



**ARTÍCULO ORIGINAL  
INVESTIGACIÓN DE OPERACIONES**

**Análisis multivariado de datos como soporte a la decisión en la selección de estudiantes en proyectos de software**

***Multivariate data analysis as decision making support in student selection in software projects***

**Hugo Arnaldo Martínez-Noriegas, Bolívar Ernesto Medrano-Broche, Lytyet Fernández-Capestany, Yunier Emilio Tejada-Rodríguez**

Universidad de las Ciencias Informáticas (UCI). La Habana, Cuba  
E-mail: hugomn@uci.cu, bolivar@uci.cu, lytyet@uci.cu, yuniere@uci.cu

*Recibido: 28/03/2011*

*Aprobado: 11/02/2012*

**RESUMEN**

La selección de personal es un proceso vital que tiene influencia directa en el éxito de cualquier organización. En este trabajo tiene como objetivo generar información de soporte a la decisión en la selección de estudiantes para su vinculación a proyectos de *software*. Para este fin se aplican técnicas del análisis multivariados a las calificaciones obtenidas por estudiantes de segundo año de la carrera Ingeniería en Ciencias Informáticas. Para reducir la cantidad de variables en estudio se utiliza el análisis de componentes principales y basado en la información resumida, se emplea el análisis de *cluster* para formar 3 grupos. A través del análisis factorial común, se lograron identificar 3 factores latentes que actúan sobre diferentes grupos de asignaturas. La información generada es utilizada como soporte a la toma de decisiones para formular estrategias en el trabajo de formación desde la producción.

**Palabras clave:** análisis de componentes principales, análisis de *cluster*, análisis factorial, selección de personal.

**ABSTRACT**

*Personnel selection is a vital process that has a direct influence on the success of any organization. This paper aims to generate information for decision support in the selection of students for software projects. The multivariate data analysis techniques are applied to the data set of academic qualifications of Computer Science Engineering's second year students. The principal component analysis is used in order to reduce the number of variables under study and based in the summarized information, it is utilized the cluster analysis to form 3 groups. Through the factor analysis, it was possible to identify 3 latent factors that act on different groups of subjects. The generated information is used as a support for the decision-making to develop strategies on the training job from production.*

**Key words:** *principal component analysis, cluster analysis, factorial analysis, personnel selection.*

## I. INTRODUCCIÓN

Los recursos humanos son de vital importancia para la industria del *software*, pues muchos de los fracasos de los proyectos de desarrollo de *software* están asociados al factor humano [1; 2]. Sin embargo, las diferentes metodologías que guían el proceso de desarrollo se centran en aspectos técnicos. Según Ampuero y Baldoquín (2010), algunos de los principales modelos de proceso de gestión de recursos humanos, como: People-CMM, el Proceso de Software Personal (PSP), el Proceso de Software en Equipo (TSP) y el Proceso Unificado de Rational (RUP); tienen poca formalización en cuanto a la selección y asignación de personal [3].

La Universidad de las Ciencias Informáticas (UCI) es una organización en la que se mezclan los procesos de formación de ingenieros y desarrollo de *software* [4]. La mayor parte del capital humano vinculados al desarrollo de *software*, son estudiantes. Su incorporación a los proyectos se realiza de forma masiva, una vez que éstos concluyen el tercer año de la carrera. Esta práctica en la gestión de recursos humanos resulta perjudicial, si no se tienen identificados qué estudiantes tienen potencial para desempeñarse en actividades importantes dentro de un proyecto. Dada la elevada matrícula existente en la UCI, se torna complejo el proceso de selección de estudiantes para la vinculación a proyectos. Por tales razones, se deben realizar estudios exploratorios de la masa estudiantil, que permitan clasificar e identificar características en los alumnos que forman los proyectos.

En la literatura aparecen reportados varios trabajos relacionados con la gestión de procesos universitarios y la selección de recursos humanos. Respecto a la gestión de procesos universitarios, Estévez y García (2007) emplean Análisis de Componentes Principales (ACP), con el fin de construir índices para medir el desarrollo en la educación superior [5]. Rodríguez y Haedo (2009), determinan la forma de sintetizar la calidad de los resultados académicos de un alumno en el primer año de la carrera Ingeniería Industrial, con un mínimo de pérdida de información, usando ACP, Análisis de clusters (AC) y Análisis Factorial Común (AFC) [6]. También Heredia y otros (2012), desarrollan una metodología basada en la utilización de técnicas de análisis multivariado, para obtener información relevante sobre las características de aprendizaje de estudiantes [7]. Las ecuaciones estructurales, los algoritmos de minería de datos para la asociación de atributos, árboles de clasificación y regresión, redes de Kohonen y redes neuronales; también han sido aplicadas con éxito en estos contextos [8].

En la selección de personal también se han empleado técnicas de análisis multivariado, como el caso del trabajo de Mays (1976), donde se aplica ACP y AFC para la obtención de factores latentes de 15 variables empleadas para la selección de inspectores fiscales y diplomáticos [9]. En la selección de personal, ha sido muy utilizada la lógica borrosa. Merigó (2008) emplea operadores OWA (*Ordering Weigthed Averaging*) para la selección de recursos humanos, inversiones e inmobiliario [10]. Canós y colaboradores (2008) utilizan subconjuntos *borrosos* discretos y la ordenación mediante distancias para la selección de personal y el diseño de plantillas laborales óptimas [11]. Una característica común en estos trabajos es el hecho que las técnicas son aplicadas a datos que presentan incertidumbre. También relacionado con el tema de selección, existen otros enfoques basados en optimización multiobjetivos para apoyar el proceso de asignación de personal en las organizaciones de *software* [3].

En este trabajo se aplican técnicas de análisis multivariado en el mismo sentido que lo hacen Rodríguez y Haedo (2009) y Heredia y otros (2012) [6; 7]. A diferencia de estos trabajos, se tiene como objetivo generar información en el proceso de selección de los estudiantes que se vinculan a proyectos de desarrollo de *software* en la UCI. A continuación se describe la secuencia de pasos tenida en cuenta para realizar el estudio exploratorio de datos. Luego se muestran y se discuten los resultados obtenidos tras la aplicación.

## II. MÉTODOS

En aras a extraer la información de los datos, se establece una secuencia lógica de pasos para la aplicación de las técnicas de análisis multivariado.

### 1. Recolección de datos

En este caso se utilizan los datos correspondientes a las notas finales de los estudiantes de un determinado año. Como el principal objetivo es contar con información sobre los estudiantes que van a ingresar a los proyectos de desarrollo de *software*, se sugiere que las asignaturas seleccionadas sean las que aporten más a la formación técnica del estudiante.

# ANÁLISIS MULTIVARIADO DE DATOS COMO SOPORTE A LA DECISIÓN EN LA SELECCIÓN DE ESTUDIANTES EN PROYECTOS DE SOFTWARE

## 2. Preparación de los datos

Para procesar los datos se debe tener en cuenta el tipo de datos en que vienen expresadas las variables (las notas del sistema educacional cubano se expresan en una escala ordinal, sin intervalo de magnitud definida e igual entre cualquier par de niveles consecutivos de la variable [6]; por tanto, se deben transformar a numérico).

## 3. Reducción

ACP: Permite la exploración y reducción de la dimensión de los datos, ayuda a los investigadores a que adquieran cierta percepción respecto a un conjunto de datos, y a encontrar la verdadera dimensión de los datos [12; 13]. A través del ACP se puede seleccionar un número menor de variables que son combinaciones lineales de las anteriores y es posible sintetizar grandes cantidades de información para una mejor comprensión de la misma. El ACP determina una forma de resumir los resultados académicos de un alumno, con la mínima pérdida de información, así como establece relaciones hipotéticas de variables y estructuras latentes; por lo que puede brindar información con carácter exploratorio y confirmatorio. Según Estévez y García (2007), "permite construir un puente entre la teoría y los datos" [5].

Análisis de Componentes Principales Robusto (ACPR): Al realizar el ACP se debe tener precaución con las observaciones atípicas, ya que éstas son capaces de incrementar artificialmente la varianza, haciendo que las componentes principales sean atraídas por ellas. Esto, sin dudas, es una característica no deseada en la reducción de la dimensión que aparece, principalmente, con la estimación clásica de las componentes principales. Por tal motivo, se sugiere el uso del ACPR el cual determina las direcciones estimando la varianza por una manera robusta en lugar de la varianza clásica [14; 15].

El ACPR puede detectar dos tipos de observaciones atípicas: los puntos de palanca (*leverage*) y las observaciones atípicas ortogonales. Una vez realizado el ACP, se escogen las componentes que explican la mayor variabilidad de los datos. La cantidad de componentes seleccionados representa la dimensión del espacio de las componentes principales, la cual será utilizada para obtener los gráficos de diagnóstico de las componentes principales del ACPR. Las observaciones atípicas, pueden ser retiradas del análisis después de ser analizadas y luego volver a aplicar el ACP. Retirar los atípicos permite que la dirección de las componentes principales se encuentre en el sentido de la mayor variabilidad de los datos. No obstante, la eliminación de observaciones en un ACP depende de varios factores que deben ser tenidos en cuenta por el investigador [12; 13; 14; 15].

## 4. Clasificación y Agrupación

Análisis de *Cluster*: permite confirmar hipótesis sobre la naturaleza de las observaciones y mediante el mismo es posible idear una clasificación o esquema de agrupación [12; 13]. Además, permite dividir los estudiantes en estudio en clases o grupos, en los cuales sus integrantes comparten similitud. El propósito de este método en el procedimiento que se propone, es identificar 3 grupos de estudiantes, cuya existencia se asume basada en la experiencia de expertos en labor docente y productiva en la universidad.

Los estudiantes se clasifican en 3 grupos: aptos, aptos con limitaciones y no aptos. Los aptos son aquellos que podrán integrarse directamente a un proyecto de desarrollo y podrán comenzar a desempeñarse en un rol afín. Los aptos con limitaciones son los que presentan limitaciones técnicas y, por tanto, deberán recibir capacitación básica para poder desempeñarse en un proyecto o tendrán responsabilidades de menor importancia. Los no aptos son aquellos que no pueden formar parte de un proyecto de producción, para los cuales se trazarán estrategias con vistas a elevar su nivel para su futura incorporación.

## 5. Identificación de estructura y reducción de los datos

Análisis Factorial Común: tiene como objetivo identificar un número relativamente pequeño de factores subyacentes (latentes), que reflejan qué es lo que las variables tienen en común. Este análisis puede asistir en la selección de un subgrupo de variables representativas o, incluso, crear nuevas variables como sustitutas de las variables originales. Además, permite establecer la relación de las variables mediante el resumen de los datos y descubrir variables subyacentes que pueden ser usadas en estudios posteriores [12; 13].

## III. RESULTADOS

En esta sección se presentan los resultados obtenidos mediante la aplicación de la secuencia de pasos que aparecen en el apartado anterior. Se tomaron las calificaciones finales de 144 estudiantes obtenidas en las asignaturas del segundo año, pues en la universidad, una vez que el

estudiante ha vencido el segundo año, debe contar con conocimientos y habilidades que le permiten integrar un proyecto de desarrollo de *software* y continuar con su formación.

Dada la cantidad de variables en estudio se hace necesario reducir su número. La reducción de variables con poca pérdida de información solo es posible realizarla si las calificaciones guardan algún tipo de relación entre sí; esto se constata a partir de la matriz de correlaciones de Spearman, que se muestra en la tabla 1. Se usaron correlaciones de Spearman y no de Pearson, teniendo en cuenta que las variables tienen escala ordinal, pues el sistema de calificaciones académicas utilizado en la educación superior cubana, representa un tipo de dato ordinal. La inspección visual de la matriz de correlación se hizo para detectar posibles relaciones no lineales que, de presentarse, inutilizarían los resultados del ACP y el AFC.

Las asignaturas que se tuvieron en cuenta para realizar el análisis fueron: Idioma Extranjero III (IE-III), Idioma Extranjero IV (IE-IV), Economía Política I (EP-I), Economía Política II (EP-II), Física I (F-I), Física II (F-II), Programación II (P-II), Máquinas Computadoras I (MC-I), Máquinas Computadoras II (MC-II), Matemática III (M-III), Matemática IV (M-IV) y Sistema Base de Datos (SBD). Se retiró del análisis la asignatura Educación Física (EF-III y EF-IV), ya que está muy poco correlacionada con el resto. Además, las calificaciones de estas materias presentan poca variabilidad.

**Tabla 1.** Matriz de Correlación.

Asignatura	EP-I	EP-II	F-I	F-II	IE-III	IE-IV	M-III	M-IV	MC-I	MC-II	P-II	PP-II	SBD
EP-I	1												
EP-II	.414	1											
F-I	.223	.254	1										
F-II	.261	.275	.456	1									
IE-III	.277	.273	.271	.204	1								
IE-IV	.239	.345	.317	.261	.786	1							
M-III	.254	.377	.445	.487	.243	.268	1						
M-IV	.313	.307	.281	.547	.173	.196	.463	1					
MC-I	.383	.392	.418	.452	.415	.409	.562	.446	1				
MC-II	.302	.287	.454	.527	.344	.444	.475	.379	.651	1			
P-II	.299	.194	.461	.446	.339	.352	.405	.382	.519	.641	1		
PP-II	.234	.064	.199	.224	.4	.441	.239	.208	.317	.372	.283	1	
SBD	.349	.369	.359	.472	.384	.368	.5	.42	.555	.562	.457	.335	1

La matriz de correlación de los datos muestra que entre las asignaturas técnicas: F-I y F-II, M-III y M-IV, MC-I y MC-II, P-II y SBD existe una correlación entre 0,375 y 0,675. Las asignaturas de IE-III y IE-IV tienen una fuerte correlación entre sí y mayor que 0,7 y con respecto a la PP, tienen correlaciones mayores que 0,4. Las asignaturas EP-I y EP-II, solo tienen correlación entre sí y es mayor que 0,4.

De forma general, todas las variables (asignaturas) seleccionadas tienen al menos un coeficiente de correlación con otra variable mayor que 0,30. Además, se realizó el test de Bartlett para un nivel de significación de 0,05; cuyos resultados condujeron al rechazo de la hipótesis nula, lo cual implica que existe multicolinealidad entre las variables.

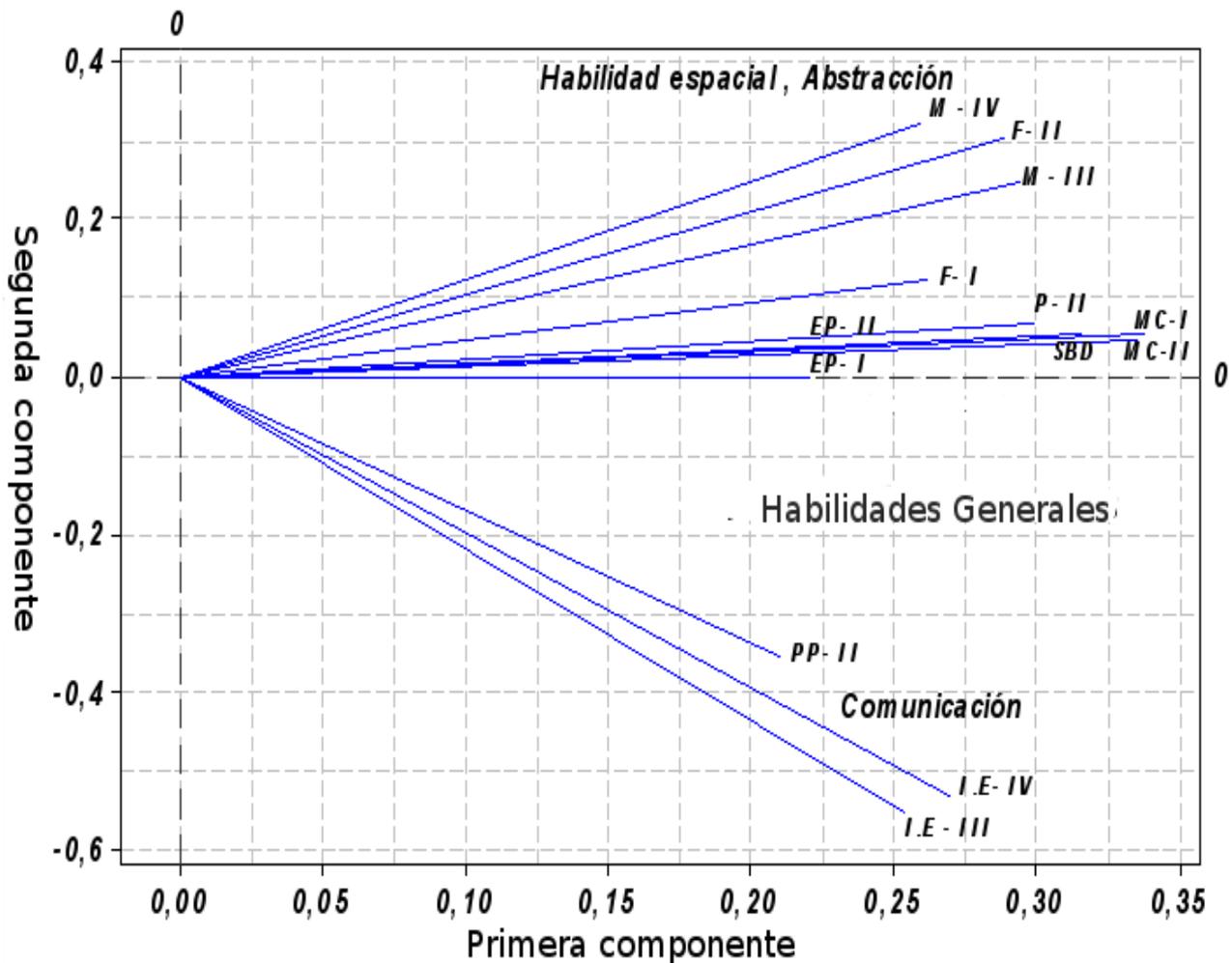
A partir del análisis de componentes principales con 13 asignaturas, se obtuvo que las 3 primeras componentes principales, tienen autovalores mayores estrictos que uno (5,516; 1,56 y 1,091; respectivamente) las cuales resumen una variabilidad total de un 61,9 % (explican 8 asignaturas aproximadamente). Para resumir más del 90 % de la variabilidad, sería necesario seleccionar 9 componentes principales, lo que complejizaría el análisis.

## ANÁLISIS MULTIVARIADO DE DATOS COMO SOPORTE A LA DECISIÓN EN LA SELECCIÓN DE ESTUDIANTES EN PROYECTOS DE SOFTWARE

En este trabajo serán tomadas solamente 3 componentes, pues para realizar estudios de esta naturaleza (Ciencias Sociales), varios autores recomiendan trabajar con más del 60 % de la variabilidad acumulada. En la tabla 2 se resumen los resultados de este análisis. Los gráficos de las cargas factoriales se muestran en las figuras 1 y 2. La discusión de los resultados del ACP se presenta en la próxima sección.

**Tabla 2.** Resumen del Análisis de Componentes Principales.

Componentes	Valor Propio	Proporción de Varianza
Primera	5,516	0,424
Segunda	1,456	0,112
Tercera	1,069	0,082
Proporción de Varianza Total = 0,619		



**Figura 1.** Análisis de las primeras componentes principales.

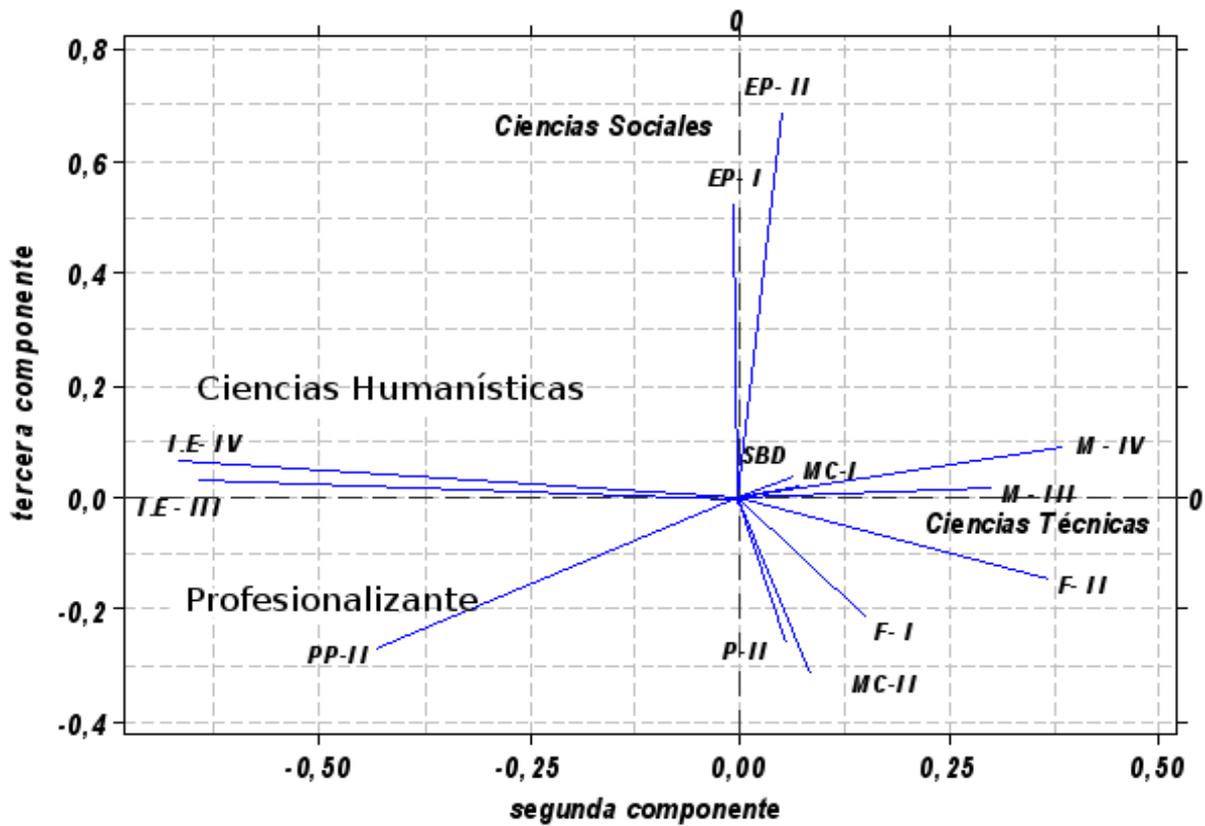


Figura 2. Análisis de la segunda y tercera componentes principales.

En las figuras 3 y 4 se muestran los gráficos de diagnóstico del ACPR creados en R a partir del paquete Chemometrics [16]. Se aprecia que solo existe una observación atípica del tipo palanca bueno. Se puede observar en la figura 4, que no existe ninguna observación con una distancia ortogonal significativa.

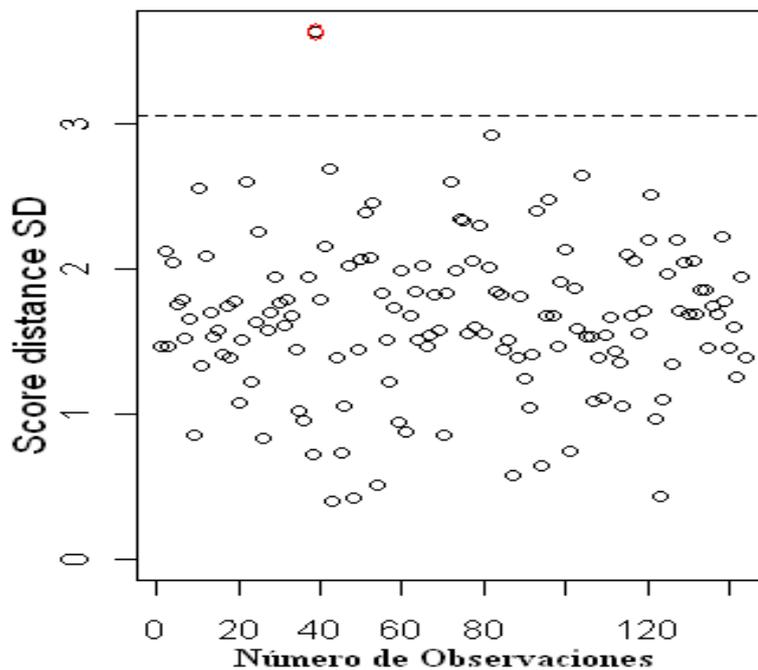
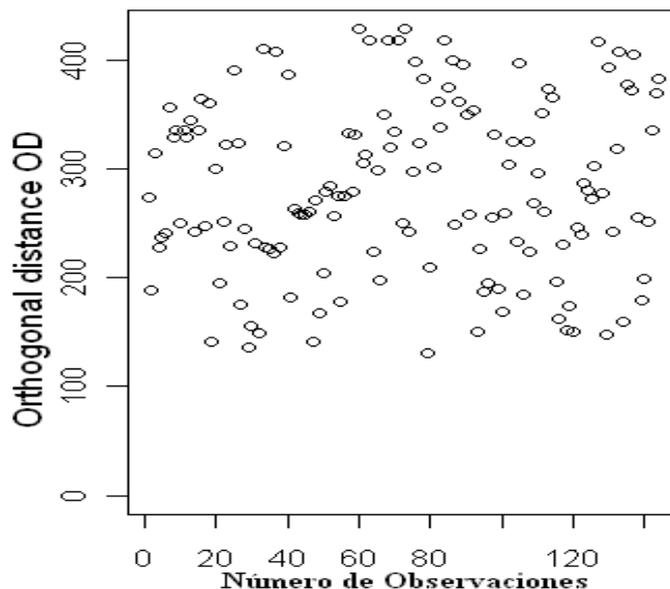


Figura 3. Gráfico de Diagnóstico.

## ANÁLISIS MULTIVARIADO DE DATOS COMO SOPORTE A LA DECISIÓN EN LA SELECCIÓN DE ESTUDIANTES EN PROYECTOS DE SOFTWARE



**Figura 4.** Gráfico de Diagnóstico.

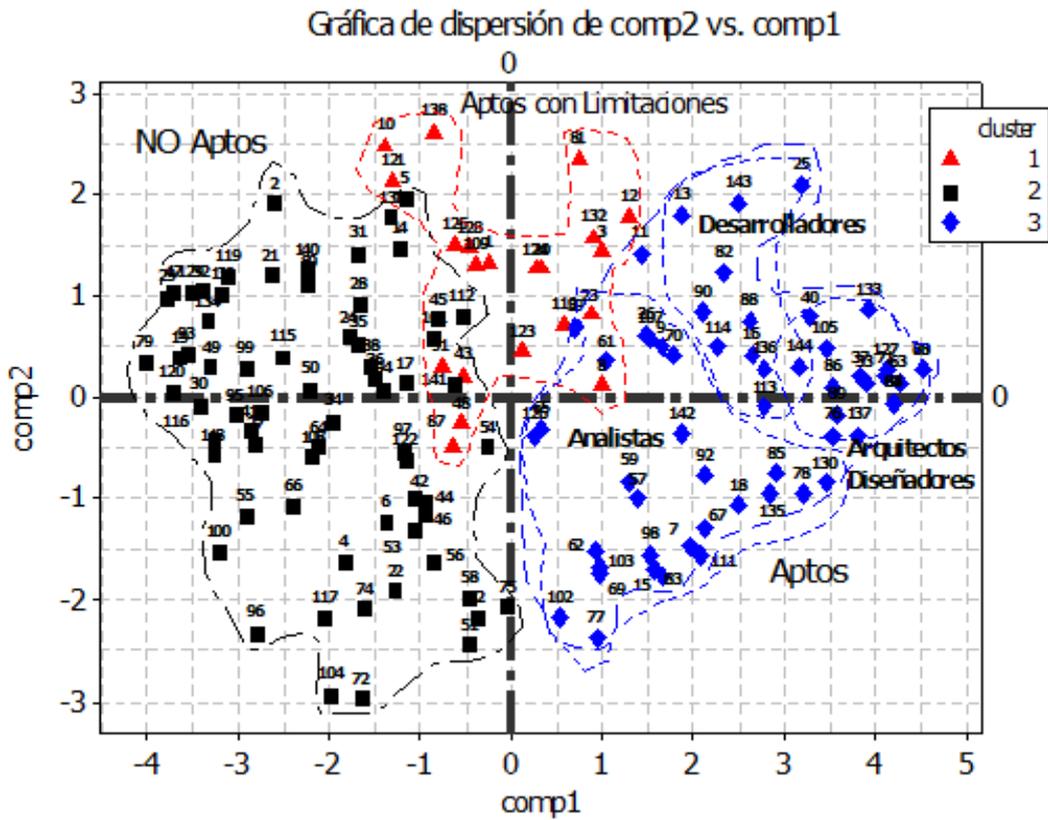
En este trabajo se asume por el conocimiento empírico del proceso, la existencia de 3 *clusters* de estudiantes con rendimientos académicos diferentes. Para formar los 3 *clusters*, se utilizaron métodos de *cluster* jerárquicos y se combinaron varios enlaces y distancias. La mejor combinación resultó ser la del enlace de McQuitty, con la distancia de Pearson Cuadrada, con una similitud entre *cluster* del 65 %. Los resultados alcanzados se resumen en la tabla 3.

**Tabla 3.** Resumen del análisis de *cluster* empleando distancia de Pearson Cuadrada y enlace de McQuitty.

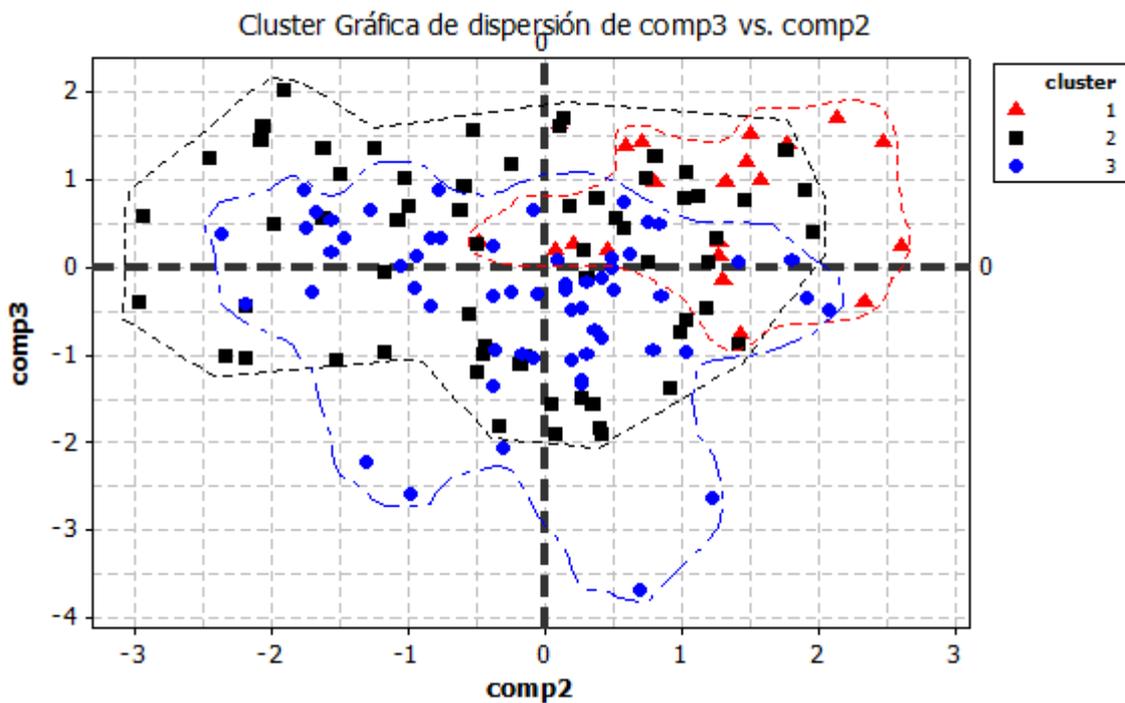
<i>Cluster</i>	Número de Observaciones		Similitud <i>Cluster</i>
1	21		73%
2	65		68,4%
3	58		67,9%
Distancia entre <i>Clusters</i>			
<i>Cluster</i>	1	2	3
1	0,0	98,5	117,76
2	98,5	0,0	165,47
3	117,76	165,47	0,0

En las figuras 5 y 6 se presentan los gráficos de dispersión de la componente 1 *versus* componente 2, y componente 2 *versus* componente 3 con los *clusters* formados. Estos gráficos permiten visualizar las características que presentan los estudiantes en los tres grupos resultantes a partir del análisis de componentes principales.

En el análisis factorial con rotación Varimax se observa que cada variable carga en un solo factor, explicando la mayoría de la variabilidad presente en los datos con un valor en cada una de sus comunales mayor que 0,50, como se observa en la tabla 4. Como puede notarse en las figuras 7 y 8, mediante la representación gráfica de los factores, se determina que las asignaturas técnicas se explican en el primer factor, las Ciencias Humanísticas en el segundo factor y en el tercer factor, se explican las asignaturas de Ciencias Sociales.



**Figura 5.** Clusters en gráfica de dispersión componente 2 versus componente 1.

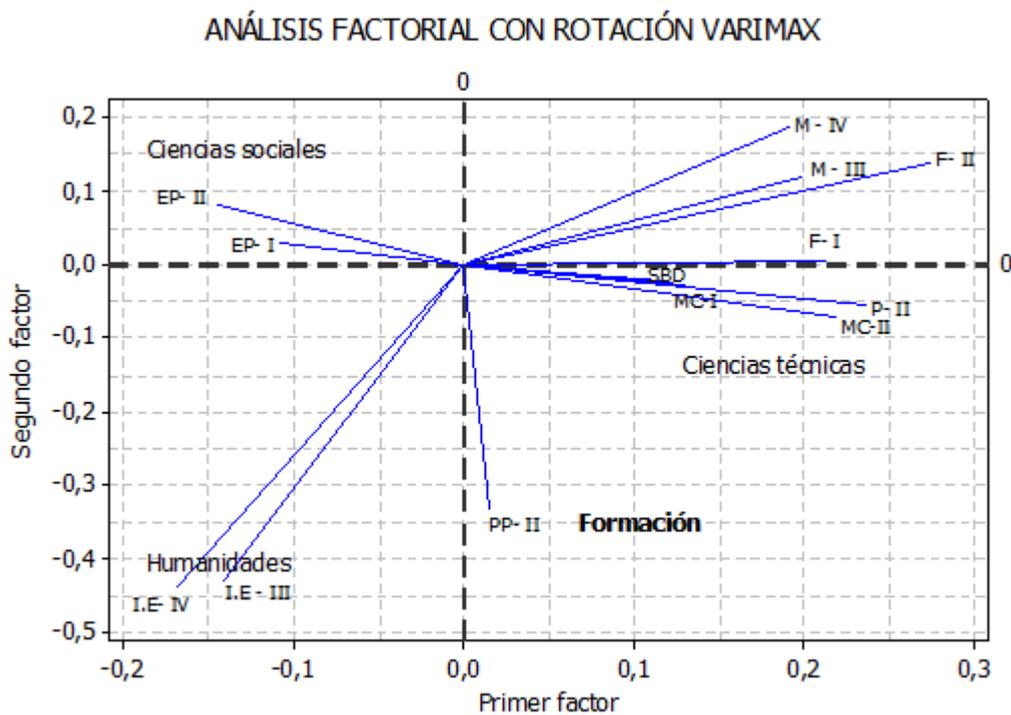


**Figura 6.** Clusters en gráfica de dispersión componente 3 versus componente 2.

## ANÁLISIS MULTIVARIADO DE DATOS COMO SOPORTE A LA DECISIÓN EN LA SELECCIÓN DE ESTUDIANTES EN PROYECTOS DE SOFTWARE

**Tabla 4.** Resumen del análisis factorial.

Análisis factorial con rotación Varimax				
Variable	Fact-1	Fact-2	Fact-3	Comunalidades
EP-I	0,21	-0,171	<b>0,688</b>	0,546
EPII	0,181	-0,105	<b>0,841</b>	0,751
F- I	<b>0,635</b>	-0,201	0,053	0,446
F- II	<b>0,769</b>	-0,035	0,153	0,617
I.E-III	0,105	<b>-0,854</b>	0,251	0,803
I.E-IV	0,162	<b>-0,853</b>	0,239	0,811
M-III	<b>0,687</b>	-0,074	0,301	0,568
M-IV	<b>0,639</b>	0,052	0,341	0,527
MC-I	<b>0,647</b>	-0,324	0,331	0,632
MC-II	<b>0,737</b>	-0,371	0,078	0,688
P-II	<b>0,707</b>	-0,315	-0,005	0,599
PP-II	0,269	<b>-0,648</b>	-0,076	0,497
SBD	<b>0,601</b>	-0,299	0,321	0,554
Varianza	3,8844	2,398	1,7581	8,0405
% Var	0,299	0,184	0,135	0,619



**Figura 7.** Rotación Varimax factor 1 versus factor 2.

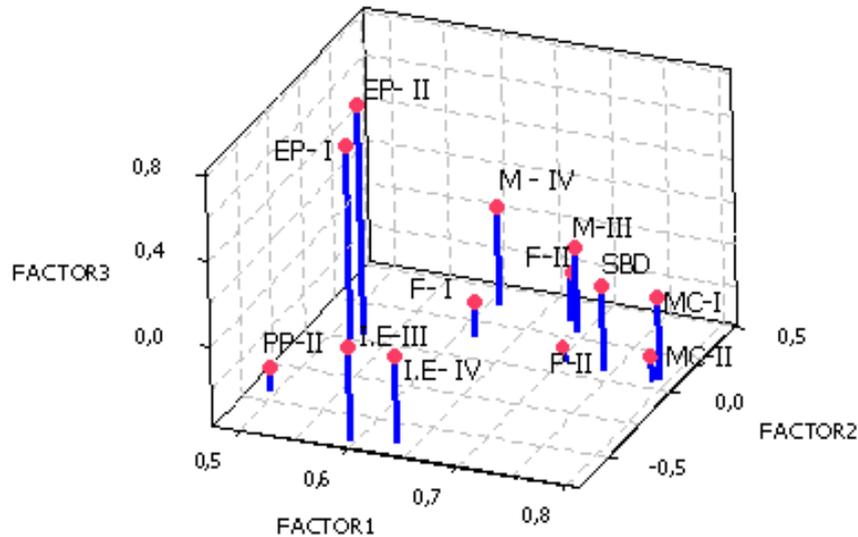


Figura 8. Gráfico de los tres factores.

## IV. DISCUSIÓN

### Análisis de componentes principales

En el gráfico de las cargas factoriales de la primera y segunda componentes, mostrado en la figura 1 se aprecia que todas las asignaturas cargan en el sentido positivo de la primera componente. Esto significa que usualmente, un estudiante con altas (bajas) calificaciones en una asignatura, también obtendrá altas (bajas) calificaciones en las demás. Este comportamiento ha sido hallado en entornos diferentes en otros estudios de rendimiento estudiantil [6; 7; 12], y se ha asociado a un factor denominado "capacidad general de aprendizaje del estudiante".

Las asignaturas: MC-I, MC-II, P-II y SBD; son las que cargan más sobre la primera componente que el resto, como se observa en la figura 1. Estas materias son las que más contribuyen a diferenciar a los estudiantes, pues son las que mayor variabilidad le introducen a los resultados. También se aprecia en la figura 1 que el grupo compuesto por las asignaturas M-III, M-IV y F-II; cargan en sentido positivo de la segunda componente. Esto se debe a que por lo general existen grupos de estudiantes que tienen un mayor desarrollo de la habilidad espacial, reflejándose ello en sus resultados y asimilación de materias tales como: física y matemática. Por otra parte, asignaturas como: PP-II, IE-III y IE-IV; cargan en sentido negativo en la segunda componente, lo que se traduce en que existen grupos de estudiantes con mayores habilidades para la comunicación y el trabajo en equipo.

Al analizar la segunda con la tercera componente, se puede apreciar la contraposición entre dos grupos de asignaturas, hacia el sentido positivo de la tercera componente cargan las asignaturas de ciencias técnicas y en el sentido negativo las asignaturas de humanidades y de carácter profesional.

### Análisis de componentes principales robustos

La observación atípica, detectada mediante el ACPR, se retiró de la muestra y se volvió a repetir el ACP donde se obtuvieron resultados similares. No obstante, se analizó de manera independiente al estudiante al cual correspondía esta observación, y resultó ser un alumno que posee un índice académico de 4,23; pero que sus calificaciones en las asignaturas que correlacionan fuerte son muy diferentes (por ejemplo: 5 puntos en P-II y 3 en MC-II, con una correlación de 0,641). El estudio de este caso particular queda pendiente a futuras investigaciones, donde se analice con profundidad este tipo de comportamiento. Por lo antes expuesto, no se toma la decisión de retirar del análisis al estudiante.

## ANÁLISIS MULTIVARIADO DE DATOS COMO SOPORTE A LA DECISIÓN EN LA SELECCIÓN DE ESTUDIANTES EN PROYECTOS DE SOFTWARE

### Análisis de Cluster

Según la interpretación realizada en el ACP, se puede observar cómo el *cluster* 3 (azul) que se forma en la parte positiva de la primera componente y hacia la parte positiva y negativa de la segunda, se encuentran estudiantes que presentan habilidades generales en todas sus asignaturas y tienen índices académicos superiores a los 4 puntos. En este *cluster* se ubican estudiantes cercanos al cero de la segunda y bien retirados en la primera componente, que son aquellos que obtienen calificaciones entre 4 y 5 en las asignaturas de ciencias técnicas, principalmente en SBD, P-II, MC-I y MC-II.

Estudiantes con tales características coinciden en ser integrales, con habilidades en todas las materias tomadas en cuenta y promedios en el intervalo de [4,45; 5]. En este *cluster* también se ubican estudiantes en la parte positiva y negativa respecto a la segunda componente y no tan retirados respecto a la primera. Los que se ubican hacia la parte positiva son aquellos que obtienen calificaciones entre 4 y 5 en las asignaturas que requieren capacidad de abstracción, como: M-III, M-IV, F-I y F-II; y poseen menos habilidades para los idiomas (IE-III, IE-IV). Hacia la parte negativa se encuentran los que tienen calificaciones de 4 y 5 puntos en idioma extranjero y mantienen 4 en las asignaturas de ciencias técnicas.

El *cluster* 1 (rojo) que se encuentra hacia la parte positiva de la segunda y en la parte positiva y negativa de la primera, es donde se encuentran los estudiantes con resultados académicos que oscilan en intervalo de [3,85; 4,30]. Las mejores calificaciones de estos estudiantes se alcanzan en EP-I, EP-II y asignaturas técnicas, como: M-III, M-IV, F-I y F-II; donde alcanzan resultados generalmente de 4 puntos.

El otro *cluster* 2 (negro) es el de los estudiantes con más bajo desempeño académico, cuyos promedios oscilan en el intervalo [3,10; 3,75]. Las calificaciones de este grupo en las asignaturas de ciencias técnicas suelen ser de 3 puntos.

El gráfico de la figura 6, que representa la dispersión de la tercera componente *versus* la segunda, permite establecer la clasificación de los estudiantes en cuanto a las asignaturas que dominan. En él se puede apreciar cómo se entremezclan los 3 *clusters*, así como se mantiene que los mejores estudiantes son los que tienen buenos resultados en las asignaturas SBD, P-II, MC-I y MC-II; y se ubican en el sentido de la proyección de estas materias. Se les debe prestar atención a algunos estudiantes que se encuentran en el *cluster* 2, que tienen buenos resultados en algunas asignaturas como SBD y P-II, ya que se muestran desmotivados por asignaturas que no son del perfil de la carrera.

### Análisis Factorial

En el gráfico de la proyección de las variables con el factor 1 *versus* factor 2, se aprecia que, respecto a la parte positiva del primer factor, se agruparon las asignaturas de ciencias técnicas (SBD, P-II, M-III, M-IV, F-I y F-II, MC-I y MC-II), siendo éstas las más representativas del primer factor con mayor correlación sobre la primera componente. En la parte negativa del primer factor, cargan las Ciencias Sociales y Humanísticas (EP-I, EP-II, IE-III e IE-IV). La asignatura PP-II como disciplina integradora (integra todas las habilidades y conocimientos adquiridos de otras asignaturas), se comporta neutra con respecto al primer factor.

Con respecto al segundo factor, se contraponen las asignaturas técnicas, las básicas (M-III, M-IV, F-I y F-II) hacia la parte positiva, y las prácticas (SBD, P-II, MC-I y MC-II) hacia la negativa. También en el segundo factor se contraponen las ciencias sociales (parte positiva) a las humanísticas (parte negativa), como se muestra en la figura 7. El AFC permitió corroborar los resultados del ACP realizado, con lo que se reafirma el hecho de que las primeras componentes permiten clasificar a los estudiantes según su desempeño docente. Esto está dado por la formación de grupos de variables como se mencionó anteriormente. Lo cual se puede considerar como un punto de partida para la toma de decisiones en el proceso de selección de personal.

### Relación de los resultados con la gestión por competencias

La gestión de competencias está directamente relacionada con el análisis de los elementos que influyen en la motivación humana y lo que las personas sean capaces de hacer mediante ésta. Los temas de su concepción se enfocan en lo que realmente se puede hacer en la actualidad y qué se puede llegar a hacer en el futuro, por lo que la productividad está estrechamente ligada a este fenómeno. Los resultados obtenidos se han relacionado con las competencias genéricas relativas al aprendizaje, según la agrupación que se realiza por González y González (2008) [17]:

- Capacidad de abstracción, análisis y síntesis: Estudiantes del grupo 3 que se encuentran en el cuadrante descrito por la parte positiva de la primera y segunda componente. Éstos muestran alto desempeño en las asignaturas MC-I, MC-II, SBD, P-II, M-III, M-IV y F-II.
- Conocimientos sobre el área de estudio y la profesión y capacidad de aplicar los conocimientos en la práctica: Estudiantes del grupo 3, cargan en la primera componente, pero cargan poco en la segunda. Éstos muestran un alto desempeño en las asignaturas MC-I, MC-II, SBD, P-II.
- Capacidad de comunicación en un segundo idioma: Estudiantes del grupo 3, ubicados en el sentido positivo de la primera componente y en el sentido negativo respecto a la segunda, que muestran alto desempeño en las asignaturas EI-III, EI-IV.

La observación en la práctica de las habilidades y conocimientos construidos es una de las mejores formas de conocer en qué nivel están las competencias desarrolladas por un individuo. En este sentido, Steedman (1994) se refiere a la insuficiencia de los exámenes escritos frente a las pruebas prácticas del ejercicio de desempeño laboral [18]. Por otra parte, Cuesta (2001) sostiene que en estudios realizados sobre el desempeño laboral de graduados universitarios, donde relacionaron sus coeficientes de inteligencia con el desempeño laboral exitoso, se llegó a la conclusión de que no necesariamente estudiantes universitarios con elevados coeficientes de inteligencia tienen un desempeño laboral satisfactorio [19].

### Utilidad de la información generada

Con la secuencia de métodos empleada en el análisis de los datos se logra generar información de apoyo para guiar el proceso de selección de estudiantes para los proyectos de *software*. El uso de la clasificación obtenida contribuye a una mejor asignación de roles en los proyectos, pues se evita, en cierta medida, la ubicación de estudiantes en roles para los cuales no tienen las competencias mínimas requeridas.

La clasificación obtenida ofrece un elemento concluyente, pues en los grupos de estudiantes 1 y 2, se encuentran aquellos con un menor desempeño docente de forma general, pero con alto rendimiento en las asignaturas P-II y SBD; relacionadas con las competencias, conocimientos sobre la profesión y la capacidad de aplicar sus conocimientos en la práctica. Existen evidencias históricas de que muchos de los estudiantes de esta naturaleza tienen buen comportamiento en los proyectos, debido a su motivación por la profesión.

Una vez en los proyectos, los estudiantes continúan su proceso de formación, por tanto la clasificación debe ser usada como apoyo a la puesta en práctica de estrategias de capacitación con los estudiantes de menor rendimiento para una incorporación futura al desarrollo de *software*.

## V. CONCLUSIONES

1. Los resultados de este trabajo conforman una guía elemental para realizar estudios exploratorios y confirmatorios, en la obtención de información útil en el proceso de selección de estudiantes para su vinculación a proyectos de desarrollo de *software*.
2. Se realizó la clasificación de los estudiantes teniendo en cuenta sus resultados docentes y se pudieron agrupar a aquellos con comportamientos similares en sus resultados docentes. Con tales argumentos, se podrán evitar incidentes en los proyectos de desarrollo de *software* causados por la inadecuada selección de los estudiantes que se vinculan a ellos como parte de su proceso de formación.
3. Se debe tener en cuenta que el análisis exploratorio solamente se efectuó con datos de estudiantes de la Facultad 3. Por tanto, no se recomiendan generalizaciones de los resultados estadísticos en otras facultades.
4. De acuerdo al carácter complejo de las competencias profesionales, para realizar análisis futuros se recomienda el empleo de datos que contengan componentes cognitivos (conocimientos, habilidades) y motivacionales (actitudes, sentimientos, valores) recogidos en el desempeño académico y profesional de los estudiantes. 📖

## VI. REFERENCIAS

1. LISTER, T., *Peopeware: Productive Projects and Teams*, New York, Dorset House, 1999, ISBN 978-0-932633-43-9.
2. WATERSON, P.; WEIBELZAHN, S.; PFAHL, D., *Software Process Modelling: Socio-Technical Perspectives* [en línea], Springer, 2005 [consulta: 2010-01-20]. ISBN 0-387-24262-7 Disponible en: <<http://www.easy-hub.org/stephan/watsonson-spm05.pdf>>

## ANÁLISIS MULTIVARIADO DE DATOS COMO SOPORTE A LA DECISIÓN EN LA SELECCIÓN DE ESTUDIANTES EN PROYECTOS DE SOFTWARE

3. AMPUERO, M. A.; BALDOQUÍN, M. G., «Un Sistema de Soporte a la Decisión para la Asignación de Recursos humanos a equipos de proyectos de Software» *Revista Cubana de Investigación Operacional*, 2010, vol. 31, no. 1, pp. 20-28, ISSN 0257-4306.
4. MARTÍNEZ, O. L., *Estrategia de caracterización. Libro blanco. Informe del Centro de Innovación y Calidad de la Educación* La Habana, Universidad de las Ciencias Informáticas, 2009 -, publ. -[consulta: 2012-03-15]. Disponible en: <<https://cice.uci.cu>>.
5. ESTÉVEZ, J. F.; GARCÍA, M. J., *Sistema de indicadores para el diagnóstico y seguimiento de la educación superior en México*, México, D. F., Editorial UNAIES, 2007, ISBN 978-904-103-X.
6. RODRÍGUEZ, A. G.; HAEDO, Y., «Análisis multivariado del proceso de aprendizaje en el 1er año de ingeniería», en *V Taller de Enseñanza de la Ingeniería Industrial* La Habana, Cuba, Instituto Superior Politécnico José Antonio Echeverría, Cujae, 2009, ISBN 978-959-261-281-5.
7. HEREDIA-RICO, J. J.; RODRÍGUEZ-HERNÁNDEZ, A. G. ; VILALTA-ALONSO, J. A., «El análisis de datos en apoyo a la gestión de la enseñanza de la carrera Ingeniería Industrial.», *Ingeniería Industrial* [en línea], 2012, vol. XXXIII, no. 1, pp. 19-30 [consulta: 2012-10-19], ISSN 1815-5936. Disponible en: <<http://rii.cujae.edu.cu/index.php/revistaind/article/view/375/439>>
8. MAYS, R., «Multivariate Analysis in Personal selection.» *Educational and psychological measurement*, 1976, vol. 36, no. 4, pp. 905-912, ISSN 1552-3888.
9. JING, L., «Data Mining and Knowledge Management in Higher Education- Potential Applications», en *Annual Forum for the Association for Institutional Research* Toronto, 2002, [consulta: 2010-10-04]. Disponible en: <<http://eric.ed.gov/PDFS/ED474143.pdf>>
10. MERIGÓ, L. M., «Nuevas extensiones de los operadores OWA, y su aplicación a los métodos de decisión», [tesis doctoral], Barcelona, Universidad de Barcelona, Facultad de Economía y Empresa, 2008.
11. CANÓS, L.; CASASÚS, T.; LARA, T.; LIERN, V.; PÉREZ, J. C., *Modelos flexibles de selección de personal basados en la valoración de competencias. Reporte Técnico* Valencia (España), Departamento de Matemáticas para la Economía y la Empresa. Universidad de Valencia, 2008 -, publ. -[consulta: 2010-10-21]. Disponible en: <[http://www.asepuma.org/recta/ordinarios/9/9\\_4.pdf](http://www.asepuma.org/recta/ordinarios/9/9_4.pdf)>.
12. JOHNSON, R. A.; WICHERN, D. W., *Applied Multivariate Statistical Analysis*, New Jersey, Prentice Hall, 1998, ISBN 0-130-41146-9.
13. HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C., *Análisis Multivariante*, Madrid, Prentice Hall Iberia, 1999, ISBN 84-8322-035-0.
14. JOLLIFFE, I. T., *Principal Component Analysis*, 2nd.ed., New York, Springer, 2002 (Series In Statistical), ISBN 0-387-95442-2.
15. VARMUZA, K.; FILZMOSER, P., *Introducion Multivariate Statistical Analysis in Chemometrics*, London, CRC Press, 2009, ISBN 978-1-4200-5947-2.
16. VARMUZA, K.; FILZMOSER, P., *Package 'Chemometrics' v-0.8 Introducion Multivariate Statistical Analysis in Chemometrics*, London, CRC Press, 2009, ISBN 978-1-4200-5947-2.
17. GONZÁLEZ, V.; GONZÁLEZ, R. M., «Competencias genéricas y formación profesional: un análisis desde la docencia universitaria» *Revista Iberoamericana de Educación*, 2008, vol. 15, no. 47, pp. 185-210, ISSN 1022-6508.
18. STEEDMAN, H., «Evaluación, certificación y reconocimiento de las destrezas y competencias profesionales» *Revista Europea de Formación del Profesional*, 1994, vol. 1, no. 1, pp. 38-45, ISSN 0258-7483.
19. CUESTA, A., *Gestión de competencias*, La Habana, Editorial Academia, 2001, ISBN 959-02-0324-8.