



Cluster y valoración de respuestas a necesidades educativas

Mirta M. Arias¹ & Ana M. Sfer²

(1) *Departamento de Matemática, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Tucumán.*

marias@herrera.unt.edu.ar

(2) *Departamento de Matemática, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Tucumán.*

asfer@herrera.unt.edu.ar

RESUMEN: El problema del pasaje de la Escuela Media a la Universidad sigue vigente. Nuestra facultad ofrece capacitación extracurricular a jóvenes que están cursando el último año del nivel secundario en contenidos que faciliten un recorrido más fluido en el tránsito al nivel superior. El propósito de este trabajo ha sido considerar al curso de ingreso a la Facultad de Ciencias Exactas y Tecnología de la UNT como una Intervención Didáctica para alumnos de la Escuela Media y validarlo utilizando el análisis de cluster de medidas pre y pos, así como los incrementos, con el objeto de poder constatar semejanzas y diferencias entre los alumnos, en cuanto a su evolución. De esta manera la técnica Clustering se utiliza como una herramienta de valoración alternativa a simples porcentajes en proyectos de enseñanza-aprendizaje, en un marco caracterizado por la diversidad, en este caso, de instituciones formadoras. Esto es de gran importancia porque muchos proyectos del área educación miden el éxito o fracaso a través de percepciones personales.

1 EXPERIENCIA

1.1 Generalidades

A través de la técnica de Clustering hemos evaluado una instancia del curso de ingreso a nuestra facultad, considerándolo como un programa de intervención didáctica que utiliza dos áreas de interés preferente: destrezas instrumentales básicas en contenidos del área matemática y estrategias de estudio y aprendizaje.

La primera fue evaluada con pruebas diagnósticas y evaluaciones parciales, y la segunda con el IEEA, inventario de Estrategias de Estudio y Aprendizaje.

De la experiencia participaron dos comisiones del curso de ingreso dictado en la ciudad capital de la provincia y una comisión del interior (ciudad de Monteros).

1.2 Pruebas diagnósticas y evaluaciones parciales

Las evaluaciones diagnósticas consistieron en dos ejercicios con varios apartados, elegidos de tal manera que permitieran detectar los conociemien-

tos previos al curso. El bajo número de ejercicios se debió a la intención de no entorpecer el normal desenvolvimiento de las clases pautadas. Esto obligaba a una elaboración concienzuda de las pruebas diagnósticas a cargo de una de las autoras de este trabajo, ex coordinadora de la asignatura matemática del ingreso durante más de diez años. En total se tomaron tres pruebas diagnósticas con las características especificadas.

Con respecto a los parciales: fueron tres pruebas elaboradas por los responsables del curso y tomadas por los respectivos profesores de las comisiones.

Cabe destacar que cada evaluación diagnóstica correspondía a los contenidos curriculares de cada parcial.

1.3 Inventario de estrategias de estudio y aprendizaje

Para medir el desarrollo metacognitivo se usó como instrumento el LASSI. (Learning and Study Strategies Inventory -1987). Es un instrumento de evaluación diseñado para medir el grado en el que los alumnos usan métodos y estrategias de estudio y aprendizaje y consiste en un cuestionario de 77 enunciados. En consecuencia sirve para recolectar

información acerca de las prácticas de estudio y aprendizaje.

El LASSI tiene 10 escalas y proporciona 10 calificaciones individuales y no hay una calificación global. Maneja percentiles, y los percentiles debajo de 50 son considerados como indicador de un uso inapropiado de las estrategias de aprendizaje. Este inventario utiliza como promedio normal percentiles de 50 a 75.

Después de usar el LASSI las respuestas se compararon con las respuestas dadas por el grupo de alumnos que se usó como base para normar el instrumento. Es decir que el LASSI también incluye normas para transformar las puntuaciones a sus equivalentes en percentiles (75% es usado como una calificación mínima en muchas universidades). Las puntuaciones directas de cada escala se transforman en percentiles y su interpretación es la siguiente: los sujetos que puntúan sobre el percentil 75 se consideran que poseen las estrategias adecuadas para ser autosuficientes en el aprendizaje, los que puntúan entre 50 y 75 se encuentran en la media del grupo y éstos deberían mejorar en el uso de estrategias para optimizar sus resultados académicos; finalmente los estudiantes que puntúan por debajo del percentil 50, requieren ayuda especial en cuanto al aprendizaje de estrategias si se desea obtener un éxito escolar. El instrumento usado, que inicialmente estaba normado con una muestra de alumnos de primer año de una universidad del sur de los Estados Unidos (Arias & Sfer, 2004), fue el Inventario de Estrategias de Estudio y Aprendizaje, ya descrito, pero normado nuevamente, después de mucho trabajo, con los alumnos de Ingeniería en Computación de la Universidad Nacional de Tucumán.

1.4 Datos

Se trabajó con la población total de la ciudad de Monteros estudiándose el grupo intacto, es decir considerando su formación natural y con dos comisiones de la capital elegidas al azar. La población del interior estaba formada por 40 alumnos y las de la ciudad por 60 alumnos. En las tres comisiones se produjo un gran desgranamiento, reduciéndose el estudio a 70 alumnos. Aunque esto puede quitar fiabilidad a los resultados, se decidió elaborar este trabajo para que sirva como muestra de una metodología novedosa a la hora evaluar proyectos de intervención didáctica.

Se trabajó con las notas de los tres parciales y las 10 escalas del IEEA:

- ACT: actitud e interés.
- MOT: motivación, diligencia, autodisciplina y voluntad para trabajar con empeño.
- ATI: administración del tiempo en la realización de actividades académicas.

- ANS: ansiedad y preocupación por el trabajo escolar.
- CON: concentración y atención en las actividades académicas.
- PIN: procesamiento de la información, adquisición de conocimiento y razonamiento.
- SIP: selección de ideas principales y reconocimiento de información importante.
- AES: ayudas (técnicas y materiales) de estudio.
- AEV: autoevaluación, repaso y preparación de clases.
- PPE: estrategias para la preparación y presentación de exámenes.

En general, las variables con mayor dispersión (es decir; grandes desviaciones estándar) tienen más impacto en el valor final de similitud, por ello se trabajó con datos estandarizados. Se usó la forma más común de estandarización que es la conversión de cada variable a una puntuación estándar.

2 ANÁLISIS CLUSTER

2.1 Introducción

Análisis cluster (o análisis de conglomerados) es un término genérico para técnicas de análisis exploratorios de un conjunto de datos. Generalmente los análisis estadísticos situados en la ortodoxia de la investigación experimental están dirigidos a estudiar individualmente cada una de las variables y por consiguiente plantean la dificultad de una visión integradora.

El análisis cluster proporciona un método objetivo y empírico para desarrollar la clasificación ya sea a efectos de simplificación, exploración o confirmación. Es la denominación de un grupo de técnicas multivariantes cuyo principal propósito es agrupar objetos basándose en las características que poseen.

Como se ha dicho, el análisis de conglomerados tiene por objeto agrupar elementos en grupos heterogéneos en función de similitudes entre ellos.

Razón por la cual se ha realizado una serie de análisis que consideran la evolución global de los sujetos a través del programa de intervención didáctica que permite explicar los diferentes grados de aprovechamiento y las mejoras en los sujetos que han recibido la intervención.

Con el objetivo de integrar la información proporcionada por la evolución de los valores de todas las variables, se ha realizado un análisis de cluster dirigido a clasificar a los individuos de acuerdo con el incremento experimentado en las variables dependientes durante el período comprendido entre las medidas previas y posteriores a la aplicación del programa.

2.2 Investigación mediante el análisis cluster

2.2.1 Selección de la medida de proximidad

La similitud entre objetos es una medida de correspondencia o parecido entre objetos que van a ser agrupados. Ésta puede medirse de varias formas: medidas de correlación, medidas de proximidad (distancia, disimilaridad, similaridad), medidas de asociación, etc., dependiendo de los objetos y del tipo de datos. Tanto las medidas de distancia como la de correlación exigen datos métricos, mientras que la medida de asociación es para datos no métricos.

Una de las medidas de correlación es el coeficiente de correlación entre un par de objetos medido sobre varias variables. Elevadas correlaciones indican similitud.

Las medidas de proximidad representan la similitud como la proximidad de las observaciones. Se denominan proximidades a las funciones distancia, similaridad y disimilaridad.

Las medidas de distancia son las más usadas, donde valores elevados indican una menor similitud. Es decir, la distancia se convierte en medida de similitud utilizando una relación inversa.

Una función $d(i, j)$ definida para los pares (i, j) de los objetos a agrupar, se dice que es una distancia si tiene las propiedades (1), (2), (3) y (4):

$$d(i, j) \geq 0 \quad (1)$$

$$d(i, i) = 0 \quad (2)$$

$$d(i, j) = d(j, i) \quad (3)$$

$$d(i, j) \leq d(i, h) + d(h, j) \quad (4)$$

La fi

Hay varios tipos de medidas de distancia: distancia euclidiana, distancia city block o distancia Manhattan, distancia de Minkowski o distancia L_q , distancia euclideana ponderada, en la cual cada variable recibe un peso de acuerdo a su importancia relativa, lo que presupone un conocimiento previo de las variables, distancia Karl Pearson (euclideana con ponderaciones), generalización de distancia euclideana ponderada, distancia de Mahalanobis (donde las ponderaciones son las componentes de la inversa de la matriz de varianza y covarianza de las variables del problema).

En la experiencia se usó la distancia euclideana definida por la ecuación (5):

$$d(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + \dots + (x_{pi} - x_{pj})^2} \quad (5)$$

donde x_{1i} a x_{pi} son las variables para el sujeto i .

2.2.2 Selección del método aglomerativo

Con las variables consideradas y la matriz de similitud o proximidad calculada, se debe seleccionar en primer lugar el algoritmo de aglomeración y luego el número de conglomerados.

Dadas las características de las variables consideradas y el objetivo pretendido, se ha realizado como primera etapa un agrupamiento de sujetos por el procedimiento de encadenamiento simple o del vecino más próximo (Florek et al. 1951 y Sneath 1957b) utilizando la distancia euclídea. Como segunda y tercera etapa se trabajó con las técnicas de agrupamiento jerárquicas: encadenamiento completo (complete linkage clustering) o método del vecino más alejado y el encadenamiento promedio (average linkage) o método del promedio grupal (Sokal y Michener). En todas las instancias se utilizaron programas del paquete estadístico Stata creado en 1985 por Stata Corp.

Teniendo en cuenta los resultados de esta tarea experimental se decidió trabajar con el segundo método mencionado. El procedimiento de encadenamiento completo se basa en la distancia máxima. La distancia máxima entre individuos de cada conglomerado representa la esfera más reducida (diámetro mínimo) que puede incluir todos los objetos de ambos conglomerados como muestra tra la figura 1.



Figura 1: Comparación de las medidas de distancia del encadenamiento simple y completo.

La figura 1 muestra cómo las distancias más cortas (encadenamiento simple) o las más largas (encadenamiento completo) representan la similitud entre grupos. Ambas medidas reflejan sólo un aspecto de los datos. El uso de la distancia más corta refleja un único par de objetos (los más cercanos) y el encadenamiento completo también refleja un único par (los más lejanos).

La distancia entre dos grupos R y Q se define usando la ecuación (6).

$$d(R, Q) = \max_{i \in R, j \in Q} d(i, j) \quad (6)$$

La distancia entre grupos: uno nuevo, producto de una fusión, y otro cualquiera (fórmula reducida) está dada por la ecuación (7).

$$d(R, Q) = \max [d(A, Q), d(B, Q)] \quad (7)$$

La fórmula para la actualización está definida por la ecuación (8).

$$d(R, Q) = \frac{1}{2} [d(A, Q) + d(B, Q)] + \left[\frac{1}{2} |d(A, Q) - d(B, Q)| \right] \quad (8)$$

2.2.3 Dendograma

El dendograma, o árbol jerárquico, es una representación gráfica del resultado del proceso de agrupamiento en forma de árbol. Se construye de la siguiente manera: en la parte inferior del gráfico se disponen los n elementos iniciales; las uniones entre elementos se indican por tres líneas rectas, dos dirigidas a los elementos que se unen, y que son perpendiculares al eje de los elementos y una paralela a este eje, que se sitúa al nivel en que se unen.

El proceso se repite hasta que todos los elementos están conectados por líneas rectas.

2.2.4 Determinación del número de grupos

En la literatura científica fueron publicados diferentes y muy variados métodos para determinar el número de cluster. El estudio comparativo más detallado y completo fue emprendido por Milligan y Cooper. Ellos analizaron y clasificaron 30 índices e investigaron el grado con el cual estos índices eran capaces de determinar el número de clusters en conjuntos simulados de datos. Las dos mejores reglas investigadas en este estudio están definidas a continuación para datos continuos en el ranking realizado en la investigación de Milligan y Cooper. El índice Calinski / Harabasz está dado por la ecuación (9).

$$CH = \frac{\frac{B}{c-1}}{\frac{W}{n-c}} \quad (9)$$

donde n es el número de individuos o unidades y c es el número de cluster en la partición. B y W denotan la suma total de las distancias al cuadrado entre cluster (considerando los centroides) y la suma de las distancias al cuadrado dentro del cluster, respectivamente. El máximo valor del índice es usado como indicador del número efectivo de cluster en el conjunto de datos.

Duda y Hart propusieron un test de hipótesis para decidir si un cluster debería ser subdividido en dos sub-cluster. El test estadístico está basado en la comparación entre W_1 , las sumas de las distancias al cuadrado dentro del cluster, y W_2 , la suma de las sumas de las distancias al cuadrado cuando el cluster es particionado de manera óptima en dos. La hipótesis nula de sólo un cluster es rechazada si:

$$J = \left(-\frac{w_2}{W_1} + 1 - \frac{2}{\pi p} \right) \left(\frac{2 \left(1 - \frac{8}{\pi^2 p} \right)}{np} \right) > Z_{1-\alpha} \quad (10)$$

donde p es la dimensión de los datos, n el número de objetos del cluster que está siendo investigado y $Z_{1-\alpha}$ una desviación estándar normal especificando el nivel de significancia del test. Varios valores para registros estándares fueron testeados.

3 RESULTADOS

3.1 Análisis Cluster. Premedidas

Los resultados obtenidos para las medidas previas a la intervención didáctica se muestran en la Figura 1 las tablas 1, 2, 3 y 4.

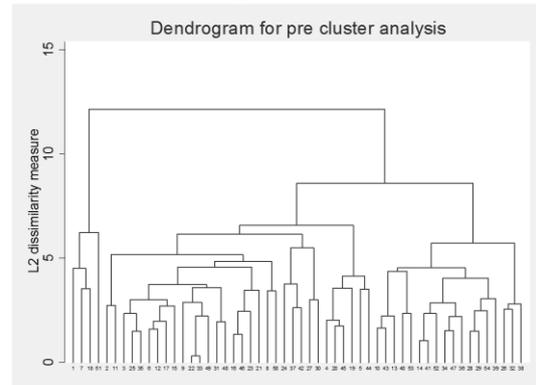


Figura 1. Dendograma usando encadenamiento completo con las pre-medidas.

Para determinar el número de grupos se usaron dos índices provistos en Stata: Calinski y Harabasz, pseudo estadístico F y Duda y Hart. Los resultados se muestran en las siguientes tablas.

Number of clusters	Calinski / Harabasz pseudo-F
2	10.46
3	14.23
4	11.54
5	9.69
6	9.58
7	9.51
8	9.00
9	8.67
10	8.30
11	8.30
12	8.33
13	8.06
14	8.45
15	8.34

Tabla 1: Valores del Índice Calinski / Harabasz.

Number of clusters	Duda/Hart	
	Je(2)/Je(1)	pseudo T-squared
1	0.8325	10.46
2	0.7533	15.72
3	0.8718	4.41
4	0.4570	2.38
5	0.8097	5.64
6	0.7349	5.77
7	0.4678	3.41
8	0.8500	3.35
9	0.8521	2.95
10	0.7612	4.71
11	0.7801	3.66
12	0.3919	1.55
13	0.2634	8.39
14	0.6136	2.52
15	0.6231	4.84

Tabla 2: Visualización para el test Duda/Hart.

Interpretando los valores obtenidos en las tablas 1 y 2 se decidió trabajar con tres grupos conformados de la siguiente manera:

grpre3	id
1	51
	7
	1
	18

Cluster 1

grpre3	id						
2	27	31	42	19	4	9	45
	21	30	46	5	35	37	11
	15	44	22	24	6	12	
	48	3	17	23	20	8	
	50	16	49	2	33	25	

Cluster 2

grpre3	id			
3	14	36	26	41
	29	34	39	47
	13	52	38	54
	53	40	43	
	28	32	10	

Cluster 3

Observando los tres clusters formados, elaboramos la tabla 3.

grpr3	Freq.	Percent	Cum.
1	4	7.41	7.41
2	18	33.33	40.74
3	32	59.26	100
Total	54	100	

Tabla 3: Porcentajes de alumnos por grupos.

El método jerárquico de encadenamiento completo agrupó aproximadamente un 7% de los alumnos en el conglomerado 1, un 33% en el conglomerado 2 y un 60% en el 3.

grpre3	mean											
	prom	act	mot	ati	ans	con	pin	sip	aes	aev	ppe	
1	15	32.25	21	21	21.25	18.75	22.75	8.5	17	24.25	18.75	
2	8.875	32.875	28.219	24.969	25.5	27.937	28.812	18.033	24.594	28.437	29.125	
3	14.722	36.833	35.222	32.556	30.667	34.778	32.556	22.167	29.611	30.667	35.778	

Tabla 4: Características de los grupos.

Ciertas características de los grupos se resumen en la tabla 4. En ella se presentan las medias de las once variables consideradas en la experiencia para los respectivos clusters obtenidos con las mediciones previas a la intervención didáctica. Haciendo una lectura metódica de la tabla 4 se pueden inferir las diferencias entre los grupos.

Si bien los tres grupos se caracterizan por una calificación aplazo, la más baja es la del cluster 2, hecho que proporciona una diferencia para el grupo mencionado.

El grupo 1 tiene los valores de media más bajos para las estrategias de estudios y aprendizajes, y el grupo 3 los más altos.

3.2 Análisis Cluster. Posmedidas

Los resultados obtenidos para las medidas posteriores a la intervención didáctica se muestran en la Figura 2, las tablas 5, 6, 7 y 8.

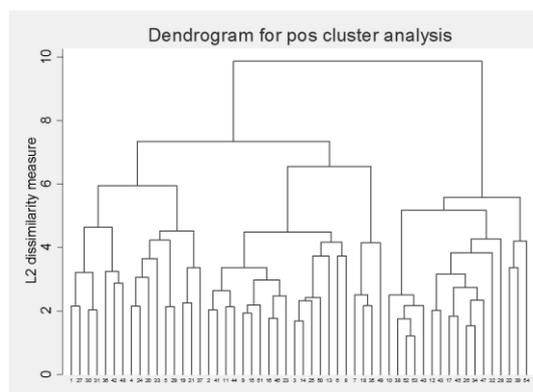


Figura 2. Dendrograma usando encadenamiento completo con las posmedidas.

Para determinar el número de grupos también se usaron los índices: Calinski y Harabasz, y Duda y Hart. Los resultados se muestran en tablas 5 y 6.

Number of clusters	Calinski / Harabasz pseudo-F
2	24.73
3	16.62
4	14.18
5	13.83
6	12.23
7	11.81
8	10.99
9	10.35
10	9.93
11	9.36
12	9.04
13	8.69
14	8.63
15	8.60

Tabla 5: Valores del Índice Calinski / Harabasz-para posmedidas.

Number of clusters	Duda/Hart	
	Je(2)/Je(1)	pseudo T-squared
1	0.6777	24.73
2	0.8586	5.76
3	0.7136	7.63
4	0.6666	7.00
5	0.8205	3.28
6	0.6342	6.92
7	0.5849	3.55
8	0.7286	2.61
9	0.8363	4.94
10	0.7388	2.48
11	0.6032	2.63
12	0.3980	1.51
13	0.6553	2.63
14	0.3478	3.75
15	0.7228	2.30

Tabla 6: Visualización para el test Duda/Hart para posmedidas.

grpos2	mean										
	prom	act	mot	ati	ans	con	pin	sip	aes	aev	ppe
1	38.108	33.189	26.324	23.081	23.622	25.270	28	16.676	23.513	26.973	27.919
2	53.235	37.059	34.471	31.941	30.882	34.882	34.471	22.882	31.059	32.235	35.471

Tabla 7: Características de los grupos generados con las posmedidas.

El grupo 1 tiene una media del promedio muy próxima al cuatro (pero sigue siendo aplazo) y el cluster 2 una que supera el 5. Lo enunciado proporciona una diferencia entre los dos grupos. Otra característica que distingue los grupos es que todas las medias de las variables que evaluaron estrategias de estudio y aprendizaje son menores en el grupo 1. Las que presentan mayores diferencias son sip, ppe, mot y aes.

3.3 Análisis Cluster.de las diferencias

El análisis cluster con las diferencias entre las medidas previas y posteriores al curso permitió validar el proyecto de intervención didáctica implementado.

Los resultados de las tablas 5 y 6 determinaron que después de la intervención didáctica se consideren dos grupos.

El método aglomerativo seleccionado con encadenamiento completo o método del vecino más alejado proporcionó los grupos que se presentan a continuación:

grpos2	id							
1	50	33	2	6	48	20	46	18
	51	42	1	29	7	5	14	19
	37	3	25	11	23	35	24	
	49	13	4	31	8	16	21	
	15	41	30	44	27	36	9	

Cluster 1

grpos2	id			
2	43	26	32	12
	52	54	39	47
	40	53	45	
	22	28	34	
	10	38	17	

Cluster 2

El 68,52% de los alumnos están agrupados en el cluster 1.

Las medias de las variables consideradas en este estudio, por grupo, se presentan en la tabla 7.

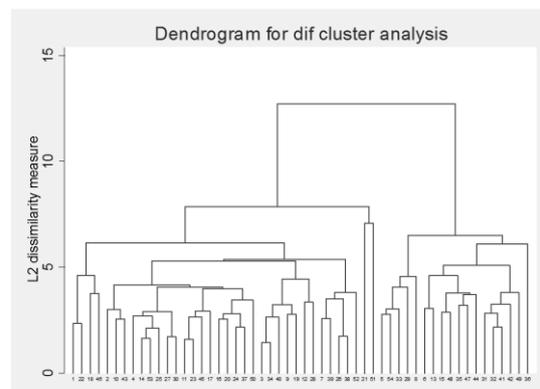


Figura 3. Dendrograma usando encadenamiento completo con las diferencias.

El número de grupos se determinó con los índices: Calinski y Harabasz y Duda y Hart que provee el paquete estadístico Stata. Los resultados provistos por los programas usados se muestran en tablas 8 y 9.

Number of clusters	Calinski / Harabasz pseudo-F
2	16.42
3	11.49
4	9.13
5	8.47
6	7.99
7	7.36
8	7.09
9	7.10
10	6.94
11	6.66
12	6.56
13	6.26
14	6.04
15	5.93

Tabla 8: Calinski / Harabasz para diferencias.

Number of clusters	Duda/Hart	
	Je(2)/Je(1)	pseudo T-squared
1	0.7600	16.42
2	0.8580	5.63
3	0.0000	
4	0.7905	4.24
5	0.8816	4.30
6	0.8152	2.49
7	0.8846	3.65
8	0.8322	4.64
9	0.7832	2.77
10	0.4492	2.45
11	0.6626	2.55
12	0.6616	1.53
13	0.6907	2.24
14	0.8755	2.28
15	0.8270	2.72

Tabla 9: Visualización para el test Duda/Hart para diferencias.

grdif2	mean										
	prom	act	mot	ati	ans	con	pin	sip	aes	aev	ppe
1	37.722	1.4167	-.6111	-1.222	.5556	1.4167	1.7222	1.5278	1.6389	1.25	.8333
2	18.5	-2.055	4.6111	6.4444	-3.222	-6	-2.167	-3.278	-2.722	-3.222	-2.5

Tabla 10: Características de los grupos generados con las diferencias entre las pre y pos mediciones.

Los alumnos que están agrupados en el cluster 1 mejoraron en promedio y en general en el desarrollo de las estrategias de estudio y aprendizaje. En particular se detecta un mejoramiento en la selección de ideas principales y reconocimiento de información importante y en menor medida (pequeñas mejorías) específicamente en: procesamiento de la información, adquisición de conocimientos y razonamiento, ayudas (técnicas y materiales) de estudio, actitudes e interés y con-

Para las diferencias no hay coincidencias entre el índice de Calinski/Harabasz y Duda/Hart. Quedando como tarea futura, estudiar las soluciones para estos casos y, de modo consecuente con Calinski y Harabasz, el número de clusters considerado para las diferencias fue dos.

El método aglomerativo seleccionado, encadenamiento completo o método del vecino más alejado, proporcionó para las diferencias los grupos que se presentan a continuación:

grdif2	id							
1	39	26	23	52	12	3	21	34
	2	16	7	25	45	1	14	
	37	9	27	53	17	4	22	
	24	30	38	11	50	19	18	
	51	43	28	40	46	10	20	

Cluster 1

grdif2	id			
2	35	31	42	44
	47	33	6	41
	29	8	49	54
	48	36	13	
	32	15	5	

Cluster 2

El 66,66% de los alumnos están agrupados en el cluster 1 y el resto en el cluster 2.

Las medias de las variables consideradas en este estudio, de cada grupo, se presentan en la tabla 10. Observando con atención los valores que se tabulan se encontraron algunas diferencias que caracterizan a los dos agrupamientos.

centración y atención en las actividades académicas.

Los alumnos agrupados en el cluster 2 mejoraron muy poco en promedio y en mayor medida sólo en dos variables que miden estrategias de estudio y aprendizaje: mot (motivación, diligencia, auto-disciplina y voluntad para trabajar con empeño) y ati (administración del tiempo en la realización de las actividades académicas). Estas variables tienen un rango de variación de 39 puntos.

Los valores negativos en las variables que se presentan en la tabla 10, excepto promedio, no se deben a una involución sino a que pudieron haber respondido el inventario sin cuidado o sin la responsabilidad que se merecía, dando una importante participación al azar en cualquiera de las dos instancias de su aplicación.

4 CONCLUSIONES

De los análisis que se hicieron se expresan las siguientes conclusiones:

- Al tener un conglomerado menos en el agrupamiento con las posmediciones, se puede inferir que la intervención didáctica fue positiva.
- Las características de un grupo generado por el método de encadenamiento completo con las posmediciones muestra que hay mejoras en cuanto a las destrezas instrumentales básicas en contenidos del área matemática y pequeños avances en el desarrollo de ciertas estrategias de estudio y aprendizaje.
- Sin embargo, teniendo en cuenta la media de las variables, de los clusters generados por las diferencias, se afirma que el proyecto implementado no alcanzó las expectativas esperadas ya que el rango de variación de la variable promedio es [0,100], el de las variables act, mot, ati, ans, pin, aev y ppe es [0,39], el de con y aes es [0,38] y el de sip [0,25].
- Las medidas previas no sólo sirven para encontrar diferencias sino que también pueden sugerir una reorientación del proyecto.
- El número tan bajo de grupos para las premedidas no mostraron la diversidad que se esperaba. Este hecho confirma la existencia de un grupo de alumnos homogéneo en general que nos plantea la necesidad de una atención educativa dirigida a enriquecer fuertemente su formación.

5 REFERENCIAS

- Arias, M. M. & A. M. Sfer, *Estrategias de estudio y aprendizaje de alumnos que ingresan a la universidad. Desarrollo metacognitivo*, IV Encuentro Nacional y I Latinoamericano. La Universidad como Objeto de Investigación, Tucumán, Argentina, 2004.
- Coleman, D. A. & D. L. Woodruff, Cluster Analysis for Large Datasets: An Effective Algorithm for Maximizing the Mixture Likelihood, *Journal of Computational and Graphical Statistics*, Volumen 9- N° 4, 672-688, 2000.
- Garré, M. , J. J. Cuadrado & M. A. Sicilia, *Comparación de Diferentes Algoritmos de Clustering en la Estimación de Coste en el Desarrollo de Software* – Universidad de Alcalá – Madrid.
- Gorosito, A. *Métodos Estadísticos de Agrupamiento y sus Aplicaciones en Geología Tesis de Magíster* – INIE – Universidad Nacional de Tucumán, 1995.
- Hair, J. F. , R. E. Anderson , Tatham, R.I. & W. C. Black, , *Multivariate Data Analysis-* Fifth Edition, Prentice Hall - New Jersey, 1999.
- Peña, D., *Análisis de Datos Multivariantes* – Mc Graw Hill – España, 2002.
- Rabe-Hesketh, S. & B. Everitt, *Handbook of Statistical Analyses using Stata*, third edition, Chapman & Hall/Crc, United States, 2004
- Symbolic Data Analysis and the SODAS Software – Editores Edwin Diday and Monique Noirhomme-Fraiture – England. 2008.
- Torres González, J. A. & A. Sánchez Palomino, El análisis cluster aplicado a la valoración de necesidades educativas especiales, *Enseñanza*, 14, 181-194, 1996.