculo validado a partir de la tabla de datos incorporada en esta versión.

DISCUSIÓN

La experiencia muestra el valor didáctico de trabajar con una base de datos ambiental común para introducir técnicas estadísticas multivariadas. Al utilizar la misma información, los equipos tuvieron condiciones de partida comparables y pudieron concentrarse en comprender qué representa cada

análisis y cómo comunicar sus resultados. La actividad permitió observar distintos niveles de interpretación: desde dificultades para reconocer patrones hasta explicaciones que relacionaron coeficientes, componentes y agrupaciones.

La correlación de Pearson ofreció una oportunidad para distinguir asociaciones positivas y negativas y discutir la intensidad de las relaciones lineales. El ACP permitió abordar la estandarización, la reducción de dimensionalidad, la varianza explicada y las cargas de las variables. El dendrograma facilitó discutir la similitud entre observaciones y la forma en que un criterio de agrupamiento representa relaciones entre puntos de muestreo.

La experiencia se presenta como una descripción didáctica y no como una evaluación experimental de eficacia. No se recuperó una rúbrica, instrumento pre/post ni criterios cuantitativos de evaluación del aprendizaje; por ello, las diferencias entre equipos se reportan como observaciones de desempeño e interpretación registradas durante la actividad, sin afirmar una mejora estadísticamente demostrada. Entre las limitaciones se encuentran el tamaño reducido del conjunto de datos (18 observaciones para 13 variables), la ausencia de los archivos originales de XLSTAT y la falta de información completa sobre los parámetros del análisis de conglomerados. En consecuencia, los resultados multivariados se consideran exploratorios y vinculados a esta experiencia concreta.

CONCLUSIONES

La experiencia didáctica permitió trabajar la interpretación de datos ambientales mediante tres herramientas estadísticas: correlación de Pearson, análisis de componentes principales y análisis de conglomerados. El propósito formativo se centró en que los estudiantes pasaran de disponer de resultados analíticos a reconocer relaciones, patrones y similitudes mediante procedimientos estadísticos.

El uso de XLSTAT integrado en Excel facilitó la aplicación de los métodos y la elaboración de tablas y representaciones gráficas. Las diferencias observadas entre los tres equipos muestran que la ejecución del análisis no garantiza por sí misma una interpretación adecuada; es necesario orientar a los estudiantes en la lectura de coeficientes, varianza explicada, cargas y dendrogramas.

El empleo de datos ambientales reales constituyó un contexto práctico para la enseñanza de la estadística aplicada en Ingeniería Ambiental. La experiencia puede servir como punto de partida para diseñar actividades posteriores que incorporen guías de interpretación y criterios explícitos de evaluación del aprendizaje.

REFERENCIAS BIBLIOGRÁFICAS

Cal, R. (2001). Métodos quimiométricos en análisis medioambiental. Universidad de Santiago de Compostela.

Carlosena, A. (1999). Aplicaciones de la quimiometría en el análisis instrumental. Universidad de Santiago de Compostela.

Cela, R. (1994). Introducción a la quimiometría: aplicaciones en química analítica. Universidad de Santiago de Compostela.

Einax, J. W., Zwanziger, H. W., & Geiß, S. (2003). Chemometrics in environmental analysis. Wiley-VCH.

Ellison, S. L. R., & Williams, A. (Eds.). (2012). EURACHEM/CITAC Guide: Quantifying uncertainty in analytical measurement (3rd ed.). EURACHEM.

Esbensen, K. H., Guyot, D., Westad, F., & Houmøller, L. P. (2018). Multivariate data analysis: An introduction to multivariate calibration and classification (6th ed.). CAMO Software.

Grupo de Quimiometría. (2001). Introducción a la quimiometría. Universidad de La Rioja.

Harris, D. C. (2020). Quantitative chemical analysis (10th ed.). W. H. Freeman and Company.

Kowalski, B. R. (Ed.). (1984). Chemometrics: Mathematics and statistics in chemistry. D. Reidel Publishing Company.

Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., et al. (1997). Handbook of chemometrics and qualimetrics: Part A & B. Elsevier.

Quintana, J. B. (1994). Aplicaciones de técnicas quimiométricas en el tratamiento de datos analíticos. Universidad de Santiago de Compostela.

Santos, M. C., Nascimento, P. A. M., Guedes, W. N., Pereira-Filho, E. R., Filletti, É. R., & Pereira, F. M. V. (2019). Chemometrics in analytical chemistry—An overview of applications from 2014 to 2018. Eclética Química, 44(2), 11–25. https://doi.org/10.26850/1678-4618eqj.v44.2.11-25