

Una aplicación del análisis de componentes principales en el área educativa

An application of principal components analysis in the educational sector

Pilar González Martín*, Amelia Díaz de Pascual*, Enrique Torres Lezama* y Elsy Garnica Olmos*

Resumen

En esta investigación se aplica el Análisis de Componentes Principales, con el objeto de observar la influencia de la metodología de enseñanza y de la preparación previa del estudiante, sobre el aprendizaje. En una muestra se miden variables referidas a pruebas psicométricas y de instrucción. Se aplican dos metodologías de enseñanza: la cognoscitiva y la conductista. Los resultados de esta investigación permiten ilustrar las aplicaciones más frecuentes del Análisis de Componentes Principales: a) análisis exploratorio; b) construcción de indicadores sintéticos y c) utilización de los componentes principales como variables independientes en el análisis de regresión. En el análisis exploratorio se obtienen dos grupos de variables: uno relacionado con inteligencia y el otro con rendimiento estudiantil. La estimación de los constructos -aprendizaje e inteligencia- se realiza mediante indicadores sintéticos, productos del análisis. Finalmente, se efectúa un análisis de regresión en el cual se comprueba la influencia significativa de la inteligencia y metodología de la enseñanza, sobre el aprendizaje.

1. Introducción

El Análisis de Componentes Principales (ACP) pertenece a un grupo de técnicas estadísticas multivariantes, eminentemente descriptivas. El enfoque francés de este análisis fue desarrollado por Benzecri (1980). Posteriormente ha sido muy difundido, especialmente en el tratamiento

* Universidad de Los Andes, Facultad de Ciencias Económicas y Sociales, Instituto de Investigaciones Económicas y Sociales

de grandes masas de datos. En el presente trabajo se aplicará el ACP según el criterio de la Escuela Francesa.

El ACP permite reducir la dimensionalidad de los datos, transformando el conjunto de p variables originales en otro conjunto de q variables incorrelacionadas ($q \leq p$) llamadas componentes principales. Las p variables son medidas sobre cada uno de los n individuos, obteniéndose una matriz de datos de orden np ($p < n$).

En el ACP existe la opción de usar la matriz de correlaciones o bien, la matriz de covarianzas. En la primera opción se le está dando la misma importancia a todas y a cada una de las variables; esto puede ser conveniente cuando el investigador considera que todas las variables son igualmente relevantes. La segunda opción se puede utilizar cuando todas las variables tengan las mismas unidades de medida y además, cuando el investigador juzga conveniente destacar cada una de las variables en función de su grado de variabilidad.

Las q nuevas variables (componentes principales) son obtenidas como combinaciones lineales de las variables originales. Los componentes se ordenan en función del porcentaje de varianza explicada. En este sentido, el primer componente será el más importante por ser el que explica mayor porcentaje de la varianza de los datos. Queda a criterio del investigador decidir cuántos componentes se elegirán en el estudio.

El análisis se realiza en el espacio de las variables y , en forma dual, en el espacio de los individuos. Se acostumbra a representar gráficamente los puntos-variables y los puntos-individuos tomando como ejes de coordenadas los componentes. A veces, puede facilitar la interpretación de los resultados, el observar la similar ubicación de los puntos en los planos respectivos. Aunque el plano de puntos-variables no se superpone al plano de puntos-individuos, es de gran utilidad “interpretar” la cercanía de un grupo de puntos-individuos, a ciertas variables.

En la práctica es frecuente que se disponga de información adicional que amplía la matriz de datos originales. Se puede tener otras medidas de los individuos de la muestra, o también nuevos individuos para los que se conozcan las variables analizadas. A estos datos adicionales se les llama suplementarios o ilustrativos porque no intervienen en la formación de los componentes. En estos casos se calculan las coordenadas de cada punto

individuo o variable respecto a los ejes y se representan en los gráficos. Esto permite analizar las relaciones de la información suplementaria con los componentes principales. Generalmente la introducción de estos datos en el análisis se hace porque facilita la interpretación de los resultados.

2. Requisitos

Los únicos requerimientos previos para la aplicación del ACP son:

- a) Continuidad en las variables.
- b) El número n de individuos o elementos observados debe ser mayor que el número p de variables originales.

Por otra parte, el ACP tiene la ventaja de no exigir supuestos tales como la normalidad u homoscedasticidad.

3. Aplicaciones

Entre los usos más frecuentes del ACP están:

- 1) Como técnica de análisis exploratorio que permite descubrir interrelaciones entre los datos y de acuerdo con los resultados, proponer los análisis estadísticos más apropiados.
- 2) Reducir la dimensionalidad de la matriz de datos con el fin de evitar redundancias y destacar relaciones. En la mayoría de los casos, tomando sólo los primeros componentes, se puede explicar la mayor parte de la variación total contenida en los datos originales.
- 3) Construir variables no observables (componentes) a partir de variables observables. Por ejemplo, la inteligencia de una persona no es observable directamente, en cambio, se puede medir distintos aspectos de ésta mediante pruebas psicométricas. Las variables que miden los distintos aspectos de la inteligencia tienden a covariar; esto sugiere que expresan la mismas características pero de diferente

forma y que sólo hay un pequeño número de rasgos no directamente medibles, que se denominan Indicadores sintéticos y que vienen estimados por los componentes.

- 4) Bajo ciertas circunstancias, es de gran utilidad usar estos componentes incorrelacionados, como datos de entrada para otros análisis. Por ejemplo, en el caso de la regresión múltiple cuando las variables independientes presentan alta colinealidad es preferible hacer la regresión sobre los componentes principales en lugar de usar las variables originales.

4. Interpretación de los gráficos

Los puntos-individuos y los puntos variables se representan en los planos formados por cada par de componentes, que vienen a constituir los ejes del sistema de coordenadas.

4.1. Representación de la nube de puntos-variables

Si se toma en cuenta el plano principal (formado por el primer y segundo componente) entonces un punto-variable viene representado por la coordenada que le corresponda a esa variable en cada uno de esos componentes. La nube de puntos-variables está situada en una superficie circular de radio 1.

Las proximidades entre los puntos-variables indican el grado de correlación que existe entre ellas. Cuando la correlación es igual a uno, los puntos coinciden.

La coordenada de un punto sobre un eje es igual a la correlación de esa variable con respecto a éste. Por lo tanto, nos indica la contribución que esta variable tiene en la formación del eje. Para facilitar la interpretación se analizan los siguientes casos extremos:

- a) En el caso de que todas las p variables están incorrelacionadas, se obtendrán p componentes igualmente importantes que serán las mismas variables.

- b) En el caso de que todas las variables tengan una correlación perfecta se genera un solo componente, que es la combinación lineal de las p variables, igualmente ponderadas, y que explica el 100% de la variación total.
- c) En el caso de que se tenga una variable incorrelacionada con las demás, uno de los componentes coincidirá con esta variable.

4.2. Representación de la nube de puntos-individuos

Para interpretar la nube de puntos-individuos en un plano principal, conviene tener en cuenta los siguientes aspectos:

- Los puntos-individuos no quedan encerrados en un círculo de radio uno (1). Los puntos-variables sí.
- Un punto-individuo situado en el extremo de uno de los ejes, significa que ese individuo está muy relacionado con el respectivo componente.
- Cuando existen puntos-individuos cercanos al origen, significa que estos individuos tienen poca o ninguna relación con los dos componentes.
- Las proximidades entre individuos se interpretan como similitud de comportamiento de éstos con respecto a las variables. Por ejemplo, dos puntos-individuos que están muy cercanos en el plano, significa que ambos individuos tienen valores próximos en cada una de las respectivas variables.
- Un punto-individuo extremadamente alejado de la nube, puede significar una de estas dos cosas:
 - a) Existe un error en la introducción del dato o en la medición.
 - b) Se trata de un individuo excepcional, el cual conviene sacar del análisis principal y usarlo como individuo suplementario, o bien, en el caso de que sean varios, analizarlos por separado.En ambos casos, se requiere realizar un nuevo ACP.

Cuando se presentan varias nubes de puntos muy diferenciadas, significa que puede haber varias sub-poblaciones de individuos. Si el

propósito del estudio es detectar grupos diferentes, el ACP ha logrado su objetivo. Pero si el objetivo es estudiar la interrelación entre las variables, la aparición de varias sub-poblaciones de individuos interfiere en este análisis; entonces conviene realizar un ACP en cada una de estas sub-poblaciones.

Se pueden comparar dos planos (uno referido a la nube de puntos-individuos y el otro, a la nube de puntos-variables), para facilitar la interpretación de los resultados.

5. Criterios para seleccionar el número de componentes

Uno de los objetivos básicos del ACP es reducir la dimensionalidad de los datos. Lo ideal sería seleccionar el primero o los dos primeros componentes, ya que de esta forma se puede analizar en una recta o en un plano las posibles interrelaciones existentes entre los puntos variables y los puntos individuos.

Los criterios que más se utilizan para decidir el número de componentes a seleccionar son:

1. *Gráfico de autovalores*: consiste en representar el porcentaje de variación explicada contra el número del componente. En el eje de las ordenadas se registra el porcentaje de variación explicada. En el eje de las abscisas se coloca el número del componente según su orden de importancia de acuerdo a la variación explicada. Por lo general, los puntos del gráfico presentan una figura similar al perfil de una bota. Al analizar este gráfico se busca el punto de quiebre, donde el cambio de la pendiente se hace mayor, y la abscisa correspondiente a este punto indica el número de componentes a retener.
2. *Promedio de autovalores*: se calcula el promedio de todos los autovalores y se eliminan aquellos autovalores que están por debajo de este promedio. Si se trabaja con la matriz de correlaciones, este promedio es uno. Mediante este criterio se tiende a retener menos componentes que en el criterio anterior y aunque es más objetivo,

puede considerarse menos flexible ya que en el primer caso el investigador puede elegir el número de componente haciendo uso de su experiencia personal.

Aparte de los criterios señalados, es importante atender otros aspectos relacionados con la naturaleza de la investigación. Por ejemplo, un componente con muy poca contribución puede estar altamente correlacionado con alguna variable importante en la investigación, entonces no sería conveniente desincorporar este componente en el análisis final.

6. Aplicación a un caso particular

Para ilustrar el uso del ACP se presenta un ejemplo con el cual se ensayarán las aplicaciones más frecuentes del ACP: a) análisis exploratorio; b) construcción de indicadores sintéticos, reduciendo la dimensionalidad de la matriz de datos y c) utilización de los componentes principales como variables independientes en el análisis de regresión.

6.1. Planteamiento del problema

Se desea analizar la influencia, en el aprendizaje, de la metodología de enseñanza y de la preparación previa del estudiante.

6.2. Descripción de los datos

Se midieron ocho (8) variables en una muestra de 54 estudiantes del primer semestre (B- 89) de la Licenciatura de Biología de la Facultad de Ciencias de la ULA. Los estudiantes están repartidos en dos secciones. Una, con 34 alumnos a los cuales se les aplicó una metodología de enseñanza con enfoque cognoscitivo, y la otra, con 20 alumnos, a los cuales se les aplicó una metodología conductista. Este procedimiento genera una primera variable a considerar:

MTD (metodología): variable dicotómica que toma el valor cero (0), cuando se aplica el enfoque conductista y el valor uno (1), cuando se aplica el enfoque cognoscitivo.

A continuación se enumeran cinco (5) variables que miden distintos aspectos de la preparación del estudiante al entrar en la Universidad:

PRB (promedio de bachillerato): es la media aritmética de las notas obtenidas en el bachillerato. El rango es de 10 a 20 puntos.

RAB (razonamiento abstracto): es la puntuación obtenida en una prueba psicométrica que trata de medir la habilidad para razonar en forma no verbal. El rango de esta puntuación es de 0 a 100.

RAV (razonamiento verbal): es la puntuación obtenida en una prueba psicométrica que trata de medir la habilidad para efectuar analogías verbales y aprecia la capacidad para abstraer, generalizar y pensar haciendo uso del lenguaje. El rango de esta puntuación es de 0 a 100.

HAN (habilidad numérica): es la puntuación obtenida en una prueba psicométrica que trata de medir la habilidad para establecer relaciones aritméticas y para razonar con material cuantitativo. El rango de esta puntuación es de 0 a 100.

CTL (test de inteligencia de Catell): es la puntuación obtenida en una prueba psicométrica que mide la inteligencia en forma general, independientemente de la cultura. El rango es de 0 a 100.

Se midieron tres tipos de aprendizaje:

API (aplicación inmediata): es la puntuación obtenida en una prueba de instrucción que evalúa la habilidad para aplicar los conceptos estudiados a problemas similares a los tratados en el material de apoyo. El rango es de 0 a 20.

TPD (transferencia a problemas difíciles): viene dada por la nota obtenida en una evaluación elaborada para medir la transferencia de conocimientos aprendidos a problemas complejos, aquellos

que exigen procesos de razonamiento relativamente superiores. El rango es de 0 a 20.

TOC (transferencia a otros contenidos): es la puntuación obtenida en una prueba de instrucción que evalúa la capacidad del estudiante para transferir lo aprendido a otro contexto. El rango es de 0 a 20.

6.3. Análisis estadístico

A continuación se desarrollan los análisis estadísticos, aplicaciones éstas del Análisis de Componentes Principales.

6.3.1. Análisis exploratorio

Este análisis se realizó con la finalidad de obtener una descripción estadística de los datos para conocer sus interrelaciones. Esta información puede sugerir la técnica estadística más adecuada a las características de las variables.

En el ACP se tomaron todas las variables como activas exceptuando la metodología, que se incluyó como una variable suplementaria. Esta elección se basa en la naturaleza de las variables: la metodología es dicotómica y las demás variables son continuas.

En el cuadro de correlaciones (Cuadro 1), uno de los resultados de la salida del ACP, se observa que las variables RAV, RAB, HAN y CTL están altamente correlacionadas entre sí y presentan bajas correlaciones con respecto a las otras variables del análisis. Por otra parte, las variables TPD y TOC presentan una alta correlación con la variable API. Debido a la naturaleza de las variables API, TPD y TOC, se esperaría que ellas estuvieran fuertemente correlacionadas, ya que las tres son medidas de distintos aspectos del aprendizaje. Sin embargo, los datos muestran una correlación relativamente baja entre dos de las variables (TPD y TOC). Puede destacarse que la variable TPD presenta muy poca relación con las variables que intentan medir la inteligencia, resultado que llama la atención porque contradice lo que teóricamente se podría esperar en tales casos.

En la representación gráfica de la nube de puntos-variables (Figura 1) se observan dos grupos de variables bien definidos y separados con

relación a sus magnitudes en el eje 2. Por un lado, las cuatro variables que miden algún aspecto de la inteligencia: razonamiento abstracto (RAB), razonamiento verbal (RAV), habilidad numérica (HAN) y Catell (CTL). Por otro lado, se encuentran las variables relacionadas con el rendimiento estudiantil: promedio de bachillerato (PRB), aprendizaje inmediato (API), transferencia a problemas difíciles (TPD) y transferencia a otro contenido (TOC). Con respecto al eje uno, que recoge el 48% de la variación total, se observa que las ocho variables del análisis forman un solo grupo; siendo la variable API la que muestra mayor correlación con este primer eje.

Cuadro 1. Matriz de correlaciones

	RAV	HAN	RAB	CTL	PRB	API	TOC	TPD
RAV	1,00							
HAN	0,53	1,00						
RAB	0,61	0,40	1,00					
CTL	0,56	0,42	0,61	1,00				
PRB	0,41	0,33	0,28	0,16	1,00			
API	0,46	0,45	0,36	0,43	0,43	1,00		
TOC	0,30	0,31	0,24	0,29	0,25	0,56	1,00	
TPD	0,25	0,36	0,40	0,35	0,33	0,62	0,38	1,00

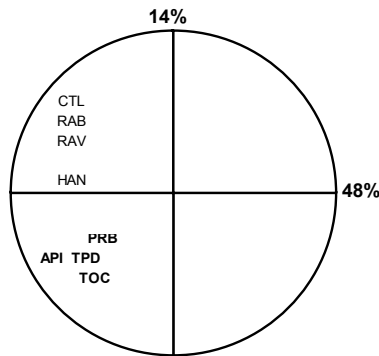


Figura 1. Círculo de correlaciones de la nube de puntos-variables en el plano principal del primer ACP

6.3.2. Construcción de indicadores sintéticos o variables no observables

Después de obtener los resultados iniciales del primer ACP (Figura 1), se ensayaron otros análisis de componentes principales. El segundo de ellos, se refería a sólo las variables que miden aspectos de la inteligencia (PRB, HAV, HAB, HAN y CTL). En este segundo análisis, el promedio de bachillerato aparece separado de las otras variables, mostrando una fuerte correlación con el eje 2 (Figura 2); mientras que las variables que miden la inteligencia están más fuertemente relacionadas con el eje 1, que recoge el 55% de la variación total. Por lo tanto, podemos concluir que el componente 1 es un indicador de la *inteligencia*, mientras que el eje 2 se puede identificar con el *rendimiento* estudiantil en bachillerato.

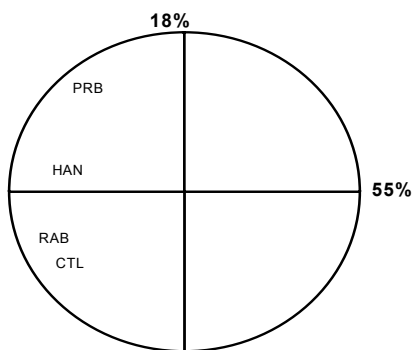


Figura 2. Círculo de correlaciones de la nube de puntos-variables en el plano principal del segundo ACP

La inteligencia de una persona está formada por tantos rasgos, que resultaría inapropiado medirla directamente con una sola prueba particular. Sin embargo, algunas manifestaciones de la inteligencia pueden ser detectadas mediante un grupo de pruebas psicológicas. Si los resultados de estas pruebas se logran resumir en una sola variable, entonces se estaría formando un indicador sintético (INT: inteligencia). Esto es justamente lo que se obtiene con el primer componente del tercer ACP (Figura 3).

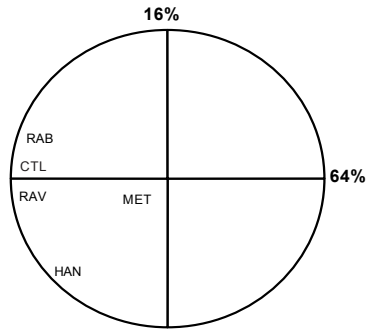


Figura 3. Círculo de correlaciones de la nube de puntos-variables en el plano principal del tercer ACP

En este ACP se introducen como variables activas las cuatro medidas de inteligencia: HAN, RAV, RAB y CATELL:

$$\text{INT} = 0,53 \text{ RAV} + 0,45 \text{ HAN} + 0,51 \text{ RAB} + 0,51 \text{ CTL}$$

De la misma forma, el aprendizaje es un concepto muy amplio, que puede ser medido parcialmente desde muchos puntos de vista. En el presente trabajo hay tres medidas del aprendizaje con las que se obtendrá, a través del cuarto ACP, un indicador sintético (APZ: aprendizaje); (Figura 4).

$$\text{APZ} = 0,62 \text{ API} + 0,57 \text{ TPD} + 0,54 \text{ TOC}$$

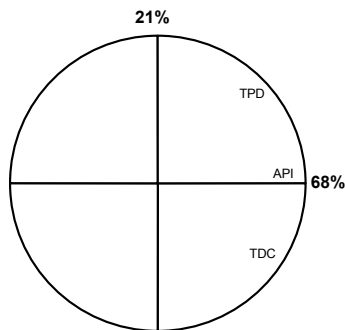


Figura 4. Círculo de correlaciones de la nube de puntos-variables en el plano principal del cuarto ACP

6.3.3. *Uso de los componentes como datos de entrada a otros análisis*

El objetivo del estudio es analizar la influencia que ejerce la metodología de enseñanza (variable MTD) y la preparación previa (variables PRB e INT), sobre el aprendizaje. Para lograrlo se hizo un análisis de regresión resultando la siguiente ecuación:

$$APZ = - 2,22 + 0,89 MTD + 0,18 PRB + 0,44 INT$$

Donde:

APZ: Aprendizaje, indicador sintético obtenido anteriormente mediante el cuarto ACP.

MTD: Metodología, variable dicotómica que toma el valor cero (0) cuando se aplica el enfoque conductista, y el valor uno (1) cuando se aplica el enfoque cognoscitivo.

PRB: Promedio de bachillerato.

INT: Inteligencia, indicador sintético obtenido anteriormente mediante el tercer ACP.

Esta ecuación explica el 42% de la variación total observada en los valores del aprendizaje. Además, según la prueba F del análisis de varianza, la regresión hallada es altamente significativa (Cuadro 2).

Cuando se observan los valores parciales de las F, para cada una de las variables independientes, se nota que la inteligencia (INT) y la metodología (MTD) son las que presentan la mayor significación (con probabilidades de 0,0002 y 0,0093). Por otra parte, la variable PRB no debería descartarse, en principio, pues su nivel de significación es de 0,0695, probabilidad considerada aceptable a un nivel del 10% (Cuadro 3).

Al observar la matriz de correlación (Cuadro 4) se detecta una correlación de 0,37 entre INT y PRB. Como existe cierto grado de relación entre estas variables independientes, sería aconsejable excluir una de ellas del análisis. De acuerdo a los resultados de la prueba F parcial (Cuadro 3), se decide eliminar la variable PRB del Modelo 1 de Regresión, pues es la que tiene mayor probabilidad (0,0695).

Cuadro 2. Análisis de varianza del modelo 1 de regresión

Fuente de Variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	F ^(3,50)	Prob(%)
Regresión	46,61	3	15,54	12,09	0,00
Residual	64,26	50	1,29		
Total	110,87	53			

Cuadro 3. Pruebas F parciales de las variables del modelo 1 de regresión

Variable	Coefficiente de regresión	Error típico	F ^(1,50)	Prob (%)
MTD	0,886	0,329	7,26	0,93
PRB	0,180	0,098	3,56	6,95
INT	0,442	0,105	17,63	0,02

Cuadro 4. Matriz de correlaciones de las variables incluidas en el modelo 1 de regresión

	MTD	PRB	INT	APZ
MTD	1,00			
PRB	0,06	1,00		
INT	0,14	0,37	1,00	
APZ	0,24	0,41	0,53	1,00

Cuadro 5. Análisis de varianza del modelo 2 de regresión

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	F ^(2,51)	Probab. (%)
Regresión	42,29	2	21,15	15,73	0,00
Residual	68,57	51	1,34		
Total	110,87	53			

La ecuación del Modelo 2 de regresión es:

$$APZ = - 0,34 + 0,51 INT + 0,96 MTD$$

Esta nueva ecuación explica el 38% de la variación total de la variable dependiente: APZ (aprendizaje). Según el cuadro del análisis de varianza de la regresión, la prueba F resulta altamente significativa ($F=15,73$) indicando que la ecuación es útil, en alguna medida, para predecir el aprendizaje en función de la metodología y de la inteligencia (Cuadro 5).

Los valores parciales de las F, para cada una de las variables también son bastante significativos, como se observa en el cuadro 6.

Cuadro 6. Pruebas F parciales de las variables del modelo 2 de regresión

Variable	Coefficiente de regresión	Error típico	$F^{(1,51)}$	Prob (%)
MTD	0,962	0,334	8,29	0,58
INT	0,515	0,100	26,81	0,00

Al observar de nuevo el cuadro de correlaciones (Cuadro 4), se puede ver que la correlación entre las variables independientes (INT y MTD) es bastante baja (0,14), y que existe una correlación relativamente alta (0,53) entre una de las variables independientes (INT) y la variable dependiente (APZ). Además, se observa que existe una correlación baja, pero no nula, entre las variables independientes (inteligencia INT y metodología MTD). Lo ideal sería que las variables independientes tengan una correlación significativamente igual a cero. Para lograr esta situación se pueden usar, como variables independientes, los componentes más importantes de un mismo ACP, situación que no es la presentada aquí ya que la metodología (MTD) es una variable dicotómica original y la inteligencia (INT) es un indicador sintético producto del tercer ACP de esta investigación.

7. Conclusiones

La presente investigación permitió ilustrar algunas aplicaciones importantes del Análisis de Componentes Principales (ACP):

1. En el análisis exploratorio, se detectan las interrelaciones entre las variables bajo estudio. Las pruebas psicométricas (razonamiento abstracto, razonamiento verbal, habilidad numérica y el test de inteligencia de Catell) constituyen un primer grupo bien diferenciado de un segundo grupo, formado por las pruebas de instrucción (aplicación inmediata, transferencia a problemas difíciles y transferencia a otros contenidos), estas últimas estrechamente relacionadas con el promedio de bachillerato. El primer grupo de variables se relacionan con inteligencia y el segundo, con el *rendimiento estudiantil*.
2. Como instrumento para la construcción de indicadores sintéticos, dos Análisis de Componentes Principales sirven para extraer dos medidas no observables o indicadores sintéticos: la *inteligencia* (INT) formada por los resultados de las cuatro pruebas psicométricas (razonamiento abstracto, razonamiento verbal, habilidad numérica y el test de Catell) y el *aprendizaje* (APZ) estructurado en base a los resultados de las tres pruebas de instrucción (aplicación inmediata, transferencia a problemas difíciles y transferencia a otros contenidos).
3. Como dato de entrada para análisis estadísticos posteriores, el ACP permite concretar una ecuación de regresión donde el aprendizaje (APZ), como indicador sintético, es la variable explicada o dependiente y las variables *inteligencia* (INT), como otro indicador sintético, y *metodología* (MTD), como variable original, son las variables explicativas:

$$APZ = - 0,34 + 0,51 INT + 0,96 MTD;$$

Esta ecuación explica el 38% de la variación del aprendizaje. Hallazgo importante que debe ser comprobado en estudios inferenciales posteriores.

8. Referencias

- Afifi, A. and S. Azen (1979). *Statistical Analysis: a Computer Oriented Approach*. 2nd ed. Academic Press, New York.
- Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Chapman-Hall, New York.
- Benzécri, J. et collaborateurs (1980). *L'analyse des donnees*. 3rd ed. Vol. 1, París.
- Chatfield, C. and A.J. Collins (1980). *Introduction to Multivariate Analysis*. Chapman and Hall, London.
- De Vega, M. (1988). *Introducción a la psicología cognitiva*. 2da. ed. Alianza Editorial, Madrid.
- Foguet, J. M. (1989). *Análisis multivariante: análisis de componentes principales*. (Colección ESADE). Editorial Hispano Europea S.A., Barcelona, España.
- Gananadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. John Wiley, New York.
- González M., P. (1991). Comparación de dos metodologías de enseñanza en un curso de Química teórica. Mérida, Universidad Nacional Abierta, Centro Local Mérida (104). Mimeografía: Tesis para Ms. Sc. en Educación Abierta y a Distancia.
- González M., P. (1987). Transferencia de los procesos de pensamiento a la resolución de problemas de Química. Mérida, ULA, Departamento de Química, Facultad de Ciencias. Mimeografía: Informe de Año Sabático.
- González M., P. (1988). "Indicadores sintéticos del rendimiento estudiantil". *Economía* (2), 69-83.
- Gujarati, D. (1981). *Econometría básica*. McGraw-Hill, Bogotá.
- Harris, R. J. (1975). *A Primer of Multivariate Statistics*. Academic Press, New York.
- Institut Technique des Céréales et des Fourrages, Ed. (1988). STAT-ITCF, manual de utilización. (Paquete estadístico) ITCF, París.
- Johnson, R. A. and D. W. Wichern (1982). *Applied Multivariate Statistical Analysis*. Prentice-Hall, Englewood Cliff, N.J.
- Jolliffe, I. (1973). "Discarding Variables in a Principal Component Analysis". *App. Stat.* 22, 21-31.

- Lebart, L., A. Morineau et N. Tabard (1977). *Techniques de la description statistique*. Dunod, París.
- Márquez M., V. A. (1989). “Apuntes sobre análisis multivariante”. Vol. 1: Análisis de componentes principales. IEAC, Facultad de Economía, ULA, Mérida. Mimeografía.
- Peñaloza, A. (1987). “Determinación de variables que influyen en el rendimiento académico de estudiantes en Química Básica”. *Investigación Educativa*. 14 (29).
- Pla, L. (1986). *Análisis multivariado: método de componentes principales*. Departamento de Asuntos Científicos y Tecnológicos, OEA. Eva V. Chesneau, SGO de la OEA, Washington, D.C.
- Tatsuoka, M. (1971). *Multivariate Analysis: Techniques for Educational and Psychological Research*. John Wiley & Sons, Inc., New York.
- Yamane, T. (1974). *Estadística*. Harla, México.